SYBASE IQ ANALYTICS SERVER
The New Generation Analytics Market Leader

#1 COLUMN-BASED ANALYTICS SERVER

- **Performance:** Industry leading performance
- **Customer deployments:** Over 3,100+ unique installations in 1800+ accounts
- **Customer acquisition:** ~200 new customer wins in 2009
- **Leadership:** Pioneering technology, v15.0 in March, 2009; v15.1 in July, 2009
**SYBASE IQ POSITIONING**
High Performance Analytics Server For Specific Use Segments

- **Predict the Business**
  - Advanced Analytics
  - Deep and complex analysis of large datasets
  - Sybase IQ=very fast complex and ad hoc query processing

- **Run the Business**
  - Data Aggregators
  - Industry information hub for large number of concurrent & special interest users
  - Sybase IQ= very fast query results + partitioning for large number of concurrent users

- **Optimize the Business**
  - DW Information Lifecycle Mgmt
  - Store and query data by storage tiers securely & efficiently
  - Sybase IQ=high compression, data partitioning and built in security

- **Manage the Business**
  - Reporting Services
  - High performance reporting and dash boarding on operational data
  - Sybase IQ=fast and mixed query, load processing for a large number of users

---

Data-Centric | User-Centric
-------------|--------------

---
**SYBASE IQ PRODUCT PROFILE**

**Outside:**
Standards based open interfaces enables best-of-breed eco-System

**Inside:**
Architected ground up for High Performance Analytics
**SYBASE IQ ARCHITECTURAL STRENGTH**

Robust Column Store Foundation

### Key Characteristics

- Data is stored vertically – Each column is stored separately
- The data is the index
- Large page sizes (128K – 512K)
- Persistent Row Identifiers
- Bitmap driven

### Benefits

- Unsurpassed concurrent, mixed workload performance, storage efficiencies
- Allows queries and updates to only access referenced columns.
- Single data type and domain per page greatly enhances the effectiveness of compression
- Large page size makes better use of modern disk and I/O subsystems
- Allows queries to evaluate multiple predicates on the same table using index-based access methods
- Allows rows to be uniquely identified without dragging all primary keys (e.g. for DELETES)
- Allows queries to read only those pages within a column store needed
- Bitmaps enable significant efficiencies – compact representation, easy to horizontally partition, intermediate results, allows perfect prefetch of rows avoiding cache misses and LRUs
### SYBASE IQ ARCHITECTURAL STRENGTH

Powerful Indexing Technology

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>TYPE</th>
<th>USAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bitmap</td>
<td>Fast Projection</td>
<td>Compressed raw data for result sets (Default)</td>
</tr>
<tr>
<td>Low Fast</td>
<td></td>
<td>Low cardinality data (up to 1000 unique values)</td>
</tr>
<tr>
<td>High Non-Group</td>
<td></td>
<td>Aggregation on the fly and range searches</td>
</tr>
<tr>
<td>Date, Time, DT</td>
<td></td>
<td>Date ranges, date part operations</td>
</tr>
<tr>
<td>Multi-Column</td>
<td></td>
<td>Concatenated indexes</td>
</tr>
<tr>
<td>Compare</td>
<td></td>
<td>Column comparisons</td>
</tr>
<tr>
<td>Traditional B-tree</td>
<td>High Group</td>
<td>Key fields and groupings for cross-tabular</td>
</tr>
<tr>
<td>Word</td>
<td></td>
<td>Key word or phrase string searches</td>
</tr>
</tbody>
</table>

- Most columns will have at least one index
- Index selection decisions based on column cardinality (number of unique values)
- Indexes and columns are stored separately
- Multiple indexes used to resolve a query
- Indexes are self maintaining
- No optimizer statistics to update
- Indexes are compressed
- Index building is a relatively simple exercise
- Index advisor tool guides users in the building of appropriate indexes based on a run of sample queries
SYBASE IQ ARCHITECTURAL STRENGTH
Key Indexing Technology – Fast Project Index

Optimized Fast Project Indexes – the column store

Unique values for a column are stored in a lookup table and the optimized FP requires 1, 2 or 3 bytes depending on cardinality

- Optimized FP Indexes covers a broad spectrum of data (> 3 bytes data)
  
  \[
  \text{FP}(1) = 2^8 = 256 \text{ Unique Values} \\
  \text{FP}(2) = 2^{16} = 65,536 \text{ Unique Values} \\
  \text{FP}(3) = 2^{24} = 16,777,216 \text{ Unique Values}
  \]

- Reduced storage for columns with cardinality between 65,536 and 16,777,216
- Reduced I/O
  - Writes during data loading & Reads during query processing
- Improved Query Processing
  - Reduced memory in query execution as the expansion of FP’s is delayed
  - More vector processing FP(3)
  - More complete optimizer statistics on column data distributions from lookup table which contains counts
SYBASE IQ ARCHITECTURAL STRENGTH
Key Indexing Technology – Low Fast Index

