



Unstructured Data Analytics

SAP Sybase IQ 16.0 SP03

DOCUMENT ID: DC01268-01-1603-01

LAST REVISED: December 2013

Copyright © 2013 by SAP AG or an SAP affiliate company. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP AG. The information contained herein may be changed without prior notice.

Some software products marketed by SAP AG and its distributors contain proprietary software components of other software vendors. National product specifications may vary.

These materials are provided by SAP AG and its affiliated companies ("SAP Group") for informational purposes only, without representation or warranty of any kind, and SAP Group shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP Group products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

SAP and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP AG in Germany and other countries. Please see <http://www.sap.com/corporate-en/legal/copyright/index.epx#trademark> for additional trademark information and notices.

Contents

| | |
|---|----------|
| Introduction to Unstructured Data Analytics | 1 |
| Audience | 1 |
| The Unstructured Data Analytics Option | 1 |
| Full Text Searching | 2 |
| Compatibility | 2 |
| Conformance to Standards | 2 |
| TEXT Indexes and Text Configuration Objects | 3 |
| TEXT Indexes | 3 |
| Comparison of WD and TEXT Indexes | 3 |
| Creating a TEXT Index Using Interactive SQL | 4 |
| Guidelines for TEXT Index Size Estimation | 5 |
| TEXT Index Restrictions | 5 |
| Displaying a List of TEXT Indexes Using Interactive SQL | 5 |
| Editing a TEXT Index Using Interactive SQL | 6 |
| Modifying the TEXT Index Location Using Interactive SQL | 6 |
| Dropping a TEXT Index Using Interactive SQL | 7 |
| TEXT Index Refresh | 7 |
| TEXT_DELETE_METHOD Database Option | 7 |
| NGRAM TEXT Index | 8 |
| Text Configuration Objects | 8 |
| Default Text Configuration Objects | 9 |
| Creating a Text Configuration Using Interactive SQL | 9 |
| Text Configuration Object Settings | 10 |
| Displaying a List of Text Configurations Using Interactive SQL | 13 |
| Altering a Text Configuration Using Interactive SQL | 13 |
| Modifying the Stoplist Using Interactive SQL | 14 |

| | |
|--|-----------|
| Dropping a Text Configuration Using Interactive SQL | 14 |
| Text Configuration Object Examples | 14 |
| MAX_PREFIX_PER_CONTAINS_PHRASE Database Option | 17 |
| External Libraries | 19 |
| Pre-Filter and Term-Breaker External Libraries | 19 |
| External Library Restrictions | 20 |
| External Libraries on Multiplex Servers | 20 |
| Enable and Disable External Libraries on Startup | 20 |
| Unload External Libraries | 21 |
| Unstructured Data Queries | 23 |
| Full Text Search | 23 |
| Types of Full Text Searches | 23 |
| NGRAM TEXT Index Searches | 55 |
| Fuzzy Search Over a TEXT Index | 55 |
| Non-fuzzy Search Over a TEXT Index | 56 |
| Queries on LONG BINARY Columns | 57 |
| Queries on LONG VARCHAR Columns | 58 |
| CONTAINS Predicate Support | 58 |
| Performance Monitoring of LONG BINARY and LONG VARCHAR Columns | 58 |
| Stored Procedure Support | 59 |
| Term Management in a TEXT Index | 59 |
| sa_char_terms System Procedure | 59 |
| sa_nchar_terms System Procedure | 60 |
| sa_text_index_stats System Procedure | 61 |
| sa_text_index_vocab System Procedure | 62 |
| External Library Identification | 63 |
| sa_external_library_unload System Procedure .. | 64 |
| sa_list_external_library System Procedure | 64 |
| Large Object Data Compression | 65 |
| sp_iqsetcompression Procedure | 65 |
| sp_iqshowcompression Procedure | 66 |
| Information About Large Object Columns | 67 |

| | |
|--|-----------|
| Size of a LONG BINARY Column | 67 |
| Size of a LONG VARCHAR Column | 68 |
| Large Object Data Load and Unload | 69 |
| Large Object Data Exports | 69 |
| BFILE Function | 69 |
| Large Object Data Loads | 71 |
| Extended LOAD TABLE Syntax | 71 |
| Large Object Data Load Example | 72 |
| Control of Load Errors | 72 |
| Load of Large Object Data with Trailing Blanks ... | 73 |
| Load of Large Object Data with Quotes | 73 |
| Truncation of Partial Multibyte Character Data ... | 73 |
| Load Support of Large Object Variables | 74 |
| Large Object Data Types | 75 |
| Large Object Data Types LONG BINARY and BLOB ... | 75 |
| LONG BINARY Data Type Conversion | 75 |
| Large Object Data Types LONG VARCHAR and CLOB | 76 |
| LONG VARCHAR Data Type Conversion | 77 |
| Large Object Variables | 78 |
| Large Object Variable Data Type Conversion | 78 |
| Index Support of Large Object Columns | 79 |
| TEXT Index Support of Large Object Columns ... | 79 |
| WD Index Support of LONG VARCHAR (CLOB) Columns | 79 |
| SQL Statement Support | 81 |
| ALTER TEXT CONFIGURATION Statement | 81 |
| ALTER TEXT INDEX Statement | 84 |
| CREATE TEXT CONFIGURATION Statement | 85 |
| CREATE TEXT INDEX Statement | 87 |
| DROP TEXT CONFIGURATION Statement | 88 |
| DROP TEXT INDEX Statement | 89 |
| Function Support | 91 |
| Summary of Function Support of Large Object Data | 91 |
| BIT_LENGTH Function | 92 |

| | |
|--|------------|
| BYTE_LENGTH Function [String] | 92 |
| BYTE_LENGTH64 Function [String] | 93 |
| BYTE_SUBSTR64 and BYTE_SUBSTR Functions [String] | 94 |
| CHAR_LENGTH Function [String] | 94 |
| CHAR_LENGTH64 Function [String] | 95 |
| CHARINDEX Function [String] | 95 |
| LOCATE Function [String] | 96 |
| OCTET_LENGTH Function [String] | 97 |
| PATINDEX Function [String] | 97 |
| SUBSTRING Function [String] | 98 |
| ANSI_SUBSTRING Option [TSQL] | 99 |
| SUBSTRING64 Function [String] | 100 |
| Aggregate Function Support of Large Object Columns | 101 |
| User-Defined Function Support of Large Object Columns | 101 |
| Index | 103 |

Introduction to Unstructured Data Analytics

Learn about unstructured data analytics in SAP® Sybase® IQ and about large object data compatibility and conformance.

Audience

This guide is for users who require reference material for working with unstructured data in SAP Sybase IQ.

Learn about available syntax, parameters, functions, stored procedures, indexes, and options related to unstructured data analytics features. Use this guide as a reference to understand storage and retrieval of unstructured data within the database.

The Unstructured Data Analytics Option

The Unstructured Data Analytics Option extends the capabilities of SAP Sybase IQ to allow storage, retrieval, and full text searching of binary large objects (BLOBs) and character large objects (CLOBs) within the database.

Note: Users must be specifically licensed to use the Unstructured Data Analytics functionality described in this product documentation.

As data volumes increase, the need to store large object (LOB) data in a relational database also increases. LOB data may be either:

- Unstructured – the database simply stores and retrieves the data, or
- Semistructured (for example, text) – the database supports the data structure and provides supporting functions (for example, string functions).

Typical LOB data sources include images, maps, documents (for example, PDF files, word processing files, and presentations), audio, video, and XML files. SAP Sybase IQ can manage individual LOB objects containing gigabytes (GB), terabytes (TB), or even petabytes (PB) of data.

By allowing relational and unstructured data in the same location, SAP Sybase IQ lets organizations access both types of data using the same application and the same interface. The full text search capability of SAP Sybase IQ supports text archival applications (text analytics) in handling unstructured and semistructured data.

Full Text Searching

Full text searching uses **TEXT** indexes to search for terms and phrases in a database without having to scan table rows.

A **TEXT** index stores positional information for terms in the indexed column. Text configuration objects control the terms that are placed in a **TEXT** index when it is built or refreshed, and how a full text query is interpreted.

Using a **TEXT** index to find rows that contain a term or phrase is generally faster than scanning every row in the table.

Compatibility

SAP Sybase SQL Anywhere® (SA) and SAP Adaptive Server® Enterprise (ASE) store large text and binary objects.

SQL Anywhere can store large objects (up to a 2GB maximum length) in columns of data type `LONG VARCHAR` or `LONG BINARY`. The support of these data types by SQL Anywhere is SQL/2003 compliant. SQL Anywhere does not support the **BYTE_LENGTH64**, **BYTE_SUBSTR64**, **BFILE**, **BIT_LENGTH**, **OCTET_LENGTH**, **CHAR_LENGTH64**, and **SUBSTRING64** functions.

Adaptive Server can store large text objects (up to a 2GB maximum length) and large binary objects (up to a 2GB maximum length) in columns of data type `TEXT` or `IMAGE`, respectively. The support of these data types by Adaptive Server is an ANSI SQL Transact-SQL® extension.

A `LONG BINARY` column of a proxy table maps to a `VARBINARY (max)` column in a Microsoft SQL Server table.

Conformance to Standards

SAP Sybase IQ `LONG BINARY` and `LONG VARCHAR` functionality conforms to the Core level of the ISO/ANSI SQL standard.

TEXT Indexes and Text Configuration Objects

Learn about working with **TEXT** indexes and text configuration objects.

A **TEXT** index stores positional information for terms in an indexed column. **TEXT** indexes are created using settings stored in a text configuration object. A text configuration object controls characteristics of **TEXT** index data, such as terms to ignore, and the minimum and maximum length of terms to include in the index.

TEXT Indexes

In a full text search, a **TEXT** index is searched, rather than table rows.

Before you can perform a full text search, you must create a **TEXT** index on the columns you want to search. A **TEXT** index stores positional information for terms in the indexed columns. Queries that use **TEXT** indexes are generally faster than those that must scan all the values in the table.

When you create a **TEXT** index, you can specify which text configuration object to use when creating and refreshing the **TEXT** index. A text configuration object contains settings that affect how an index is built. If you do not specify a text configuration object, the database server uses a default configuration object.

You can create **TEXT** indexes on these types of columns: `CHAR`, `VARCHAR`, and `LONG VARCHAR`, as well as `BINARY`, `VARBINARY`, and `LONG BINARY`. `BINARY`, `VARBINARY`, and `LONG BINARY` columns require that the **TEXT** index use a text configuration with an external prefilter library.

Comparison of WD and TEXT Indexes

A comparison of **WD** and **TEXT** indexes in terms of syntax and capability.

Table 1. WD versus TEXT Index

| Feature | Supported by WD index? | Supported by TEXT index? |
|----------------------|---|--|
| Conjunction of terms | Yes, expressed in the form: <code>tbl.col CONTAINS('great', 'white', 'whale')</code> | Yes, expressed in the form: <code>CON- TAINS(tbl.col, 'great white whale')</code> |

| Feature | Supported by WD index? | Supported by TEXT index? |
|---|---|--|
| General boolean expressions | Yes, expressed in the form: tbl.col CONTAINS ('great') AND (tbl.col CONTAINS('white) OR tbl.col CONTAINS('whale') AND NOT tbl.col CONTAINS('ship')) | Yes, expressed in the form: CONTAINS (tbl.col, 'great AND (white OR whale AND NOT ship)') |
| Search for terms matching prefix | No | Yes, for example: CONTAINS (tbl.col, 'whale*') |
| Acceleration of LIKE predicates | Yes, for example: tbl.col LIKE 'whale%' | Yes, if the TEXT index is an NGRAM TEXT index without an external document pre-filter. For example: c LIKE '%apple%fruit' |
| Searches for terms in proximity | No | Yes, for example: CONTAINS (tbl.col, 'white BEFORE whale') CONTAINS (tbl.col, 'whale NEAR white') CONTAINS (tbl.col, ' "white whale" ') |
| Ordering of results based on search scoring | No | Yes |

In **TEXT** index, searching for terms matching a prefix and searching for a **LIKE** expression have different semantics and may return very different results depending on the text configuration. The specification of minimum length, maximum length and a stoplist will govern the prefix processing but does not affect **LIKE** semantics.

Note: Meaning of boolean expressions will differ between **WD** index and **TEXT** index when term dropping occurs, because the effect of dropped terms in **TEXT** index processing has no equivalent in the **WD** index.

Creating a **TEXT** Index Using Interactive SQL

Before you can perform a full text search, you must create a **TEXT** index on the columns you want to search.

A **TEXT** index stores positional information for terms in the indexed columns.

1. Connect to the database as a user with `CREATE ANY INDEX` or `CREATE ANY OBJECT` system privilege, or as the owner of the underlying table.
2. Execute a **CREATE TEXT INDEX** statement.

This example creates a **TEXT** index, `myTxtIdx`, on the `CompanyName` column of the `Customers` table in the `iqdemo` database. The `default_char` text configuration object is used.

```
CREATE TEXT INDEX myTxtIdx ON Customers
  ( CompanyName ) CONFIGURATION default_char
```

Guidelines for TEXT Index Size Estimation

Formula estimates **TEXT** index main store size.

$$\text{Number of bytes} = (15+L)*U + U*PAGESIZE * R + T$$

where:

- L = average term length for the vocabulary
- U = number of unique terms in the vocabulary
- R = number of millions of documents
- T = total number of all terms in all documents

The temporary space required in bytes for the **TEXT** index is $(M+20)*T$, where:

- M = the maximum term length for the text configuration in bytes

Note: The temporary space required is subject to compressibility of the sort data.

TEXT Index Restrictions

Text configuration objects and **TEXT** indexes have limitations by design.

- SAP Sybase IQ does not provide support for **TEXT** indexes spanning multiple columns.
- **TEXT** index manual refresh or automatic refresh options are not supported.
- `sp_iqrebuildindex` cannot be used to build **TEXT** indexes.
- You cannot create **TEXT** indexes within **BEGIN PARALLEL IQ...END PARALLEL IQ**.
- **NGRAM** term breaker is built on **TEXT** indexes, so use text configuration object settings to define whether to use an **NGRAM** or **GENERIC TEXT** index.
- **NGRAM TEXT** index search is mainly useful when words are misspelled. SAP Sybase IQ does not support searches like synonyms and antonyms.

Displaying a List of TEXT Indexes Using Interactive SQL

View a list of all the **TEXT** indexes in the database.

1. Connect to the database as a user with any of the following system privileges:
 - `CREATE ANY INDEX`

TEXT Indexes and Text Configuration Objects

- ALTER ANY INDEX
- DROP ANY INDEX
- CREATE ANY OBJECT
- ALTER ANY OBJECT
- DROP ANY OBJECT
- MANAGE ANY DBSPACE

2. Execute a **SELECT** statement.

To list all **TEXT** indexes:

```
SELECT * FROM sp_iqindex() WHERE index_type = 'TEXT';
```

To list all **TEXT** indexes, including those on catalog tables:

```
SELECT index_name, table_name, name FROM SYSIDX, SYSTEXTIDX,  
SYSTABLE, SYSUSERS  
WHERE SYSIDX.object_id=SYSTEXTIDX.index_id  
AND SYSIDX.table_id=SYSTABLE.table_id  
AND SYSTABLE.creator=SYSUSERS.uid;
```

Editing a TEXT Index Using Interactive SQL

Change the settings for the **TEXT** index, including the dbspace and **TEXT** index name.

1. Connect to the database as a user with ALTER ANY INDEX or ALTER ANY OBJECT system privilege or as the owner of the underlying table.
2. Execute an **ALTER TEXT INDEX** statement.

To rename the **TEXT** index myTxtIdx to MyTextIndex:

```
ALTER TEXT INDEX MyTxtIdx  
ON Customers  
RENAME AS MyTextIndex;
```

Modifying the TEXT Index Location Using Interactive SQL

Change the dbspace where the **TEXT** index is stored.

1. Connect to the database as a user with MANAGE ANY DBSPACE system privilege.
2. Execute an **ALTER TEXT INDEX** statement with the **MOVE TO** clause.

To move the **TEXT** index MyTextIndex to a dbspace named tispac:

```
ALTER TEXT INDEX MyTextIndex ON  
GROUPO.customers MOVE TO tispac;
```

Dropping a TEXT Index Using Interactive SQL

Drop a **TEXT** index from the database.

1. Connect to the database as a user with **DROP ANY INDEX** or **DROP ANY OBJECT** system privilege or as the owner of the underlying table.
2. Execute a **DROP TEXT INDEX** statement.

To drop the `MyTextIndex` **TEXT** index:

```
DROP TEXT INDEX MyTextIndex ON Customers;
```

TEXT Index Refresh

The only supported refresh type for **TEXT** indexes on SAP Sybase IQ tables is Immediate Refresh, which occurs when data in the underlying table changes.

Immediate-refresh **TEXT** indexes on tables support isolation level 3. They are populated at creation time and every time the data in the column is changed using an **INSERT**, **UPDATE**, or **DELETE** statement. An exclusive lock is held on the table during the initial refresh.

TEXT_DELETE_METHOD Database Option

Specifies the algorithm used during a delete in a **TEXT** index.

Allowed Values

0 – 2

0 – the delete method is selected by the cost model.

1 – forces small method for deletion. Small method is useful when the number of rows being deleted is a very small percentage of the total number of rows in the table. Small delete can randomly access the index, causing cache thrashing with large data sets.

2 – forces large method for deletion. This algorithm scans the entire index searching for rows to delete. Large method is useful when the number of rows being deleted is a high percentage of the total number of rows in the table.

Default

0

Scope

Requires the **SET ANY PUBLIC OPTION** system privilege to set this option for **PUBLIC** or for other user or role. Can be set temporary only for the **PUBLIC** role. Takes effect immediately.

Description

TEXT_DELETE_METHOD specifies the algorithm used during a delete operation in a **TEXT** index. When this option is not set or is set to 0, the delete method is selected by the cost model.

TEXT Indexes and Text Configuration Objects

The cost model considers the CPU-related costs as well as I/O-related costs in selecting the appropriate delete algorithm. The cost model takes into account:

- Rows deleted
- Index size
- Width of index data type
- Cardinality of index data
- Available temporary cache
- Machine-related I/O and CPU characteristics
- Available CPUs and threads

Example

To force the large method for deletion from a **TEXT** index:

```
SET TEMPORARY OPTION TEXT_DELETE_METHOD = 2
```

NGRAM TEXT Index

NGRAM TEXT index stores the text in the column by breaking the text into n-grams of text value N, where N is the value given by a user.

You can perform a search over an **NGRAM TEXT** index by matching the n-grams of the text value in the **CONTAINS** clause of the query against the stored n-grams in the index.

NGRAM TEXT index accommodates fuzzy searching capability over the text for both European and non-European languages.

Note: **NGRAM TEXT** index search is mainly useful when words are misspelled. SAP Sybase IQ does not support searches like synonyms and antonyms.

NGRAM term breaker is built on **TEXT** indexes, so use text configuration object settings to define whether to use an **NGRAM** or **GENERIC TEXT** index.

Text Configuration Objects

Text configuration objects control the terms that are placed in a **TEXT** index when it is built or refreshed, and how a full text query is interpreted.

When the database server creates or refreshes a **TEXT** index, it uses the settings for the text configuration object specified when the **TEXT** index was created. If a text configuration object is not specified, the database server chooses one of the default text configuration objects, based on the type of data in the columns being indexed. In an SAP Sybase IQ database, the `default_char` text configuration object is always used.

Text configuration objects specify which prefilter library and which term breaker are used to generate terms from the documents to be indexed. They specify the minimum and maximum

length of terms to be stored within the **TEXT** index, along with the list of terms that should not be included. Text configuration objects consist of these parameters:

- Document pre-filter – removes unnecessary information, such as formatting and images, from the document. The filtered document is then picked up by other modules for further processing. The document pre-filter is provided by a third-party vendor.
- Document term-breaker – breaks the incoming byte stream into terms separated by term separators or according to specified rules. The document term-breaker is provided by the server or a third-party vendor.
- Stoplist processor – specifies the list of terms to be ignored while building the **TEXT** index.

Default Text Configuration Objects

SAP Sybase IQ provides default text configuration objects.

The default text configuration object `default_char` is used with non-NCHAR data. This configuration is created the first time you create a text configuration object or **TEXT** index.

The text configuration object `default_nchar` is supported for use with NCHAR for **TEXT** indexes on **IN SYSTEM** tables; you cannot use `default_nchar` text configuration for **TEXT** indexes on tables.

The table "Default text configuration settings" shows the default settings for `default_char` and `default_nchar`, which are best suited for most character-based languages. We strongly suggest that you do not change the settings in the default text configuration objects.

Table 2. Default text configuration settings

| Setting | Installed value |
|---------------------|-----------------|
| TERM BREAKER | GENERIC |
| MINIMUM TERM LENGTH | 1 |
| MAXIMUM TERM LENGTH | 20 |
| STOPLIST | (empty) |

If you delete a default text configuration object, it is automatically re-created with default values the next time you create a **TEXT** index or text configuration object.

Creating a Text Configuration Using Interactive SQL

Create a text configuration to specify how **TEXT** indexes dependent on the text configuration process handle terms within the data.

1. Connect to the database as a user with **CREATE TEXT CONFIGURATION** or **CREATE ANY TEXT CONFIGURATION** or **CREATE ANY OBJECT** system privilege.

