



SYBASE®

# Keys to Balancing Explosive Data Growth

JOSEPH SHAFFNER  
DIRECTOR OF TECHNICAL SALES, SYBASE

May 5, 2009

# With Explosive Data Growth, Challenges Arise

Can't get the insights about the business

What you don't know is costing you....in risk, resources, and knowledge

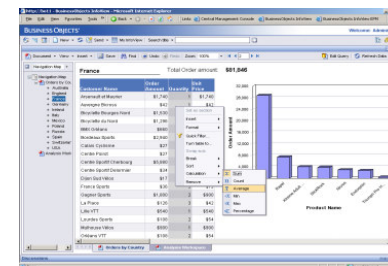
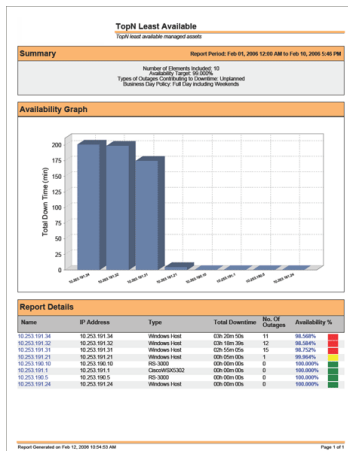


Source: *The Toxic Terabyte*, IBM

**SYBASE**<sup>®</sup>

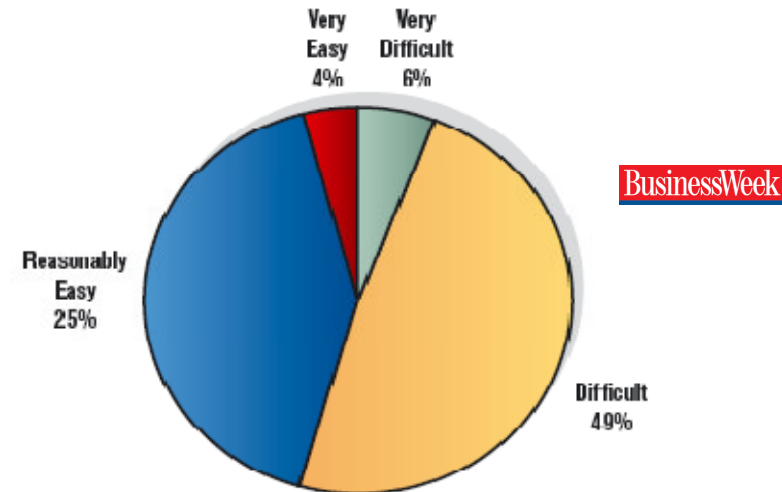
# Need For Knowledge

- Business Intelligence ranked top technology priority in 2009, 4th year in a row. (Source: Gartner Executive Programs survey 1,500+ CIOs)
- Leverage & expand existing BI investments
  - More pervasive
  - Actionable

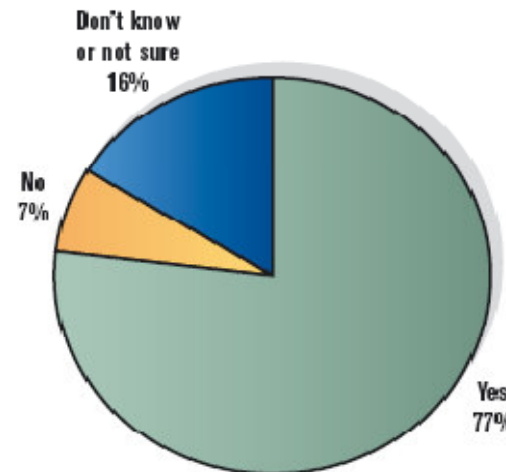


# Why are Businesses Not Fixing the Problem

- Cost of delivery is too high
  - Resources
  - Infrastructure
- Inconsistency in information across the data silos
- Using manual processes to integrate data into one version of truth
- Time to deliver is too slow



