MICROSTRATEGY World 2008

Evolving Business Intelligence Using Column Store Databases

Sybase IQ™

Presented by: Bill Jacobs, Evangelism Manager,
Sr. Mgr. of Evangelism
Sybase, Inc.

Date: 15 January, 2008

Pervasive BI:

News Flash: BI Goes Mission Critical

But We Already Knew That…

Growing Application Areas

- Algorithmic Trading
- Risk Analytics
- Real-Time Fraud Detection
- "Instant" Customer Scoring and Behavior
- Web Analytics
- Traffic Analysis

New Pressures

- Exploding Volumes
- Demand for "Right-Time"
- Continuous Loading
- 24x7 Availability
- Fast Disaster Recovery
- Security & Privacy
- Exploding User Populations
- Divergent Usage Profiles
- Not-so-Growing Budgets

The Breaking Point for Some?

How Does Your Platform Stack Up?

- Most OLTP systems are stretched to beyond capacity:
  - Optimized for transactions, not reporting or analytics
  - DBA tuning expertise needed
  - Data volumes explode from indices & aggregate tables
  - Load window margins are shrinking
- DW hardware appliances, while fast for many jobs, remain expensive, complex and can be inflexible
- OLAP engines are fast but somewhat inflexible; often needing tuned, bounded queries.

Analytics Architecture Alternatives

Data management technologies for high-demand analytics

TX OLTP database

Flexibility in querying, but large data stores, complex queries and large user loads will adversely impact system performance.

Examples: ASE, Microsoft SQL Server, Oracle, IBM

PC Parallel computing

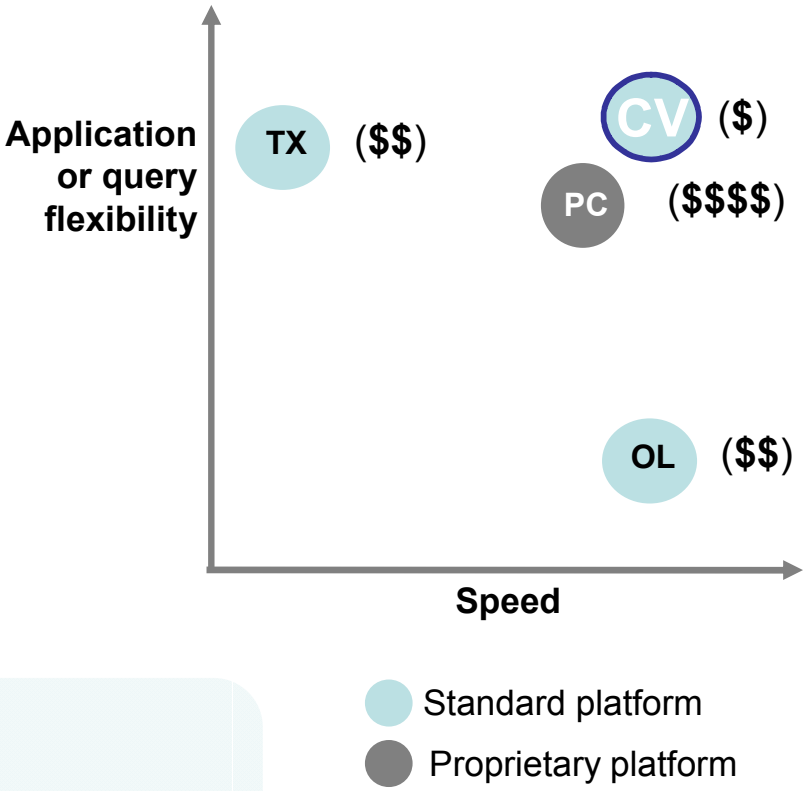
Speed and flexibility in querying, but will require significant hardware expense for proprietary platform. Complex query from one user will impact all.

Examples: Teradata, Netezza

OL OLAP

Speed in querying, but low flexibility - design requires advance understanding of what will be asked.

Examples: Hyperion, Microsoft, Cognos, Business Objects



CV Column Vector Databases

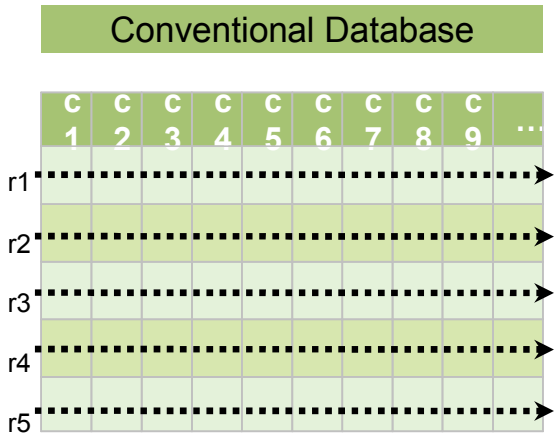
Speed and flexibility in querying. Commodity Hardware & Operating Systems. Near linear scalability in support of large user populations.

Column Vector Databases Are All The Rage

> The column-based approach is up to 100 times faster than traditional row-based databases. Sybase IQ is the most notable of these and, with almost 1,000 customers, the most experienced of the data warehousing providers to support this approach.

