Getting Started with SAP Sybase IQ Column Store Analytics Server

Lesson 3: Create Schema and Load Data



Table of Contents

1. In	ntroduction	1
2. P	reparation: Unzip Data and Queries File	2
3. U	sers and Security Policies	4
3.1	Users, Roles and Privileges	4
3.2	Login Management	
3.3		
3. C	reate a DBSpace to Contain User Data	6
4. In	nstalling the Schema	10
5. SA	AP Sybase IQ Read and Write Operations	14
6. L	oading Data into the Tables	15
6.1	The LOAD TABLE Command	15
	TPCH Data Files	
7. St	ummary	19

1. Introduction

This chapter discusses creating users in the database you created in the previous chapter, installing the TPCH schema, and loading data into the tables.



Getting ready to put data in the database.

2. Preparation: Unzip Data and Queries File

This course comes with a zip file of data, metadata, and queries to make your life a little easier as you go through this course. This file is called "Data_Queries.tar.gz". After you acquire this file onto your machine, unzip it with the following commands:

```
gunzip Data_Queries.tar.gz
tar xvf Data Queries.tar
```

You will now have a subdirectory called "TPCH" with the following contents:

```
sybase@evangengl-dbc-node001:~/TPCH_Data_Qu

File Edit View Search Terminal Help

[sybase@evangengl-dbc-node001 TPCH]$ ls -l

total 16

drwxrwxr-x 2 sybase sybase 4096 Dec 31 13:34 Data

drwxrwxr-x 2 sybase sybase 4096 Jan 2 15:07 Metadata

drwxrwxr-x 2 sybase sybase 4096 Jan 2 14:14 Queries

drwxrwxr-x 3 sybase sybase 4096 Dec 31 17:27 UDF

[sybase@evangengl-dbc-node001 TPCH]$
```

The "Data" directory contains the raw data files to load into your TPCH schema:

```
sybase@evangengl-dbc-node001:~/TPCH Data Queries/TPCH/Data
File Edit View Search Terminal Help
[sybase@evangengl-dbc-node001 Data]$ ls -l
total 1074804
-rw-rw-r-- 1 sybase sybase 24357279 Apr 15 2010 customer.tbl
-rw-rw-r-- 1 sybase sybase 759861366 Apr 15 2010 lineitem.tbl
-rw-rw-r-- 1 sybase sybase
                               3013 Dec 27 15:39 LoadTables.sql
-rw-rw-r-- 1 sybase sybase
                               2909 Jun 24 2010 LoadTables.sql~
-rw-rw-r-- 1 sybase sybase
                               2103 Apr 15
                                           2010 nation.tbl
-rw-rw-r-- 1 sybase sybase 171915496 Apr 15
                                            2010 orders.tbl
-rw-rw-r-- 1 sybase sybase 118812986 Apr 15 2010 partsupp.tbl
-rw-rw-r-- 1 sybase sybase 24207151 Apr 15
                                            2010 part.tbl
-rw-rw-r-- 1 sybase sybase
                                396 Apr 15 2010 region.tbl
-rw-rw-r-- 1 sybase sybase
                            1408788 Dec 31 10:44 supplier.tbl
-rw-rw-r-- 1 sybase sybase
                                230 Aug 24 2010 TruncateTables.sql
[sybase@evangengl-dbc-node001 Data]$
```

The "Metadata" directory contains SQL files to create tables and indexes, and also a SAP Sybase IQ database configuration file which you can use to start up your TPCH database (tpch.cfg):

sybase@evangengl-dbc-node001:~/TPCH_Data_Queries/TPCH/Metadata _ File Edit View Search Terminal Help [sybase@evangengl-dbc-node001 Metadata]\$ ls -l total 36 -rw-rw-r-- 1 sybase sybase 77 Jan 2 14:40 CreateShipdateHGIndex.sql -rw-rw-r-- 1 sybase sybase 9147 Dec 27 15:23 CreateTables.sql -rw-rw-r-- 1 sybase sybase 147 Dec 27 15:23 CreateUserDBSpace.sql -rw-rw-r-- 1 sybase sybase 3733 May 17 2010 DropTables.sql -rw-rw-r-- 1 sybase sybase 447 Jan 2 10:13 SetDefaultOptions.sql -rw-rw-r-- 1 sybase sybase 437 Jan 2 10:12 SetOptions.sql -rw-rw-r-- 1 sybase sybase 636 Jan 2 14:38 tpch.cfg [sybase@evangengl-dbc-node001 Metadata]\$ ■

