# Getting Started with SAP Sybase IQ Column Store Analytics Server

Lesson 1: Introduction to SAP Sybase IQ



### **Table of Contents**

1.	Introduction	1
2.	Traditional Relational Databases vs. SAP Sybase IQ	2
3.	SAP Sybase IQ Components	4
4.	SAP Sybase IQ Additional Features	6
5.	SAP Sybase IQ Indexes	7
	Snapshot versioning	
7.	Summary	11

#### 1. Introduction

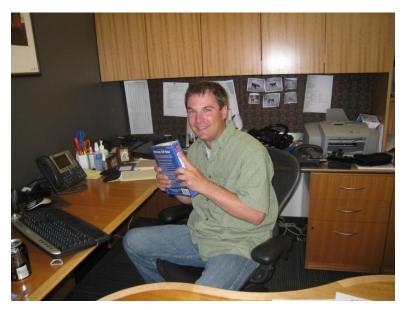
Welcome to "Getting started with SAP Sybase IQ Column Store Analytics Server". This book is set up in a way that will put the subject areas most pertinent to you right at your fingertips. This is not a reference book. Instead, it is a how-to book where the information is presented with the goal of getting you started with SAP Sybase IQ as quickly as possible.

This course is designed to present information in a readily graspable manner, so that you can learn the basics of SAP Sybase IQ quickly. You do not need to go through all of the lessons. The first three: "Introduction to SAP Sybase IQ", "Product Installation and Database Creation", and "Create Schema and Load Data" will get you to the point of having a running SAP Sybase IQ database (the course comes with a database schema and data ready to load). Then you can pick and choose from the remaining lessons, based on your interest:

- Optimize a Query using the SAP Sybase IQ Query Plan
- Scaling Out with Multiplex and Distributed Query Processing
- Monitoring SAP Sybase IQ with Sybase Control Center
- Row-Level Versioning

To benefit from this book, you should be familiar with the following concepts:

- Relational database systems
- Database schemas
- SQL



Getting ready to begin the course!

#### 2. Traditional Relational Databases vs. SAP Sybase IQ

All of us are familiar with databases, and most of us have direct experience with the most ubiquitous type – relational databases.

Relational databases present data to the user as a set of tables. Tables contain rows of data, comprised of columns of attributes. Most conventional relational databases are *row-oriented*. A row-oriented database serializes all of the values in a row together. When a query retrieves a set of rows, all columns of the resulting rows are returned.

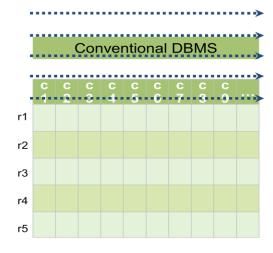
Row-oriented databases are efficient when many columns of the same row are accessed at the same time. In practice, a row-oriented architecture is well suited for OLTP (online transaction processing) workloads, which are more heavily loaded with interactive transactions. OLTP involves data entry and retrieval transactions that are fundamental to business operations in a number of industries - banking, airline reservations, mail order, supermarkets, and manufacturing.

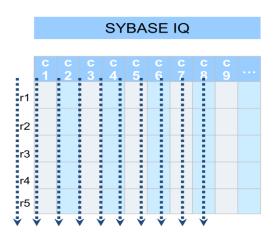
OLTP is only one class of application that interacts with databases, however. Another class is analytic applications. Analytic applications include:

- traditional reporting: dashboards of basic performance indicators within a business, such as quarterly sales figures, and customer satisfaction levels
- complex advanced analytics: predictive models, analysis of real time data streams, and unstructured text analytics
- data aggregation: systematic gathering of data from various sources (questionnaires, interviews, observation, electronic devices, etc.) for the purpose of statistical analysis
- data mining: the process of analyzing data from different perspectives by finding correlations or patterns and summarizing it into useful information

Analytic applications require fast query response across a large volume of data. Typically, the questions asked by an analytics application access only a subset of the columns in a table. For these types of applications, a row-oriented database tends to be inefficient, returning more data than is being requested.

An alternative to row-oriented architectures are column-oriented architectures. With this approach, data is organized and stored by columns. Because each column can be stored separately, for any query, the system can evaluate which columns are being accessed, and retrieve only the values requested from the specific columns. SAP Sybase IQ was designed as a high performance analytics platform, and utilizes this column based architecture:





When row-oriented databases are used for analytics, they need to be tuned with additional indexes and pre-aggregated data to deliver data more quickly. This requires storage resources for the indexes and maintenance overhead to keep the pre-aggregated data current. Also, the system won't respond to ad hoc queries without additional tuning.