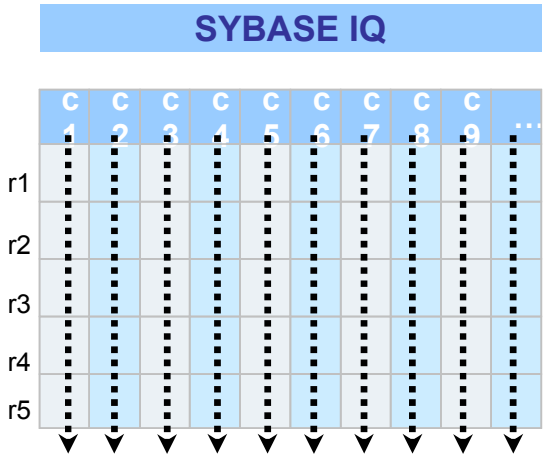
–Philip Howard
Bloor Research

Comparing Column Vector RDBMSs with OLTP RDBMSs



Traditional RDBMSs

- Data is stored horizontally
- Querying without indexes and views is extremely I/O intensive
- Building indexes and views is a huge time and resource drain
- Database footprint must be dramatically expanded to make the environment efficient for querying



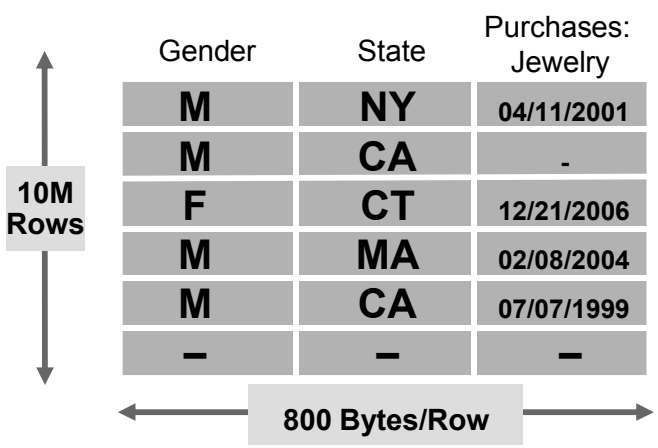
Column Vector Databases

- Data is stored vertically – Each column is stored separately
- The data is the index
- Retrieve only columns used in the query – Reduce System I/O dramatically
- Allocate a thread for each column individually – *Process the query in parallel*

The Column-Vector Advantage: Potential for Reduced IO

An online shopping service has millions of subscribers. Valentine's Day is coming. How many male customers in New York have purchased jewelry in the last three years?

Row-based Database

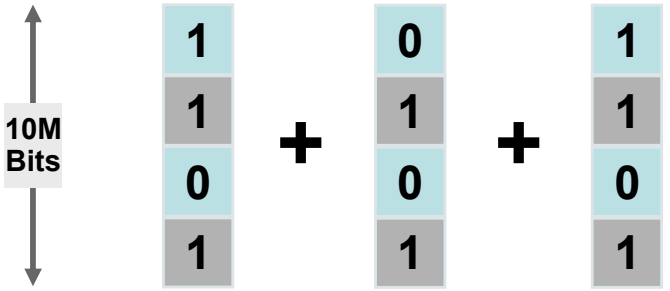


$$\frac{\text{800 Bytes x 10M}}{\text{16K Page}} = \textbf{500,000 I/Os}$$

- Process large amounts of unused data
- Often requires full table scan

Column-based Sybase IQ

$$\frac{\text{10M Bits x 3 Column/8 Bits}}{\text{16K Page}} = \textbf{234 I/Os}$$



Introducing Sybase IQ

Sybase IQ is an analytics and BI server optimized fast query response, fast loading, linear scalability, and powerful data compression that runs on commodity hardware and operating systems

- Designed from the start for analytics and reporting
- Unsurpassed query performance
- Lowest TCO for analytical and reporting environments
- Column-oriented storage and patented indexing
- Ranked in Gartner's Magic Quadrant as "Visionary."[1]

[1] Magic Quadrant for Data Warehouse Database Management Systems 2007. Gartner Group, October, 2007.

Sybase IQ: Designed for Analytics

Data is Stored "Vertically" in column vectors.

- Each column is stored separately
  - Bit-Mapped Indices
  - Index every column
  - Fast for both query & load

Optimized Storage

- Input data is compressed
  - Usually = 40%-70%
- Database size < raw input data
- Fewer Drives = Less Power, A/C and Failures

Query Engine Retrieves Only Columns Used in the Query

- Dramatically reduces I/O
  - Average 90% < OLTP
- Permits better data manipulation
  - Easy to alter and manage

Schema Not Restricted

- Flat, Star, Relational, Snowflake
- Nomalized or denormalized
- Effective ad hoc query solution

Sybase IQ™

Technical Innovations

Key Innovations in Sybase IQ

- Column-Vector Storage
- Rich Indexing Including Patented Bit-Maps
- Powerful Compression
- Unlimited Scalability - IQ Multiplex

Indexing: Designed for Analytics

9 Specialized Index Types

- Low or High Cardinality Data
- Bitmapped Comparisions
- Date & Time Columns
- Text Including Tokenization
- Fast Aggregations

Index Advisor Helps in Selection

Advantages:

- Index sizes remain small
- Index tuning is data dependent, not query dependent.
- Entire DB is indexed
- Queries resolve using only req'd indices – reducing I/O