2. Execute a **CREATE TEXT CONFIGURATION** statement.

To create a text configuration object called `myTxtConfig` using the `default_char` text configuration object as a template:

```
CREATE TEXT CONFIGURATION myTxtConfig FROM default_char;
```

Text Configuration Object Settings

Learn about text configuration object settings, how they affect what is indexed, and how a full text search query is interpreted.

See also

- *Text Configuration Object Setting Interpretations* on page 14

Term Breaker Algorithm (TERM BREAKER)

The **TERM BREAKER** setting specifies the algorithm to use for breaking strings into terms.

SAP Sybase IQ **GENERIC** (the default) or **NGRAM** for storing terms.

Note: **NGRAM** term breakers store n-grams. An n-gram is a group of characters of length n where n is the value of **MAXIMUM TERM LENGTH**.

Regardless of the term breaker you specify, the database server records in the **TEXT** index the original positional information for terms when they are inserted into the **TEXT** index. In the case of n-grams, the positional information of the n-grams is stored, not the positional information for the original terms.

Table 3. TERM BREAKER impact

| To TEXT index | To query terms |
|---|--|
| <ul style="list-style-type: none"> • GENERIC TEXT index – when building a GENERIC TEXT index (the default), groups of alphanumeric characters appearing between non-alphanumeric characters are processed as terms by the database server. After the terms have been defined, terms that exceed the term length settings, and terms found in the stoplist, are counted but not inserted in the TEXT index. <p>Performance on GENERIC TEXT indexes can be faster than NGRAM TEXT indexes. However, you cannot perform fuzzy searches on GENERIC TEXT indexes.</p> | <ul style="list-style-type: none"> • GENERIC TEXT index – when querying a GENERIC TEXT index, terms in the query string are processed in the same manner as if they were being indexed. Matching is performed by comparing query terms to terms in the TEXT index. |

| To TEXT index | To query terms |
|--|--|
| <ul style="list-style-type: none"> • NGRAM TEXT index – when building an NGRAM TEXT index, the database server treats as a term any group of alphanumeric characters between non-alphanumeric characters. Once the terms are defined, the database server breaks the terms into n-grams. In doing so, terms shorter than n, and n-grams that are in the stoplist, are discarded. <p>For example, for an NGRAM TEXT index with MAXIMUM TERM LENGTH 3, the string 'my red table' is represented in the TEXT index as these n-grams: red tab abl ble.</p> | <ul style="list-style-type: none"> • NGRAM TEXT index – when querying an NGRAM TEXT index, terms in the query string are processed in the same manner as if they were being indexed. Matching is performed by comparing n-grams from the query terms to n-grams from the indexed terms. |

Minimum Term Length Setting (MINIMUM TERM LENGTH)

The **MINIMUM TERM LENGTH** setting specifies the minimum length, in characters, for terms inserted in the index or searched for in a full text query.

MINIMUM TERM LENGTH is not relevant for **NGRAM TEXT** indexes.

MINIMUM TERM LENGTH has special implications on prefix searching. The value of **MINIMUM TERM LENGTH** must be greater than 0. If you set it higher than **MAXIMUM TERM LENGTH**, then **MAXIMUM TERM LENGTH** is automatically adjusted to be equal to **MINIMUM TERM LENGTH**.

The default for **MINIMUM TERM LENGTH** is taken from the setting in the default text configuration object, which is typically 1.

Table 4. MINIMUM TERM LENGTH impact

| To TEXT index | To query terms |
|--|--|
| <ul style="list-style-type: none"> • GENERIC TEXT index – for GENERIC TEXT indexes, the TEXT index will not contain words shorter than MINIMUM TERM LENGTH. | <ul style="list-style-type: none"> • GENERIC TEXT index – when querying a GENERIC TEXT index, query terms shorter than MINIMUM TERM LENGTH are ignored because they cannot exist in the TEXT index. |
| <ul style="list-style-type: none"> • NGRAM TEXT index – for NGRAM TEXT indexes, this setting is ignored. | <ul style="list-style-type: none"> • NGRAM TEXT index – the MINIMUM TERM LENGTH setting has no impact on full text queries on NGRAM TEXT indexes. |

Maximum Term Length Setting (MAXIMUM TERM LENGTH)

The **MAXIMUM TERM LENGTH** setting specifies the maximum length, in characters, for terms inserted in the index or searched for in a full text query.

The **MAXIMUM TERM LENGTH** setting is used differently, depending on the term breaker algorithm. The value of **MAXIMUM TERM LENGTH** must be less than or equal to 60. If you set **MAXIMUM TERM LENGTH** lower than the **MINIMUM TERM LENGTH**, then **MINIMUM TERM LENGTH** is automatically adjusted to be equal to **MAXIMUM TERM LENGTH**.

The default for this setting is taken from the setting in the default text configuration object, which is typically 20.

Table 5. MAXIMUM TERM LENGTH impact

| To TEXT index | To query terms |
|---|--|
| <ul style="list-style-type: none"> GENERIC TEXT index – for GENERIC TEXT indexes, MAXIMUM TERM LENGTH specifies the maximum length, in characters, for terms inserted in the TEXT index. | <ul style="list-style-type: none"> GENERIC TEXT index – for GENERIC TEXT indexes, query terms longer than MAXIMUM TERM LENGTH are ignored because they cannot exist in the TEXT index. |
| <ul style="list-style-type: none"> NGRAM TEXT index – for NGRAM TEXT indexes, MAXIMUM TERM LENGTH determines the length of the n-grams that terms are broken into. An appropriate choice of length for MAXIMUM TERM LENGTH depends on the language. Typical values are 4 or 5 characters for English, and 2 or 3 characters for Chinese. | <ul style="list-style-type: none"> NGRAM TEXT index – for NGRAM TEXT indexes, query terms are broken into n-grams of length n, where n is the same as MAXIMUM TERM LENGTH. The database server uses the n-grams to search the TEXT index. Terms shorter than MAXIMUM TERM LENGTH are ignored because they do not match the n-grams in the TEXT index. |

Stoplist Setting (STOPLIST)

The stoplist setting specifies terms that are not indexed.

The default for the stoplist setting is taken from the setting in the default text configuration object, which typically has an empty stoplist.

Table 6. STOPLIST impact

| To TEXT index | To query terms |
|--|---|
| <ul style="list-style-type: none"> GENERIC TEXT index – for GENERIC TEXT indexes, terms that are in the stoplist are not inserted into the TEXT index. | <ul style="list-style-type: none"> GENERIC TEXT index – for GENERIC TEXT indexes, query terms that are in the stoplist are ignored because they cannot exist in the TEXT index. |

| To TEXT index | To query terms |
|--|---|
| <ul style="list-style-type: none"> • NGRAM TEXT index – for NGRAM TEXT indexes, the TEXT index does not contain the n-grams formed from the terms in the stoplist. | <ul style="list-style-type: none"> • NGRAM TEXT index – terms in the stoplist are broken into n-grams and the n-grams are used for the stoplist. Likewise, query terms are broken into n-grams and any that match n-grams in the stoplist are dropped because they cannot exist in the TEXT index. |

Consider carefully whether to put terms in to your stoplist. In particular, do not include words that have non-alphanumeric characters in them such as apostrophes or dashes. These characters act as term breakers. For example, the word you'll (which must be specified as 'you'll') is broken into you and ll and stored in the stoplist as these two terms. Subsequent full text searches for 'you' or 'they'll' are negatively impacted.

Stoplists in **NGRAM TEXT** indexes can cause unexpected results because the stoplist that is stored is actually in n-gram form, not the actual stoplist terms you specified. For example, in an **NGRAM TEXT** index where **MAXIMUM TERM LENGTH** is 3, if you specify **STOPLIST** 'there', these n-grams are stored as the stoplist: the her ere. This impacts the ability to query for any terms that contain the n-grams the, her, and ere.

Displaying a List of Text Configurations Using Interactive SQL

View a list of all the text configurations in the database.

1. Connect to the database as a user with **CREATE TEXT CONFIGURATION** or **CREATE ANY TEXT CONFIGURATION** or **CREATE ANY OBJECT** system privilege.
2. Execute a **SELECT** statement.

To list all text configuration objects:

```
SELECT * FROM SYSTEXTCONFIG;
```

Altering a Text Configuration Using Interactive SQL

Change the settings of the text configuration object, including the dbspace and permitted term lengths range.

You can alter only text configuration objects that are not being used by a **TEXT** index.

1. Connect to the database as a user with **ALTER ANY TEXT CONFIGURATION** or **ALTER ANY OBJECT** system privilege, or as the owner of the text configuration object.
2. Execute an **ALTER TEXT CONFIGURATION** statement.

To alter the minimum term length for the myTxtConfig text configuration object:

```
ALTER TEXT CONFIGURATION myTxtConfig
MINIMUM TERM LENGTH 2;
```

Modifying the Stoplist Using Interactive SQL

Modify the stoplist, which contains a list of terms to ignore when building a **TEXT** index with this text configuration.

You can alter only text configuration objects that are not being used by a **TEXT** index.

1. Connect to the database as a user with **ALTER ANY TEXT CONFIGURATION** or **ALTER ANY OBJECT** system privilege, or as the owner of the text configuration object.
2. Execute an **ALTER TEXT CONFIGURATION** statement with the **STOPLIST** clause.

To add a stoplist to the myTxtConfig configuration object:

```
ALTER TEXT CONFIGURATION myTxtConfig  
  STOPLIST 'because about therefore only';
```

Dropping a Text Configuration Using Interactive SQL

Remove an unnecessary text configuration from the database.

Only text configurations that are not being used by a **TEXT** index can be dropped.

1. Connect to the database as a user with **DROP ANY TEXT CONFIGURATION** or **DROP ANY OBJECT** system privilege.
2. Execute a **DROP TEXT CONFIGURATION** statement.

To drop the text configuration object myTxtConfig:

```
DROP TEXT CONFIGURATION myTxtConfig;
```

Text Configuration Object Examples

Review the samples to understand how text configuration settings impact the **TEXT** index, and how the index is interpreted.

Text Configuration Object Setting Interpretations

Examples that show the settings for different text configuration objects, how the settings impact what is indexed, and how a full text query string is interpreted.

All the examples use the string 'I'm not sure I understand'.

Table 7. Text configuration setting interpretations

| Configuration settings | Terms that are indexed | Query interpretation |
|---|---------------------------|---|
| TERM BREAKER: GENERAL MINIMUM TERM LENGTH: 1 MAXIMUM TERM LENGTH: 20 STOPLIST: "" | I m not sure I understand | '("I m" AND not sure) AND I AND understand' |
| TERM BREAKER: GENERAL MINIMUM TERM LENGTH: 2 MAXIMUM TERM LENGTH: 20 STOPLIST: 'not and' | sure understand | 'understand' |
| TERM BREAKER: GENERAL MINIMUM TERM LENGTH: 1 MAXIMUM TERM LENGTH: 20 STOPLIST: 'not and' | I m sure I understand | '"I m" AND sure AND I AND understand' |

Text Configuration Object CONTAINS Query String Interpretations

Examples of how the settings of the text configuration object strings are interpreted for CONTAINS queries.

The parenthetical numbers in the Interpreted string column in the table "CONTAINS string interpretations" reflect the position information stored for each term. The numbers are for illustration purposes in the documentation. The actual stored terms do not include the parenthetical numbers.

Note: The maximum number of positions for a text document is 4294967295.

The interpretations in this table are only for **CONTAINS** queries. When data is parsed, **AND**, **NOT**, and **NEAR** are considered regular tokens; symbols like *, !, and others are dropped as they are not alphanumeric.

Table 8. CONTAINS string interpretations

| Configuration settings | String | Interpreted string |
|--|---|--|
| TERM BREAKER: GENERAL MINIMUM TERM LENGTH: 3 MAXIMUM TERM LENGTH: 20 | 'w*' | '"w*(1) "' |
| | 'we*' | '"we*(1) "' |
| | 'wea*' | '"wea*(1) "' |
| | 'we* -the' | '"we*(1) " -"the(1) "' |
| | 'for* wonderl*' | '"for*(1) " "wonderl*(1) "' |
| | 'wonderlandwonderlandwonderland*' | ' ' |
| | '"tr* weather"' | '"weather(1) "' |
| | '"tr* the weather"' | '"the(1) weather(2) "' |
| | '"wonderlandwonderlandwonderland* wonderland"' | '"wonderland(1) "' |
| | '"wonderlandwonderlandwonderland* weather"' | '"weather(1) "' |
| | '"the_wonderlandwonderlandwonderland* weather"' | '"the(1) weather(3) "' |
| | 'the_wonderlandwonderlandwonderland* weather' | '"the(1) " & "weather(1) "' |
| | '"light_a* the end" & tunnel' | '"light(1) the(3) end(4) " & "tunnel(1) "' |
| | light_b* the end" & tunnel' | '"light(1) the(3) end(4) " & "tunnel(1) "' |
| | '"light_at_b* end"' | '"light(1) end(4) "' |
| 'and-te*' | '"and(1) te*(2) "' | |

| Configuration settings | String | Interpreted string |
|------------------------|---------------------------|--|
| | 'a_long_and_t* & journey' | '"long(2) and(3)t*(4)" & "journey(1)'' |

MAX_PREFIX_PER_CONTAINS_PHRASE Database Option

Specifies the number of prefix terms allowed in a text search expression.

Allowed Values

0 – 300

0 – no limit for prefix terms in search phrase

300 – upper limit (this is the overall limit for total number of terms allowed in a phrase)

Default

1

Scope

Requires the SET ANY PUBLIC OPTION system privilege to set this option. Can be set temporary, for an individual connection, or for the PUBLIC role. Takes effect immediately.

Description

MAX_PREFIX_PER_CONTAINS_PHRASE specifies the threshold used to disallow more than one prefix in an expression for a text search.

When this option is set to 0, any number is allowed. SAP Sybase IQ detects and reports an error, if the query has any **CONTAINS** expressions with a phrase having more prefix terms than specified by this option.

Examples

With the default MAX_PREFIX_PER_CONTAINS_PHRASE setting:

```
SET MAX_PREFIX_PER_CONTAINS_PHRASE = 1
```

this **CONTAINS** clause is valid:

```
SELECT ch1 FROM tabl
WHERE CONTAINS(ch1, '"concord bed* in mass"')
```

With the default MAX_PREFIX_PER_CONTAINS_PHRASE setting of 1, this **CONTAINS** clause returns a syntax error:

```
SELECT ch1 FROM tabl
WHERE CONTAINS (ch1, '"con* bed* in mass"')
```

When MAX_PREFIX_PER_CONTAINS_PHRASE is set equal to 0 (no limit) or 2, this **CONTAINS** clause is valid.

External Libraries

Learn about using external libraries to supply prefiltering and term breaking for documents.

Pre-Filter and Term-Breaker External Libraries

SAP Sybase IQ can use external pre-filter and term-breaker libraries written in C or C++ to prefilter and tokenize the documents during index creation or query processing. These libraries can be dynamically loaded into the process space of the database server.

Note: External pre-filter and term-breaker libraries must be provided by an SAP Sybase-certified partner. For information on certified vendor solutions, see the *Partner Certification Reports web site* and then filter the certification reports to show SAP Sybase IQ certifications.

The external dynamically loadable pre-filter and term-breaker libraries are specified in the text configuration, and need to be loaded by the database server. Each library contains an exported symbol that implements the external function specified in the text configuration object. This function returns a set of function descriptors that are used by the caller to perform the necessary tasks.

The external pre-filter and term-breaker libraries are loaded by the database server with the first **CREATE TEXT INDEX** request, when a query for a given column is received that requires the library to be loaded, or when the **TEXT** index needs to be updated.

The libraries are not loaded when an **ALTER TEXT CONFIGURATION** call is made, nor are they automatically unloaded when a **DROP TEXT CONFIGURATION** call is made. The external pre-filter and term-breaker libraries are not loaded if the server is started with the startup option to disallow the loading of external libraries.

Because these external C/C++ libraries involve the loading of non-server library code into the process space of the server, there are potential risks to data integrity, data security, and server robustness from poorly or maliciously written functions. To manage these risks, each server can explicitly enable or disable this functionality.

The `ISYSTEXTCONFIG` system view stores information about the external libraries associated with a text configuration object.

See also

- *Enable and Disable External Libraries on Startup* on page 20

External Library Restrictions

Text configuration objects and **TEXT** indexes using external libraries have limitations by design.

- For **TEXT** indexes on binary columns, you must use external libraries provided by external vendors for document conversion. SAP Sybase IQ does not implicitly convert documents stored in binary columns.
- N-gram based text searches are not supported if an external term-breaker is used to tokenize the document.

External Libraries on Multiplex Servers

All multiplex servers must have access to the pre-filter and term-breaker external libraries.

Users must ensure that each external pre-filter and term-breaker library is copied to the machine hosting a multiplex server and placed in a location where the server is able to load the library.

Each multiplex server works independently of other servers when calling the external pre-filter and term-breaker. Each process space can have the external libraries loaded and perform its own executions. It is assumed that the implementation of the pre-filter and term-breaker functions is the same on each server and will return the same result.

Unloading an external library from one server process space does not unload the library from other server process spaces.

Enable and Disable External Libraries on Startup

SAP Sybase IQ provides the **-sf** startup switch to enable or disable loading of external third-party libraries.

You can specify this switch in either the server startup command line or the server configuration file.

To enable the loading of external third-party libraries:

```
-sf -external_library_full_text
```

To disable the loading of external third-party libraries:

```
-sf external_library_full_text
```

To view a list of the libraries currently loaded in the server, use the **sa_list_external_library** stored procedure.

Unload External Libraries

Use the system procedure **dbo.sa_external_library_unload** to unload an external library when the library is not in use.

dbo.sa_external_library_unload takes one optional parameter, a `LONG VARCHAR`. The parameter specifies the name of the library to unload. If you do not specify a parameter, all external libraries that are not in use are unloaded.

Unload an external function library:

```
call sa_external_library_unload('library.dll')
```

External Libraries

Unstructured Data Queries

Learn about querying large object data, including the full text search capability that handles unstructured and semistructured data.

Full Text Search

Full text search uses **TEXT** indexes to quickly find all instances of a term (word) in a database without having to scan table rows.

TEXT indexes store positional information for terms in the indexed columns. Using a **TEXT** index to find rows that contain a term is faster than scanning every row in the table.

Full text search uses the **CONTAINS** search condition, which differs from searching using predicates such as **LIKE**, **REGEXP**, and **SIMILAR TO**, because the matching is term-based and not pattern-based.

String comparisons in full text search use all the normal collation settings for the database. For example, you configure the database to be case-insensitive, then full text searches are also case-insensitive.

See also

- *CONTAINS Conditions* on page 47

Types of Full Text Searches

Using full text search, you can search for terms, prefixes, or phrases (sequences of terms). You can also combine multiple terms, phrases, or prefixes into Boolean expressions, or use proximity searches to require that expressions appear near to each other.

Perform a full text search using a **CONTAINS** clause in either a **WHERE** clause, or a **FROM** clause of a **SELECT** statement.

Term and phrase search

When performing a full text search for a list of terms, the order of terms is not important unless they are within a phrase. If you put the terms within a phrase, the database server looks for those terms in exactly the same order, and same relative positions, in which you specified them.

When performing a term or phrase search, if terms are dropped from the query because they exceed term length settings or because they are in the stoplist, you can get back a different number of rows than you expect. This is because removing the terms from the query is equivalent to changing your search criteria. For example, if you search for the phrase `'"grown cotton"'` and grown is in the stoplist, you get every indexed row containing cotton.

Unstructured Data Queries

You can search for the terms that are considered keywords of the CONTAINS clause grammar, as long as they are within phrases.