With SAP Sybase IQ, the data values within a column are stored as an index – the index is not separate from the data. This approach minimizes storage requirements, and speeds data retrieval. Also, when columns are stored separately from each other, multiple processing threads can be applied to concurrently access different areas of storage to return data more quickly. Another benefit of independent column layout is that compression algorithms can be optimized to operate on the single data type of a column. Improved compression reduces data storage requirements even further. SAP Sybase IQ customers regularly see 40% to 70% compression compared to other systems.

SAP Sybase IQ offers additional index types that can be added to columns for even faster query performance. These different index types are tuned for particular data types, operators and search conditions. You can place multiple indexes on a column, and the SAP Sybase IQ query engine can make use of multiple indexes in order to optimize a query.

Although SAP Sybase IQ has a very different architecture from a row-oriented database, to the user, SAP Sybase IQ looks like a regular relational database. The data is logically organized as tables, rows, and columns. There is a standard SQL interface to create and access data. You can write stored procedures using the Transact-SQL dialect. You can execute Java code within user defined functions. You won't know that IQ is different from any other type of database – that is until you are impressed with how well it performs when you run reports, aggregate data, and execute ad hoc and complex queries!

#### 3. SAP Sybase IQ Components

SAP Sybase IQ includes an embedded SQL Anywhere DBMS engine. SQL Anywhere is Sybase's lightweight relational database, often used on mobile devices. The SQL Anywhere layer of SAP Sybase IQ maintains the database catalog (metadata and stored procedures), and handles connectivity between client and server, query parsing and security. The SAP Sybase IQ layer incorporates patented storage technology and a high performance query engine.

Access to your SAP Sybase IQ database must be made through a SAP Sybase IQ server, which provides a communications channel for the database. A SAP Sybase IQ server starts and stops a database, accepts connections from applications or users running on the same machine or on other computers across a network. Users have connection rights to a database, not to the server. Generally, the SAP Sybase IQ server manages a single database.

You can combine multiple SAP Sybase IQ servers (some executing queries, and some executing data loads) running on multiple machines, accessing the same SAP Sybase IQ database on shared storage, in order to improve throughput and responsiveness to a larger number of concurrent clients. This Multiple Parallel Processing (MPP) based on a shared everything architecture is branded as the PlexQ<sup>TM</sup> Distributed Query Platform.

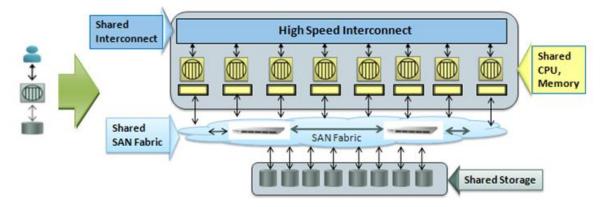


Figure 1 − Single SAP Sybase IQ deployment (left) can scale to PlexQ<sup>TM</sup> grid (right)

SAP Sybase IQ can scale from a single server to an MPP grid containing dozens of servers working together. SAP Sybase IQ with PlexQ<sup>TM</sup> technology scales out in a far more flexible manner than shared-nothing approaches by separating compute resources from data storage. Queries can be processed in parallel on a PlexQ<sup>TM</sup>grid without concern for how, or whether, data is partitioned. Furthermore, compute resources can be dynamically added to support more users or workloads without impacting data storage. Similarly, data capacity can be added without having to rebalance data stored on compute nodes, simplifying maintenance. Unlike shared-nothing MPP architectures, PlexQ<sup>TM</sup> dynamically manages query workloads across all the compute nodes in the cluster. The automatic workload rebalancer aggressively works to avoid contention among users for system memory and CPU resources, thereby providing predictable high performance and resource efficiency for a broad spectrum of workloads.

SAP Sybase IQ offers database drivers for web programming environments, such as PHP, Perl, Python and Ruby. SAP Sybase IQ is integrated with the Hadoop distributed processing environment in several ways: you can load data from the Hadoop File System (HDFS) into IQ for deeper analysis, perform data and query federation with Hadoop, and there is an API so that you can write MapReduce like functions that operate completely within IQ.