Type of Query Usage	Recommended Index Type
In a **SELECT** projection list	Default
In calculation expressions such as **SUM**(A+B)	Default
As **AVG**/**SUM** argument	**HNG, LF, HG**, Default
As **MIN**/**MAX** argument	**LF, HG, HNG**
As **COUNT** argument	**LF, HG**
As **COUNT DISTINCT, SELECT DISTINCT** or **GROUP BY** argument	**LF, HG**, Default
As analytical function argument	**LF**, Default
If field does not allow duplicates	**HG**
Columns used in ad hoc join	Default, **HG, LF,**
Columns used in a join index	**HG, LF**
As **LIKE** argument in a **WHERE** clause	Default
As **IN** argument	**HG, LF**
In equality or inequality (=, !=)	**HG, LF**; also **CMP**
In range predicate in **WHERE** clause (>, <, >=, <=, **BETWEEN**	**LF, HG**, or **HNG**; also **CMP, DATE, TIME, DTTM**
In **DATEPART** equality, range, and IN list predicates	**DATE, TIME, DTTM**

Index Advisor: Aggregated Indexing Advice

Retrieves aggregated index advice messages

- Displays Advice, Number of Instances and Last Date issued

SQL Statements

```
exec dbo.sp_iqindexadvice
```

Results

	Advice	NInst	LastDT
1	Add LF or HG index on DBA.loutest.col_1 Predicate: (loutest.col_1 = 'abc')	2	2006-07-18 10:54:31.000000
2	Add a HG index to join key column DBA.SPLIT_EVENT.EFFECTIVE_DATE	4	2006-07-18 11:06:12.000000
3	Add a LF index to grouping column DBA.hist6_temp.TRADE_DATE	2	2006-07-18 11:06:36.000000
4	Add a LF index to grouping column DBA.hist_temp.TRADE_DATE	4	2006-07-18 11:06:23.000000
5	Add a unique HG index to join key column DBA.#temp_tick3a.INSTRUMENT_ID	2	2006-07-18 11:00:37.000000
6	Add a unique HG index to join key column DBA.#temp_tick3b.INSTRUMENT_ID	2	2006-07-18 11:00:37.000000
7	Convert HG index on DBA.DIVIDEND_EVENT.INSTRUMENT_ID to a unique HG	1	2006-07-18 11:05:14.000000
8	Convert HG index on DBA.INDEX_CMPSTN.INSTRUMENT_ID to a unique HG	3	2006-07-18 11:05:52.000000
9	Convert HG index on DBA.hist_temp.row_nbr to a unique HG	8	2006-07-18 11:06:23.000000

Indexing: Compressed Bit-Mapped Indexes

Row-based RDBMS

Index on Price

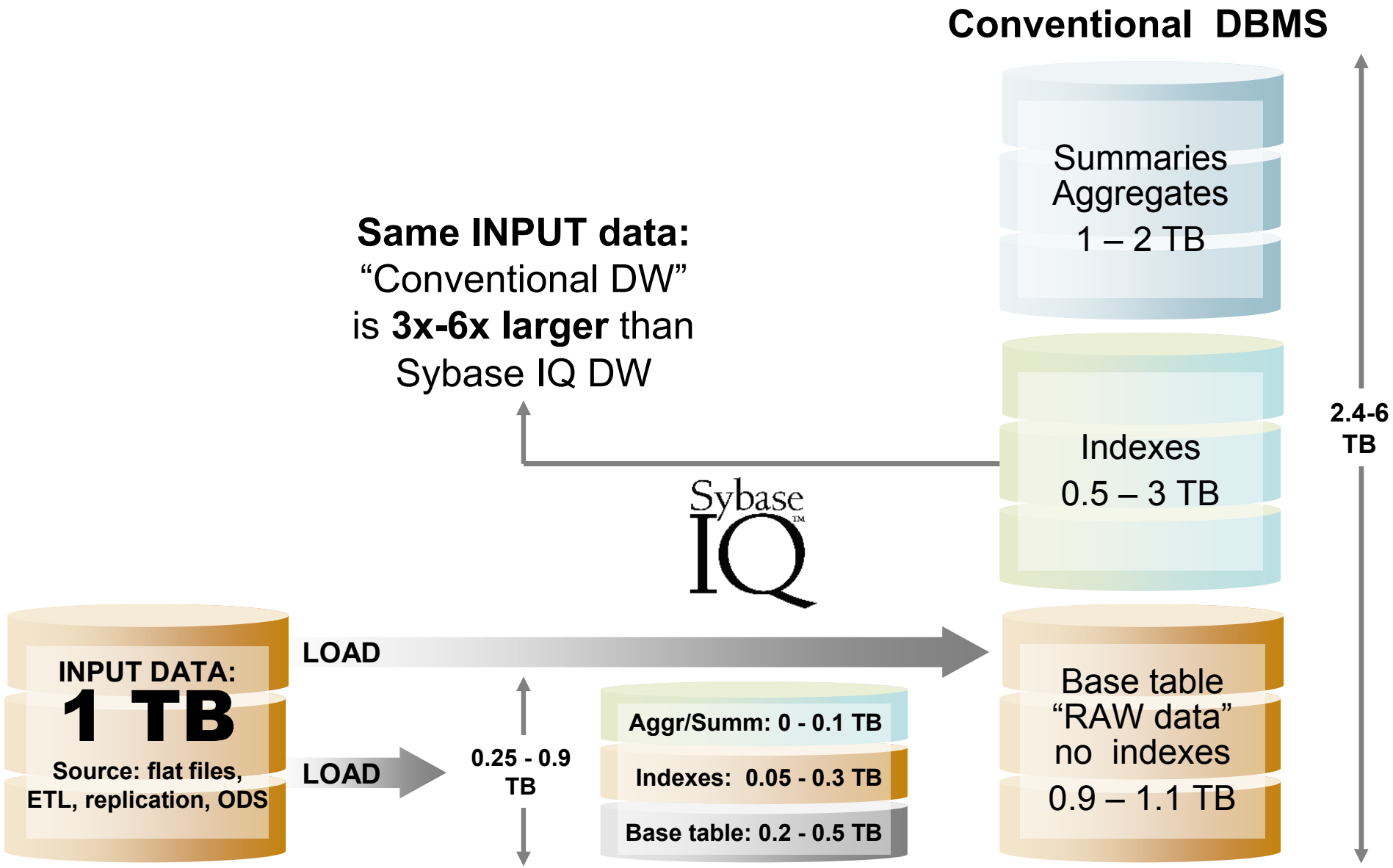
Gender	Price	Approved	Comments
Male	500	Yes	Hello W..
Male	20	No	Needs work
Male	30	No	Best Regards ..
Female	1000	Yes	Lots of accidents ..