Term searching

In the sample database, a text index called MarketingTextIndex has been built on the Description column of the MarketingInformation table. The following statement queries the MarketingInformation.Description column and returns the rows where the value in the Description column contains the term cotton.

```
SELECT ID, Description
FROM MarketingInformation
WHERE CONTAINS ( Description, 'cotton' );
```

| ID | Description |
|-----|---|
| 906 | <html><head><meta http-equiv=Content-Type content="text/html; charset=windows-1252"><title>Visor</title></head><body lang=EN-US><p>Lightweight 100% organically grown cotton construction. Shields against sun and precipitation.cotton Metallic ions in the fibers inhibit bacterial growth, and help neutralize odor.</p></body></html> |

| ID | Description |
|-----|--|
| 908 | <pre><html><head><meta http-equiv=Content-Type content="text/html; charset=windows-1252"><title>Sweatshirt</title></head><body lang=EN-US><p>Lightweight 100% organically grown cotton hooded sweatshirt with taped neck seams. Comes pre-washed for softness and to lessen shrinkage.</p></body></html></pre> |
| 909 | <pre><html><head><meta http-equiv=Content-Type content="text/html; charset=windows-1252"><title>Sweatshirt</title></head><body lang=EN-US><p>Top-notch construction includes durable topstitched seams for strength with low-bulk, resilient rib-knit collar, cuffs and bottom. An 80% cotton/20% polyester blend makes it easy to keep them clean.</p></body></html></pre> |

Unstructured Data Queries

| ID | Description |
|-----|---|
| 910 | <pre><html><head><meta http-equiv=Content-Type content="text/html; charset=windows-1252"><title>Shorts</title></head><body lang=EN-US><p>These quick-drying cotton shorts provide all day comfort on or off the trails. Now with a more comfortable and stretchy fabric and an adjustable drawstring waist.</p></body></html></pre> |

The following example queries the MarketingInformation table and returns a single value for each row indicating whether the value in the Description column contains the term cotton.

```
SELECT ID, IF CONTAINS ( Description, 'cotton' )
      THEN 1
      ELSE 0
      ENDIF AS Results
FROM MarketingInformation;
```

| ID | Results |
|-----|---------|
| 901 | 0 |
| 902 | 0 |
| 903 | 0 |
| 904 | 0 |
| 905 | 0 |
| 906 | 1 |
| 907 | 0 |
| 908 | 1 |
| 909 | 1 |

| ID | Results |
|-----|---------|
| 910 | 1 |

The next example queries the MarketingInformation table for items that have the term cotton the Description column, and shows the score for each match.

```
SELECT ID, ct.score, Description
  FROM MarketingInformation CONTAINS
 ( MarketingInformation.Description, 'cotton' ) as ct
 ORDER BY ct.score DESC;
```

| ID | score | Description |
|-----|--------------------|---|
| 908 | 0.9461597363521859 | <html><head><meta http-equiv=Content-Type content="text/html; charset=windows-1252"><title>Sweatshirt</title></head><body lang=EN-US><p>Lightweight 100% organically grown cotton hooded sweatshirt with taped neck seams. Comes pre-washed for softness and to lessen shrinkage.</p></body></html> |

Unstructured Data Queries

| ID | score | Description |
|-----|--------------------|--|
| 910 | 0.9244136988525732 | <pre><html><head><meta http-equiv=Con- tent-Type con- tent="text/html; charset=win- dows-1252"><ti- tle>Shorts</ti- tle></head><body lang=EN- US><p>These quick-drying cotton shorts provide all day comfort on or off the trails. Now with a more com- fortable and stret- chy fabric and an adjustable draw- string waist.</ span></p></body></ html></pre> |

| ID | score | Description |
|-----|--------------------|--|
| 906 | 0.9134171046194403 | <pre><html><head><meta http-equiv=Con- tent-Type con- tent="text/html; charset=win- dows-1252"><ti- tle>Visor</ti- tle></head><body lang=EN- US><p>Light- weight 100% organi- cally grown cotton construction. Shields against sun and precipitation. Metallic ions in the fibers inhibit bacterial growth, and help neutralize odor.</p></ body></html></pre> |

| ID | score | Description |
|-----|--------------------|--|
| 909 | 0.8856420222728282 | <pre><html><head><meta http-equiv=Con- tent-Type con- tent="text/html; charset=win- dows-1252"><ti- tle>Sweatshirt</ title></head><body lang=EN- US><p>Top- notch construction includes durable topstitched seams for strength with low-bulk, resilient rib-knit collar, cuffs and bottom. An 80% cotton/20% polyester blend makes it easy to keep them clean.</ span></p></body></ html></pre> |

Phrase searching

When performing a full text search for a phrase, you enclose the phrase in double quotes. A column matches if it contains the terms in the specified order and relative positions.

You cannot specify CONTAINS keywords, such as AND or FUZZY, as terms to search for unless you place them inside a phrase (single term phrases are allowed). For example, the statement below is acceptable even though NOT is a CONTAINS keyword.

```
SELECT * FROM table-name CONTAINS ( Remarks, '"NOT"' );
```

With the exception of asterisk, special characters are not interpreted as special characters when they are in a phrase.

Phrases cannot be used as arguments for proximity searches.

The following statement queries MarketingInformation.Description for the phrase "grown cotton", and shows the score for each match:

```
SELECT ID, ct.score, Description
  FROM MarketingInformation CONTAINS
 ( MarketingInformation.Description, '"grown cotton"' ) as ct
 ORDER BY ct.score DESC;
```

| ID | score | Description |
|-----|--------------------|---|
| 908 | 1.6619019465461564 | <html><head><meta http-equiv=Content-Type content="text/html; charset=windows-1252"><title>Sweatshirt</title></head><body lang=EN-US><p>Lightweight 100% organically grown cotton hooded sweatshirt with taped neck seams. Comes pre-washed for softness and to lessen shrinkage.</p></body></html> |

| ID | score | Description |
|-----|--------------------|--|
| 906 | 1.6043904700786786 | <pre><html><head><meta http-equiv=Con- tent-Type con- tent="text/html; charset=win- dows-1252"><ti- tle>Visor</ti- tle></head><body lang=EN- US><p>Light- weight 100% organi- cally grown cotton construction. Shields against sun and precipitation. Metallic ions in the fibers inhibit bacterial growth, and help neutralize odor.</p></ body></html></pre> |

Prefix search

The full text search feature allows you to search for the beginning portion of a term. This is called a **prefix search**. To perform a prefix search, you specify the prefix you want to search for, followed by an asterisk. This is called a **prefix term**.

Keywords for the CONTAINS clause cannot be used for prefix searching unless they are in a phrase.

You also can specify multiple prefix terms in a query string, including within phrases (for example, '"shi* fab"').

The following example queries the MarketingInformation table for items that start with the prefix shi:

```
SELECT ID, ct.score, Description
FROM MarketingInformation CONTAINS
( MarketingInformation.Description, 'shi*' ) AS ct
ORDER BY ct.score DESC;
```

| ID | score | Description |
|-----|-------------------|--|
| 906 | 2.295363835537917 | <pre><html><head><meta http-equiv=Con- tent-Type con- tent="text/html; charset=win- dows-1252"><ti- tle>Visor</ti- tle></head><body lang=EN- US><p>Light- weight 100% organi- cally grown cotton construction. Shields against sun and precipitation. Metallic ions in the fibers inhibit bacterial growth, and help neutralize odor.</p></ body></html></pre> |

Unstructured Data Queries

| ID | score | Description |
|-----|--------------------|---|
| 901 | 1.6883275743936228 | <pre><html><head><meta http-equiv=Con- tent-Type con- tent="text/html; charset=win- dows-1252"><ti- tle>Tee Shirt</ti- tle></head><body lang=EN- US><p>We've im- proved the design of this perennial favorite. A sleek and technical shirt built for the trail, track, or sidewalk. UPF rat- ing of 50+.</ span></p></body></ html></pre> |

| ID | score | Description |
|-----|--------------------|---|
| 903 | 1.6336529491832605 | <pre><html><head><meta http-equiv=Con- tent-Type con- tent="text/html; charset=win- dows-1252"><ti- tle>Tee Shirt</ti- tle></head><body lang=EN- US><p>A sporty, casual shirt made of recycled water bottles. It will serve you equally well on trails or around town. The fabric has a wick- ing finish to pull perspiration away from your skin.</ span></p></body></ html></pre> |

| ID | score | Description |
|-----|--------------------|---|
| 902 | 1.6181703448678983 | <pre><html><head><meta http-equiv=Con- tent-Type con- tent="text/html; charset=win- dows-1252"><ti- tle>Tee Shirt</ti- tle></head><body lang=EN- US><p>This sim- ple, sleek, and lightweight techni- cal shirt is de- signed for high-in- tensity workouts in hot and humid weather. The recy- cled polyester fab- ric is gentle on the earth and soft against your skin.</p></ body></html></pre> |

ID 906 has the highest score because the term shield occurs less frequently than shirt in the text index.

Prefix searches on GENERIC text indexes

On GENERIC text indexes, the behavior for prefix searches is as follows:

- If a prefix term is longer than the MAXIMUM TERM LENGTH, it is dropped from the query string since there can be no terms in the text index that exceed the MAXIMUM TERM LENGTH. So, on a text index with MAXIMUM TERM LENGTH 3, searching for 'red appl*' is equivalent to searching for 'red'.
- If a prefix term is shorter than MINIMUM TERM LENGTH, and is not part of a phrase search, the prefix search proceeds normally. So, on a GENERIC text index where MINIMUM TERM LENGTH is 5, searching for 'macintosh a*' returns indexed rows that contain macintosh and any terms of length 5 or greater that start with a.
- If a prefix term is shorter than MINIMUM TERM LENGTH, but is part of a phrase search, the prefix term is dropped from the query. So, on a GENERIC text index where

MINIMUM TERM LENGTH is 5, searching for `'"macintosh appl* turnover"'` is equivalent to searching for macintosh followed by any term followed by turnover. A row containing `"macintosh turnover"` is not found; there must be a term between macintosh and turnover.

Prefix searches on NGRAM text indexes

On NGRAM text indexes, prefix searching can return unexpected results since an NGRAM text index contains only n-grams, and contains no information about the beginning of terms. Query terms are also broken into n-grams, and searching is performed using the n-grams not the query terms. Because of this, the following behaviors should be noted:

- If a prefix term is shorter than the n-gram length (MAXIMUM TERM LENGTH), the query returns all indexed rows that contain n-grams starting with the prefix term. For example, on a 3-gram text index, searching for `'ea*'` returns all indexed rows containing n-grams starting with ea. So, if the terms weather and fear were indexed, the rows would be considered matches since their n-grams include eat and ear, respectively.
- If a prefix term is longer than n-gram length, and is not part of a phrase, and not an argument in a proximity search, the prefix term is converted to an n-grammed phrase and the asterisk is dropped. For example, on a 3-gram text index, searching for `'purple blac*'` is equivalent to searching for `"pur urp rpl ple" AND "bla lac"`.
- For phrases, the following behavior also takes place:
 - If the prefix term is the only term in the phrase, it is converted to an n-grammed phrase and the asterisk is dropped. For example, on a 3-gram text index, searching for `'"purpl*"'` is equivalent to searching for `"pur urp rpl"`.
 - If the prefix term is in the last position of the phrase, the asterisk is dropped and the terms are converted to a phrase of n-grams. For example, on a 3-gram text index, searching for `'"purple blac*"'` is equivalent to searching for `"pur urp rpl ple bla lac"`.
 - If the prefix term is not in the last position of the phrase, the phrase is broken up into phrases that are ANDed together. For example, on a 3-gram text index, searching for `'"purp* blac*"'` is equivalent to searching for `"pur urp" AND "bla lac"`.
- If a prefix term is an argument in a proximity search, the proximity search is converted to an AND. For example, on a 3-gram text index, searching for `'red NEAR[1] appl*'` is equivalent to searching for `'red AND "app ppl"'`.

Proximity search

The full text search feature allows you to search for terms that are near each other in a single column. This is called a **proximity search**. To perform a proximity search, you specify two terms with either the keyword NEAR or BEFORE between them, or the tilde (~).

You can use proximity-expression to search for terms that are near each other. For example, `b NEAR[2, 5] c` searches for instances of `b` and `c` that are at most 5 and at least 2 terms away

from each other. The order of terms is not significant; *b* NEAR *c* is equivalent to *c* NEAR *b*. If NEAR is specified without distance, a default of ten terms is applied. You can specify a tilde (~) instead of NEAR. This is equivalent to specifying NEAR without a distance so a default of 10 terms is applied. NEAR expressions cannot be chained together (for example, *a* NEAR [1] *b* NEAR [1] *c*).

BEFORE is like NEAR except that the order of terms is significant. *b* BEFORE *c* is not equivalent to *c* BEFORE *b*; in the former, the term *b* must precede *c*, while in the latter, the term *b* must follow *c*. BEFORE accepts both minimum and maximum distances (like NEAR). The default minimum distance is 1. The minimum distance, if given, must be less than or equal to the maximum distance; otherwise, an error is returned.

If you do not specify a distance, the database server uses 10 as the default distance.

You can also specify a tilde (~) instead of the NEAR keyword. For example, ' *term1* ~ *term2* '. However, you cannot specify a distance when using the tilde form; the default of ten terms is applied.

You cannot specify a phrase as an argument in proximity searches.

In a proximity search using an NGRAM text index, if you specify a prefix term as an argument, the proximity search is converted to an AND expression. For example, on a 3-gram text index, searching for 'red NEAR[1] appl*' is equivalent to searching for 'red AND "app pp1" '. Since this is no longer a proximity search, the search is no longer restricted to a single column in the case where multiple columns are specified in the CONTAINS clause.

Examples

Suppose you want to search MarketingInformation.Description for the term fabric within 10 terms of the term skin. You can execute the following statement.

```
SELECT ID, "contains".score, Description
      FROM MarketingInformation CONTAINS ( Description, 'fabric ~
skin' );
```

| ID | score | Description |
|-----|--------------------|---|
| 902 | 1.5572371866083279 | <pre><html><head><meta http-equiv=Con- tent-Type con- tent="text/html; charset=win- dows-1252"><ti- tle>Tee Shirt</ti- tle></head><body lang=EN- US><p>This sim- ple, sleek, and lightweight techni- cal shirt is de- signed for high-in- tensity workouts in hot and humid weather. The recy- cled polyester fab- ric is gentle on the earth and soft against your skin.</p></ body></html></pre> |

Since the default distance is 10 terms, you did not need to specify a distance. By extending the distance by one term, however, another row is returned:

```
SELECT ID, "contains".score, Description
  FROM MarketingInformation CONTAINS ( Description, 'fabric NEAR[11]
skin' );
```

Unstructured Data Queries

| ID | score | Description |
|-----|--------------------|---|
| 903 | 1.5787803210404958 | <pre> <html><head><meta http-equiv=Con- tent-Type con- tent="text/html; charset=win- dows-1252"><ti- tle>Tee Shirt</ti- tle></head><body lang=EN- US><p>A sporty, casual shirt made of recycled water bottles. It will serve you equally well on trails or around town. The fabric has a wick- ing finish to pull perspiration away from your skin.</ span></p></body></ html> </pre> |

| ID | score | Description |
|-----|-------------------|---|
| 902 | 2.163125855043747 | <pre><html><head><meta http-equiv=Con- tent-Type con- tent="text/html; charset=win- dows-1252"><ti- tle>Tee Shirt</ti- tle></head><body lang=EN- US><p>This sim- ple, sleek, and lightweight techni- cal shirt is de- signed for high-in- tensity workouts in hot and humid weather. The recy- cled polyester fab- ric is gentle on the earth and soft against your skin.</p></ body></html></pre> |

The score for ID 903 is higher because the terms are closer together.

Boolean search

You can specify multiple terms separated by Boolean operators when performing full text searches. SAP Sybase IQ supports the following Boolean operators when performing a full text search: AND, OR, and AND NOT.

Using the AND operator in full text searches

The AND operator matches a row if it contains both of the terms specified on either side of the AND. You can also use an ampersand (&) for the AND operator. If terms are specified without an operator between them, AND is implied.

The order in which the terms are listed is not important.

Unstructured Data Queries

For example, each of the following statements finds rows in MarketingInformation.Description that contain the term `fabric` and a term that begins with `ski`:

```
SELECT *
  FROM MarketingInformation
 WHERE CONTAINS ( MarketingInformation.Description, 'ski* AND
fabric' );
SELECT *
  FROM MarketingInformation
 WHERE CONTAINS ( MarketingInformation.Description, 'fabric &
ski*' );
SELECT *
  FROM MarketingInformation
 WHERE CONTAINS ( MarketingInformation.Description, 'ski*
fabric' );
```

Using the OR operator in full text searches

The OR operator matches a row if it contains at least one of the specified search terms on either side of the OR. You can also use a vertical bar (`|`) for the OR operator; the two are equivalent.

The order in which the terms are listed is not important.

For example, either statement below returns rows in the MarketingInformation.Description that contain either the term `fabric` or a term that starts with `ski`:

```
SELECT *
  FROM MarketingInformation
 WHERE CONTAINS ( MarketingInformation.Description, 'ski* OR
fabric' );
SELECT *
  FROM MarketingInformation
 WHERE CONTAINS ( MarketingInformation.Description, 'fabric |
ski*' );
```

Using the AND NOT operator in full text searches

The AND NOT operator finds results that match the left argument and do not match the right argument. You can also use a hyphen (`-`) for the AND NOT operator; the two are equivalent.

For example, the following statements are equivalent and return rows that contain the term `fabric`, but do not contain any terms that begin with `ski`.

```
SELECT *
  FROM MarketingInformation
 WHERE CONTAINS ( MarketingInformation.Description, 'fabric AND NOT
ski*' );
SELECT *
  FROM MarketingInformation
 WHERE CONTAINS ( MarketingInformation.Description, 'fabric -
ski*' );
SELECT *
  FROM MarketingInformation
 WHERE CONTAINS ( MarketingInformation.Description, 'fabric & -
ski*' );
```


Combining different boolean operators

The boolean operators can be combined in a query string. For example, the following statements are equivalent and search the MarketingInformation.Description column for items that contain fabric and skin, but not cotton:

```
SELECT *
  FROM MarketingInformation
  WHERE CONTAINS ( MarketingInformation.Description, 'skin fabric -
cotton' );
SELECT *
  FROM MarketingInformation
  WHERE CONTAINS ( MarketingInformation.Description, 'fabric -cotton
AND skin' );
```

The following statements are equivalent and search the MarketingInformation.Description column for items that contain fabric or both cotton and skin:

```
SELECT *
  FROM MarketingInformation
  WHERE CONTAINS ( MarketingInformation.Description, 'fabric |
cotton AND skin' );
SELECT *
  FROM MarketingInformation
  WHERE CONTAINS ( MarketingInformation.Description, 'cotton skin OR
fabric' );
```

Grouping terms and phrases

Terms and expressions can be grouped with parentheses. For example, the following statement searches the MarketingInformation.Description column for items that contain cotton or fabric, and that have terms that start with ski.

```
SELECT ID, Description FROM MarketingInformation
  WHERE CONTAINS( MarketingInformation.Description, '( cotton OR
fabric ) AND shi*' );
```

| ID | Description |
|-----|---|
| 902 | <pre><html><head><meta http-equiv=Content-Type content="text/html; charset=windows-1252"><title>Tee Shirt</title></head><body lang=EN-US><p>This simple, sleek, and lightweight technical shirt is designed for high-intensity workouts in hot and humid weather. The recycled polyester fabric is gentle on the earth and soft against your skin.</p></body></html></pre> |
| 903 | <pre><html><head><meta http-equiv=Content-Type content="text/html; charset=windows-1252"><title>Tee Shirt</title></head><body lang=EN-US><p>A sporty, casual shirt made of recycled water bottles. It will serve you equally well on trails or around town. The fabric has a wicking finish to pull perspiration away from your skin.</p></body></html></pre> |

| ID | Description |
|-----|--|
| 906 | <pre><html><head><meta http- equiv=Content-Type con- tent="text/html; charset=win- dows-1252"><title>Visor</ti- tle></head><body lang=EN- US><p>Lightweight 100% organically grown cotton con- struction. Shields against sun and precipitation. Metal- lic ions in the fibers inhib- it bacterial growth, and help neutralize odor.</p></ body></html></pre> |

Searching across multiple columns

You can perform a full text search across multiple columns in a single query, as long as the columns are part of the same text index.

```
SELECT *
  FROM t
 WHERE CONTAINS ( t.c1, t.c2, 'term1|term2' );
```

```
SELECT *
  FROM t
 WHERE CONTAINS( t.c1, 'term1' )
    OR CONTAINS( t.c2, 'term2' );
```

The first query matches if *t1.c1* contains *term1*, or if *t1.c2* contains *term2*.

The second query matches if either *t1.c1* or *t1.c2* contains either *term1* or *term2*. Using the contains in this manner also returns scores for the matches.

Fuzzy search

Fuzzy searching can be used to search for misspellings or variations of a word. To do so, use the FUZZY operator followed by a string in double quotes to find an approximate match for the string. For example, `CONTAINS (Products.Description, 'FUZZY "cotton" ')` returns cotton and misspellings such as coton or cotten.

Note: You can only perform fuzzy searches on text indexes built using the NGRAM term breaker.

Using the FUZZY operator is equivalent to breaking the string manually into substrings of length *n* and separating them with OR operators. For example, suppose you have a text index

configured with the NGRAM term breaker and a MAXIMUM TERM LENGTH of 3. Specifying 'FUZZY "500 main street"' is equivalent to specifying '500 OR mai OR ain OR str OR tre OR ree OR eet'.

The FUZZY operator is useful in a full text search that returns a score. This is because many approximate matches may be returned, but usually only the matches with the highest scores are meaningful.

View search

To use a full text search on a view or derived table, you must build a text index on the columns in the base table that you want to perform a full text search on. The following statements create a view on the MarketingInformation table in the sample database, which already has a text index name, and then perform a full text search on that view.

To create a view on the MarketingInformation base table, execute the following statement:

```
CREATE VIEW MarketingInfoView AS
SELECT MI.ProductID AS ProdID,
       MI."Description" AS "Desc"
FROM GROUPO.MarketingInformation AS MI
WHERE MI."ID" > 3
```

Using the following statement, you can query the view using the text index on the underlying table.

```
SELECT *
FROM MarketingInfoView
WHERE CONTAINS ( "Desc", 'Cap OR Tee*' )
```

You can also execute the following statement to query a derived table using the text index on the underlying table.

```
SELECT *
FROM (
    SELECT MI.ProductID, MI."Description"
    FROM MarketingInformation AS MI
    WHERE MI."ID" > 4 ) AS dt ( P_ID, "Desc" )
WHERE CONTAINS ( "Desc", 'Base*' )
```

Note: The columns on which you want to run the full text search must be included in the SELECT list of the view or derived table.