The "Queries" directory contains SQL files to run the standard queries against your TPCH database:

sybase@evangengl-dbc-node001:~/TPCH_Data_Queries/TPCH/Queries _ □ File Edit View Search Terminal Help

```
[sybase@evangengl-dbc-node001 Queries]$ ls -l
total 104
-rw-rw-r-- 1 sybase sybase 13433 May 18
                                         2010 allqueries.txt
-rw-rw-r-- 1 sybase sybase
                             563 May 18
                                         2010 querv10.sal
-rw-rw-r-- 1 sybase sybase
                                         2010 queryll.sql
                             597 May 18
-rw-rw-r-- 1 sybase sybase
                             623 May 18
                                         2010 query12.sql
-rw-rw-r-- 1 sybase sybase
                             407 May 18
                                         2010 query13.sql
-rw-rw-r-- 1 sybase sybase
                             357 May 18
                                         2010 query14.sql
-rw-rw-r-- 1 sybase sybase
                             554 May 18
                                         2010 query15.sql
-rw-rw-r-- 1 sybase sybase
                             544 May 18
                                         2010 query16.sql
-rw-rw-r-- 1 sybase sybase
                             349 May 18
                                         2010 query17.sql
                                         2010 query18.sql
-rw-rw-r-- 1 sybase sybase
                             518 May 18
-rw-rw-r-- 1 sybase sybase 1031 May 18
                                         2010 query19.sql
-rw-rw-r-- 1 sybase sybase
                             556 May 18
                                         2010 query1.sql
-rw-rw-r-- 1 sybase sybase
                             642 May 18
                                         2010 query20.sql
-rw-rw-r-- 1 sybase sybase
                             737 May 18
                                         2010 query21.sql
-rw-rw-r-- 1 sybase sybase
                             702 May 18
                                         2010 query22.sql
                                         2010 query2.sql
-rw-rw-r-- 1 sybase sybase
                             787 May 18
-rw-rw-r-- 1 sybase sybase
                             463 May 18
                                         2010 query3.sql
-rw-rw-r-- 1 sybase sybase
                             383 May 18
                                         2010 query4.sql
-rw-rw-r-- 1 sybase sybase
                             544 May 18
                                         2010 query5.sql
-rw-rw-r-- 1 sybase sybase
                             266 May 18
                                         2010 query6.sql
                             868 May 18
-rw-rw-r-- 1 sybase sybase
                                         2010 query7.sql
-rw-rw-r-- 1 sybase sybase
                             860 May 18
                                         2010 query8.sql
-rw-rw-r-- 1 sybase sybase
                             673 May 18
                                         2010 query9.sql
[sybase@evangengl-dbc-node001 Queries]$
```

3. Users and Security Policies

Proper management of users and permissions is essential in a database. It allows users to carry out their jobs effectively, while maintaining the security of information within the database.

Note: Security Management has been changed in SAP Sybase IQ 16. If you have a previous implementation of IQ, consult the "Upgrading to Role-Based Security" section in the Migration document for further details.

3.1 Users, Roles and Privileges

SAP Sybase IQ defines user permissions and privileges using a *role-based security* model.

Roles and privileges can be *granted* to users or *revoked* from users. A role is a set of privileges that can, as a group, be assigned to a user. Then, as the role's privileges change, the user's privileges change accordingly. Roles can be broken down as follows:

- User-Defined Roles are a custom collection, often created to group privileges and tasks
- System Roles are build-in and automatically created with a new database
- Compatibility Roles are created for backwards compatibility with versions of IQ earlier than 16.0

A privilege provides the ability to perform an operation on the system. A permission, on the other hand, is that ability in the given environment. A user may not have *permission* to perform a task if they have the privilege, but not on the currently acted on object. Privileges are broken down as follows:

- System Privileges give you the right to perform the action
- Object-level Privileges restrict your right to perform the action to the specified objects, on which the privilege is granted.

A user can be extended to be used as a role, defined as a User-Extended Role. This can be used to define all of a user's current privileges as a role and then assign them to another user. A user-defined role cannot be converted to a user-extended role, or vice versa, but can be used similarly.

When a new database is created, the default user is DBA, with an initial password of "sql". This user is automatically granted all system and object-level privileges in the database.

A new user is automatically granted the *public* system role, which gives view data and general execution of system stored procedure access. From there, roles and privileges can be granted to the user accordingly.

When a new object is created, the *owner* can be defined, otherwise, the creator becomes the owner. This gives privileges to modify the structure of the table and grant other privileges to other database users.