Sybase IQ Column Based Storage (Bit-Wise)

Low Cardinality | High Group | Low Cardinality | Word

Gender	Price	Approved	Comments
1,1, 1	01001	1,1	[Hello] [W] [..]
2	10010	0,0	[Needs] [work] [..]
	11000		[Best] [Regards] [..]
	010101		[Lot's] [of] [accidents] [..]

Compression: Delivering the Lowest TCO



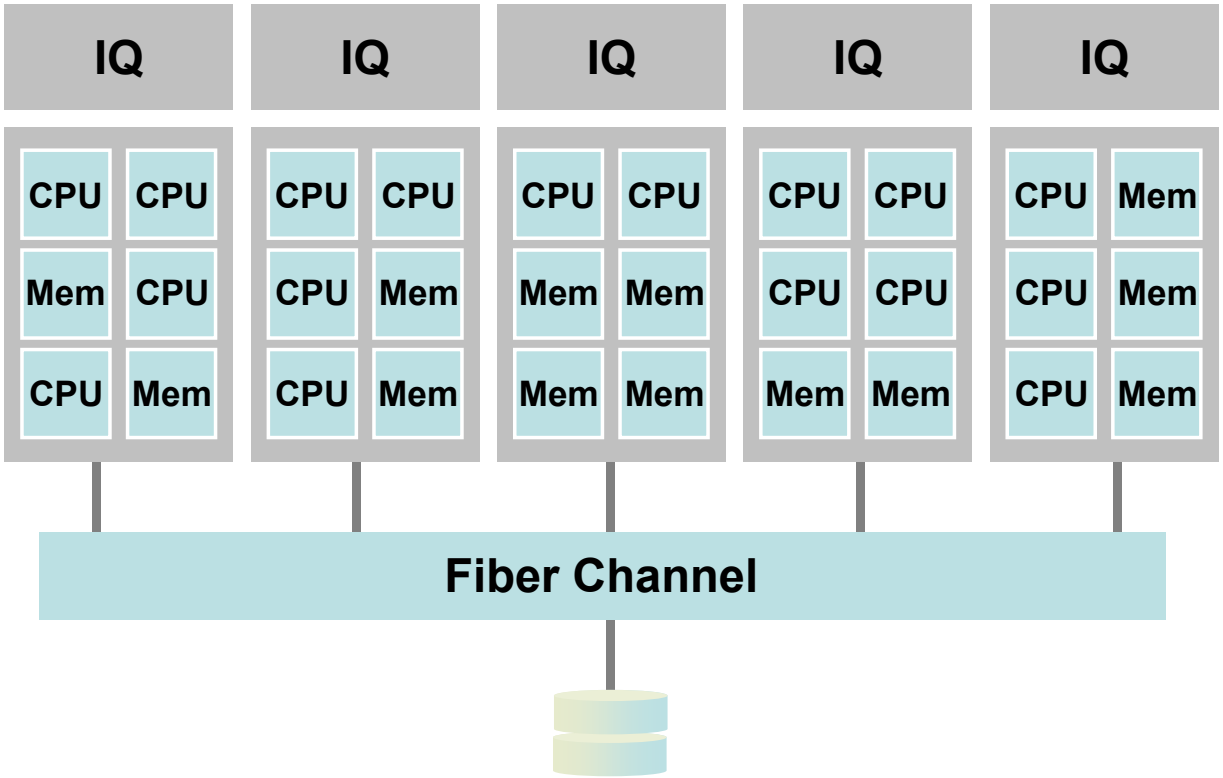
Compression Results Across Our Customer Base
Proven Storage Savings

4 – 10 times less storage

Save USD$1 million per terabyte of input data

SYBASE IQ DATA COMPRESSION EXAMPLES	Raw Data Loaded	Sybase IQ Compressed	Estimated Competitor's Data Explosion
VLDW Ref. Architecture (InfoSizing – June 2004)	155 TB	55 TB	500 TB to 1,000 TB
Telefonica	70 TB	15 TB	210 TB to 490TB
comScore Networks	40 TB	16 TB	120 TB to 280 TB
Health Insurance Review Agency	27 TB	12 TB	81 TB to 189 TB
Samsung Card	15 TB	7 TB	45 TB to 105 TB
Nielsen Media Research	12 TB	12 TB	36 TB to 84 TB
Large Credit Card Company	10 TB	4 TB	30 TB to 70 TB