Searching a view using a text index on the underlying base table is restricted as follows:

- The view cannot contain a TOP, FIRST, DISTINCT, GROUP BY, ORDER BY, UNION, INTERSECT, EXCEPT clause, or window function.
- The view cannot contain aggregate functions.
- A CONTAINS query can refer to a base table inside a view, but not to a base table inside a view that is inside another view.

FROM Clause

Specifies the database tables or views involved in a **SELECT** statement.

Syntax

```
... FROM table-expression [, ...]
```

Parameters

table-expression: { *table-spec* | *table-expression* *join-type* *table-spec* [**ON** *condition*] | (*table-expression* [, ...]) }

table-spec: { [*userid.*] *table-name* [[**AS**] *correlation-name*] | *select-statement* [**AS** *correlation-name* (*column-name* [, ...])] }

contains-expression: { *table-name* | *view-name* } **CONTAINS** (*column-name* [, ...], *contains-query*) [[**AS**] *score-correlation-name*]

Usage

contains-expression—use the **CONTAINS** clause after a table name to filter the table, and return only those rows matching the full text query specified with *contains-query*.

Every matching row of the table is returned, along with a score column that can be referred to using *score-correlation-name*, if it is specified. If *score-correlation-name* is not specified, then the score column can be referred to by the default correlation name, *contains*.

With the exception of the optional correlation name argument, the **CONTAINS** clause takes the same arguments as the **CONTAINS** search condition. There must be a **TEXT** index on the columns listed in the **CONTAINS** clause.

See also

- *CONTAINS Conditions* on page 47

CONTAINS Conditions

Perform a full text query using the **CONTAINS** clause in the **FROM** clause of a **SELECT** statement, or by using the **CONTAINS** search condition (predicate) in a **WHERE** clause.

Both methods return the same rows; however, the **CONTAINS** clause also returns scores for the matching rows.

Syntax

```
CONTAINS ( column-name [, ...], contains-query-string )
```

```
contains-query-string:  
    simple-expression | or-expression
```

```
simple-expression:  
    primary-expression | and-expression
```

Unstructured Data Queries

```
or-expression:
  simple-expression { OR | | } contains-query-string

primary-expression:
  basic-expression
  | FUZZY " fuzzy-expression "
  | and-not-expression

and-expression:
  primary-expression [ AND | & ] simple-expression

and-not-expression:
  primary-expression [ AND | & ]
  { NOT | - } basic-expression

basic-expression:
  term
  | phrase
  | ( contains-query-string )
  | proximity-expression

fuzzy-expression:
  term | fuzzy-expression term

term:
  simple-term | prefix-term

prefix-term:
  simple-term*

phrase:
  " phrase-string "

proximity-expression:
  term ( BEFORE | NEAR )
  [ minimum distance, | maximum distance ] term | term
  { BEFORE | NEAR | ~ } term

phrase-string:
  term | phrase-string term
```

Parameters

- **simple-term** – a string separated by white space and special characters that represents a single indexed term (word) for which to search.
- **distance** – a positive integer.
- **and-expression** – to specify that both *primary-expression* and *simple-expression* must be found in the **TEXT** index. By default, if no operator is specified between terms or expressions, an and-expression is assumed. For example, 'a b' is interpreted as 'a AND b'. An ampersand (&) can be used instead of AND, and can abut the expressions or terms on either side (for example, 'a & b').

- **and-not-expression** – to specify that *primary-expression* must be present in the **TEXT** index, but that *basic-expression* must not be found in the **TEXT** index. This is also known as a negation. When you use a hyphen for negation, a space must precede the hyphen, and the hyphen must abut the subsequent term. For example, 'a -b' is equivalent to 'a AND NOT b'; whereas for 'a – b', the hyphen is ignored and the string is equivalent to 'a AND b'. 'a-b' is equivalent to the phrase '"a b"'.
- **or-expression** – to specify that at least one of *simple-expression* or *contains-query-string* must be present in the **TEXT** index. For example, 'a|b' is interpreted as 'a OR b'.
- **fuzzy-expression** – to find terms that are similar to what you specify. Fuzzy matching is only supported on **NGRAM TEXT** indexes.
- **proximity-expression** – to search for terms that are near each other. For example, 'b NEAR [2, 5] c' searches for instances of b and c that are at most five and at least 2 terms away from each other. The order of terms is not significant; 'b NEAR c' is equivalent to 'c NEAR b'. If **NEAR** is specified without *distance*, a default of 10 terms is applied. You can specify a tilde (~) instead of **NEAR**. This is equivalent to specifying **NEAR** without a distance so a default of 10 terms is applied. **NEAR** expressions cannot be chained together (for example, 'a NEAR [1] b NEAR [1] c').

BEFORE is like **NEAR**, except that the order of terms is significant. 'b BEFORE c' is not equivalent to 'c BEFORE b'; in the former, the term 'b' must precede 'c' while in the latter the term 'b' must follow 'c'. **BEFORE** accepts both minimum and maximum distances like **NEAR**. The default minimum distance is 1. The minimum distance, if given, must be less than or equal to the maximum distance; otherwise, an error is returned.

- **prefix-term** – to search for terms that start with the specified prefix. For example, 'datab*' searches for any term beginning with *datab*. This is also known as a prefix search. In a prefix search, matching is performed for the portion of the term to the left of the asterisk.

Usage

The **CONTAINS** search condition takes a column list and *contains-query-string* as arguments.

The **CONTAINS** search condition can be used anywhere a search condition (also referred to as predicate) can be specified, and returns TRUE or FALSE. *contains-query-string* must be a constant string, or a variable, with a value that is known at query time.

If multiple columns are specified, then they must all refer to a single base table; a **TEXT** index cannot span multiple base tables. You can reference the base directly in the **FROM** clause, or use it in a view or derived table, provided that the view or derived table does not use **DISTINCT**, **GROUP BY**, **ORDER BY**, **UNION**, **INTERSECT**, **EXCEPT**, or a row limitation.

Queries using ANSI join syntax are supported (**FULL OUTER JOIN**, **RIGHT OUTER JOIN**, **LEFT OUTER JOIN**), but may have suboptimal performance. Use outer joins for **CONTAINS** in the **FROM** clause only if the `score` column from each of the **CONTAINS** clauses is required. Otherwise, move **CONTAINS** to an **ON** condition or **WHERE** clause.

These types of queries are unsupported:

Unstructured Data Queries

- Remote queries using a SQL Anywhere table with a full **TEXT** index that is joined to a remote table.
- Queries using SAP Sybase IQ and SQL Anywhere tables, where the full **TEXT** index to be used is on the SQL Anywhere table.
- Queries using TSQL style outer join syntax (***=***, **=*** and ***=**).

If you use a SQL variable less than 32KB in length as a search term and the type of variable is `LONG VARCHAR`, use **CAST** to convert the variable to `VARCHAR` data type. For example:

```
SELECT * FROM tab1 WHERE CONTAINS(c1, cast(v1 AS VARCHAR(64)))
```

The following warnings apply to the use of non-alphanumeric characters in query strings:

- An asterisk in the middle of a term returns an error.
- Avoid using non-alphanumeric characters (including special characters) in fuzzy-expression, because they are treated as white space and serve as term breakers.
- If possible, avoid using non-alphanumeric characters that are not special characters in your query string. Any non-alphanumeric character that is not a special character causes the term containing it to be treated as a phrase, breaking the term at the location of the character. For example, 'things we've done' is interpreted as 'things "we ve" done'.

Within phrases, the asterisk is the only special character that continues to be interpreted as a special character. All other special characters within phrases are treated as white space and serve as term breakers.

Interpretation of *contains-query-string* takes place in two main steps:

- Step 1: Interpretation of operators and precedence: During this step, keywords are interpreted as operators, and rules of precedence are applied.
- Step 2: Application of text configuration object settings: During this step, the text configuration object settings are applied to terms. Any query terms that exceed the term length settings, or that are in the stop list, are dropped.

See also

- *Fuzzy Search Over a TEXT Index* on page 55

Operator Precedence in a CONTAINS Search Condition

During query evaluation, expressions are evaluated using an order of precedence.

The order of precedence for evaluating query expressions is:

1. **FUZZY, NEAR**
2. **AND NOT**
3. **AND**
4. **OR**

Allowed Syntax for Asterisk ()*

The asterisk is used for prefix searching in a query.

An asterisk can occur at the end of the query string, or be followed by a space, ampersand, vertical bar, closing bracket, or closing quotation mark. Any other usage of asterisk returns an error.

The table "Asterisk interpretations" shows allowable asterisk usage:

Table 9. Asterisk interpretations

| Query string | Equivalent to | Interpreted as |
|-------------------|-------------------------------|--|
| 'th*&best' | 'th* AND best' and 'th* best' | Find any term beginning with th, and the term best. |
| 'th* best' | 'th* OR best' | Find either any term beginning with th, or the term best. |
| 'very&(best th*)' | 'very AND (best OR th*)' | Find the term very, and the term best or any term beginning with th. |
| '"fast auto*"' | | Find the term fast, immediately followed by a term beginning with auto. |
| '"auto* price"' | | Find a term beginning with auto, immediately followed by the term price. |

Note: Interpretation of query strings containing asterisks varies depending on the text configuration object settings.

Allowed Syntax for Hyphen (-)

The hyphen can be used in a query for term or expression negation, and is equivalent to NOT.

Whether a hyphen is interpreted as a negation depends on its location in the query string. For example, when a hyphen immediately precedes a term or expression, it is interpreted as a negation. If the hyphen is embedded within a term, it is interpreted as a hyphen.

A hyphen used for negation must be preceded by a white space, and followed immediately by an expression.

When used in a phrase of a fuzzy expression, the hyphen is treated as white space and used as a term breaker.

The table "Hyphen interpretations" shows the allowed syntax for hyphen:

Table 10. Hyphen interpretations

| Query string | Equivalent to | Interpreted as |
|---------------------|--|---|
| 'the - best' | 'the AND NOT best', 'the AND -best', 'the & -best', 'the NOT best' | Find the term the, and not the term best. |
| 'the - (very best)' | 'the AND NOT (very AND best)' | Find the term the, and not the terms very and best. |
| 'the -"very best"' | 'the AND NOT "very best"' | Find the term the, and not the phrase very best. |
| 'alpha- numerics' | "alpha numerics" | Find the term alpha, immediately followed by the term numerics. |
| 'wild - west' | 'wild west', and 'wild AND west' | Find the term wild, and the term west. |

Allowed Syntax for Special Characters

The table "Special character interpretations" shows the allowed syntax for all special characters, except asterisk and hyphen.

The asterisk and hyphen characters are not considered special characters, if they are found in a phrase, and are dropped.

Note: The restrictions on specifying string literals also apply to the query string. For example, apostrophes must be within an escape sequence.

Table 11. Special character interpretations

| Character or syntax | Usage examples and remarks |
|---------------------|---|
| ampersand (&) | <p>The ampersand is equivalent to AND, and can be specified as follows:</p> <ul style="list-style-type: none"> • 'a & b' • 'a &b' • 'a& b' • 'a&b' |

| Character or syntax | Usage examples and remarks |
|---------------------|--|
| vertical bar () | The vertical bar is equivalent to OR , and can be specified as follows: <ul style="list-style-type: none"> 'a b' 'a b' 'a b' 'a b' |
| double quotes ("") | Double quotes are used to contain a sequence of terms where order and relative distance are important. For example, in the query string 'learn "full text search"', "full text search" is a phrase. In this example, learn can come before or after the phrase, or exist in another column (if the TEXT index is built on more than one column), but the exact phrase must be found in a single column. |
| parentheses () | Parentheses are used to specify the order of evaluation of expressions, if different from the default order. For example, 'a AND (b c)' is interpreted as a, and b or c. |
| tilde (~) | The tilde is equivalent to NEAR , and has no special syntax rules. The query string 'full~text' is equivalent to 'full NEAR text', and is interpreted as: the term full within ten terms of the term text. |
| square brackets [] | Square brackets are used in conjunction with the keyword NEAR to contain <i>distance</i> . Other uses of square brackets return an error. |

Effect of Dropped Terms

A **TEXT** index may exclude terms that meet certain conditions.

TEXT indexes are built according to the settings defined for the text configuration object used to create the **TEXT** index. A **TEXT** index excludes terms that meet any of the following conditions:

- The term is included in the stop list.
- The term is shorter than the minimum term length (**GENERIC** only).
- The term is longer than the maximum term length.

The same rules apply to query strings. The dropped term can match zero or more terms at the end or beginning of the phrase. For example, suppose the term 'the' is in the stop list:

- If the term appears on either side of an **AND**, **OR**, or **NEAR**, then both the operator and the term are removed. For example, searching for 'the AND apple', 'the OR apple', or 'the NEAR apple' are equivalent to searching for 'apple'.
- If the term appears on the right side of an **AND NOT**, both the **AND NOT** and the term are dropped. For example, searching for 'apple AND NOT the' is equivalent to searching for 'apple'.
- If the term appears on the left side of an **AND NOT**, the entire expression is dropped. For example, searching for 'the AND NOT apple' returns no rows. Another example:

Unstructured Data Queries

'orange and the AND NOT apple' is the same as 'orange AND (the AND NOT apple)' which, after the **AND NOT** expression is dropped, is equivalent to searching for 'orange'. Contrast this with the search expression '(orange and the) and not apple', which is equivalent to searching for 'orange and not apple'.

- If the term appears in a phrase, the phrase is allowed to match with any term at the position of the dropped term. For example, searching for 'feed the dog' matches 'feed the dog', 'feed my dog', 'feed any dog', and so on.

Note: If all of the terms for which you are searching are dropped, SAP Sybase IQ returns the error `CONTAINS has NULL search term`. SQL Anywhere reports no error and returns zero rows.

Query Match Score

You can sort query results using the score that indicates the closeness of a match.

When you include a **CONTAINS** clause in the **FROM** clause of a query, each match has a score associated with it. The score indicates how close the match is, and you can use score information to sort the data. Two main criteria determine score:

- The number of times a term appears in the indexed row. The more times a term appears in an indexed row, the higher its score.
- The number of times a term appears in the **TEXT** index. The more times a term appears in a **TEXT** index, the lower its score.

Depending on the type of full text search, other criteria affect scoring. For example, in proximity searches, the proximity of search terms impacts scoring. By default, the result set of a **CONTAINS** clause has the correlation name `contains` that has a single column in it called `score`. You can refer to `"contains".score` in the **SELECT** list, **ORDER BY** clause, or other parts of the query. However, because `contains` is a SQL reserved word, you must remember to put it in double quotes. Alternatively, you can specify another correlation name, for example, `CONTAINS (expression) AS ct`. The examples for full text search refer to the `score` column as `ct.score`.

This statement searches `MarketingInformation.Description` for terms starting with "stretch" or terms starting with "comfort":

```
SELECT ID, ct.score, Description
FROM MarketingInformation
CONTAINS ( MarketingInformation.Description,
           'stretch* | comfort*' )
AS ct ORDER BY ct.score DESC;
```

NGRAM TEXT Index Searches

Fuzzy and non-fuzzy search capability over a **TEXT** index is possible for **TEXT** indexes of type **NGRAM**.

Fuzzy Search Over a TEXT Index

Fuzzy search capability over a **TEXT** index is possible only if the **TEXT** index is of type **NGRAM**. The **GENERIC TEXT** index cannot handle the fuzzy search.

Fuzzy searching can be used to search for misspellings or variations of a word. To do so, use the **FUZZY** operator followed by a string in double quotes to find an approximate match for the string.

Using the **FUZZY** operator is equivalent to breaking the string manually into substrings of length n and separating them with **OR** operators. For example, if you have a text index configured with the **NGRAM** term breaker and a **MAXIMUM TERM LENGTH** of 3, specifying 'FUZZY "500 main street" ' is equivalent to specifying '500 OR mai OR ain OR str OR tre OR ree OR eet'.

The **FUZZY** operator is useful in a full text search that returns a score. Many approximate matches may be returned, but usually only the matches with the highest scores are meaningful.

Note: Fuzzy search does not support prefix or suffix searching. For example, the search clause cannot be “v*” or “*vis”.

Example 1: Fuzzy search over an NGRAM TEXT index

Create a table and an **NGRAM TEXT** index:

```
CREATE TEXT CONFIGURATION NGRAMTtxtcfg
  FROM default_char;
ALTER TEXT CONFIGURATION NGRAMTtxtcfg TERM BREAKER      NGRAM;
ALTER TEXT CONFIGURATION NGRAMTtxtcfg maximum term     length 3;
CREATE TABLE t_iq(a int, b varchar(100));
CREATE TEXT INDEX TXT_IQ on t_iq(b) CONFIGURATION      NGRAMTtxtcfg
```

Insert this data into the table:

```
INSERT INTO t_iq values (1,'hello this is hira ');
INSERT INTO t_iq values(2, ' book he ookw worm okwo
kwor');
INSERT INTO t_iq values(3,'Michael is a good person');
INSERT INTO t_iq values(4,'hello this is evaa');
INSERT INTO t_iq values(5,'he is a bookworm');
INSERT INTO t_iq values (6,'boo ook okw kwo wor orm');
```

After inserting the data, execute this query to perform fuzzy searching over an **NGRAM TEXT** index:

Unstructured Data Queries

```
SELECT * FROM t_iq WHERE CONTAINS (b, 'FUZZY "bookerm"');
```

The results of the query are:

```
a      b
2      book he ookw worm okwo kwor
5      he is a bookworm
6      boo ook okw kwo wor orm
```

Example 2: Additional letter in the fuzzy search clause

This query illustrates an additional letter in the fuzzy search clause:

```
SELECT * FROM t_iq WHERE CONTAINS (b, 'FUZZY "hellow"');
```

The results of the query are:

```
a      b
1      hello this is hira
4      hello this is evaa
```

Example 3: Letter removed from the fuzzy search clause

In this query, a letter is removed from the fuzzy search clause:

```
SELECT * FROM t_iq WHERE CONTAINS(b, 'FUZZY "hlllo"');
```

The results of the query are:

```
a      b
1      hello this is hira
4      hello this is evaa
```

Non-fuzzy Search Over a TEXT Index

Non-fuzzy search on **NGRAM** breaks the term into corresponding n-grams and searches for the n-grams in the **NGRAM TEXT** index.

The query `CONTAINS (M.Description, 'ams') ct;` illustrates a non-fuzzy **NGRAM** search over a 2GRAM index, which is semantically equal to searching query `CONTAINS(M.Description, '"am ms"') ct;`

If you search for a 'v*' TERM on a 2GRAM index, then v followed by any alphabet is considered as a matching 2GRAM for the searching term and is output as a result.

The query `CONTAINS (M.Description, 'white whale') ct;` illustrates a non-fuzzy **NGRAM** search over a 3GRAM index and is semantically equal to searching query `CONTAINS (M.Description, '"whi hit ite wha hal ale"');`

The difference between **NGRAM** fuzzy and non-fuzzy search is that fuzzy search is a disjunction over individual GRAMS. Non-fuzzy search is a conjunction over the individual GRAMS. When **GENERIC** and **NGRAM TEXT** indexes are created on the same column, then the **GENERIC TEXT** index is used for a query with non-fuzzy search and the **NGRAM TEXT** index is used for fuzzy search.

Example 1: non-fuzzy search after creating a **GENERIC TEXT** index on the same column

This query illustrates non-fuzzy search after creating a **GENERIC TEXT** index on the same column:

```
SELECT * FROM t_iq WHERE CONTAINS (b, 'bookworm');
```

The results of the query are:

| a | b |
|---|------------------|
| 5 | he is a bookworm |

Example 2: fuzzy search with both **NGRAM** and **GENERIC TEXT** indexes on the same column

This query illustrates fuzzy search with both **NGRAM** and **GENERIC TEXT** indexes on the same column:

```
SELECT * FROM t_iq
WHERE CONTAINS (b, 'FUZZY "bookworm"');
```

The results of the query are:

| a | b |
|---|-----------------------------|
| 2 | book he ookw worm okwo kwor |
| 5 | he is a bookworm |
| 6 | boo ook okw kwo wor orm |

Example 3: fuzzy search phrase in a non-fuzzy search clause

This query illustrates the behavior of a fuzzy search phrase in a non-fuzzy search clause:

```
SELECT * FROM t_iq WHERE CONTAINS (b, 'bookworm');
```

No result is returned for this query.

Queries on **LONG BINARY** Columns

In **WHERE** clauses of the **SELECT** statement, you can use **LONG BINARY** columns only in **IS NULL** and **IS NOT NULL** expressions, in addition to the **BYTE_LENGTH64**, **BYTE_SUBSTR64**, **BYTE_SUBSTR**, **BIT_LENGTH**, **OCTET_LENGTH**, **CHARINDEX**, and **LOCATE** functions.

You cannot use **LONG BINARY** columns in the **SELECT** statement clauses **ORDER BY**, **GROUP BY**, and **HAVING** or with the **DISTINCT** keyword.

SAP Sybase IQ does not support **LIKE** predicates on **LONG BINARY** (BLOB) columns or variables. Searching for a pattern in a **LONG BINARY** column using a **LIKE** predicate returns the error `Invalid data type comparison in predicate`.