Ownership of a table is not sufficient to load the table with data. The user must also have INSERT permission on the table (for convenience, this permission can be set globally across users with the "-gl" server start up option).

3.2 Login Management

A *login policy* defines the rules to be followed when establishing a user's connection to the database. Login policies are assigned to users. A login policy lets the database administrator limit the number of active logins for a single user, limit the number of failed login attempts, lock out particular users, and set rules for changing and expiring passwords.

3.3 TPCD User

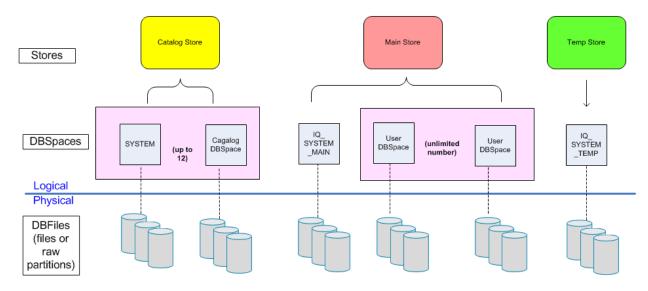
In your database, you will be creating the user TPCD (password "sql", as defined in "CreateTables.sql" in the Metadata folder of the unzipped data queries.), who will be the owner of all the TPCH tables and indexes.

3. Create a DBSpace to Contain User Data

Before you install the schema, you need to allocate some space to create the tables where the user data will be loaded. You already created your database with some available space in the IQ_SYSTEM_MAIN DBSpace, but that space is intended to be used for SAP Sybase IQ system management. You need to define a separate area to contain your user data. Let's first explain how SAP Sybase IQ manages storage.

In SAP Sybase IQ, data is divided into logical storage areas called *DBSpaces*. A DBSpace is a logical container that is composed of one of more operating system files, called *DBFiles*. A *store* is one or more DBSpaces that store persistent or temporary data for a special purpose. SAP Sybase IQ has three stores:

- The catalog store contains the metadata for the database. It is composed of the SYSTEM DBSpace and up to twelve additional catalog DBSpaces.
- The IQ main store contains user data and indexes. It is composed of the IQ_SYSTEM_MAIN DBSpace and other user defined DBSpaces.
- The IQ temporary store is used to house the data and indexes for temporary tables. Also, during loads, the temporary store is used to cache the intermediate data to be loaded into HG (High Group), WD (word) and text indexes (a few of the index types available to you with SAP Sybase IQ). During queries, the optimizer uses the temporary store for sorting and other operations that generate temporary data, such as updatable cursors, intermediate result sets, and synchronization of join indexes. The IQ temporary store contains the IQ_SYSTEM_TEMP DBSpace.



Each type of storage has an associated cache in IQ memory: catalog store has a catalog cache, main store has a main cache, and temporary store has a temporary cache.

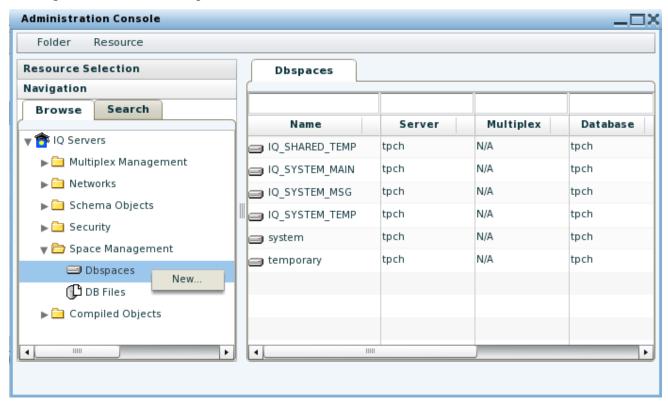
SAP Sybase IQ needs more memory for caches than for any other purpose. All user connections share these buffer caches.

When you created your SAP Sybase IQ database, some of the DBSpaces shown in the picture above were automatically created, except the "User DBSpaces" portion. You are going to create a DBSpace and a DBFile to contain your TPCH data. The raw TPCH data is approximately 1 GB in size. You are going to create a DBSpace of 2GB for the data and additional indexes.