Scalability - IQ Multiplex
Dynamically Respond To Changing Loads



Can start with one server and add CPUs and memory as needed

Can add servers and CPUs with little or no loss in scalability

Can add terabytes of disk to the SAN and Sybase IQ will manage it efficiently

With Sybase IQ, you can support multiple users per CPU

64-bit architecture limits:

Max DB size: 192 petabyte
Max row per table: 286 trillion rows
Max nodes:12,000 nodes
Max CPUs: 1.2 million CPUs (+12,000*106)

The Sybase IQ Advantages

Unsurpassed Performance

- Query Speed – up to 100X OLTP
- Mixed Workloads
- Support 1000's of Concurrent Users

Low Cost

- Compresses Raw Data up to 70%
- Enhanced Reliability from lower Disk & CPU count
- Lower Data Center Costs for Power, Cooling & Footprint
- Easy to Learn and Use

Flexibility

- Schema Independence
- On-Demand Scalability via Multiplex

Capability

- VLDB Capacity
- Real-Time Latencies

Enough Vendor Speak. Introducing Peter Dobler.

Introducing…

Peter Dobler

Director of Database Technology
Nielsen Media Research
www.neilsenmediaresearch.com

And…

Peter Dobler

President
Dobler Consulting Inc
www.doblerconsulting.com

MICROSTRATEGY World 2008

Nielsen Media Research
Sybase IQ
Real World Example

Presented by: Peter Dobler

Date: January 15th 2008

Sybase IQ at Nielsen Media Research

Overview

- Sybase ASE
  - Main database engine used for all mission critical OLTP applications and some reporting engines.
- Sybase ASA
  - 2,700 fields reps keeping Nielsen's sample data up-to-date with this remote database solution and real time sync with the master database.
- Sybase Replication Server
  - Real time data movement tool of choice. Used for data synchronization and DR activities.
- Sybase IQ
  - Several dozen data warehouses with database sizes up to 22TB to power the main reporting systems.