- Sample query
  
  ```sql
  select count(*) from customers
  where state = 'NY' and class = 'A'
  ```

- Filter low cardinality fields with bitmaps
  - Bit position correspond to fixed row ID

- Bitmaps further reduce the amount of data read
  - Small number of bits rather than entire field
  - ANDing and ORing bitmaps is very efficient with today’s processors
  - Note that even vertically stored data is not read
SYBASE IQ ARCHITECTURAL STRENGTH
Key Indexing Technology – High Non Group Index

- Data with large number of values stored in binary form
- Data sliced vertically so each bit position can be manipulated separately
- Many bit positions are either all on or all off so no storage space is required
  - System only needs to store mixed bitmaps (1s and 0s)
  - Typical storage is 10-20% of size of raw data
SYBASE IQ ARCHITECTURAL STRENGTH
Key Indexing Technology – Usage of Multiple Indices

• Example:
  select sum(sales)
  from customers
  where state = 'NY'
  and class = 'A'

• Sybase IQ will use the LF indexes to filter rows and then apply to HNG to compute the sum

• Minimal amount of data is read to resolve the query
SYBASE IQ ARCHITECTURAL STRENGTH
Key Indexing Technology – Other Indexes

• Word Index
  • The `like` query operator will call the Word Index
  • Both predicates below would use the Word Index

    Where company_name `contains` 'Sybase'
    Or
    Where company_name `like` '%Sybase%'

• Compare Index is an index on relationship b/w two columns
  • Stores comparison bitmap of (<, >, or =) of its two columns

• Date/Time/DateTime indexes
  • Range searches
  • Datepart searches
Compression

Page level
- LZW Compression with implicit dictionary

Enumerated FP
- Distinct values stored in vector
- Column stores vector ordinals

Bitmaps
- ROWIDs stored in multiple formats
- Format based on locality of ROWIDs

<table>
<thead>
<tr>
<th>Ordinal</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>“Wide Data”</td>
</tr>
<tr>
<td>2</td>
<td>“Wider Data”</td>
</tr>
</tbody>
</table>

Single | Ranges | Bit vector

1, {3-99}, {200-999}, {1295: 10001110101}
SYBASE IQ ARCHITECTURAL STRENGTH
Query Processing

Query Engine
- Highly parallel plans: tuple streams segregated, data flows produce parallel streams, termination of parallel streams
- Many access paths to the indexes and columns
- Concurrent querying aware, elastic CPU/memory usage, delayed projection
- Can use the vertical projection layer above the columns and indexes, can push projections, aggregation
- Join types supported
  - Nested Loop, Hash, Sort Merge, Nested Loop Push Down, Hash Push Down, Sort Merge Pushdown (Bloom filters)
  - Joins reordered based on arity, size, join ratio, connectness, substitutions applied
SYBASE IQ ARCHITECTURAL STRENGTH
Query Language and Stored Procedure Support

Pure ANSI SQL based
- Covers SQL-99, SQL-2003, SQL-2008 specifications with few restrictions
- Many useful vendor extensions
- Extensive OLAP support
  - Windowing aggregation
  - Ranking functions
  - Statistical functions
  - Distribution functions
  - Numeric functions

Stored Procedures
- Extensive support – both Watcom SQL and T-SQL
- Security enabled: access control and execution context
- Used for customization of key functionalities such as login password verification logic
- Used for user events -
  - fully configurable, scheduled or system triggered execution of active content
SYBASE IQ ARCHITECTURAL STRENGTH
Query Processing – XML/Graphical Plan Tracker

Double Bar Indicates Parallel Dataflow

1 row
#07 Root
1 row
#08 Group By Single
88 rows
#102 Parallel Combiner (ordered)
88 rows
#06 Group By Single
1 row (est.)
7484260 rows
#03 Join (Sort-Merge)
7484260 rows 19999806 rows
#106 Order By (Parallel inputs and outputs)
7484260 rows
#02 Leaf <lineitem>
#117 Order By (Parallel inputs and outputs)
20000000 rows
#01 Leaf <part>
Max Threads: 14
Max Threads: 7

Connector Width Varies With Row Count

ToolTip for Estimated Row Count

ToolTip for Max Thread Count

Node Depth Varies with Max Thread Count
SYBASE IQ ARCHITECTURAL STRENGTH
In Database Analytics

- Data NEVER leaves the database until results are materialized
- Models are SHAREABLE and allow AD-HOC analysis
- Models applicable to the LATEST data set
- PRIVACY protection is ensured
- STANDARDS based access
- PERFORMANCE and SCALABILITY is a given
- AVERAGE developer able to build models