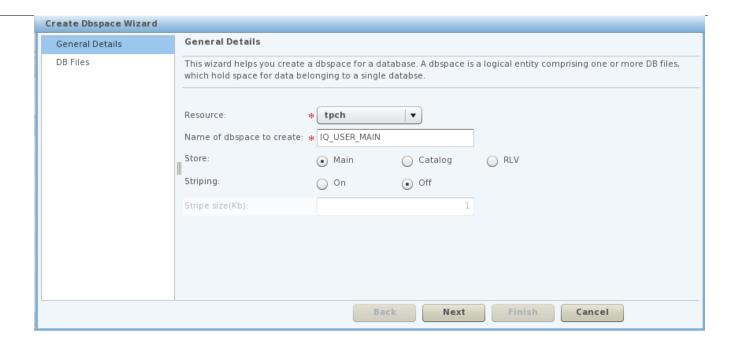
To create a DBSpace and DBFile, your database needs to be running. If it is not, then from a command window, go to the directory where your TPCH database resides (in this example, "/opt/sybase/TPCHDB", and enter the following command:

start iq @tpch.cfg tpch.db

Once the server has started, connect to it with Sybase Control Center by registering the resource, as described in the previous chapter "Product Installation and Database Creation". After you have connected, click on the drop-down for the tpch server and open the Administration Console. Then, click the arrow to the left of "IQ Servers" to expand the view. Expand "Space Management". Then, click on the drop-down arrow for Dbspaces and click "New..."



Fill in the fields as shown below. Note that "Striping" is turned off. IQ has a striping feature that helps you balance I/O by striping data across all the DBFiles in a DBSpace. Many hardware I/O systems also have striping features to allow you to set up RAID disks, and write data in parallel across those disks. The written data is spread out over the disks, and consolidated when it is read back into memory. For the purposes of this exercise, you are going to create a single DBFile, so striping is irrelevant. Click "Next".

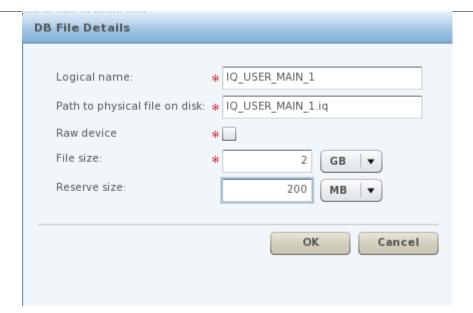


Click on "Add" to add a DBFile to your new DBSpace.



Specify the logical name for the DBFile, and its physical file location. We are not going to use a raw device for this storage, which allows us to specify the size of the file to be stored on the file system.

Create a 2GB file, for the raw data and indexes. The "Reserve Size" controls the amount of space set aside for some small, but critical data structures used during checkpoint and commit operations. Set it to 200 (for 200MB). Click "OK".



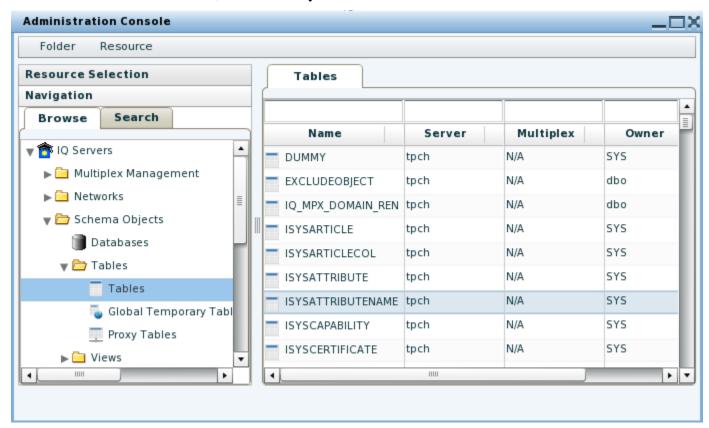
You will see a summary of the properties of the newly created DBFile.



Click on "Finish" to complete creation of the DBFile.

4. Installing the Schema

Now you are going to load the schema. First, expand the "Schema Objects" followed by "Tables" in the Administration Console in SCC, connected to your server. Click "Tables".



In the "Tables" folder, you will see that all currently existing tables are system tables. You will now create the table definitions.

To create tables, we will run a SQL command file created to define tables and schemas.