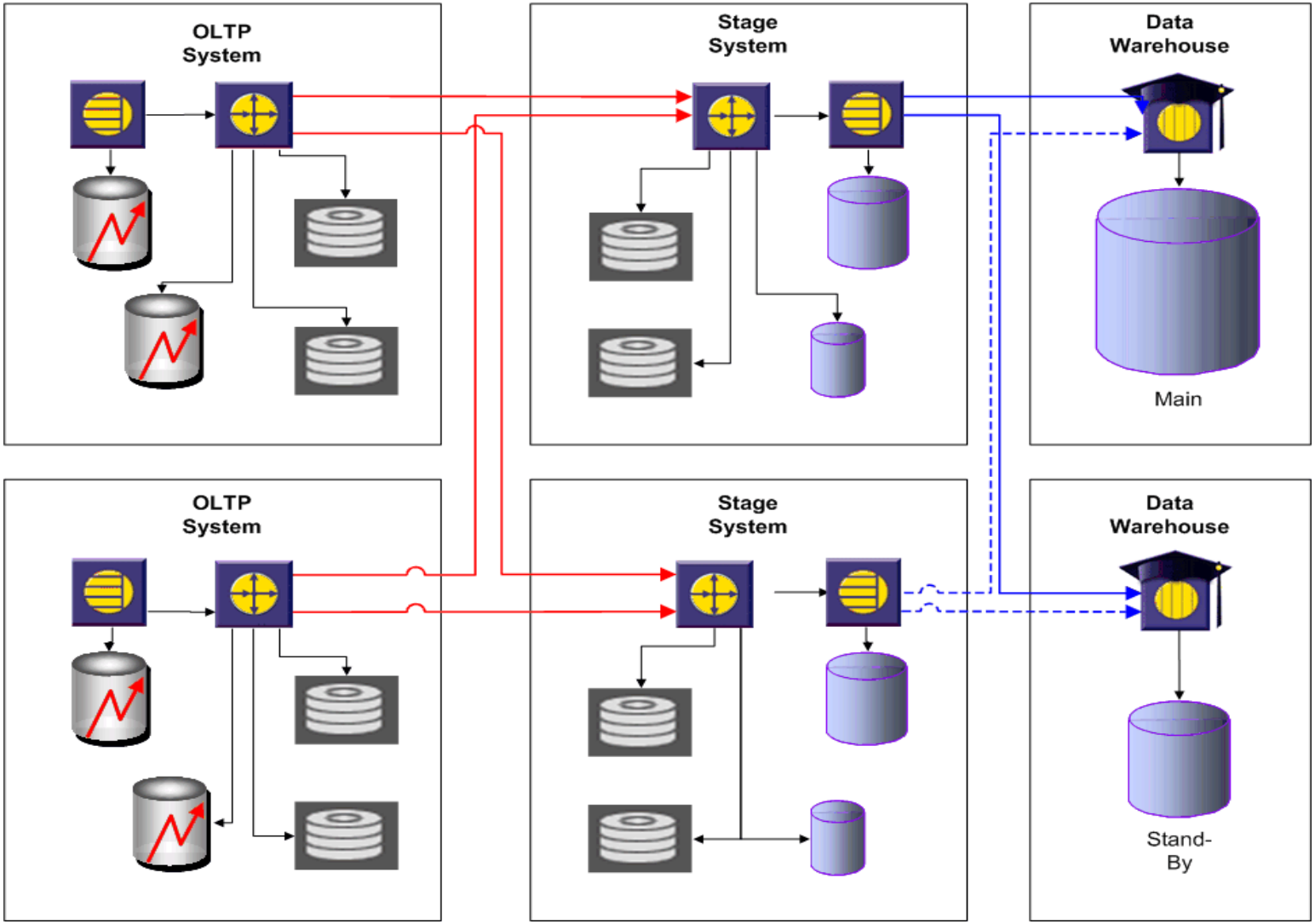
Sybase IQ at Nielsen Media Research

Data Warehouse System Criteria

- The decision to use Sybase IQ was made in 2000 based on the following criteria;
  - High speed load
    - Sybase IQ achieved 300,000 rows a second and a SUN 450.
    - Today the load speed is around 3 million rows a second.
  - Data Compression
    - The POC provided 60% data compression on our data.
    - Today compression rates over 85% are possible.
  - Low TOC
    - With the enormous data compression the storage footprint and cost was significantly reduced. And the high speed data load on inexpensive hardware saves cost on server hardware.

Sybase IQ at Nielsen Media Research

System Architecture of an Sybase IQ Environment



Sybase IQ at Nielsen Media Research

Summary

- Nielsen Media Research is using Sybase IQ since 2000.
- Several mission critical applications are running on Sybase IQ. Including our main client data delivery system.
- Thanks to Sybase IQ's massive data compression and column oriented data structure we are able to run Sybase IQ on a much lower TCO than other data warehouse systems. Less CPU power and less storage needed. This reduces the footprint in the data center.
- Sybase IQ's Multiplex enables us to add computing power at peak times like premier season, Super Bowl, Olympics, etc. and redeploy this computing power when no longer needed.

Sybase IQ is a Proven Solution

Who else can claim this kind of experience?

- > 1900 customer projects at 900 sites worldwide
- 145 new enterprise application wins in last 12 months

Pick-n-Pay:
Reporting Solution for Major Grocery Retailer

Background / Issues

- Increased data volumes and concurrent user query and reporting demands
- 32-processor Teradata warehouse could not provide acceptable performance at an acceptable cost
- Needed a highly cost-effective, scalable reporting solution that could integrate data from many operational systems

Results of the Sybase IQ Implementation

- Reduces annual costs
- Improves efficiency and access to live operational data
- Enhances performance and query response times
- Integrates easily with SAP ERP System

Pick'n Pay

"Even in the early stages of development, IQ is already delivering improved performance in corporate reporting."

Harold Dawson,
IT Director
Pick 'n Pay

Barclays Global Investors
Sybase IQ Analytics Accelerator Revitalized Apps

Business Issues

- Multi-axis investment simulations integrate historical time-series financial data points across a large number of samples for each point
- Some queries exceeded 30 hours; others used so many database resources and took so long they did not complete in a reasonable timeframe for relevance
- Daily loads taking 25 hrs – catching up on weekends

Results

- Query performance increased between 70% and an extreme case of 148,000% – average increase was 800%
- In several cases, queries that took 20 minutes or more on the production system took less than a second on Sybase IQ
- Sybase IQ reduced daily load times from 25 hours to 1½ hours
- Two applications running on IQ; planning to add three more

" I think many of the human productivity gains are immeasurable because the applications we were dealing with on OLTP were struggling or stalling. Sybase IQ not only performed better, it literally rejuvenated those applications."

Tom Lu
Lead DBA
Barclays Global Investors

BARCLAYS

comScore:
eCommerce Customer Behavior for 1M Web Users

Background / Issues:

- A Provider of customer behavior and preferences data for the eCommerce marketplace.
- They track the surfing and buying behavior of more than 1 million Web users.
- They selected Sybase IQ as the most cost-effective way to get the business performance it needed.

Results

- “The speed of Sybase IQ improves our ability to mine the data and produce results for our customers much more quickly. That helps them market more effectively and generate more business.”
- “We need to make sure we can economically scale to large amounts of data and support data-intense reporting while not incurring significant costs,” “Sybase IQ’s multi-node capabilities make this possible and give us easy-to-manage query, data-load and data-mining functions.”

- Ric Elert, VP of Engineering, comScore

Neos Banca:
Financial Services for Millions of Customers

Background / Issues

- System serving more than four million customers across Italy
- Broad range of financial products
- In 2006, generated three billion euros worth of consumer loans
- Needed to optimize analytic performance without impacting transactional performance

Results of the Sybase IQ Implementation

- Improves query response time
- Avoids significant data storage costs
- Integrates with existing technology assets easily
- Enhances ability to comply with Basel II and other regulatory standards
- Enables production of existing monthly reports on a weekly or daily basis

"Sybase IQ is a unique tool. It was designed from the ground up to support analytics and its query response time is extremely fast…Also, Sybase IQ compresses raw input data up to 70 percent compared to conventional transactional-based data warehouse or reporting systems. This frees up valuable amounts of disk space and eliminates the need to invest in additional storage capacity. Without Sybase IQ, we would have had to spend a great deal of money to purchase new storage hardware."