See also

- *Function Support* on page 91

Queries on LONG VARCHAR Columns

In **WHERE** clauses of the **SELECT** statement, you can use `LONG VARCHAR` columns only in **IS NULL** and **IS NOT NULL** expressions, in addition to the **BIT_LENGTH**, **CHAR_LENGTH**, **CHAR_LENGTH64**, **CHARINDEX**, **LOCATE**, **OCTET_LENGTH**, **PATINDEX**, **SUBSTRING64**, and **SUBSTRING** functions.

You can use the **LIKE** predicate to search for a pattern on a `LONG VARCHAR` column. All patterns of 126 or fewer characters are supported. Patterns longer than 254 characters are not supported. Some patterns between 127 and 254 characters in length are supported, depending on the contents of the pattern.

The **LIKE** predicate supports **LONG VARCHAR (CLOB)** variables of any size of data. Currently, a **SQL** variable can hold up to 2GB - 1 in length.

You cannot use `LONG VARCHAR` columns in the **SELECT** statement clauses **ORDER BY**, **GROUP BY**, and **HAVING** or with the **DISTINCT** keyword (**SELECT DISTINCT** and **COUNT DISTINCT**).

See also

- *Function Support* on page 91

CONTAINS Predicate Support

You can create a **WORD (WD)** index on a `LONG VARCHAR (CLOB)` column and use the **CONTAINS** predicate to search the column for string constants of maximum length 255 characters.

The **CONTAINS** predicate is not supported on **LONG BINARY (BLOB)** columns using **WD** indexes. If you attempt to search for a string in a `LONG BINARY` column with a **WD** index using a **CONTAINS** predicate, an error is returned. **TEXT** indexes that use an external library support **CONTAINS** on binary data.

Performance Monitoring of LONG BINARY and LONG VARCHAR Columns

The performance monitor displays performance data for `LONG BINARY` and `LONG VARCHAR` columns.

Stored Procedure Support

Learn about stored procedure support for the `LONG BINARY (BLOB)` and `LONG VARCHAR (CLOB)` data type columns and full text searching.

Term Management in a TEXT Index

You can use stored procedures to break strings into terms, to discover how many terms are in the **TEXT** index and their position, and to display statistical information about the **TEXT** indexes.

sa_char_terms System Procedure

Breaks a `CHAR` string into terms and returns each term as a row along with its position.

Syntax

```
sa_char_terms( 'char-string' [, 'text-config-name'
[, 'owner' ] ] )
```

Parameters

- **char-string** – the `CHAR` string you are parsing.
- **text-config-name** – the text configuration object to apply when processing the string. The default value is 'default_char'.
- **owner** – the owner of the specified text configuration object. The default value is `DBA`.

Privileges

None

Remarks

You can use **sa_char_terms** to find out how a string is interpreted when the settings for a text configuration object are applied. This can be helpful when you want to know what terms would be dropped during indexing or from a query string.

Example

Return the terms in the `CHAR` string 'the quick brown fox jumped over the fence':

```
CALL sa_char_terms
( 'the quick brown fox jumped over the fence' );
```

Table 12. CHAR string interpretation

| Term | Position |
|--------|----------|
| the | 1 |
| quick | 2 |
| brown | 3 |
| fox | 4 |
| jumped | 5 |
| over | 6 |
| the | 7 |
| fence | 8 |

sa_nchar_terms System Procedure

Breaks an NCHAR string into terms and returns each term as a row along with its position.

Syntax

```
sa_nchar_terms ( 'char-string' [ , 'text-config-
name' [, 'owner' ] ] )
```

Parameters

- **char-string** – the NCHAR string you are parsing.
- **text-config-name** – the text configuration object to apply when processing the string. The default value is 'default_nchar'.
- **owner** – the owner of the specified text configuration object. The default value is DBA.

Privileges

You must have EXECUTE privilege on the system procedure.

Remarks

You can use **sa_nchar_terms** to find out how a string is interpreted when the settings for a text configuration object are applied. This can be helpful when you want to know what terms would be dropped during indexing or from a query string.

The syntax for **sa_nchar_terms** is similar to the syntax for the **sa_char_terms** system procedure.

Note: The NCHAR data type is supported only for **IN SYSTEM** tables.

sa_text_index_stats System Procedure

Returns statistical information about the TEXT indexes in the database.

Syntax

```
sa_text_index_stats( )
```

Privileges

You must have EXECUTE privilege on the system procedure. You must also have one of the following system privileges:

- MANAGE ANY STATISTICS
- CREATE ANY INDEX
- ALTER ANY INDEX
- DROP ANY INDEX
- CREATE ANY OBJECT
- ALTER ANY OBJECT
- DROP ANY OBJECT

Remarks

Use **sa_text_index_stats** to view statistical information for each TEXT index in the database.

Table 13. Statistical information for TEXT indexes returned by sa_text_index_stats

| Column name | Type | Description |
|--------------------|-----------------|---|
| owner_id | UNSIGNED INT | ID of the owner of the table |
| table_id | UNSIGNED INT | ID of the table |
| index_id | UNSIGNED INT | ID of the TEXT index |
| text_config_id | UNSIGNED BIGINT | ID of the text configuration referenced by the TEXT index |
| owner_name | CHAR(128) | Name of the owner |
| table_name | CHAR(128) | Name of the table |
| index_name | CHAR(128) | Name of the TEXT index |

| Column name | Type | Description |
|------------------|-----------------|---|
| text_config_name | CHAR(128) | Name of the text configuration object |
| doc_count | UNSIGNED BIGINT | Total number of indexed column values in the TEXT index |
| doc_length | UNSIGNED BIGINT | Total length of data in the TEXT index |
| pending_length | UNSIGNED BIGINT | Total length of the pending changes |
| deleted_length | UNSIGNED BIGINT | Total length of the pending deletions |
| last_refresh | TIMESTAMP | Date and time of the last refresh |

The `pending_length`, `deleted_length`, and `last_refresh` values are NULL for IMMEDIATE REFRESH TEXT indexes.

Example

Return statistical information for each **TEXT** index in the database:

```
CALL sa_text_index_stats( );
```

sa_text_index_vocab System Procedure

Lists all terms that appear in a **TEXT** index, and the total number of indexed values in which each term appears.

Syntax

```
sa_text_index_vocab (
    'text-index-name',
    'table-name',
    'table-owner'
)
```

Parameters

- **text-index-name** – use this CHAR(128) parameter to specify the name of the **TEXT** index.
- **table-name** – use this CHAR(128) parameter to specify the name of the table on which the **TEXT** index is built.
- **table-owner** – use this CHAR(128) parameter to specify the owner of the table.

Privileges

You must have EXECUTE privilege on the system procedure. You must also have one of the following:

- SELECT ANY TABLE system privilege
- SELECT privilege on the indexed table

Remarks

sa_text_index_vocab returns all terms that appear in a **TEXT** index, and the total number of indexed values in which each term appears (which is less than the total number of occurrences, if the term appears multiple times in some indexed values).

Parameter values cannot be host variables or expressions. The arguments *text-index-name*, *table-name*, and *table-owner* must be constraints or variables.

Example

Execute **sa_text_index_vocab** to return all the terms that appear in the **TEXT** index MyTextIndex on table Customers owned by GROUPO:

```
sa_text_index_vocab
('MyTextIndex', 'Customers', 'GROUPO');
```

Table 14. Terms in the index

| term | freq |
|-------------|------|
| a | 1 |
| Able | 1 |
| Acres | 1 |
| Active | 5 |
| Advertising | 1 |
| Again | 1 |
| ... | ... |

External Library Identification

The **sa_list_external_library** stored procedure lists the libraries that are currently loaded in the server. Once identified, use **sa_external_library_unload** to unload the library from the server.

sa_external_library_unload System Procedure

Unloads an external library.

Syntax

```
sa_external_library_unload ( [ 'external-library' ] )
```

Parameters

- **external-library** – optionally use this LONG VARCHAR parameter to specify the name of a library to be unloaded. If no library is specified, all external libraries that are not in use are unloaded.

Privileges

You must have EXECUTE privilege on the system procedure. You must also have the MANAGE ANY EXTERNAL OBJECT system privilege.

Remarks

If an external library is specified, but is in use or is not loaded, an error is returned. If no parameter is specified, an error is returned if no loaded external libraries are found.

Example

Unload an external library called myextlib.dll:

```
CALL sa_external_library_unload( 'myextlib.dll' );
```

Unload all libraries that are not currently in use:

```
CALL sa_external_library_unload();
```

sa_list_external_library System Procedure

Lists the external libraries currently loaded in the server.

Syntax

```
sa_list_external_library ( )
```

Privileges

You must have EXECUTE privilege on the system procedure. You must also have the MANAGE ANY EXTERNAL OBJECT system privilege.

Remarks

Returns a list of external libraries loaded in the engine along with their reference count.

The reference count is the number of instances of the library in the engine. An external library can be unloaded by executing the procedure **sa_external_library_unload**, only if its reference count is 0.

Example

List the external libraries and their reference count:

```
CALL sa_list_external_library()
```

Large Object Data Compression

The **sp_iqsetcompression** stored procedure controls the compression of large object columns.

sp_iqsetcompression controls the compression of columns of data type `LONG BINARY` and `LONG VARCHAR` when writing database buffers to disk. You can also use

sp_iqsetcompression to disable compression. This functionality saves CPU cycles, because certain data formats stored in a `LONG BINARY` or `LONG VARCHAR` column (for example, JPG files) are already compressed and gain nothing from additional compression.

The **sp_iqshowcompression** stored procedure displays the compression setting of large object columns.

sp_iqsetcompression Procedure

Sets compression of data in columns of `LONG BINARY (BLOB)` and `LONG VARCHAR (CLOB)` data types.

Syntax

```
sp_iqsetcompression ( owner, table, column, on_off_flag )
```

Permissions

You must have EXECUTE privilege on the system procedure. You must also have one of the following system privilege:

- ALTER ANY TABLE
- ALTER ANY OBJECT

Remarks

sp_iqsetcompression provides control of compression of `LONG BINARY (BLOB)` and `LONG VARCHAR (CLOB)` data type columns. The compression setting applies only to base tables.

A side effect of **sp_iqsetcompression** is that a **COMMIT** occurs after you change the compression setting.

Table 15. sp_iqsetcompression parameters

| Name | Description |
|--------------------|---|
| <i>owner</i> | Owner of the table for which you are setting compression |
| <i>table</i> | Table for which you are setting compression |
| <i>column</i> | Column for which you are setting compression |
| <i>on_off_flag</i> | Compression setting: ON enables compression, OFF disables compression |

Example

Assume this table definition:

```
CREATE TABLE USR.pixTable (picID INT NOT NULL,
picJPG LONG BINARY NOT NULL);
```

To turn off compression on the LOB column picJPG, call **sp_iqsetcompression**:

```
CALL sp_iqsetcompression('USR', 'pixTable', 'picJPG',
'OFF') ;
```

This command returns no rows.

sp_iqshowcompression Procedure

Displays compression settings for columns of LONG BINARY (BLOB) and LONG VARCHAR (CLOB) data types.

Syntax

```
sp_iqshowcompression ( owner, table, column )
```

Privileges

You must have EXECUTE privilege on the system procedure. You must also have one of the following system privileges:

- ALTER ANY TABLE
- ALTER ANY OBJECT

Remarks

Returns the column name and compression setting. Compression setting values are 'ON' (compression enabled) and 'OFF' (compression disabled).

Table 16. sp_iqshowcompression parameters

| Name | Description |
|--------------|--|
| <i>owner</i> | Owner of the table for which you are setting compression |

| Name | Description |
|---------------|--|
| <i>table</i> | Table for which you are setting compression |
| <i>column</i> | Column for which you are setting compression |

Example

Assume this table definition:

```
CREATE TABLE USR.pixTable (picID INT NOT NULL,
picJPG LONG BINARY NOT NULL);
```

To check the compression status of the columns in the `pixTable` table, call **sp_iqshowcompression**:

```
CALL sp_iqshowcompression('USR', 'pixTable',
'picJPG') ;
```

This command returns one row:

```
'picJPG', 'ON'
```

Information About Large Object Columns

The stored procedure `sp_iqindexsize` displays the size of an individual `LONG BINARY` and `LONG VARCHAR` column.

Size of a LONG BINARY Column

`sp_iqindexsize` output that shows a `LONG BINARY` column with approximately 42GB of data.

The page size is 128KB. The `largelob` Info type is in the last row.

| Username | Indexname | Type Info | KBytes | Pages | Compressed Pages |
|----------|--------------------------------|-------------|----------|--------|------------------|
| DBA | test10.DBA.ASIQ_IDX_T128_C3_FP | FP Total | 42953952 | 623009 | 622923 |
| DBA | test10.DBA.ASIQ_IDX_T128_C3_FP | FP vdo | 0 | 0 | 0 |
| DBA | test10.DBA.ASIQ_IDX_T128_C3_FP | FP bt | 0 | 0 | 0 |
| DBA | test10.DBA.ASIQ_IDX_T128_C3_FP | FP garray | 0 | 0 | 0 |
| DBA | test10.DBA.ASIQ_IDX_T128_C3_FP | FP bm | 136 | 2 | 1 |
| DBA | test10.DBA.ASIQ_IDX_T128_C3_FP | FP barray | 2312 | 41 | 40 |
| DBA | test10.DBA.ASIQ_IDX_T128_C3_FP | FP dpstore | 170872 | 2551 | 2549 |
| DBA | test10.DBA.ASIQ_IDX_T128_C3_FP | FP largelob | 42780632 | 620415 | 620333 |

In this example, the compression ratio is $42953952 / (623009 * 128) = 53.9\%$.

Size of a LONG VARCHAR Column

`sp_iqindexsize` output that shows a LONG VARCHAR column with approximately 42GB of data.

The page size is 128KB. The `largelob` Info type is in the last row.

| Username | Indexname | Type | Info | KBytes | Pages | Compressed Pages |
|----------|--------------------------------|------|----------|----------|--------|------------------|
| DBA | test10.DBA.ASIQ_IDX_T128_C3_FP | FP | Total | 42953952 | 623009 | 622923 |
| DBA | test10.DBA.ASIQ_IDX_T128_C3_FP | FP | vdo | 0 | 0 | 0 |
| DBA | test10.DBA.ASIQ_IDX_T128_C3_FP | FP | bt | 0 | 0 | 0 |
| DBA | test10.DBA.ASIQ_IDX_T128_C3_FP | FP | garray | 0 | 0 | 0 |
| DBA | test10.DBA.ASIQ_IDX_T128_C3_FP | FP | bm | 136 | 2 | 1 |
| DBA | test10.DBA.ASIQ_IDX_T128_C3_FP | FP | barray | 2312 | 41 | 40 |
| DBA | test10.DBA.ASIQ_IDX_T128_C3_FP | FP | dpstore | 170872 | 2551 | 2549 |
| DBA | test10.DBA.ASIQ_IDX_T128_C3_FP | FP | largelob | 42780632 | 620415 | 620333 |

In this example, the compression ratio is $42953952 / (623009 * 128) = 53.9\%$.

Large Object Data Load and Unload

Learn how to export and load large object data.

Large Object Data Exports

The SAP Sybase IQ data extraction facility includes the **BFILE** function, which allows you to extract individual `LONG BINARY` and `LONG VARCHAR` cells to individual operating system files on the server.

You can use **BFILE** with or without the data extraction facility.

BFILE Function

Extracts individual `LONG BINARY` and `LONG VARCHAR` cells to individual operating system files on the server.

Syntax

```
BFILE( file-name-expression, large-object-column )
```

Parameters

- **file-name-expression** – the name of the output file into which the `LONG BINARY` or `LONG VARCHAR` data is written. This file name can be up to (32K - 1) bytes in length, but must be a valid path name that is supported by the file system.
- **large-object-column** – the name of the `LONG BINARY` or `LONG VARCHAR` column.

Usage

BFILE returns:

- 1, if the file is successfully written
- 0, if the file is not successfully opened or written
- NULL, if the `LONG BINARY` or `LONG VARCHAR` cell value is NULL

If the `LONG BINARY` or `LONG VARCHAR` cell value is NULL, no file is opened and no data is written.

The file path is relative to where the server was started and the open and write operations execute with the permissions of the server process. Tape devices are not supported for the **BFILE** output file.

`LONG BINARY` and `LONG VARCHAR` cells retrieved other than with the **BFILE** function (that is, retrieved through the client/server database connection later) are limited in size to a maximum length of 2GB. Use **SUBSTRING64** or **BYTE_SUBSTR64** to retrieve `LONG`

Large Object Data Load and Unload

BINARY cells greater than 2GB using a **SELECT (SELECT, OPEN CURSOR)**. Use **SUBSTRING64** to retrieve LONG VARCHAR cells greater than 2GB using a **SELECT (SELECT, OPEN CURSOR)**. Some connection drivers, for example ODBC, JDBC™, and Open Client™, do not allow more than 2GB to be returned in one **SELECT**.

You can use **BFILE** with or without the data extraction facility.

BFILE Function Example

Use **BFILE** to extract and reload LOB data.

Create table LobA:

```
create table LobA
  (rowid  int primary key,
   col1   clob null,
   col2   blob null)
```

Assume LobA has two rows of data.

Extract the non-LOB data and the paths to the files into which the LOB data is extracted:

```
BEGIN
  SET TEMPORARY OPTION
    Temp_Extract_Name1 = LobA_data.txt;
  SELECT rowid,
    'row' + string(rowid) + '.' + 'col1',
    'row' + string(rowid) + '.' + 'col2'
  FROM LobA;
END
```

The file LobA_data.txt is created and contains this non-LOB data and these filenames:

```
1,row1.col1,row1.col2,
2,row2.col1,row2.col2,
```

Perform the LOB data extraction:

```
SELECT
  BFILE('row' + string(rowid) + '.' + 'col1',col1),
  BFILE('row' + string(rowid) + '.' + 'col2',col2)
FROM LobA;
```

After the extraction, there is a file for each cell of LOB data extracted. For example, if table LobA contains two rows of data with rowid values of 1 and 2, you have these files:

- row1.col1
- row1.col2
- row2.col1
- row2.col2

Reload the extracted data:

```
LOAD TABLE LobA
(rowid,
 col1 ASCII FILE ('') NULL('NULL'),
```

```
col2 BINARY FILE ('') NULL('NULL'))
FROM Loba_data.txt'
DELIMITED BY ','
ROW DELIMITED BY '\n'
ESCAPES OFF;
```

Large Object Data Loads

Load LONG BINARY and LONG VARCHAR data using the extended syntax of the **LOAD TABLE** statement.

You can load large object data of unlimited size, unless restricted by the operating system, from a primary file in ASCII or BCP format. The maximum length of fixed-width data loaded from a primary file into large object columns is 32K - 1.

You can also specify a secondary load file in the primary load file. Each individual secondary data file contains exactly one LONG BINARY or LONG VARCHAR cell value.

Extended LOAD TABLE Syntax

LOAD TABLE has extended syntax for loading large object data.

```
LOAD [ INTO ] TABLE [ owner ].table-name
... ( column-name load-column-specification [, ...] )
... FROM 'filename-string' [, ...]
... [ QUOTES { ON | OFF } ]
... ESCAPES OFF
... [ FORMAT { ascii | binary | bcp } ]
... [ DELIMITED BY 'string' ]
...
```

```
load-column-specification:
...
| { BINARY | ASCII } FILE ( integer )
| { BINARY | ASCII } FILE ( 'string' )
```

The keywords **BINARY FILE** (for LONG BINARY) or **ASCII FILE** (for LONG VARCHAR) specify to the load that the primary input file for the column contains the path of the secondary file (which contains the LONG BINARY or LONG VARCHAR cell value), rather than the LONG BINARY or LONG VARCHAR data itself. The secondary file pathname can be either fully qualified or relative. If the secondary file pathname is not fully qualified, then the path is relative to the directory in which the server was started. Tape devices are not supported for the secondary file.

SAP Sybase IQ supports loading LONG BINARY and LONG VARCHAR values of unlimited length (subject to operating system restrictions) in the primary load file. When binary data of hexadecimal format is loaded into a LONG BINARY column from a primary file, SAP Sybase IQ requires that the total number of hexadecimal digits is an even number. The error "Odd length of binary data value detected on column" is reported, if the cell

Large Object Data Load and Unload

value contains an odd number of hexadecimal digits. Input files for `LONG BINARY` loads should always contain an even number of hexadecimal digits.

SAP Sybase IQ does not support loading large object columns from primary files using **LOAD TABLE...FORMAT BINARY**. You can load large object data in binary format from secondary files.

For **LOAD TABLE FORMAT BCP**, the load specification may contain only column names, **NULL**, and **ENCRYPTED**. This means that you cannot use secondary files when loading `LONG BINARY` and `LONG VARCHAR` columns using the **LOAD TABLE FORMAT BCP** option.

Large Object Data Load Example

Create and load a table with `LONG BINARY` data.

```
CREATE TABLE ltab (c1 INT, filename CHAR(64),
  ext CHAR(6), lobcol LONG BINARY NULL);

LOAD TABLE ltab (
  c1,
  filename,
  ext NULL('NULL'),
  lobcol BINARY FILE ('.') NULL('NULL')
)
FROM 'abc.inp'
QUOTES OFF ESCAPES OFF;
```

The primary file `abc.inp` contains this data:

```
1,boston,jpg,/s1/loads/lobs/boston.jpg,
2,map_of_concord,bmp,/s1/loads/maps/concord.bmp,
3,zero length test,NULL,,
4,null test,NULL,NULL,
```

After the `LONG BINARY` data is loaded into table `tab`, the first and second rows for column `lobcol` contain the contents of files `boston.jpg` and `concord.bmp`, respectively. The third and fourth rows contain a zero-length value and `NULL`, respectively.