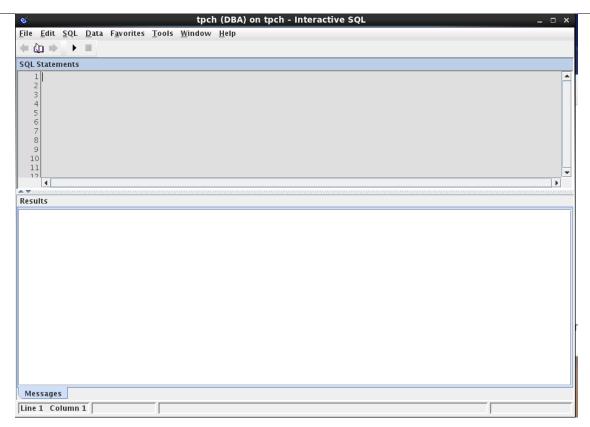
From an open terminal window, run a command similar to the following to create an ODBC connection called "tpchdb". Specify the id and password as created for the server. Specify the dbf as the path to the .db file where the database is stored.

iqdsn -w "tpchdb" -c "uid=DBA; pwd=sql; dbf=/opt/sybase/TPCHDB/tpch.db"

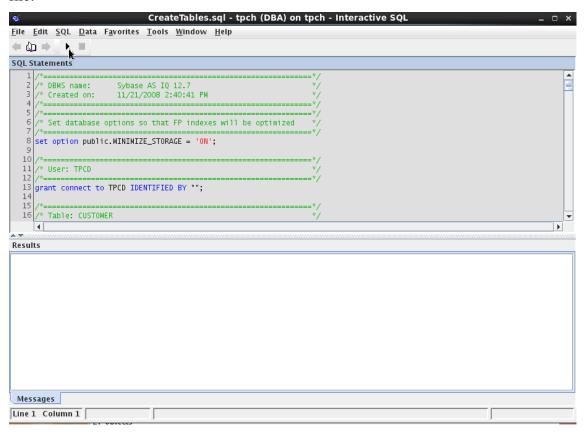
Run the following command to open Interactive SQL.

dbisql -datasource "tpchdb"

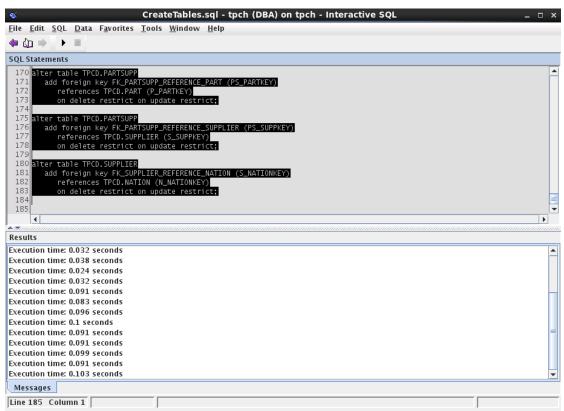
The program will open up connected to the database through the ODBC connection.



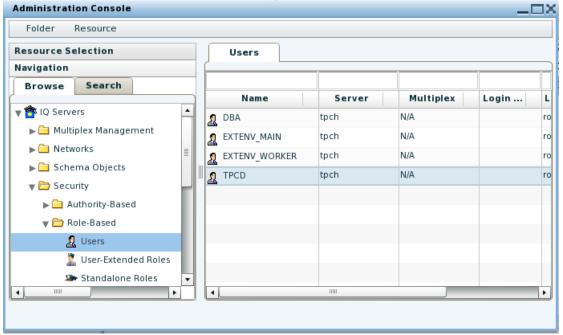
You can enter SQL commands in the top pane, and results will show up in the lower pane. Open up a SQL command file that has already been created to define the tables and indexes for the TPCH schema. Choose "File...Open...", browse to "\$TPCHROOT/TPCH/Metadata/CreateTables.sql", and open the file:



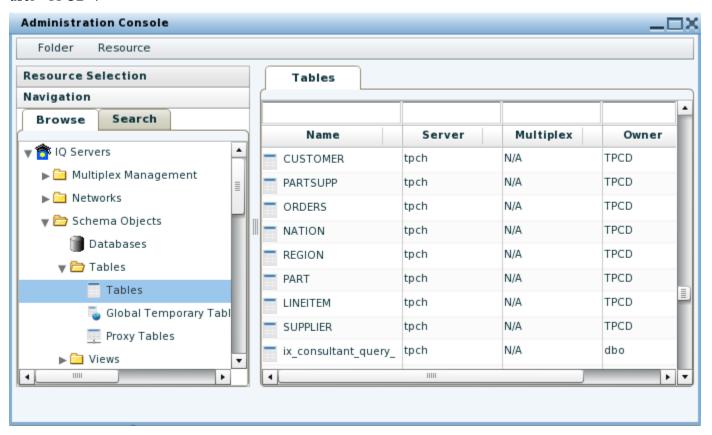
Click on the "Execute all SQL statement(s)" icon (the black triangle):



Leave the interactive SQL window open, for future steps. Look at the Administration Console in SCC. In the left pane, expand IQ Servers, Security, Role-Based and click on Users. There is a "TPCD" user that was created by the "CreateTables.sql" script.