SAP Sybase IQ includes a set of applications to help you manage your database. These utilities are:

- **DBISQL** or **Interactive SQL**: an application that lets you issue an SQL statement and send it to the database.
- **Command line utilities**: a set of utilities for carrying out administrative activities such as monitoring, and backing up and restoring a database.
- **Sybase Control Center (SCC):** a web-based administration tool that helps with managing and monitoring. It provides intelligent tools for analyzing trends and performance as well as can be used to perform general administrative tasks. This tool now encompasses both SCC and Sybase Central, from previous versions (15.4 and earlier), and Sybase Central has been deprecated.

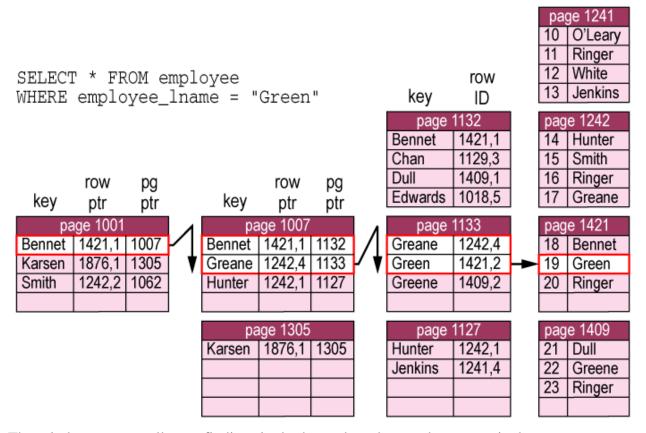
#### 4. SAP Sybase IQ Additional Features

In addition to the base product, SAP Sybase IQ offers enhanced feature sets to extend the power of SAP Sybase IQ and meet specific performance, security, scalability and more advanced analytics. These SAP Sybase IQ options include:

- Advanced Security Option: Provides the highest levels of security capability for SAP Sybase IQ environments; both for data in-flight and at rest in the database. It enhances SAP Sybase IQ security protection with FIPS certified encryption algorithms, integration with LDAP and column-based encryption.
- Unstructured Data Analytics Option: Supports binary large object (BLOB) and character large object (CLOB) storage and retrieval. Uses the same applications and interface to access all structured and unstructured data in the repository. Enables search for words and phrases within text data, boolean and proximity searches and score results from text queries based on relevance. Allows high performance indexing, search, and analysis on textual data using the SQL language.
- In-Database Analytics Option: Provides an application programming interface (API) that allows building (in C/C++) and deployment of User-Defined Functions (UDF) that run inside the database process space close to the data. This permits implementation of proprietary algorithms securely inside SAP Sybase IQ facilitating performance. Further, a simulation environment allows testing of Table Parameterized UDFs executing inside SAP Sybase IQ facilitating development.
- Very Large Database Management (VLDB) Option: Enhances the manageability of large data sets with table partitions and multiple DBSpaces. Range, hash and has-range partitioning schemes are now supported.
- **Multiplex Grid Option:** As described in the component section above the PlexQ<sup>TM</sup> option enables creation of multiple IQ servers operating against the same IQ database scaling out an analytics environment in a far more flexible manner than shared-nothing approaches.

#### 5. SAP Sybase IQ Indexes

Many databases use a balanced-tree (B-tree) for indexing data:

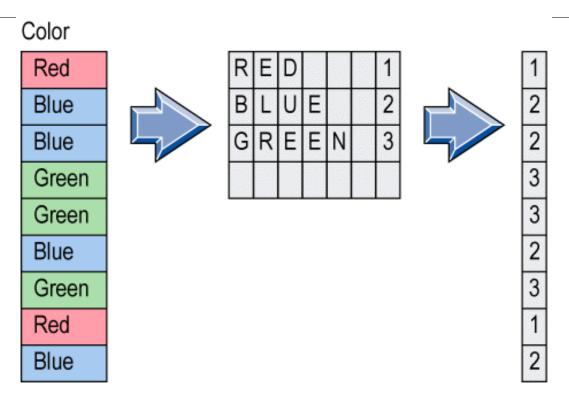


These indexes are excellent at finding single-data values, but can become quite large.