Control of Load Errors

The database option `SECONDARY_FILE_ERROR` allows you to specify the action of the load if an error occurs while opening or reading from a secondary **BINARY FILE** or **ASCII FILE**.

If `SECONDARY_FILE_ERROR` is `ON`, the load rolls back if an error occurs while opening or reading from a secondary **BINARY FILE** or **ASCII FILE**.

If `SECONDARY_FILE_ERROR` is `OFF` (the default), the load continues, regardless of any errors that occur while opening or reading from a secondary **BINARY FILE** or **ASCII FILE**. The `LONG BINARY` or `LONG VARCHAR` cell is left with one of these values:

- `NULL`, if the column allows nulls
- Zero-length value, if the column does not allow nulls

Any user can set `SECONDARY_FILE_ERROR` for the `PUBLIC` role or temporary; the option setting takes effect immediately.

When logging integrity constraint violations to the load error **ROW LOG** file, the information logged for a `LONG BINARY` or `LONG VARCHAR` column is:

- Actual text as read from the primary data file, if the logging occurs within the first pass of the load operation
- Zero-length value, if the logging occurs within the second pass of the load operation

Load of Large Object Data with Trailing Blanks

The **LOAD TABLE...STRIP** option has no effect on `LONG VARCHAR` data.

Trailing blanks are not stripped from `LONG VARCHAR` data, even if the **STRIP** option is on.

Load of Large Object Data with Quotes

The **LOAD TABLE...QUOTES** option does not apply to loading `LONG BINARY` (BLOB) or `LONG VARCHAR` (CLOB) data from the secondary file, regardless of its setting,

A leading or trailing quote is loaded as part of CLOB data. Two consecutive quotes between enclosing quotes are loaded as two consecutive quotes with the **QUOTES ON** option.

Truncation of Partial Multibyte Character Data

Partial multibyte `LONG VARCHAR` data is truncated during the load according to the value of the `TRIM_PARTIAL_MBC` database option.

- If `TRIM_PARTIAL_MBC` is **ON**, a partial multibyte character is truncated for both primary data and the **LOAD** with **ASCII FILE** option.
- If `TRIM_PARTIAL_MBC` is **OFF**, the **LOAD** with **ASCII FILE** option handles the partial multibyte character according to the value of the `SECONDARY_FILE_ERROR` database option.

The table "Partial multibyte character on loading `LONG VARCHAR` with **ASCII FILE** option" lists how a trailing multibyte character is loaded, depending on the values of `TRIM_PARTIAL_MBC` and `SECONDARY_FILE_ERROR`.

Table 17. Partial multibyte character on loading `LONG VARCHAR` with **ASCII FILE option**

| <code>TRIM_PARTIAL_MBC</code> | <code>SECONDARY_FILE_ERROR</code> | Trailing Partial Multibyte Character Found |
|-------------------------------|-----------------------------------|--|
| ON | ON/OFF | Trailing partial multibyte character truncated |

| TRIM_PARTIAL_MBC | SECONDARY_FILE_ERROR | Trailing Partial Multibyte Character Found |
|------------------|----------------------|---|
| OFF | ON | Cell — null, if null allowed LOAD error — roll back, if null not allowed |
| OFF | OFF | Cell — null, if null allowed Cell — zero-length, if null not allowed |

Load Support of Large Object Variables

Large object variables are supported by the **LOAD TABLE**, **INSERT...VALUES**, **INSERT...SELECT**, **INSERT...LOCATION**, **SELECT...INTO**, and **UPDATE SQL** statements.

See also

- *Large Object Data Types* on page 75
- *Large Object Variables* on page 78

Large Object Data Types

Learn about the characteristics of the large object `LONG BINARY` and `LONG VARCHAR` data type columns and index support of large object data.

Large Object Data Types `LONG BINARY` and `BLOB`

Binary large object (BLOB) data in SAP Sybase IQ is stored in columns of data type `LONG BINARY` or `BLOB`.

An individual `LONG BINARY` data value can have a length ranging from zero (0) to 512TB (terabytes) for an IQ page size of 128KB, or 2PB (petabytes) for an IQ page size of 512KB. (The maximum length is equal to 4GB multiplied by the database page size.) To accommodate a table with `LONG BINARY` data, an IQ database must be created with an IQ page size of at least 128KB (131072 bytes).

A table or database can contain any number of `LONG BINARY` columns up to the supported maximum columns per table and maximum columns per database, respectively.

`LONG BINARY` columns can be either `NULL` or `NOT NULL` and can store zero-length values. The domain `BLOB` is a `LONG BINARY` data type that allows `NULL`.

You cannot construct a non-FP index on a `LONG BINARY` column.

Prefetching is disabled, if the result set contains `BLOB` columns.

Modify `LONG BINARY` columns using the **UPDATE**, **INSERT**, **LOAD TABLE**, **DELETE**, **TRUNCATE**, **SELECT...INTO** and **INSERT...LOCATION** SQL statements. Positioned updates and deletes are not supported on `LONG BINARY` columns.

You can insert an Adaptive Server `IMAGE` column into a `LONG BINARY` column using the **INSERT...LOCATION** command. All `IMAGE` data inserted is silently right-truncated at 2147483648 bytes (2GB).

`LONG BINARY` Data Type Conversion

There are limited implicit data type conversions to and from the `LONG BINARY` data type and non-`LONG BINARY` data types.

There are no implicit data type conversions from the `LONG BINARY` data type to another non-`LONG BINARY` data type, except to the `BINARY` and `VARBINARY` data types for **INSERT** and **UPDATE**. There are implicit conversions to `LONG BINARY` data type from `TINYINT`, `SMALLINT`, `INTEGER`, `UNSIGNED INTEGER`, `BIGINT`, `UNSIGNED BIGINT`, `CHAR`, and `VARCHAR` data types. There are no implicit conversions from `BIT`,

Large Object Data Types

REAL, DOUBLE, or NUMERIC data types to LONG BINARY data type. Implicit conversion can be controlled using the CONVERSION_MODE database option.

The currently supported byte substring functions for the LONG BINARY data type are accepted as input for implicit conversion for the **INSERT** and **UPDATE** statements.

The LONG BINARY data type can be explicitly converted to BINARY or VARBINARY. No other explicit data type conversions (for example, using the **CAST** or **CONVERT** function) exist either to or from the LONG BINARY data type.

Truncation of LONG BINARY data during conversion of LONG BINARY to BINARY or VARBINARY is handled the same way the truncation of BINARY and VARBINARY data is handled. If the STRING_RTRUNCATION option is ON, any right-truncation (of any values, not just non-space characters) on **INSERT** or **UPDATE** of a binary column results in a truncation error and the transaction is rolled back.

See also

- *Function Support* on page 91

Large Object Data Types LONG VARCHAR and CLOB

Character large object (CLOB) data in SAP Sybase IQ is stored in columns of data type LONG VARCHAR or CLOB.

An individual LONG VARCHAR data value can have a length ranging from zero (0) to 512TB (terabytes) for an IQ page size of 128KB, or 2PB (petabytes) for an IQ page size of 512KB. (The maximum length is equal to 4GB multiplied by the database page size.) To accommodate a table with LONG VARCHAR data, an IQ database must be created with an IQ page size of at least 64KB (65536 bytes).

A table or database can contain any number of LONG VARCHAR columns up to the supported maximum columns per table and maximum columns per database, respectively.

SAP Sybase IQ supports both single-byte and multibyte LONG VARCHAR data.

LONG VARCHAR columns can be either NULL or NOT NULL, and can store zero-length values. The domain CLOB is a LONG VARCHAR data type that allows NULL. To create a non-null LONG VARCHAR column, explicitly specify NOT NULL in the column definition.

You can create a LONG VARCHAR column using the domain CLOB, when you create a table or add a column to an existing table. For example:

```
CREATE TABLE lvtab (c1 INTEGER, c2 CLOB,  
                   c3 CLOB NOT NULL);
```

```
ALTER TABLE lvtab ADD c4 CLOB;
```

You can create a **WORD (WD)** index on a `LONG VARCHAR` column. You cannot construct other non-**FP** index types and join indexes on a `LONG VARCHAR` column.

You can modify a `LONG VARCHAR` column using the **UPDATE**, **INSERT...VALUES**, **INSERT...SELECT**, **LOAD TABLE**, **DELETE**, **TRUNCATE**, **SELECT...INTO** and **INSERT...LOCATION** SQL statements. Positioned updates and deletes are not supported on `LONG VARCHAR` columns.

You can insert an Adaptive Server `TEXT` column into a `LONG VARCHAR` column using the **INSERT...LOCATION** command. All `TEXT` data inserted is silently right-truncated at 2147483648 bytes (2GB).

LONG VARCHAR Data Type Conversion

There are limited implicit data type conversions to and from the `LONG VARCHAR` data type and non-`LONG VARCHAR` data types.

There are no implicit data type conversions from the `LONG VARCHAR` data type to another non-`LONG VARCHAR` data type, except `LONG BINARY`, and `CHAR` and `VARCHAR` for **INSERT** and **UPDATE** only. There are implicit conversions to `LONG VARCHAR` data type from `CHAR` and `VARCHAR` data types. There are no implicit conversions from `BIT`, `REAL`, `DOUBLE`, `NUMERIC`, `TINYINT`, `SMALLINT`, `INT`, `UNSIGNED INT`, `BIGINT`, `UNSIGNED BIGINT`, `BINARY`, `VARBINARY`, or `LONG BINARY` data types to `LONG VARCHAR` data type. Implicit conversion can be controlled using the `CONVERSION_MODE` database option.

The currently supported string functions for the `LONG VARCHAR` data type are accepted as input for implicit conversion for the **INSERT** and **UPDATE** statements.

The `LONG VARCHAR` data type can be explicitly converted to `CHAR` and `VARCHAR`. No other explicit data type conversions (for example, using the **CAST** or **CONVERT** function) exist either to or from the `LONG VARCHAR` data type.

Truncation of `LONG VARCHAR` data during conversion of `LONG VARCHAR` to `CHAR` is handled the same way the truncation of `CHAR` data is handled. If the `STRING_RTRUNCATION` option is `ON` and string right-truncation of non-spaces occurs, a truncation error is reported and the transaction is rolled back. Trailing partial multibyte characters are replaced with spaces on conversion.

Truncation of `LONG VARCHAR` data during conversion of `LONG VARCHAR` to `VARCHAR` is handled the same way the truncation of `VARCHAR` data is handled. If the `STRING_RTRUNCATION` option is `ON` and string right-truncation of non-spaces occurs, a truncation error is reported and the transaction is rolled back. Trailing partial multibyte characters are truncated on conversion.

See also

- *Function Support* on page 91

Large Object Variables

SAP Sybase IQ supports large object variables.

Inbound `LONG BINARY` and `LONG VARCHAR` variables (host variables or SQL variables used by IQ) have no maximum length.

Outbound `LONG BINARY` and `LONG VARCHAR` variables (variables set by IQ) have a maximum length of 2GB - 1.

The **LOAD TABLE**, **INSERT...VALUES**, **INSERT...SELECT**, **INSERT...LOCATION**, **SELECT...INTO**, and **UPDATE SQL** statements accept `LONG BINARY` and `LONG VARCHAR` variables of any size of data. Currently, a SQL variable can hold up to 2GB - 1 in length.

The **BIT_LENGTH**, **BYTE_LENGTH**, **BYTE_LENGTH64**, **BYTE_SUBSTR**, **BYTE_SUBSTR64**, **CHARINDEX**, **LOCATE**, **OCTET_LENGTH**, and **SUBSTRING64** functions support `LONG BINARY` and `LONG VARCHAR` variables of any size data that the SQL variable can hold. In addition, the **CHAR_LENGTH**, **CHAR_LENGTH64**, **PATINDEX**, **SUBSTR**, and **SUBSTRING** functions support `LONG VARCHAR` variables of any size data that the SQL variable can hold.

Large Object Variable Data Type Conversion

The database option `ENABLE_LOB_VARIABLES` controls the data type conversion of large object variables.

ENABLE LOB VARIABLES Option

Controls the data type conversion of large object variables.

Allowed Values

ON, OFF

Default

OFF

Scope

Option can be set at the database (PUBLIC) or user level. When set at the database level, the value becomes the default for any new user, but has no impact on existing users. When set at the user level, overrides the PUBLIC value for that user only. No system privilege is required to set option for self. System privilege is required to set at database level or at user level for any user other than self.

Requires the SET ANY PUBLIC OPTION system privilege to set this option. Can be set temporary for an individual connection or for the PUBLIC role. Takes effect immediately.

Remarks

ENABLE_LOB_VARIABLES controls the data type conversion of large object variables.

When **ENABLE_LOB_VARIABLES** is OFF, large object variables less than 32K are implicitly converted; an error is reported if a large object variable is greater than or equal to 32K. A `LONG VARCHAR` variable is implicitly converted to a `VARCHAR` data type and truncated at 32K. A `LONG BINARY` variable is implicitly converted to a `VARBINARY` data type and truncated at 32K.

When **ENABLE_LOB_VARIABLES** is ON, large object variables of any size retain their original data type and size.

Example

Retain the data type and size of large object variables greater than 32K:

```
SET TEMPORARY OPTION ENABLE_LOB_VARIABLES = ON
```

Index Support of Large Object Columns

SAP Sybase IQ supports the **TEXT** index on `LONG BINARY` and `LONG VARCHAR` columns, and the **WORD (WD)** index on `LONG VARCHAR` columns.

TEXT Index Support of Large Object Columns

The **TEXT** index supports `LONG BINARY` and `LONG VARCHAR` columns.

See also

- *SQL Statement Support* on page 81
- *TEXT Indexes and Text Configuration Objects* on page 3

WD Index Support of LONG VARCHAR (CLOB) Columns

SAP Sybase IQ offers limited support for the **WORD (WD)** index on `LONG VARCHAR (CLOB)` columns.

- The widest column supported by the **WD** index is the maximum width for a LOB column. (The maximum length is equal to 4GB multiplied by the database page size.) The maximum word length supported by SAP Sybase IQ is 255 bytes.
- All **sp_iqcheckdb** options for **WD** indexes over `CHAR` and `VARCHAR` columns are also supported for `LONG VARCHAR (CLOB)` columns, including allocation, check, and verify modes.
- You can use **sp_iqrebuildindex** to rebuild a **WD** index over a `LONG VARCHAR (CLOB)` column.

Large Object Data Types

Chinese text or documents in a binary format require ETL preprocessing to locate and transform words into a form that can be parsed by the **WD** index.

SQL Statement Support

Learn about the SQL statements and syntax that support working with **TEXT** indexes and text configurations.

ALTER TEXT CONFIGURATION Statement

Alters a text configuration object.

Note: This statement requires the Unstructured Data Analytics (IQ_UDA) license.

Quick Links:

Go to Parameters on page 81

Go to Examples on page 82

Go to Usage on page 83

Go to Permissions on page 83

Syntax

```
ALTER TEXT CONFIGURATION [ owner.] config-name
  STOPLIST stoplist
  | DROP STOPLIST
  | { MINIMUM | MAXIMUM } TERM LENGTH integer
  | TERM BREAKER
  | { GENERIC
      | [ EXTERNAL NAME library-and-entry-point-name-string ]
      | NGRAM }
  | PREFILTER EXTERNAL NAME library-and-entry-point-name-string
  | DROP PREFILTER
```

Parameters

(back to top) on page 81

- **stoplist** – a string-expression used to create or replace the list of terms to ignore when building a TEXT index. Terms specified in this list are also ignored in a query. Separate stoplist terms with spaces.

Stoplist terms cannot contain whitespace and should not contain non-alphanumeric characters. Non-alphanumeric characters are interpreted as spaces and break the term into multiple terms. For example, “and/or” is interpreted as the two terms “and” and “or”. The maximum number of stoplist terms is 7999.

- **DROP STOPLIST** – use to drop the stoplist for a text configuration object.

- **MINIMUM TERMLENGTH** – specifies the minimum length, in characters, of a term to include in the TEXT index. The value specified in the MINIMUM TERM LENGTH clause is ignored when using NGRAM TEXT indexes.

Terms that are shorter than this setting are ignored when building or refreshing the TEXT index. The value of this option must be greater than 0. If you set this option to be higher than MAXIMUM TERM LENGTH, the value of MAXIMUM TERM LENGTH is automatically adjusted to be the same as the new MINIMUM TERM LENGTH value.

- **MAXIMUM TERM LENGTH** – with GENERIC TEXT indexes, specifies the maximum length, in characters, of a term to include in the TEXT index. Terms that are longer than this setting are ignored when building or refreshing the TEXT index.

The value of MAXIMUM TERMLENGTH must be less than or equal to 60. If you set this option to be lower than MINIMUM TERM LENGTH, the value of MINIMUM TERM LENGTH is automatically adjusted to be the same as the new MAXIMUM TERM LENGTH value.

- **TERM BREAKER** – specifies the name of the algorithm to use for separating column values into terms. The choices for IN SYSTEM tables are GENERIC (the default) or NGRAM. The GENERIC algorithm treats any string of one or more alphanumerics, separated by non-alphanumerics, as a term.

The NGRAM algorithm breaks strings into n-grams. An n-gram is an n-character substring of a larger string. The NGRAM term breaker is required for fuzzy (approximate) matching, or for documents that do not use whitespace or non-alphanumeric characters to separate terms. NGRAM is supported for IN SYSTEM tables.

NGRAM term breaker is built on TEXT indexes, so use text configuration object settings to define whether to use an NGRAM or GENERIC TEXT index.

TERM BREAKER can include the specification for the external term breaker library using EXTERNAL NAME and the library entry point.

- **library-and-entry-point-name-string** – *[operating-system:]function-name@library*
- **PREFILTER EXTERNAL NAME** – specifies the entry_point and the library name of the external pre-filter library provided by external vendors.
- **DROP PREFILTER** – drops the external prefilter and sets NULL to the prefilter columns in ISYSTEXTCONFIG table.

Examples

(back to top) on page 81

- **Example 1** – creates a text configuration object, maxTerm16, and then change the maximum term length to 16:


```
CREATE TEXT CONFIGURATION maxTerm16 FROM default_char;
ALTER TEXT CONFIGURATION maxTerm16 MAXIMUM TERM LENGTH 16;
```

- **Example 2** – adds stoplist terms to the maxTerm16 configuration object:

```
ALTER TEXT CONFIGURATION maxTerm16
STOPLIST 'because about therefore only';
```

- **Example 3** – updates the text configuration object, my_text_config, to use the entry point my_term_breaker in the external library mytermbreaker.dll for breaking the text:

```
CREATE TEXT CONFIGURATION my_text_config FROM default_char;
ALTER TEXT CONFIGURATION my_text_config
TERM BREAKER GENERIC EXTERNAL NAME
'platform:my_term_breaker@mytermbreaker';
```

- **Example 4** – updates the text configuration object, my_text_config, to use the entry point my_prefilter in the external library myprefilter.dll for prefiltering the documents:

```
ALTER TEXT CONFIGURATION my_text_config
PREFILTER EXTERNAL NAME 'platform:my_prefilter@myprefilter';
```

Usage

(back to top) on page 81

TEXT indexes are dependent on a text configuration object. SAP Sybase IQ TEXT indexes use immediate refresh, and cannot be truncated; you must drop the indexes before you can alter the text configuration object. To view the settings for text configuration objects, query the SYSTEXTCONFIG system view.

Side Effects:

- Automatic commit.

Permissions

(back to top) on page 81

TERM BREAKER or **PREFILTER EXTERNAL NAME** clause – Requires the CREATE ANY EXTERNAL REFERENCE system privilege, along with one of:

- ALTER ANY TEXT CONFIGURATION system privilege.
- ALTER ANY OBJECT system privilege.
- You own the text configuration object.

All other clauses – Requires the ALTER ANY TEXT CONFIGURATION system privilege, regardless of whether the user is the owner of the configuration object.

ALTER TEXT INDEX Statement

Renames, moves or alters the definition of a TEXT index.

Note: This statement requires the Unstructured Data Analytics (IQ_UDA) license.

Quick Links:

Go to Parameters on page 84

Go to Examples on page 84

Go to Usage on page 85

Go to Permissions on page 85

Syntax

```
ALTER TEXT INDEX [owner.] text-index-name
  ON [owner.] table-name
  alter-clause
```

alter-clause - (back to Syntax)
rename-object | **move-object**

rename-object - (back to alter-clause)
RENAME { AS | TO } new-name

move-object - (back to alter-clause)
MOVE TO dbspace-name

Parameters

(back to top) on page 84

- **RENAME** – rename the TEXT index.
- **MOVE** – move the TEXT index to the specified dbspace.