Now, expand Schema Objects, Tables and click on Tables. You will see some new tables, owned by user "TPCD":



Now you are ready to load data into the tables.

5. SAP Sybase IQ Read and Write Operations

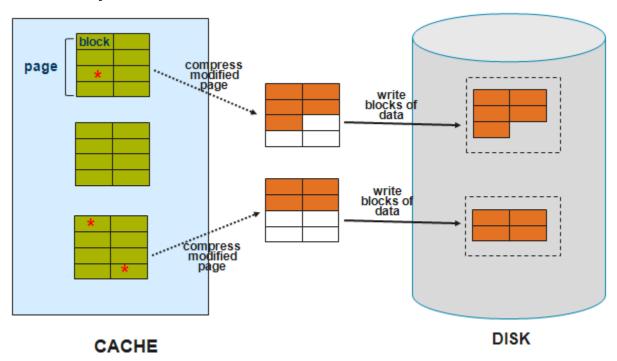
Before you load data, let's talk about how SAP Sybase IQ performs I/O against the database.

SAP Sybase IQ makes all read and write operations directly to cache. Data is flushed to disk from the cache when:

- The cache "dirty" pages exceed a threshold
- At commit time, if the transaction is committing a modification to a table

SAP Sybase IQ stores data on disk in compressed form. It swaps data in and out of memory in units of *pages*. When you create a database you specify the IQ page size. This cannot be changed once it is set. SAP Sybase IQ is biased towards readers. It likes to have data available in cache, and tries to minimize reads from disk. For SAP Sybase IQ, a large page size for reading and writing provides a performance advantage – I/O setup time is proportionally smaller with a larger page size, so I/O is more efficient. The default page size is 128KB.

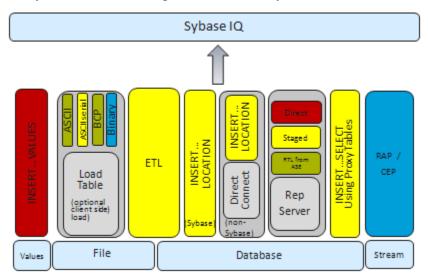
A SAP Sybase IQ page is divided into an integral number of *blocks*. When IQ writes a modified page to disk (the picture shows an asterisk on blocks in a page that have been modified), it first compresses it, then writes out just those blocks that contain actual data:



When IQ reads a page from disk, it reads all the compressed blocks of the page into memory, and then un-compresses the data to regenerate a complete page.

6. Loading Data into the Tables

There are many choices for loading data into SAP Sybase IQ. The following picture depicts the choices:



Reading the image from bottom to top, let's start with the data sources and move into the various load methods. The load methods are placed above valid data sources. For example, the "INSERT...LOCATION" is a SQL statement that loads data directly from databases. ETL (Extract, Transform and Load) is an engine that can load data from either files or databases. It extracts data from a source, transforms it to the format and content you specify, and loads it into SAP Sybase IQ. We recommend SAP Data Services as a full featured ETL product for loading data into SAP Sybase IQ. SAP Sybase IQ also works with commercial ETL products such as Pentaho, Talend, and Informatica. The color of the box indicates the relative speeds of the loading method – red is suitable for smaller tables, yellow is faster, green is very fast, and blue is the fastest.

Sybase RAP/CEP (Real-time Analytics Platform/Complex Event Processing) is a product designed for the low latency requirements of financial services markets. It consolidates streamed financial market data into a repository for consumption by automated trading applications and various users with analytical needs. It is the fastest of the options, but works only with streamed data.

Data replication (the Rep Server boxes above), tracks changes to your source database, and is used for applying incremental changes to your SAP Sybase IQ target database. You can configure data replication in a few different ways to apply data to SAP Sybase IQ. RTL stands for Real Time Loading. It is an edition of Replication Server, and is optimized for continuous, fast loads into SAP Sybase IQ.

Your best option for performance is the "LOAD TABLE" SQL bulk loader, which can handle thousands to millions of rows per second. That is the method you are going to use to load data into your TPCH database.

6.1 The LOAD TABLE Command

The LOAD TABLE SQL statement imports data from an ASCII or binary formatted file into an existing database table. An ASCII source file may be in fixed width format (each field occupies a fixed length of bytes) or delimited format (each field is terminated by a specified delimiter). The LOAD TABLE statement loads both the data and all indexes that have been created, and does not require any further table reorganizations or index rebuilds. There are a lot of statement options that specify the formatting

of the input file, how to handle errors, and how to notify the user of loading progress. Look at the SAP Sybase IQ user documentation at http://sybooks.sybase.com/nav/base.do for a detailed explanation.