Paolo Simonini
Sr. Manager, New Technologies
Neos Banca

Allianz Insurance:
Sybase IQ Report Server in Record Time

Business Issues

- Meet a government-imposed deadline for implementing claims processing analytical reporting functionality
- Off-load analytics and reporting from existing claims processing system to protect transaction response times

Results

- Completely new analytical reporting system implemented in just eight weeks using Sybase IQ.
- Claims processing system's transactional response times reduced
- Opened possibilities for new software system upgrade initiatives. Initiatives, previously considered too time-consuming and costly are now in progress.
- Oracle site license not able to deliver results in the right time for customer's government regulation

"What Sybase IQ did was raise the awareness that a lot of the reporting and analytics … could be moved onto a Sybase IQ report server, dramatically speeding up those original processes."

Scott Wyld
IT Project Manager
Allianz Australia Insurance Limited

Shopzilla.com
Faster Web Analytics—Better Customer Service

Business Issues

- Well-known, shopping comparison site
- delivers analysis of over 30 million products from more than 60,000 merchants in 2,500 categories
- Offer merchants ability to completely customize their surveys
- Offer merchants ad-hoc analysis of 18 -24 months of survey results – Microsoft could not scale for this
- Offer real-time access to data

Business Results

- No query – not even the most complex, ad-hoc query – takes more than 10 – 15 seconds
- Maintained performance levels – even as data volumes tripled
- Flexible reporting system for merchants – columns added on-the-fly
- Scaleable solution that will grow – access to 18 to 36 months of survey data
- No significant increase in hardware investment – keeping TCO low

"Sybase packs an extraordinary amount of data processing and analytical power into a small footprint that represents a realistic investment for small and mid-sized firms"

Henri Asseily
CTO and Founder
Shopzilla

Yapi Kredi Bank: Sybase IQ Delivers Better Price/Performance

Business Issues:

- Manage and analyze immense amounts of data produced in their daily operations
- Oracle performance not meeting business needs – some queries taking up to a day to get results
- Develop a smooth, fast migration path
- Deploy before new credit card products and services were offered and advertised

Results:

- 8-week deployment
- 154% return on investment
- Queries that used to take a day are now completed in minutes
- 63% data compression – used 3 CPUs, compared to Oracle's 12 CPUs
- Overall Price/Performance far superior to Oracle

We selected Sybase IQ because it provided faster query results, reduced data storage requirements, and gave lab test performance results that were sometimes 10 times as good as the competitor.

Alpaslan Ozlu
General Manager of Technology Management
Yapi Kredi Bank

COMPUTERWORLD BUSINESS INTELLIGENCE PERSPECTIVES
Best Practices IN BUSINESS INTELLIGENCE
AWARDS PROGRAM

YAPI KREDI World 2008

U.S. Securities and Exchange Commission Mission-Critical Analytics—Disaster Recovery

Business Issues

- Needed a more efficient means to provide disclosure of important securities information
- Needed a more efficient means to enforce securities laws

Results

- New data warehouse built in less than 12 months
- 25% reduction in "off hour" data loads
- 35% improvements in query response time
- The ability to create and run more complex queries
- Disaster recovery capabilities with under 1 minute failover
- No unplanned down time
- Now incorporates documents and emails (unstructured data) into analysis with structured data
- Accommodates future growth

"Our success is directly due to the support we received from Sybase and the quick and easy implementation. As a result, this architecture is presented as a model for other divisions and agencies for data warehousing."

Samuel Foster
President, **FosterSoft**

FSI FOSTERSOFT INCORPORATED

Sybase IQ Highlights

Business Strengths

- 35% year over year growth for many years
- 1700 production installations
- Strong presence among companies that offer data analysis as a business service
- Customers in finance, healthcare, telco, government, retail, manufacturing