How difficult to get relevant corporate information to make business decisions

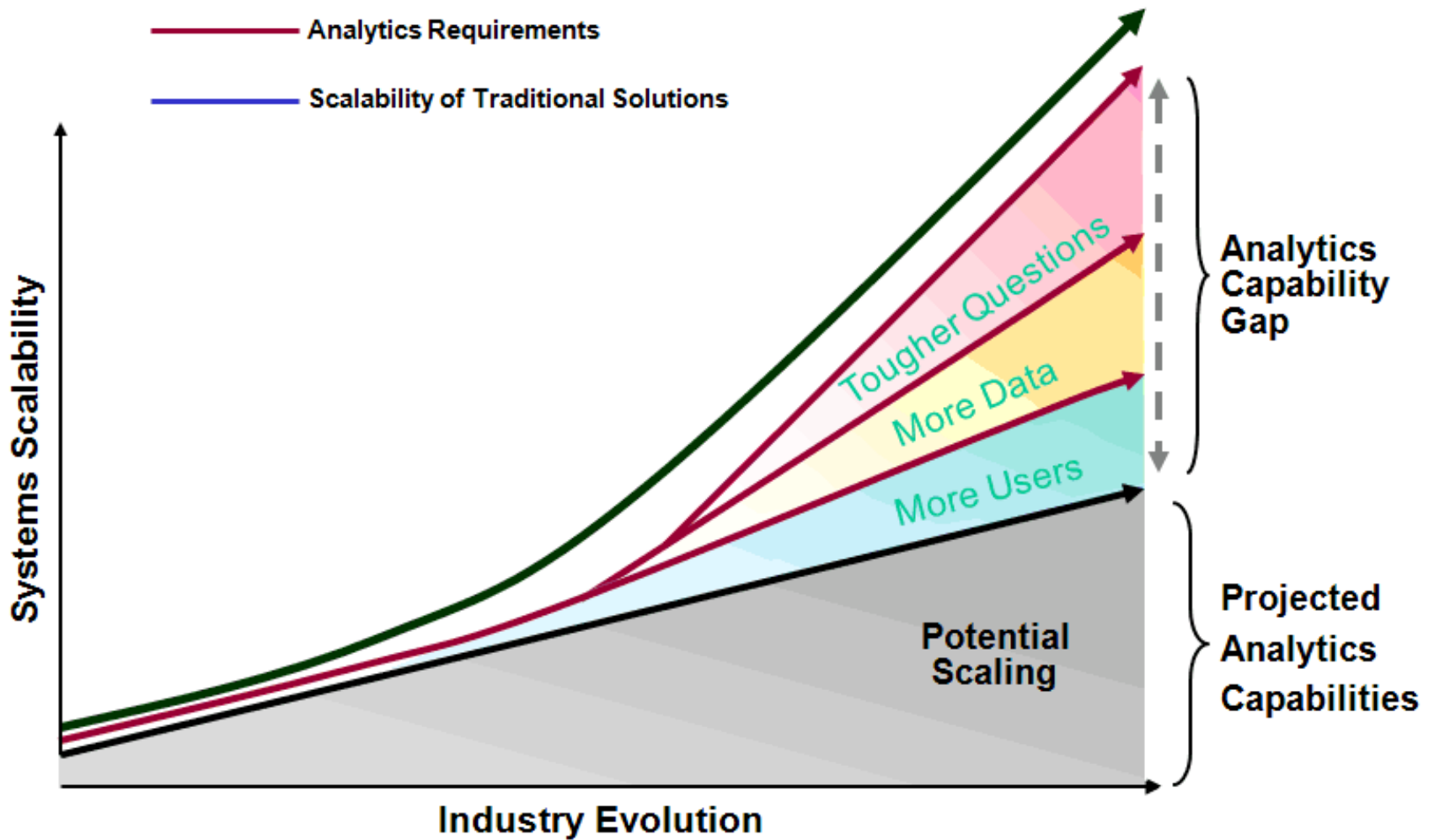


Aware of bad decisions managers have made due to insufficient information

SYBASE®



# The Gap



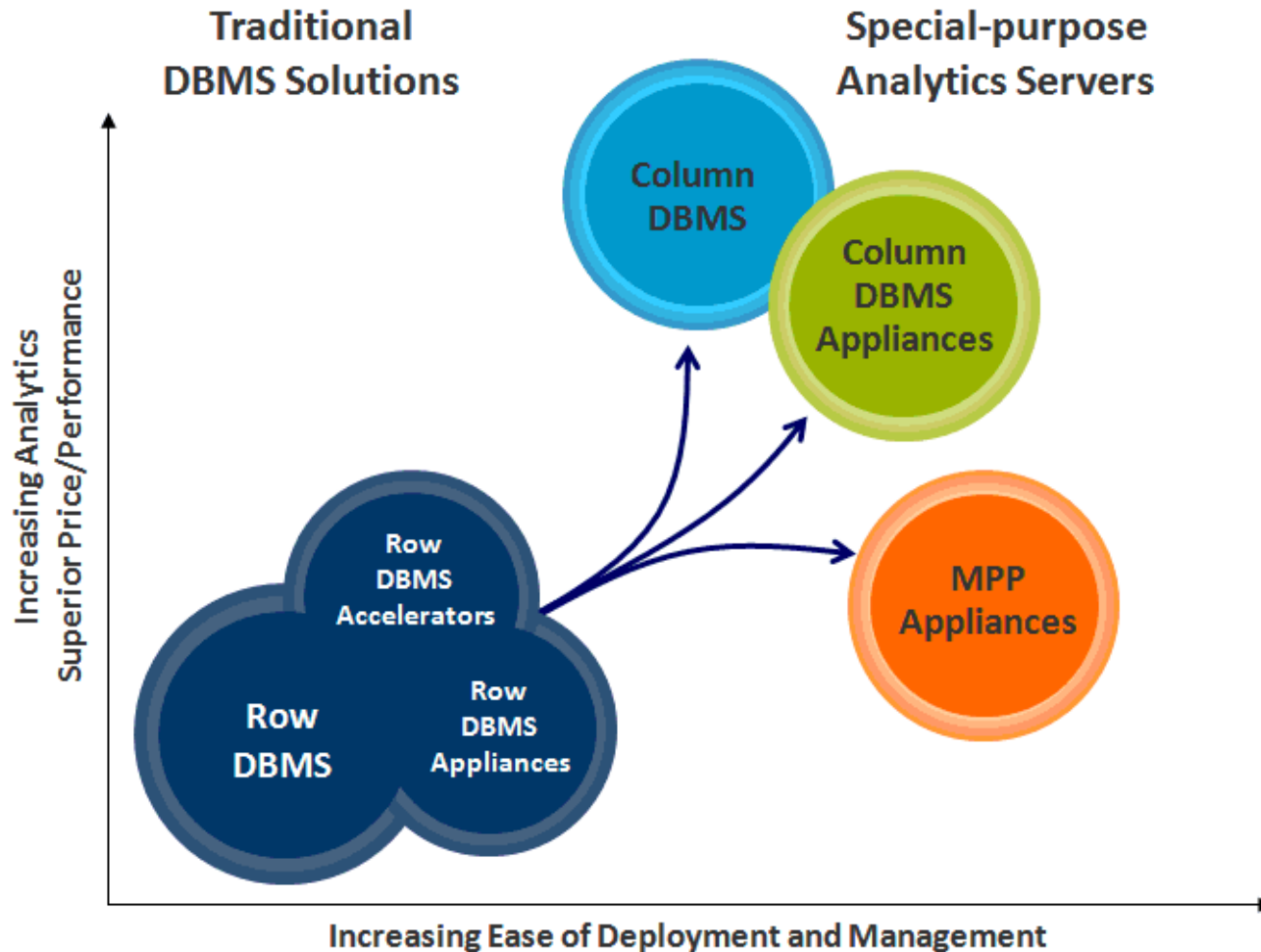
# Bridging the Gap: Something Has to Change



- Relational emerged in 1980 and mainstream 1990
  - Transactional
  - Data warehouse, Business Intelligence and Analytics
- Relational builds on transactional architecture for DW
  - More infrastructure
  - More complexity
  - More skills needs
  - “one size fits all”
- OLAP Emerge
  - Still requires specialized skills
  - Hard to scale

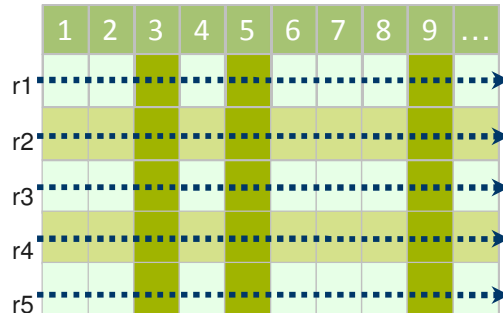
# Emerging Column-based DBMS can Bridge the Gap