Examples

(back to top) on page 84

- **Example** – creates a TEXT index, MyTextIndex, defining it as IMMEDIATE REFRESH, rename the TEXT index to Text_index_daily, and move the TEXT index to a dbspace named tispac:

```
CREATE TEXT INDEX MyTextIndex ON Customers ( CompanyName )
IMMEDIATE REFRESH;
ALTER TEXT INDEX MyTextIndex ON Customers RENAME AS
```

```
Text_index_daily;
ALTER TEXT INDEX Text_Index_Daily ON Customers MOVE TO tspace;
```

Usage

(back to top) on page 84

Side Effects:

- Automatic commit.

Permissions

(back to top) on page 84

move-object clause – Requires one of:

- ALTER ANY INDEX system privilege.
- ALTER ANY OBJECT system privilege.
- REFERENCES privilege the underlying table.
- You own the underlying table.

rename-object clause – Requires one of:

- ALTER ANY INDEX system privilege.
- ALTER ANY OBJECT system privilege.
- MANAGE ANY DBSPACE.
- One of the following:
 - You own the underlying table being indexed.
 - REFERENCES privilege on the table along with one of the following:
 - CREATE ANY OBJECT system privilege.
 - CREATE privilege on the target dbspace.

CREATE TEXT CONFIGURATION Statement

Creates a text configuration object.

Note: This statement requires the Unstructured Data Analytics (IQ_UDA) license.

Quick Links:

Go to Parameters on page 86

Go to Examples on page 86

Go to Usage on page 86

Go to Permissions on page 86

Syntax

```
CREATE TEXT CONFIGURATION [ owner.] new-config-name  
FROM [ owner.] existing-config-name
```

Parameters

(back to top) on page 85

- **FROM** – specifies the name of a text configuration object to use as the template for creating the new text configuration object. The names of the default text configuration objects are DEFAULT_CHAR and DEFAULT_NCHAR. DEFAULT_CHAR is supported for SAP Sybase IQ tables only; DEFAULT_NCHAR is supported on SQL Anywhere tables only.

Examples

(back to top) on page 85

- **Example 1** – creates a text configuration object, max_term_sixteen, using the default_char text configuration object, then use **ALTER TEXT CONFIGURATION** to change the maximum term length for max_term_sixteen to 16:

```
CREATE TEXT CONFIGURATION max_term_sixteen FROM default_char;
```

```
ALTER TEXT CONFIGURATION max_term_sixteen MAXIMUM TERM LENGTH 16;
```

Usage

(back to top) on page 85

Create a text configuration object using another text configuration object as a template, then alter the options as needed using the **ALTER TEXT CONFIGURATION** statement.

To view the list of all text configuration objects and their settings in the database, query the SYSTEXTCONFIG system view.

Side Effects:

- Automatic commit.

Permissions

(back to top) on page 85

Text configuration object to be owned by self –

- Requires CREATE TEXT CONFIGURATION system privilege.

Text configuration object to be owned by any user – Requires one of:

- CREATE ANY TEXT CONFIGURATION system privilege.

- CREATE ANY OBJECT system privilege.

All text configuration objects have PUBLIC access. Any user with privilege to create a **TEXT** index can use any text configuration object.

CREATE TEXT INDEX Statement

Creates a TEXT index and specifies the text configuration object to use.

Note: This statement requires the Unstructured Data Analytics (IQ_UDA) license.

Quick Links:

Go to Parameters on page 87

Go to Examples on page 87

Go to Usage on page 88

Go to Permissions on page 88

Syntax

```
CREATE TEXT INDEX text-index-name
  ON [ owner.] table-name( column-name, ...)
  [ IN dbspace-name ]
  [ CONFIGURATION [ owner.] text-configuration-name ]
  [ IMMEDIATE REFRESH ]
```

Parameters

(back to top) on page 87

- **ON** – specifies the table and column on which to build the TEXT index.
- **IN** – specifies the dbspace in which the TEXT index is located. If this clause is not specified, then the TEXT index is created in the same dbspace as the underlying table.
- **CONFIGURATION** – specifies the text configuration object to use when creating the TEXT index. If this clause is not specified, the default_char text configuration object is used.
- **IMMEDIATE REFRESH** – (default) refreshes the TEXT index each time changes in the underlying table impact data in the TEXT index. Only permitted value for tables in SAP Sybase IQ main store. Once created, the IMMEDIATE REFRESH clause cannot be changed.

Examples

(back to top) on page 87

SQL Statement Support

- **Example 1** – creates a TEXT index, myTxtIdx, on the CompanyName column of the Customers table in the iqdemo database, using the max_term_sixteen text configuration object:

```
CREATE TEXT INDEX myTxtIdx ON Customers (CompanyName );  
  
CONFIGURATION max_term_sixteen;
```

Usage

(back to top) on page 87

You cannot create a TEXT index on views or temporary tables, or on an IN SYSTEM materialized view. The **BEGIN PARALLEL IQ...END PARALLEL IQ** statement does not support **CREATE TEXT INDEX**.

Side Effects:

- Automatic commit.

Permissions

(back to top) on page 87

Requires one of:

- CREATE ANY INDEX system privilege along with CREATE privilege on the dbspace where the index is being created.
- CREATE ANY OBJECT system privilege.

DROP TEXT CONFIGURATION Statement

Drops a text configuration object.

Note: This statement requires the Unstructured Data Analytics (IQ_UDA) license.

Quick Links:

Go to Examples on page 88

Go to Usage on page 89

Go to Permissions on page 89

Syntax

```
DROP TEXT CONFIGURATION [ owner. ] text-config-name
```

Examples

(back to top) on page 88

- **Example 1** – creates and drops the `mytextconfig` text configuration object:

```
CREATE TEXT CONFIGURATION mytextconfig FROM default_char;
DROP TEXT CONFIGURATION mytextconfig;
```

Usage

(back to top) on page 88

Use **DROP TEXT CONFIGURATION** to drop a text configuration object.

Attempting to drop a text configuration object with dependent TEXT indexes results in an error. You must drop the dependent TEXT indexes before dropping the text configuration object.

Text configuration objects are stored in the `ISYSTETEXTCONFIG` system table.

Side Effects:

- Automatic commit.

Permissions

(back to top) on page 88

Text configuration object owned by self – None required.

Table configuration object owned by any user – Requires one of:

- `DROP ANY TEXT CONFIGURATION` system privilege.
- `DROP ANY OBJECT` system privilege.

DROP TEXT INDEX Statement

Removes a TEXT index from the database.

Note: This statement requires the Unstructured Data Analytics (IQ_UDA) license.

Quick Links:

Go to Parameters on page 90

Go to Examples on page 90

Go to Usage on page 90

Go to Permissions on page 90

Syntax

```
DROP TEXT INDEX text-index-name
ON [ owner ] table-name
```

Parameters

(back to top) on page 89

- **ON** – specifies the table on which the TEXT index is built.

Examples

(back to top) on page 89

- **Example 1** – creates and drops the TextIdx TEXT index:

```
CREATE TEXT INDEX TextIdx ON Customers ( Street );  
DROP TEXT INDEX TextIdx ON Customers;
```

Usage

(back to top) on page 89

You must drop dependent TEXT indexes before you can drop a text configuration object.

Side Effects:

- Automatic commit.

Permissions

(back to top) on page 89

Requires one of:

- DROP ANY INDEX system privilege.
- DROP ANY OBJECT system privilege.
- REFERENCES privilege on the table being indexed.
- You own the underlying table.

Function Support

Learn about the functions that support the `LONG BINARY` and `LONG VARCHAR` data types.

Summary of Function Support of Large Object Data

A summary of function support of `LONG BINARY` (BLOB) and `LONG VARCHAR` (CLOB) data types and `LONG BINARY` and `LONG VARCHAR` variables.

In addition to the functions listed in this table, you can use the **BFIELD** function to extract LOB data.

Scalar and aggregate user-defined functions support large object data types as input parameters. . .

Table 18. Function Support of LOB Data Types and Variables

| Function | BLOB data supported? | BLOB variables supported? | CLOB data supported? | CLOB variables supported? |
|-------------------------------------|----------------------|---------------------------|----------------------|---------------------------|
| <code>BIT_LENGTH()</code> | Yes | Yes | Yes | Yes |
| <code>BYTE_LENGTH()</code> | Yes* | Yes* | Yes* | Yes* |
| <code>BYTE_LENGTH64()</code> | Yes | Yes | Yes | Yes |
| <code>BYTE_SUBSTR()</code> | Yes | Yes | Yes | Yes |
| <code>BYTE_SUBSTR64()</code> | Yes | Yes | Yes | Yes |
| <code>CHAR_LENGTH()</code> | No | No | Yes | Yes |
| <code>CHAR_LENGTH64()</code> | No | No | Yes | Yes |
| <code>CHARINDEX()</code> | Yes | Yes | Yes | Yes |
| <code>LOCATE()</code> | Yes | Yes | Yes | Yes |
| <code>OCTET_LENGTH()</code> | Yes | Yes | Yes | Yes |
| <code>PATINDEX()</code> | No | No | Yes | Yes |
| <code>SUBSTR() / SUBSTRING()</code> | No | No | Yes | Yes |
| <code>SUBSTRING64()</code> | Yes | Yes | Yes | Yes |

*The **BYTE_LENGTH** function supports both `LONG BINARY` columns and variables and `LONG VARCHAR` columns and variables, only if the query returns less than 2GB. If the byte

Function Support

length of the returned `LONG BINARY` or `LONG VARCHAR` data is greater than 2GB, `BYTE_LENGTH` returns an error that says you must use the `BYTE_LENGTH64` function.

See also

- *User-Defined Function Support of Large Object Columns* on page 101
- *Large Object Data Exports* on page 69

BIT_LENGTH Function

The `BIT_LENGTH` function returns an unsigned 64-bit value containing the bit length of the large object column or variable parameter. If the argument is `NULL`, `BIT_LENGTH` returns `NULL`.

Syntax

```
BIT_LENGTH( large-object-column )
```

Parameters

large-object-column – the name of a `LONG VARCHAR` or `LONG BINARY` column or variable.

Usage

`BIT_LENGTH` supports all SAP Sybase IQ data types and `LONG BINARY` and `LONG VARCHAR` variables of any size of data. Currently, a SQL variable can hold up to 2GB - 1 in length.

BYTE_LENGTH Function [String]

Returns the number of bytes in a string.

Syntax

```
BYTE_LENGTH ( string-expression )
```

Parameters

| Parameters | Description |
|--------------------------------|--|
| <code>string-expression</code> | The string whose length is to be calculated. |

Returns

INT

Remarks

Trailing white space characters are included in the length returned.

The return value of a NULL string is NULL.

If the string is in a multibyte character set, the **BYTE_LENGTH** value differs from the number of characters returned by **CHAR_LENGTH**.

If you are licensed to use the Unstructured Data Analytics functionality, you can use this function with large object data:

- The **BYTE_LENGTH** function supports both `LONG BINARY` columns and variables and `LONG VARCHAR` columns and variables, only if the query returns less than 2GB. If the byte length of the returned `LONG BINARY` or `LONG VARCHAR` data is greater than or equal to 2GB, **BYTE_LENGTH** returns an error that says you must use the **BYTE_LENGTH64** function.

Standards and Compatibility

- SQL—Vendor extension to ISO/ANSI SQL grammar.
- SAP Sybase—Not supported by Adaptive Server.

Example

Returns the value 12:

```
SELECT BYTE_LENGTH( 'Test Message' ) FROM iq_dummy
```

BYTE_LENGTH64 Function [String]

The **BYTE_LENGTH64** function returns an unsigned 64-bit value containing the byte length of the large object column or variable parameter.

Syntax

```
BYTE_LENGTH64( large-object-column )
```

Parameters

large-object-column – the name of a `LONG VARCHAR` or `LONG BINARY` column or variable.

Usage

The **BYTE_LENGTH64** function supports both `LONG BINARY` and `LONG VARCHAR` columns and `LONG BINARY` and `LONG VARCHAR` variables of any size of data. Currently, a SQL variable can hold up to 2GB - 1 in length.

BYTE_SUBSTR64 and BYTE_SUBSTR Functions [String]

The **BYTE_SUBSTR64** and **BYTE_SUBSTR** functions return the byte substring of the large object column or variable parameter.

Syntax

```
BYTE_SUBSTR64( large-object-column, start, length )
```

```
BYTE_SUBSTR( large-object-column, start, length )
```

Parameters

large-object-column – the name of a LONG VARCHAR or LONG BINARY column or variable.

start – an integer expression indicating the start of the substring. A positive integer starts from the beginning of the string, with the first byte at position 1. A negative integer specifies a substring starting from the end of the string, with the final byte at position -1.

length – an integer expression indicating the length of the substring. A positive length specifies the number of bytes to return, starting at the *start* position. A negative length specifies the number of bytes to return, ending at the *start* position.

Usage

- Nested operations of the functions **BYTE_LENGTH64**, **BYTE_SUBSTR64**, and **BYTE_SUBSTR** do not support large object columns or variables.
- The **BYTE_SUBSTR64** and **BYTE_SUBSTR** functions support both LONG BINARY and LONG VARCHAR columns and LONG BINARY and LONG VARCHAR variables of any size of data. Currently, a SQL variable can hold up to 2GB - 1 in length.

CHAR_LENGTH Function [String]

The **CHAR_LENGTH** function returns a signed 32-bit value containing the character length of the LONG VARCHAR column or variable parameter, including the trailing blanks.

Syntax

```
CHAR_LENGTH( long-varchar-object )
```

Parameters

- **long-varchar-object** – the name of a LONG VARCHAR column or LONG VARCHAR variable.

Usage

- **CHAR_LENGTH** supports LONG VARCHAR columns and LONG VARCHAR variables of any size of data. Currently, a SQL variable can hold up to 2GB - 1 in length.
- If the argument is NULL, **CHAR_LENGTH** returns NULL.
- If the character length exceeds 2GB - 1 (2147483647), an error is returned.

CHAR_LENGTH64 Function [String]

The **CHAR_LENGTH64** function returns an unsigned 64-bit value containing the character length of the LONG VARCHAR column or variable parameter, including the trailing blanks.

Syntax

```
CHAR_LENGTH64( long-varchar-object )
```

Parameters

long-varchar-object – the name of a LONG VARCHAR column in a table or a LONG VARCHAR variable.

Usage

- **CHAR_LENGTH64** supports LONG VARCHAR columns and LONG VARCHAR variables of any size of data. Currently, a SQL variable can hold up to 2GB - 1 in length.
- If the argument is NULL, **CHAR_LENGTH64** returns NULL.

CHARINDEX Function [String]

The **CHARINDEX** function returns a 64-bit signed integer containing the position of the first occurrence of the specified string in the large object column or variable parameter. For CHAR and VARCHAR columns, **CHARINDEX** returns a 32-bit signed integer position.

Syntax

```
CHARINDEX( string-expression, large-object-column )
```

Parameters

string-expression – the string of up to 255 bytes, for which you are searching.

large-object-column – the name of the LONG VARCHAR or LONG BINARY column or variable.

Usage

- All the positions or offsets, returned or specified, in the **CHARINDEX** function are always character offsets and may be different from the byte offset for multibyte data.
- If the large object cell being searched contains more than one instance of *string-expression*, **CHARINDEX** returns only the position of the first instance.
- If the column does not contain the string, the **CHARINDEX** function returns zero (0).
- Searching for a string longer than 255 bytes returns NULL.
- Searching for a zero-length string returns 1.
- If any of the arguments is NULL, the result is NULL.
- **CHARINDEX** supports searching LONG VARCHAR and LONG BINARY columns and LONG VARCHAR and LONG BINARY variables of any size of data. Currently, a SQL variable can hold up to 2GB - 1 in length.

LOCATE Function [String]

The **LOCATE** function returns a 64-bit signed integer containing the position of the specified string in the large object column or variable parameter. For CHAR and VARCHAR columns, **LOCATE** returns a 32-bit signed integer position.

Syntax

```
LOCATE( large-object-column, string-expression  
[, numeric-expression ] )
```

Parameters

large-object-column – the name of the LONG VARCHAR or LONG BINARY column or variable to search.

string-expression – the string of up to 255 bytes, for which you are searching.

numeric-expression – the character position or offset at which to begin the search in the string. The *numeric-expression* is a 64-bit signed integer for LONG VARCHAR and LONG BINARY columns and is a 32-bit signed integer for CHAR, VARCHAR, and BINARY columns. The first character is position 1. If the starting offset is negative, **LOCATE** returns the last matching string offset, rather than the first. A negative offset indicates how much of the end of the string to exclude from the search. The number of characters excluded is calculated as $(-1 * \text{offset}) - 1$.

Usage

- All the positions or offsets, returned or specified, in the **LOCATE** function are always character offsets and may be different from the byte offset for multibyte data.
- If the large object cell being searched contains more than one instance of the string:

- If *numeric-expression* is specified, **LOCATE** starts the search at that offset in the string.
- If *numeric-expression* is not specified, **LOCATE** returns only the position of the first instance.
- If the column does not contain the string, **LOCATE** returns zero (0).
- Searching for a string longer than 255 bytes returns NULL.
- Searching for a zero-length string returns 1.
- If any of the arguments is NULL, the result is NULL.
- **LOCATE** supports searching LONG VARCHAR and LONG BINARY columns and LONG VARCHAR and LONG BINARY variables of any size of data. Currently, a SQL variable can hold up to 2GB - 1 in length.

OCTET_LENGTH Function [String]

The **OCTET_LENGTH** function returns an unsigned 64-bit value containing the byte length of the large object column or variable parameter.

Syntax

```
OCTET_LENGTH ( column-name )
```

Parameters

large-object-column – the name of a LONG VARCHAR or LONG BINARY column or variable.

Usage

- If the argument is NULL, **OCTET_LENGTH** returns NULL.
- **OCTET_LENGTH** supports all SAP Sybase IQ data types and LONG VARCHAR and LONG BINARY variables of any size of data. Currently, a SQL variable can hold up to 2GB - 1 in length.

PATINDEX Function [String]

The **PATINDEX** function returns a 64-bit unsigned integer containing the position of the first occurrence of the specified pattern in a LONG VARCHAR column or variable. For CHAR and VARCHAR columns, **PATINDEX** returns a 32-bit unsigned integer position.

Syntax

```
PATINDEX ( '%pattern%' , long-varchar-column )
```

Parameters

- **pattern** – the pattern for which you are searching. This string is limited to 126 bytes for patterns with wildcards. If you omit the leading percent wildcard, **PATINDEX** returns one

(1) if the pattern occurs at the beginning of the column value, and zero (0) if the pattern does not occur at the beginning of the column value. Similarly, if you omit the trailing percent wildcard, the pattern should occur at the end of the column value. The pattern uses the same wildcards as the **LIKE** comparison.

Patterns without wildcards—percent (%) and underscore (—) can be up to 255 bytes in length.

- **long-varchar-column** – the name of the `LONG VARCHAR` column or variable.

Usage

- All the positions or offsets, returned or specified, in the **PATINDEX** function are always character offsets and may be different from the byte offset for multibyte data.
- If the `LONG VARCHAR` cell being searched contains more than one instance of the string pattern, **PATINDEX** returns only the position of the first instance.
- If the column does not contain the pattern, **PATINDEX** returns zero (0).
- Searching for a pattern longer than 126 bytes returns `NULL`.
- Searching for a zero-length pattern returns 1.
- If any of the arguments is `NULL`, the result is zero (0).
- **PATINDEX** supports `LONG VARCHAR` variables of any size of data. Currently, a `SQL` variable can hold up to 2GB - 1 in length. **PATINDEX** does not support `LONG BINARY` variables or searching `LONG BINARY` columns.

SUBSTRING Function [String]

The **SUBSTRING** function returns a variable-length character string of the `LONG VARCHAR` column or variable parameter. If any of the arguments are `NULL`, **SUBSTRING** returns `NULL`.

Syntax

```
{ SUBSTRING | SUBSTR } ( long-varchar-column, start [, length ] )
```

Parameters

long-varchar-column – the name of a `LONG VARCHAR` column or variable.

start – an integer expression indicating the start of the substring. A positive integer starts from the beginning of the string, with the first character at position 1. A negative integer specifies a substring starting from the end of the string, with the final character at position -1.

length – an integer expression indicating the character length of the substring. A positive length specifies the number of characters to return, starting at the *start* position. A negative length specifies the number of characters to return, ending at the *start* position.

Usage

SUBSTRING supports `LONG VARCHAR` variables of any size of data. Currently, a `SQL` variable can hold up to 2GB - 1 in length. **SUBSTRING** does not support `LONG BINARY` variables or searching `LONG BINARY` columns.

When the **ansi_substring** database option is set to ON (default), negative values are invalid.

ANSI_SUBSTRING Option [TSQL]

Controls the behavior of the `SUBSTRING` (`SUBSTR`) function when negative values are provided for the start or length parameters.

Allowed Values

ON, OFF

Default

ON

Scope

Option can be set at the database (`PUBLIC`) or user level. When set at the database level, the value becomes the default for any new user, but has no impact on existing users. When set at the user level, overrides the `PUBLIC` value for that user only. No system privilege is required to set option for self. System privilege is required to set at database level or at user level for any user other than self.

Requires the `SET ANY PUBLIC OPTION` system privilege to set this option. Can be set temporary for an individual connection or for the `PUBLIC` role. Takes effect immediately.