Partner plug-in library via C++ UDF offers a rich set of functionalities to solve problems such as
- Classification e.g. Neural Networks,
- Clustering e.g. K-Means Clustering
- Statistical Simulation e.g. Monte Carlo
- Continuous/Categorical Predictions e.g. Linear / Logistic Regression
- Many more

Sybase IQ 15.1 In-database Analytics:
Logic to data = FAST + EFFICIENT

Enables concurrent high performance advanced analytics on large data sets on consistent models

External high performance C++ analytics libraries can be registered and invoked from Sybase IQ
SYBASE IQ ARCHITECTURAL STRENGTH
Load Engine

- Loading can be in multiple modes
  - Bulk (Load from files, remote databases)
  - Incremental Bulk
  - Continuous / Trickle feed via microbatching (Change Data Capture)
- Page level snapshot versioning – no locks required (just table lock in-memory catalog) – allows concurrent loads and queries with no blockage
- Load from client machines
- ELT interface into Sybase IQ with load balancer, transactionality and web based monitoring support
SYBASE IQ ARCHITECTURAL STRENGTH

Multiplex Grid – Independent Scale Up/Scale Out for Concurrent Performance

- IQ Database Size is a function of available Storage and is not tied to the Number of Nodes or CPU’s
- Single copy of the IQ Database shared through the SAN across multiple computer nodes
- All data and indexes are stored in the IQ Database
- Additional CPUs scale linearly when added to existing nodes
- IQ automatically spreads data and indexes across all SAN devices
- Individual nodes can have different configurations (CPUs, memory)
- Each node manages its own local temp space and catalog
- The IQ Writer Nodes are used for loading data into the IQ Database No data redistribution required
- Start small and grow HUGE
- Load balancing can be used to spread out users across available nodes
SYBASE IQ ARCHITECTURAL STRENGTH
Multiplex Grid And Virtual Backup – Foundation for HA-DR

- IQ Multiplex for Scalability and High Availability
  - SAN Storage enables IQ Multiplex - multiple nodes access SAN LUNs
  - No software required between EMC Storage and Sybase IQ
  - Ability to script adding database storage and creation of Multiplex nodes
- IQ Virtual Backup
  - IQ provides easy integration with EMC Storage Software
    - Tight coordination b/w database backup and storage commands
    - Able to use ATA grade storage for database copies
  - EMC Clariion (Snapview, SAN Copy)
    - Many Sybase IQ Customers use this Methodology
    - Joint Sybase & EMC Whitepaper
  - EMC Symmetrix (SRDF, Timefinder, SAN Copy)
    - Largest Sybase IQ deployments on EMC Storage
  - Fast database restore of Storage Copies
    - Verify Backup
- IQ Disaster Recovery
  - Storage Level Replication in conjunction with Sybase IQ Virtual Backup
SYBASE IQ ARCHITECTURAL STRENGTH
Virtual Backup – a brief outline

- Verify Backups
  - Quick Restore
  - Integrity Checking
- Testing Upgrades
  - IQ Major
  - IQ Major
  - Application Upgrades
- Development Copy
  - Avoid developing against Production
  - Test against full size data
- End User Playground
  - Run any query
SYBASE IQ ARCHITECTURAL STRENGTH
Information Lifecycle Management

Manage large data sets to ensure peak performance, cost savings and regulatory safeguards

Define, generate, track, administer object lifecycle policy (table in this release)

Optimized for storage efficiencies and regulatory retention

Configurable table spaces, range partitioning lower TCO for VLDB Mgmt
Partitioning of Data
- Tables can be partitioned
- Objects – tables, columns, partitions and indexes can be placed (into dbspaces)
- Dbspaces can be placed in different tiers and backed up / restored independently, marked read only
- Partitioning is by range only - purely targeted at Information lifecycle management

4 DBSpaces
Now assume that DSpace_4 is the default

create table emp_id (id int) partition by range(id)
(DBSpace_1 values <= (1000),
DBSpace_2 values > (1000))

Note: unless otherwise specified
All indexes will be placed on the default DBSpace (DBSpace_4)
SYBASE IQ ARCHITECTURAL STRENGTH
Overview – High Performance, Scalable, Concurrent Analytics

- High Velocity Real-Time Loads
- High Performance Batch ETLs
- Super Fast Parallel Analytics
- High Concurrency Reports
- Flexible, Direct Client Loads and Queries

Kerberos – Authenticated
ECC/RSA/FIPS-Encrypted
Connectivity

High Speed Interconnect

Active Store
Near Line Store
Historical Store

Shared, compressed, partitioned columnar store

GUI-based System Administration Console
SYBASE IQ ANALYTICS SERVER

Summary

- Market leading column store analytics product with proven success in several segments
- Very healthy new customer acquisition and product adoption trend
- Strong technical foundation for dependable, large scale data warehousing and analytics