Column-based Analytic Engines Can Manage the Explosive Data Growth

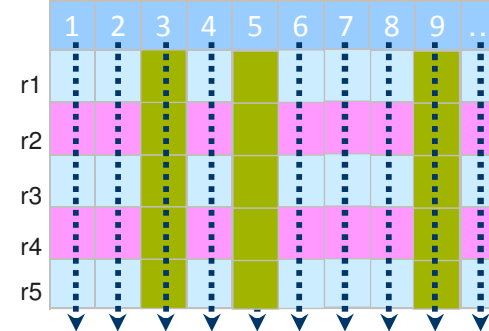


# Why Column-based DBMS for Analytics

Relational Database



Columnar Database

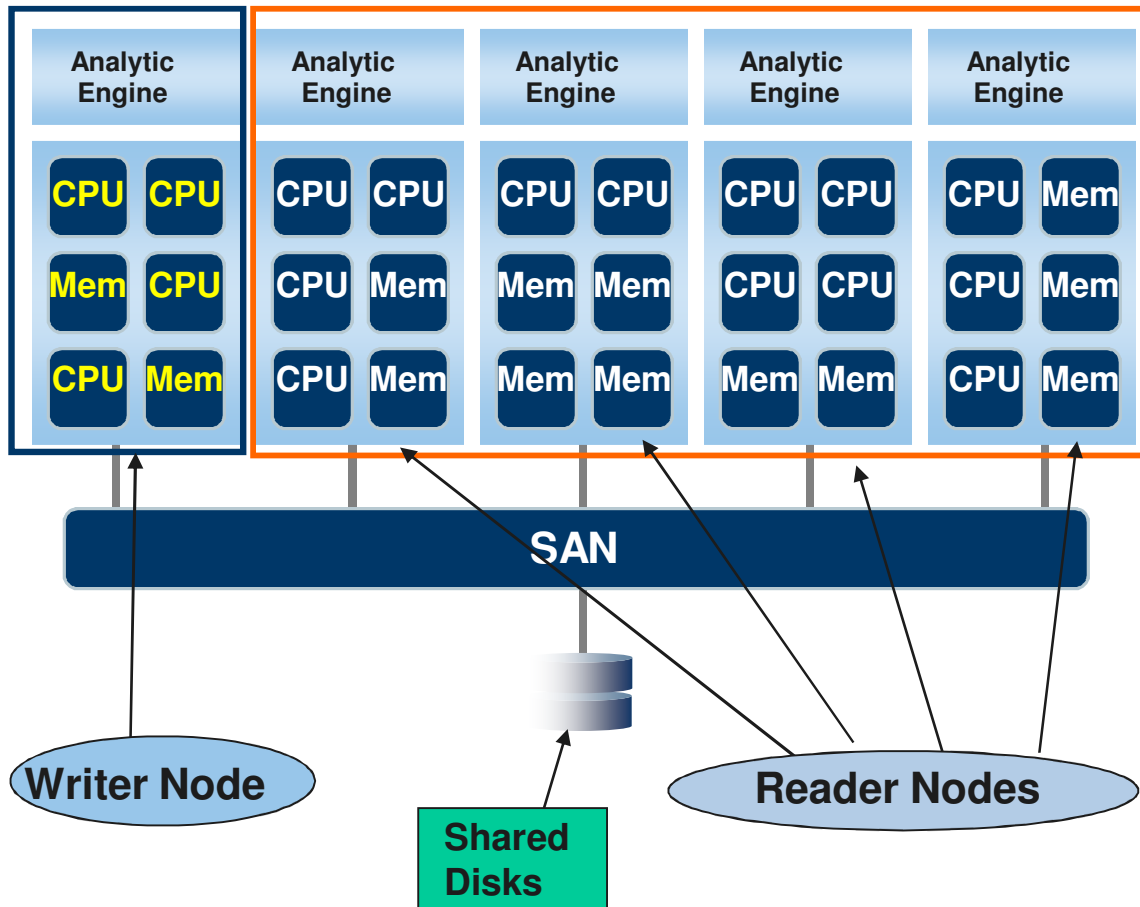


- Data is stored in rows (horizontally)
- Need indexes, partitions, etc for performance
- Much of the database space is taken up for indexes & growth
- 3NF vs. Dimensional
- Aggregations, Materialized Views & Denormalization

- Data is stored in columns (vertically)
- The data is the indexes
  - **Little to no resources required**
- Retrieve only columns used in the query
  - Reduce I/O
  - **Data Compression up to 70%**

