SYBASE IQ

SYBASE*

ANALYTICS SERVER

Sybase Inc

March, 2010

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

Strategic

Predict the Business Advanced Analytics

Deep and complex analysis of large datasets

Sybase IQ=very fast complex and ad hoc query processing

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

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

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



3 – Sybase Inc – March 8, 2010

Operational

SYBASE IQ PRODUCT PROFILE



Inside:

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 indexbased 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

CATEGORY	ТҮРЕ	USAGE	• Most columns will have at least one index
Bitmap	Fast Projection	Compressed raw data for result sets (Default)	 Index selection decisions based on column cardinality (number of unique values)
	Low Fast	Low cardinality data (up to 1000 unique values)	 Indexes and columns are stored separately Multiple indexes used to resolve a
	High Non- Group	Aggregation on the fly and range searches	 query Indexes are self maintaining No optimizer statistics to update
	Date, Time, DT	Date ranges, date part operations	 Indexes are compressed Index building is a relatively simple
	Multi-Column	Concatenated indexes	exerciseIndex advisor tool guides users in
	Compare	Column comparisons	the building of appropriate indexes based on a run of sample queries
Traditional B-tree	High Group	Key fields and groupings for cross- tabular	
	Word	Key word or phrase string searches	SYBASE

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)

 $FP(1) = 2^8 = 256$ Unique Values

FP(2) = 2^16 = 65,536 Unique Values

FP(3) = 2^24 = 16,777,216 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

Date	Store	State	Class	Sales	Date	Store	State	Cla	ss Sales	s = NY	Class =A
3/1	32	NY	А	6	3/1	32	NY	Α	6	1	1
3/1	36	MA	Α	9	3/1	36	MA	Α	9	0	1
3/1	38	NY	В	5	3/1	38	NY	В	5	1	0
3/1	41	CT	Α	11	3/1	41	CT	Α	11	0	1
3/1	43	NY	Α	9	3/1	43	NY	Α	9	1	1
3/1	46	RI	В	3	3/1	46	RI	в	3	0	0
3/1	47	CT	В	7	3/1	47	CT	В	7	0	0
3/1	49	NY	Α	12	3/1	49	NY	Α	12	1	1

• Sample query

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

32	NY	Α	6
36	MA	Α	9
38	NY	В	5
41	CT	Α	11
43	NY	А	9
46	RI	В	3
47	CT	В	7
49	NY	Α	12
	32 36 38 41 43 46 47 49	32 NY 36 MA 38 NY 41 CT 43 NY 46 RI 47 CT 49 NY	32 NY A 36 MA A 38 NY B 41 CT A 43 NY A 46 RI B 47 CT B 49 NY A

Store State Class Sales

Date Store State Class Sales 32 3/1NY А 6 3/136 MA Α 9 3/138 NY В 5 3/141 Α 11 CT 3/143 NY 9 Α 3/1RI 46 В 3 3/147 В 7 CT 3/149 NY 12



Sales in binary form 8bit 4bit 2bit 1bit



• Example:

Date

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



SYBASE IQ ARCHITECTURAL STRENGTH

Leading Compression Technology

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



Ordinal	Value
1	"Wide Data"
2	"Wider Data"

Со	lumn
2	
1	
1	
2	
1	
1	



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





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



DMX-3

Storagescope SanCopy SRDF Timefinder Mirrorview Snapview



CLARIION



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

eneral	Phases	ables Notes	X A V	wes Preview	
	Name 🕶	Lifecycle	Start date	Data movement	~
+	Table_1 Table_3	Lifecycle_1 Lifecycle_1	2009-3-7 2009-3-16	1	
					×
					>

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



SYBASE IQ ARCHITECTURAL STRENGTH Information Lifecycle Management

- 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 lifecyle management



SYBASE IQ ARCHITECTURAL STRENGTH Overview – High Performance, Scalable, Concurrent Analytics





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