Here is an example of a LOAD TABLE statement that loads ASCII, pipe delimited data from a file on a Windows platform:

```
load table TPCD.SUPPLIER (
    S_SUPPKEY,
    S_NAME,
    S_ADDRESS,
    S_NATIONKEY,
    S_PHONE,
    S_ACCTBAL,
    S_COMMENT
)
from '/opt/sybase/TPCH_Data_Queries/TPCH/Data/supplier.tbl'
quotes off
escapes off
format ascii
delimited by '|'
row delimited by '\n'
```

Note that when you refer to a table in an SQL statement, and you are not logged in as the "TPCD" user, then you should identify it with the prefix "TPCD".

Loading involves three major steps:

- 1. Reading the data out of the file and placing it into data structures in memory
- 2. Building indexes from the data structures in memory
 - a. For FP, LF, HNG, CMP, DATE, DTTM and TIME indexes, data is inserted directly into the index
 - b. For HG, WD and text indexes, data is sorted during a first pass, and then inserted during a second pass
- 3. Writing the data and indexes into the database

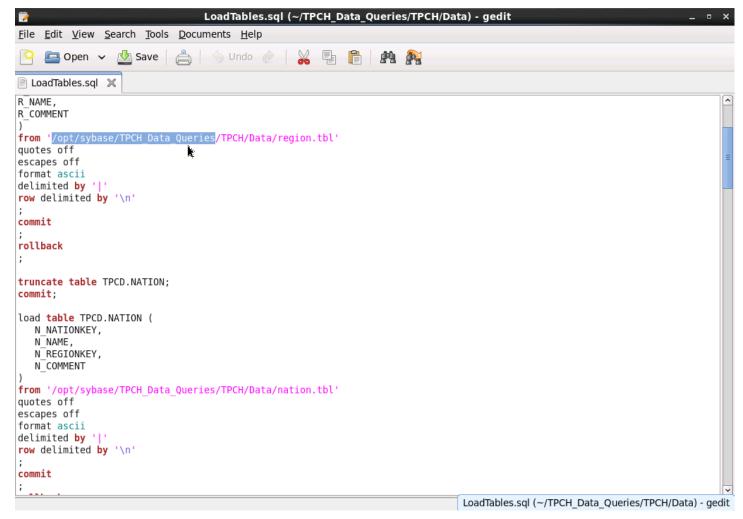
6.2 TPCH Data Files

There are a set of data files in the directory "\$TPCHROOT/TPCH/Data":

```
[sybase@evangengl-dbc-node001 Data]$ ls -l
total 1074804
-rw-rw-r-- 1 sybase sybase 24357279 Apr 15 2010 customer.tbl
-rw-rw-r-- 1 sybase sybase 759861366 Apr 15 2010 lineitem.tbl
-rw-rw-r-- 1 sybase sybase
                               3013 Dec 27 15:39 LoadTables.sql
-rw-rw-r-- 1 sybase sybase
                               2909 Jun 24 2010 LoadTables.sgl~
-rw-rw-r-- 1 sybase sybase
                               2103 Apr 15 2010 nation.tbl
-rw-rw-r-- 1 sybase sybase 171915496 Apr 15 2010 orders.tbl
-rw-rw-r-- 1 sybase sybase 118812986 Apr 15
                                            2010 partsupp.tbl
-rw-rw-r-- 1 sybase sybase 24207151 Apr 15
                                            2010 part.tbl
-rw-rw-r-- 1 sybase sybase
                                396 Apr 15 2010 region.tbl
-rw-rw-r-- 1 sybase sybase 1408788 Dec 31 10:44 supplier.tbl
-rw-rw-r-- 1 sybase sybase
                                230 Aug 24 2010 TruncateTables.sql
[sybase@evangengl-dbc-node001 Data]$
```

Each data file contains the data for one table.