# Shared Disk vs. MPP Architecture

Multiplexing architecture – concurrency and scalability



## Shared Disk Benefits

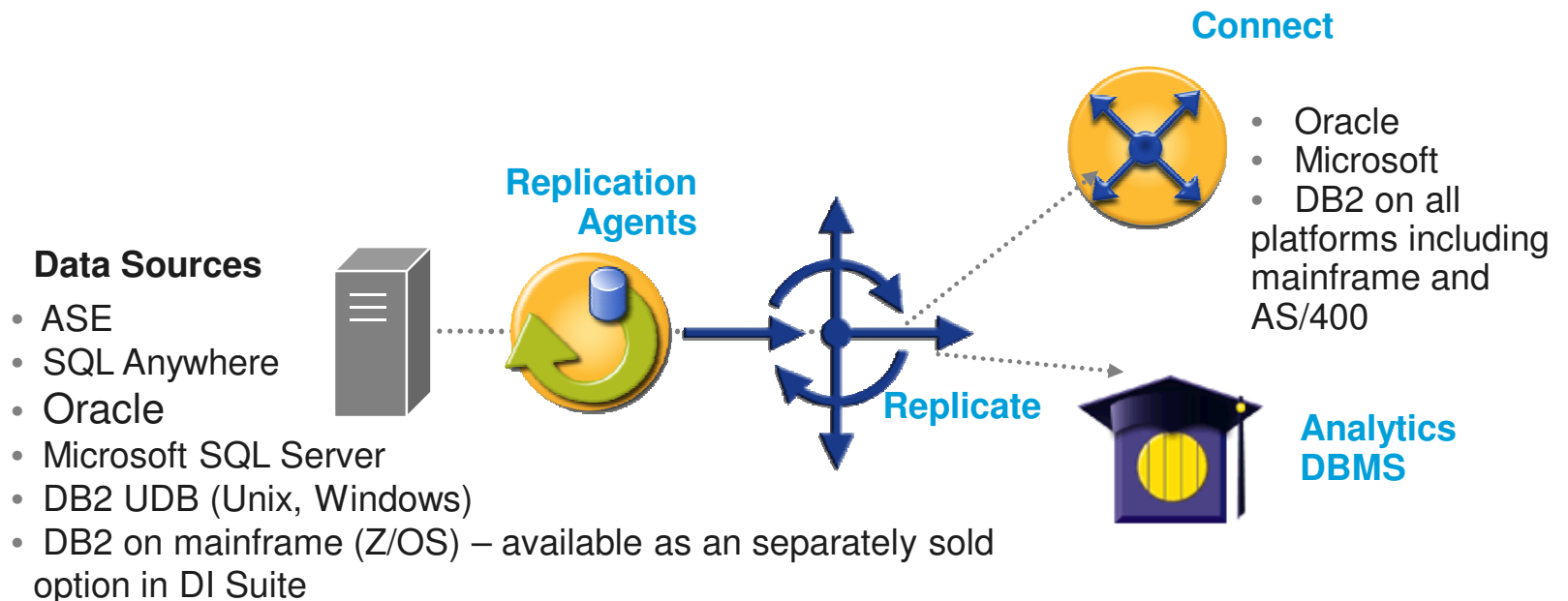
- One data store
- Shared Across Multiple Computer Nodes
- No Partitioning of Data
- No Distributed Lock Management
- Individual nodes are Independent of Other Nodes
- Add servers and CPUs – terabytes of disk – with little or no loss in performance

**One copy of data, backup is compressed, large savings in power, cooling and storage costs.**

# Replicate Data Assets Without Intensive Resources and Additional Storage

Look for...

- Non-intrusive transaction capture
- Flexible transformation of data
- Efficient routing across networks
- Real-time synchronization across heterogeneous DBs



# TELSTRACLEAR

## Business Issues

- Huge data volume – 3 billion rows of information, growing by 6 million rows per day
- Goal – Develop a reporting system across the enterprise to provide on demand reporting and business analytics against enough historical data to reflect historical trends

## Results

- Not able to solve need with existing technology
- Call reporting system stored and managed 13 months of data – at no additional storage cost
- Complex **report times reduced from two hours to two minutes**
- Business analysts able to create and request their own reports – without the need for IT intervention
- Increased capacity without additional hardware
- Low maintenance and DBA costs

*“Everybody was pretty amazed. We were hoping for a 50% decrease in query time. But we saw many times more than that when we started up the system. The Sybase IQ implementation far exceeded our expectations. The results were stunning!”*

Simon Falconer  
DBA  
TelstraClear



SYBASE



# Remarkable Gains

**“We have seen remarkable improvements in productivity, and big decreases in training, support, and hardware costs. Sybase IQ allows the IRS to better analyze data, and because of that, we are able to administer the tax code more fairly and equitably.”**



Jeff Butler  
Director of Research Databases  
Office of Research, US Internal Revenue Service



# Apply A New Approach to Analytics

## **BE SMARTER about the problem**

- Leverage existing investments by using replication for heterogeneous integration
- Continue using relational databases for your transactional systems
- Think about using columnar databases for analytics and reporting
- Evaluate compression technology to help manage the data volume



**THANK YOU!**