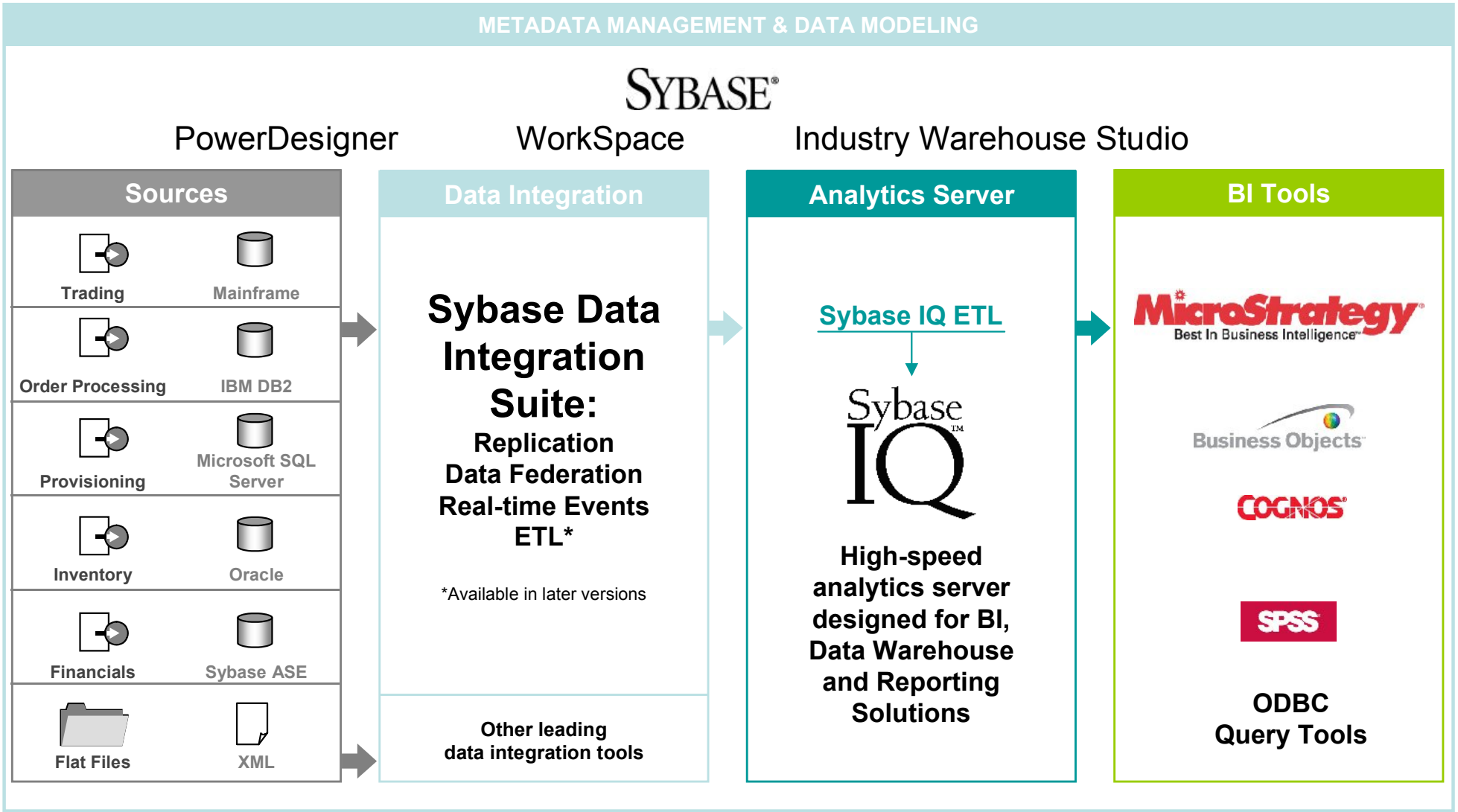
Technology Highlights

- #1 TPC-H price–performance in 100GB and 300GB categories
- Customer DB sizes – numerous multi-terabyte warehouses in production
- Audited 1000 concurrent complex query users against 10TB
- Audited benchmark with 1 trillion rows, 155TB input data

Customer Experience

- Euskaltel Data Centre
  - 600+ users
  - 10,000x faster queries
  - 75% less storage
- LoanPerformance
  - 100 x faster reports
- S&H Solutions
  - Query times accelerated 500%
  - 50% compression
- IRS ROI 995%, payback period 1 month, 5 year net benefit $76M

Sybase IQ Part of a Total Environment for Analytics



MicroStrategy and Sybase

Successful Partners, For Sure.

Existing Users? Many.

Links to Sybase IQ Customers Who Also Use MicroStrategy:

- **Nielsen:** http://www.sybase.com/detail?id=1035802
- **Pick n Pay:** http://www.sybase.com/detail?id=1036050
- **Chohung Bank:** http://www.sybase.com/detail?id=1025509
- **AOPC:** http://www.sybase.com/detail?id=1055477
- **HIRA:** http://www.sybase.com/detail?id=1033785
- **Redecard:** http://www.sybase.com/detail?id=1033829
- **Samsung Life:** http://www.sybase.com/detail?id=1034404
- **Sistema 4B:** http://www.sybase.com/detail?id=1038539

For More Information: Sybase.com

http://www.sybase.com/products/datawarehousing/sybaseiq

Questions?

IQ Release History

October 1994 – Sybase acquired Expressway Technology IQ

- Released version 11.0 in 1995
- During a 4 year period versions of the IQ Version 11.x product were released
- Concurrently, an engineering effort was underway to fold the IQ technology into the Sybase framework

February 1999 – Released IQ 12.0

- Integration with SQL Anywhere for language and interface capabilities
- Query optimizer and indexes redesigned
- Added JDBC and Open Client Support
- Added more datatypes

September 1999 – Released IQ 12.4

- Added more maintenance/DBA functionality
- Updated ODBC support to ODBC 3.5.1
- Improved join performance
- Improved query performance through optimizer enhancements

June 2000 – Released IQ 12.4.2

- Added new default (FastProjection) index types
- Improved performance of maintenance utilities
- Added character and binary functions
- Added OLAP functions
- Full implementation of the Multiplex (multi-node) capability
- Added update command support

May 2001 – Released IQ 12.4.3

- Added detailed query plan outputs including HTML output, pre- and post-execution plans, query naming
- Added parallel processing of Group By statements
- Added new CMP and WD indexes
- Improved union over view processing (view with UNION ALL statement)
- Multi-column primary keys
- Added new datatypes
- XML data extraction added
- Statistical functions added
- Update command now supports joins
- High speed data extraction utilities added
- OLEDB support

October 2002 – Released IQ 12.5

- Full referential integrity support (primary and foreign key relationships) with parent/child support