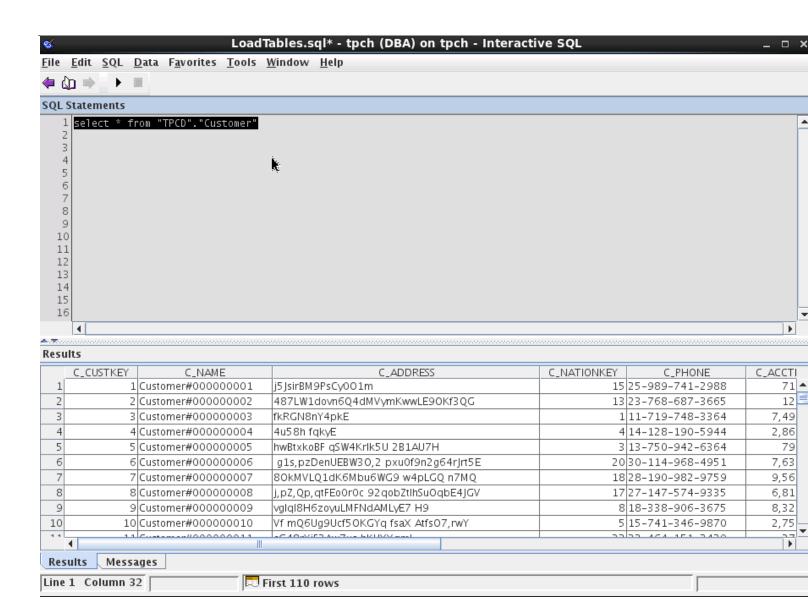
A script has already been created to run LOAD TABLE to load data from these files into your database. It needs to be modified, however, with the correct paths to the data files within your environment. Open up the file "\$TPCHROOT/TPCH/Data/LoadTables.sql" with a text editor:



Note the highlighted text "/opt/sybase/TPCH_Data_Queries". In this environment, the "\$TPCHROOT" directory is "opt/sybase/TPCH_Data_Queries". If you have unzipped the queries and data files into a different directory, you need to substitute every instance of "opt/sybase/TPCH_Data_Queries" with the proper directory from your environment. Find and replace all instances. Save the modified file.

Now, return to your open Interactive SQL window (if it is closed, reopen with the command as before).

Browse to your modified script "\$TPCHROOT/TPCH/Data/LoadTables.sql" and execute it (this will take a couple of minutes as you load the larger tables). Then you can view the data in the tables. Here is the data from the CUSTOMER table:



7.

Summary
In this lesson, you have created a schema, and loaded data into the tables. Now you can start running queries against your data.

Copyright

© Copyright 201H SAP AG. 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.

Microsoft, Windows, Excel, Outlook, and PowerPoint are registered trademarks of Microsoft Corporation.

IBM, DB2, DB2 Universal Database, System i, System i5, System p, System p5, System x, System z10, System z9, z10, z9, iSeries, pSeries, zSeries, eServer, z/VM, z/OS, i5/OS, S/390, OS/390, OS/400, AS/400, S/390 Parallel Enterprise Server, PowerVM, Power Architecture, POWER6+, POWER6, POWER5+, POWER5, POWER, OpenPower, PowerPC, BatchPipes, BladeCenter, System Storage, GPFS, HACMP, RETAIN, DB2 Connect, RACF, Redbooks, OS/2, Parallel Sysplex, MVS/ESA, AIX, Intelligent Miner, WebSphere, Netfinity, Tivoli and Informix are trademarks or registered trademarks of IBM Corporation.

Linux is the registered trademark of Linus Torvalds in the U.S. and other countries.

Adobe, the Adobe logo, Acrobat, PostScript, and Reader are either trademarks or registered trademarks of Adobe Systems Incorporated in the United States and/or other countries.

Oracle is a registered trademark of Oracle Corporation.

UNIX, X/Open, OSF/1, and Motif are registered trademarks of the Open Group.

Citrix, ICA, Program Neighborhood, MetaFrame, WinFrame, VideoFrame, and MultiWin are trademarks or registered trademarks of Citrix Systems, Inc.

HTML, XML, XHTML and W3C are trademarks or registered trademarks of W3C®, World Wide Web Consortium, Massachusetts Institute of Technology.

Java is a registered trademark of Sun Microsystems, Inc.

JavaScript is a registered trademark of Sun Microsystems, Inc., used under license for technology invented and implemented by Netscape.

SAP, R/3, SAP NetWeaver, Duet, PartnerEdge, ByDesign, SAP Business ByDesign, 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.

Business Objects and the Business Objects logo, BusinessObjects, Crystal Reports, Crystal Decisions, Web Intelligence, Xcelsius, and other Business Objects products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of Business Objects S.A. in the United States and in other countries. Business Objects is an SAP company.

All other product and service names mentioned are the trademarks of their respective companies. Data contained in this document serves informational purposes only. National product specifications may vary.

These materials are subject to change without notice. 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.