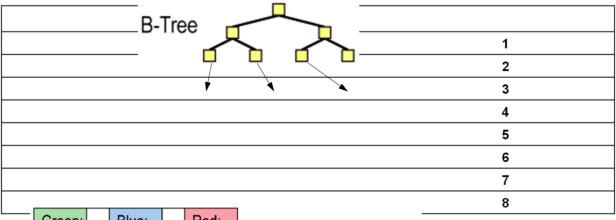
SAP Sybase IQ indexes consist of:

- Lookup table indexes
- Enhanced bitmap indexes
- Other types, such as fast text search indexes

A lookup table stores the distinct values in a column as a lookup table with the actual value and an associated encoding (integer). The column data is then stored as the encoded values that point into the lookup table:



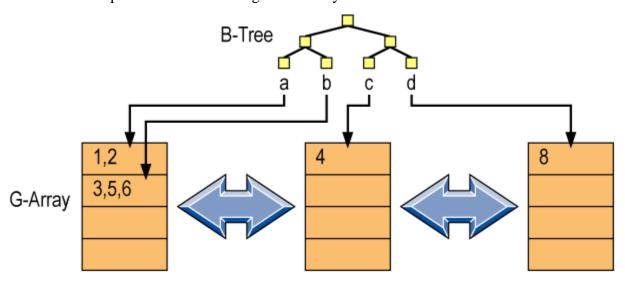
A bitmap index is comprised of bitmaps for each distinct value within a column. A bitmap is a range of bits, one bit for each row in the database, and a bit has the value 1 when the corresponding column of the corresponding row has the particular value and a 0 when it does not. SAP Sybase IQ includes a variety of bitmap indexes, some of which are enhanced with B-Trees and group arrays for faster access. For example, the Low-Fast (LF) index is enhanced with a B-Tree to quickly locate a bitmap for a particular value:



Green:	Blue:	Red:
0	0	1
0	1	0
0	1	0
1	0	0
1	0	0
0	1	0
1	0	0
0	0	1

SELECT \*
FROM colortable
WHERE color = 'red'

The High-Group (HG) index is structured as a B-Tree that points to a group array that points to the individual bitmaps. This is used for high cardinality data:



Here are the 10 index types in SAP Sybase IQ:

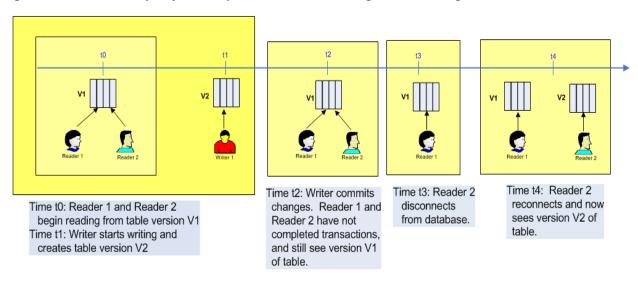
Index type	Purpose		
Fast Projection (FP)	Default index generated by a CREATE TABLE statement. In SAP		
	Sybase IQ, the data is stored as an FP index.		
Compare (CMP)	Stores the binary comparison $(<, >, =, <=, >=, \text{ or } !=)$ of two columns.		
DATE	An index on columns of data type DATE.		
Datetime (DTTM)	An index on columns of data type DATETIME or TIMESTAMP.		
High Group (HG)	Used to process equality and group by operations on high cardinality		
	data.		
High Non Group (HNG)	Used for high cardinality data involving ranges or aggregates.		
Low Fast (LF)	Similar to the HG index, but for low cardinality data.		
TIME	An index on columns of data type TIME.		
WD	Used to index distinct terms within the contents of a CHAR,		
	VARCHAR, or LONG VARCHAR columns.		
TEXT	Stores positional information for the terms in a character LOB data		
	type.		

#### 6. Snapshot versioning

SAP Sybase IQ is focused on readers, but your IQ database needs to be updated regularly, too. SAP Sybase IQ has a method for allowing writes to occur with good performance, without affecting the active readers querying the database.

SAP Sybase IQ supports concurrency by ensuring that all database operations occur within a transaction, and that these operations do not interfere with each other. It does so by setting access restrictions at the table level, and by using a technique called *snapshot versioning*. On a given table, IQ allows concurrent processing of multiple read transactions, but only one write transaction. You can have multiple concurrent writers, but they must be writing to different tables. SAP Sybase IQ readers are not interrupted as database updates occur, and writers have sole access to tables, simplifying the data locking mechanism.

This is how snapshot versioning works. When a SAP Sybase IQ client executes a connect, commit, or rollback operation, IQ takes a snapshot of the state of the objects in your database. As objects are modified in the database, IQ makes a copy of the database pages that have changed. Database pages that have not changed are shared among all active versions in the database. Readers see the unchanged versions until they perform another connect, commit or rollback. At that point, they will now see the updated version of any objects they access. Here is a depiction of the process:



## 7. Summary

This chapter has given you a basic introduction to SAP Sybase IQ: what it is, how it is different from traditional relational databases, and some of its key features. Now let's move on to product installation, and hands on use.

#### 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.