Remarks

When the **ANSI_SUBSTRING** option is set to ON, the behavior of the **SUBSTRING** function corresponds to ANSI/ISO SQL/2003 behavior. A negative or zero start offset is treated as if the string were padded on the left with noncharacters, and gives an error if a negative length is provided.

When this option is set to OFF, the behavior of the **SUBSTRING** function is the same as in earlier versions of SAP Sybase IQ: a negative start offset means an offset from the end of the string, and a negative length means the desired substring ends length characters to the left of the starting offset. Using a start offset of 0 is equivalent to a start offset of 1.

Avoid using nonpositive start offsets or negative lengths with the **SUBSTRING** function. Where possible, use the **LEFT** or **RIGHT** functions instead.

Example

These examples show the difference in the values returned by the **SUBSTRING** function based on the setting of the **ANSI_SUBSTRING** option:

```
SUBSTRING( 'abcdefgh',-2,4 );
ansi_substring = Off ==> 'gh'
// substring starts at second-last character
ansi_substring = On ==> 'gh'
// takes the first 4 characters of
// ???abcdefgh and discards all ?

SUBSTRING( 'abcdefgh',4,-2 );
ansi_substring = Off ==> 'cd'
ansi_substring = On ==> value -2 out of range
for destination

SUBSTRING( 'abcdefgh',0,4 );
ansi_substring = Off ==> 'abcd'
ansi_substring = On ==> 'abcd'
```

SUBSTRING64 Function [String]

The **SUBSTRING64** function returns a variable-length character string of the large object column or variable parameter.

Syntax

```
SUBSTRING64 ( large-object-column, start [, length ] )
```

Parameters

large-object-column – the name of a LONG VARCHAR or LONG BINARY column or variable.

start – an 8-byte integer indicating the start of the substring. **SUBSTRING64** interprets a negative or zero *start* offset as if the string were padded on the left with "non-characters." The first character starts at position 1.

length – an 8-byte integer indicating the length of the substring. If *length* is negative, an error is returned.

Example

Values returned by **SUBSTRING64**, given a column named `col1` that contains the string ("ABCDEFGH"):

```
SUBSTRING64( col1, 2, 4 ) returns the string "BCDE"
```

```
SUBSTRING64( col1, 1, 3 ) returns the string "ABC"
```

```
SUBSTRING64( col1, 0, 3 ) returns the string "AB"
```

```
SUBSTRING64( col1, -1, 3 ) returns the string "A"
```

Usage

- If any of the arguments are NULL, **SUBSTRING64** returns NULL.

- Nested operations of the functions **SUBSTRING64**, **SUBSTRING**, **SUBSTR**, **BYTE_SUBSTR**, and **BYTE_SUBSTR64** do not support large object columns or variables.
- **SUBSTRING64** supports searching `LONG VARCHAR` and `LONG BINARY` columns and `LONG VARCHAR` and `LONG BINARY` variables of any size of data. Currently, a SQL variable can hold up to 2GB - 1 in length.

Aggregate Function Support of Large Object Columns

Only the aggregate function **COUNT (*)** is supported for `LONG BINARY` and `LONG VARCHAR` columns.

The **COUNT DISTINCT** parameter is not supported. An error is returned if a `LONG BINARY` or `LONG VARCHAR` column is used with the **MIN**, **MAX**, **AVG**, or **SUM** aggregate functions.

User-Defined Function Support of Large Object Columns

Scalar and aggregate user-defined functions support large object data types `LONG VARCHAR` (CLOB) and `LONG BINARY` (BLOB) up to 4GB (gigabytes) as input parameters. LOB data types are not supported as output parameters.

You must be specifically licensed to use the User-defined Function support functionality. .

Index

A

- Adaptive Server
 - inserting IMAGE data 75
 - inserting TEXT data 77
- adding
 - text configuration object 9
 - TEXT index 4, 5
- ALTER TEXT CONFIGURATION 13, 14
 - syntax 81
- ALTER TEXT INDEX 6
 - syntax 84
- altering
 - text configuration object 13, 81
 - TEXT index 6, 84
- analytics
 - Unstructured Data Analytics Option 1
- ANSI_SUBSTRING option 99

B

- BEGIN PARALLEL IQ statement 87
- BFILE function 69
 - example 70
 - extraction example 70
 - extraction facility 69
 - syntax 69
- binary large object
 - aggregate function support 101
 - BIT_LENGTH function 92
 - BLOB 75
 - BYTE_LENGTH64 function 93
 - BYTE_SUBSTR function 94
 - BYTE_SUBSTR64 function 94
 - columns 75
 - data type 75
 - data type conversions 75
 - description 75
 - in queries 23, 57
 - index support 79
 - indexes 75, 79
 - inserting IMAGE data 75
 - LONG BINARY 75
 - modifying 75
 - monitoring performance 58
 - OCTET_LENGTH function 97
 - prefetching 75
 - size 75
 - sp_iqindexsize 67
 - stored procedure support 59
 - SUBSTRING64 function 100
 - TEXT index 79
 - updating data 75
 - user-defined function support 101
- binary large object variable
 - data type conversions 78
- BIT_LENGTH function
 - description 92
 - syntax 92
- BLOB
 - aggregate function support 101
 - binary large object 75
 - BIT_LENGTH function 92
 - BYTE_LENGTH64 function 93
 - BYTE_SUBSTR function 94
 - BYTE_SUBSTR64 function 94
 - CHARINDEX function 95
 - columns 75
 - data type 75
 - data type conversions 75
 - description 75
 - exporting data 69
 - function support 91
 - in queries 23, 57
 - index support 79
 - indexes 75, 79
 - inserting data 75
 - inserting IMAGE data 75
 - loading data 71
 - LOCATE function 96
 - LONG BINARY 75
 - modifying 75
 - monitoring performance 58
 - OCTET_LENGTH function 97
 - prefetching 75
 - size 75
 - sp_iqindexsize 67
 - stored procedure support 59
 - SUBSTRING64 function 100
 - TEXT index 79
 - updating data 75
 - user-defined function support 101

Index

- variable function support 91
- variables 78
- BLOB variable
 - data type conversions 78
- breaking
 - terms 59, 60
- BYTE_LENGTH function 92
- BYTE_LENGTH64 function
 - description 93
 - syntax 93
- BYTE_SUBSTR function
 - description 94
 - syntax 94
- BYTE_SUBSTR64 function
 - description 94
 - syntax 94

C

- char
 - breaking into terms¶ 59
- CHAR_LENGTH function
 - description 94
 - syntax 94
- CHAR_LENGTH64 function
 - description 95
 - syntax 95
- character large object
 - aggregate function support 101
 - BIT_LENGTH function 92
 - BYTE_LENGTH64 function 93
 - BYTE_SUBSTR function 94
 - BYTE_SUBSTR64 function 94
 - CHAR_LENGTH function 94
 - CHAR_LENGTH64 function 95
 - CHARINDEX function 95
 - CLOB 76
 - columns 76
 - data type 76
 - data type conversions 77
 - description 75
 - in queries 23, 58
 - index support 75, 79
 - indexes 77, 79
 - inserting TEXT data 77
 - LOCATE function 96
 - LONG VARCHAR 76
 - modifying 77
 - OCTET_LENGTH function 97
 - PATINDEX function 97

- size 76
- sp_iqindexsize 68
- stored procedure support 59
- SUBSTRING function 98
- SUBSTRING64 function 100
- TEXT index 79
- variables 78
- WD index 77, 79
- WORD index 77, 79
- character large object variable
 - data type conversions 78
- CHARINDEX function
 - description 95
 - syntax 95
- CLOB
 - aggregate function support 101
 - BIT_LENGTH function 92
 - BYTE_LENGTH64 function 93
 - BYTE_SUBSTR function 94
 - BYTE_SUBSTR64 function 94
 - CHAR_LENGTH function 94
 - CHAR_LENGTH64 function 95
 - character large object 76
 - CHARINDEX function 95
 - columns 76
 - data type 76
 - data type conversions 77
 - description 75
 - exporting data 69
 - function support 91
 - in queries 23, 58
 - index support 75, 79
 - indexes 77, 79
 - inserting data 77
 - inserting TEXT data 77
 - loading data 71
 - LOCATE function 96
 - LONG VARCHAR 76
 - modifying 77
 - OCTET_LENGTH function 97
 - PATINDEX function 97
 - size 76
 - sp_iqindexsize 68
 - stored procedure support 59
 - SUBSTRING function 98
 - SUBSTRING64 function 100
 - TEXT index 79
 - updating data 77
 - user-defined function support 101

- variable function support 91
- variables 78
- WD index 77, 79
- WORD index 77, 79
- CLOB variable
 - data type conversions 78
- compatibility
 - with Adaptive Server 2
 - with ASE 2
 - with SA 2
 - with SQL Anywhere 2
- compression of LOB data 65
 - changing settings 65
 - displaying settings 66
- CONTAINS
 - table expressions 47, 54
- CONTAINS clause
 - determining close matches 54
- CONTAINS conditions
 - TEXT index 47
- CONTAINS examples
 - text configuration object 15
- contains-expression
 - FROM clause 47
- CREATE TEXT CONFIGURATION 9
 - syntax 85
- CREATE TEXT INDEX 4, 5
 - syntax 87
- creating
 - text configuration object 9, 85
 - TEXT index 4, 5, 87

D

- data compression of LOB 65
 - changing settings 65
 - displaying settings 66
- data type
 - BLOB 75
 - CLOB 76
 - LONG BINARY 75
 - LONG VARCHAR 76
- data type conversion
 - LONG BINARY to BINARY 75
 - LONG BINARY to VARBINARY 75
 - LONG BINARY variables 78
 - LONG VARCHAR to CHAR 77
 - LONG VARCHAR to VARCHAR 77
- database option
 - ENABLE_LOB_VARIABLES 78

- MAX_PREFIX_PER_CONTAINS_PHRASE
 - 17
- TEXT_DELETE_METHOD 7
- dbspace
 - modifying 6
 - TEXT indexes 6
- default_char 9
- default_nchar 9
- deleting
 - text configuration object 14
 - TEXT index 7
- disabling
 - compression 65
 - external library 20
- DROP TEXT CONFIGURATION 14
 - syntax 88
- DROP TEXT INDEX 7
 - syntax 89
- dropping
 - text configuration object 14, 88
 - TEXT index 7, 89

E

- editing
 - text configuration object 13
 - TEXT indexes 6
- ENABLE_LOB_VARIABLES option 78
- enabling
 - compression 65
 - external library 20
- END PARALLEL IQ
 - CREATE TEXT INDEX 87
- examples
 - text configuration object 14, 15
- exporting
 - BFILE example 70
 - BFILE function 69
 - BLOB 69
 - CLOB 69
 - large object data 69
 - LOB 69
 - LONG BINARY 69
- exporting LONG VARCHAR 69
- external libraries 19
 - disabling 20
 - enabling 20
 - identifying 63
 - listing 64
 - multiplex servers 20

Index

- restrictions 20
 - unloading 21, 64
- extraction facility
- BFILE function 69

F

FROM clause

- CONTAINS 47, 54
- contains-expression 47
- syntax 47

- full text searching 2, 3, 23
- types 23

functions

- BFILE 69
- BFILE example 70
- BIT_LENGTH 92
- BYTE_LENGTH function 92
- BYTE_LENGTH64 93
- BYTE_SUBSTR 94
- BYTE_SUBSTR64 94
- CHAR_LENGTH 94
- CHAR_LENGTH64 95
- CHARINDEX 95
- for BLOB 91
- for BLOB variables 91
- for CLOB 91
- for CLOB variables 91
- for LOB 91
- LOCATE 96
- LONG BINARY aggregate support 101
- LONG BINARY user-defined function support 101
- LONG VARCHAR aggregate support 101
- LONG VARCHAR user-defined function support 101
- OCTET_LENGTH 97
- PATINDEX 97
- SUBSTRING 98
- SUBSTRING64 100

- fuzzy search 55

I

identifying

- external libraries 63, 64

IMAGE data

- inserting from ASE 75
- inserting into LONG BINARY 75

- immediate refresh 7, 61, 62

index

- listing 5

indexes

- binary large object 75, 79
- BLOB 75, 79
- character large object 77, 79
- CLOB 77, 79
- Containment 77, 79
- full text searching 2, 23
- large object data 79
- LOB 79
- LONG BINARY 75, 79
- LONG VARCHAR 77, 79
- refreshing 7
- TEXT 3, 79
- WD 77, 79
- WORD 77, 79

inserting

- BLOB 75
- CLOB 77
- large object data 75, 77
- LOB 75, 77
- LONG BINARY 75
- LONG VARCHAR 77

instances

- external libraries¶ 64

L

large object data

- exporting 69
- index support 79
- inserting 75, 77
- loading 71
- updating 75, 77

libraries, external 8, 9

licensing 1

list

- external libraries 64
- text configuration object 13
- TEXT index 5

LOAD TABLE

- example 72
- extended syntax 71
- primary load file 71
- secondary load file 71

loading

- BLOB 71
- CLOB 71

- controlling errors 72
- large object data 71
- LOAD TABLE example 72
- LOB 71
- LONG BINARY 71
- LONG VARCHAR 71
- SECONDARY_FILE_ERROR option 72
- stripping trailing blanks 73
- TRIM_PARTIAL_MBC option 73
- truncating character data 73
- LOB
 - exporting data 69
 - function support 91
 - index support 79
 - inserting data 75, 77
 - introduction 1
 - loading data 71
 - typical sources 1
 - updating data 75, 77
 - user-defined function support 101
- LOB compression
 - changing settings 65
 - disabling 65
 - displaying settings 66
 - enabling 65
- LOB variables
 - data type conversion 78
- LOCATE function
 - description 96
 - syntax 96
- LONG BINARY
 - aggregate function support 101
 - binary large object 75
 - BIT_LENGTH function 92
 - BLOB 75
 - BYTE_LENGTH64 function 93
 - BYTE_SUBSTR function 94
 - BYTE_SUBSTR64 function 94
 - CHARINDEX function 95
 - columns 75
 - data type conversions 75
 - DELETE 75
 - exporting data 69
 - in queries 23, 57
 - index support 79
 - indexes 75, 79
 - INSERT 75
 - inserting data 75
 - inserting IMAGE data 75
 - LOAD TABLE 75
 - loading data 71
 - LOCATE function 96
 - modifying 75
 - monitoring performance 58
 - OCTET_LENGTH function 97
 - SELECT...INTO 75
 - size 75
 - sp_iqindexsize 67
 - stored procedure support 59
 - SUBSTRING64 function 100
 - TEXT index 79
 - TRUNCATE 75
 - UPDATE 75
 - updating data 75
 - user-defined function support 101
 - variables 78
- LONG BINARY variable
 - data type conversions 78
- LONG VARCHAR
 - aggregate function support 101
 - BIT_LENGTH function 92
 - BYTE_LENGTH64 function 93
 - BYTE_SUBSTR function 94
 - BYTE_SUBSTR64 function 94
 - CHAR_LENGTH function 94
 - CHAR_LENGTH64 function 95
 - character large object 76
 - CHARINDEX function 95
 - CLOB 76
 - columns 76
 - data type conversions 77
 - DELETE 77
 - exporting data 69
 - in queries 23, 58
 - index support 79
 - indexes 77, 79
 - INSERT 77
 - inserting data 77
 - inserting TEXT data 77
 - LOAD TABLE 77
 - loading data 71
 - LOCATE function 96
 - modifying 77
 - OCTET_LENGTH function 97
 - PATINDEX function 97
 - SELECT...INTO 77
 - size 76
 - sp_iqindexsize 68

Index

- stored procedure support 59
- SUBSTRING function 98
- SUBSTRING64 function 100
- TEXT index 79
- TRUNCATE 77
- UPDATE 77
- updating data 77
- user-defined function support 101
- variables 78
- WD index 77, 79
- WORD index 77, 79
- LONG VARCHAR variable
 - data type conversions 78

M

- MAX_PREFIX_PER_CONTAINS_PHRASE
 - option 17
- modifying
 - dbspace 6
 - stoplist 14
- multibyte characters
 - TRIM_PARTIAL_MBC option 73
 - trimming partial 73
 - truncating on load 73
- multiplex servers
 - external libraries 20

N

- nchar
 - breaking into terms¶ 60
- NGRAM
 - TEXT index search 55, 56
- NGRAM TEXT index 8
 - fuzzy search 8
- non-fuzzy search 55, 56

O

- OCTET_LENGTH function
 - description 97
 - syntax 97
- option
 - ENABLE_LOB_VARIABLES 78
 - MAX_PREFIX_PER_CONTAINS_PHRASE 17
 - TEXT_DELETE_METHOD 7
 - Unstructured Data Analytics 1

P

- PATINDEX function
 - description 97
 - syntax 97
- performance monitor
 - binary large object 58
 - BLOB 58
 - LONG BINARY 58
- prefetching 75
- prefilter library 8–10, 19
 - maximum term length 12
 - minimum term length 11
 - restrictions 20
- prefix
 - term limit 17

Q

- queries
 - binary large object 23, 57
 - BLOB 23, 57
 - character large object 23, 58
 - CLOB 23, 58
 - LONG BINARY 23, 57
 - LONG VARCHAR 23, 58

R

- refresh
 - immediate¶ 61, 62
- refreshing
 - immediate 7
 - TEXT index 7
- restrictions
 - external library 20
 - TEXT index 5

S

- sa_char_terms stored procedure 59
- sa_external_library_unload stored procedure 21, 64
- sa_list_external_library stored procedure 20, 64
- sa_nchar_terms stored procedure 60
- sa_text_index_stats stored procedure 61
- sa_text_index_vocab stored procedure 62

- search conditions
 - CONTAINS clause 54
 - CONTAINS conditions 47
 - searching
 - CONTAINS clause 54
 - CONTAINS conditions 47
 - full text 2, 23
 - fuzzy 55
 - NGRAM TEXT index 55, 56
 - non-fuzzy 55, 56
 - prefix term limit 17
 - TEXT indexes 3
 - SECONDARY_FILE_ERROR option 72
 - SELECT statement
 - FROM clause syntax 47
 - sorting query results 54
 - sp_iqindexsize
 - binary large object 67
 - BLOB 67
 - character large object 68
 - CLOB 68
 - LONG BINARY 67
 - LONG VARCHAR 68
 - sp_iqindexsize stored procedure 67
 - sp_iqsetcompression stored procedure 65
 - sp_iqshowcompression stored procedure 66
 - standards 2
 - statistics
 - TEXT indexes 61, 62
 - stoplist 8, 9, 12
 - modifying 14
 - stored procedures
 - binary large object 59
 - BLOB 59
 - character large object 59
 - CLOB 59
 - LONG BINARY 59
 - LONG VARCHAR 59
 - sa_char_terms 59
 - sa_external_library_unload 64
 - sa_list_external_library 64
 - sa_nchar_terms 60
 - sa_text_index_stats 61
 - sa_text_index_vocab 62
 - sp_iqindexsize 67
 - sp_iqsetcompression 65
 - sp_iqshowcompression 66
 - string
 - length 92
 - string functions
 - BYTE_LENGTH 92
 - STRING_RTRUNCATION option 76, 77
 - SUBSTRING function
 - description 98
 - syntax 98
 - SUBSTRING64 function
 - description 100
 - syntax 100
- ## T
- term breaker
 - algorithm 10
 - setting 10
 - text configuration object 10
 - term breaker library 8–12, 19
 - restrictions 20
 - terms
 - breaking 59, 60
 - full text searching 23
 - ignoring 12
 - maximum length 12
 - minimum length 11
 - row position 59, 60
 - stoplist 14
 - TEXT index 59
 - text configuration object 8
 - altering 13, 81
 - CONTAINS examples 15
 - creating 9, 85
 - defaults 9
 - dropping 14, 88
 - examples 14, 15
 - listing 13
 - settings 10–12
 - term breaker 10
 - TEXT data
 - inserting from ASE 77
 - inserting into LONG VARCHAR 77
 - TEXT index 3, 79
 - altering 84
 - changing dbspace 6
 - comparison with WD index 3
 - CONTAINS conditions 47
 - creating 4, 5, 87
 - deleting rows 7
 - dropping 7, 89
 - editing 6
 - fuzzy search 55

Index

- listing 5
- NGRAM 8
- non-fuzzy search 55, 56
- refreshing 7
- restrictions 5
- statistics 61, 62
- terms 59
- text configuration object 8, 9
- text search
 - FROM contains-expression 47
- TEXT_DELETE_METHOD option 7
- TRIM_PARTIAL_MBC option 73

U

- unloading
 - external libraries¶¶ 64
 - external library 21
- Unstructured Data Analytics Option 1
 - licensing 1
- updating
 - BLOB 75
 - CLOB 77
 - large object data 75, 77
 - LOB 75, 77
 - LONG BINARY 75
 - LONG VARCHAR 77

- upgrading
 - existing LONG BINARY columns 75
 - LONG BINARY 75

V

- variables
 - binary large object 78
 - binary large object conversion 78
 - BLOB 78
 - BLOB conversion 78
 - character large object 78
 - character large object conversion 78
 - CLOB 78
 - CLOB conversion 78
 - function support for BLOB 91
 - function support for CLOB 91
 - LONG BINARY 78
 - LONG BINARY conversion 78
 - LONG VARCHAR 78
 - LONG VARCHAR conversion 78

W

- WD index 79
 - comparison with TEXT index 3