




Solving Performance Issues in BI through Purpose-built Analytical Databases

Silicon India BI Conference, Sept 24, 2011, Hyderabad

Vivek Bhatnagar
Action Corporation



Agenda

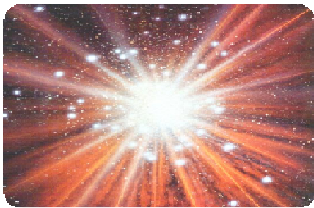
- ❑ **Today's Biggest Challenge in BI - Performance**
 - ❑ **Common Approaches Used For Database Performance**
 - ❑ **New Breakthrough: On-Chip Computing/ Columnar Design**
 - ❑ **VectorWise – A Purpose-Built Analytical Database**
 - ❑ **Illustrative Use Cases**
- 

Today's Challenges in BI



Make informed decisions faster

Analysis in seconds not minutes, minutes not hours



Data explosion

Collecting more information, less resources
>44x growth in next 10 years



Existing Tools:

Too Slow. Too complex. Too expensive.

Analytical databases designed in 80s & 90s do not take advantage of today's modern hardware

Biggest BI Challenge - SPEED

What Problem will eventually drive you to replace your current data warehouse platform?

1. Poor Query Response **45%**

2. Can't Support Advanced Analytics **40%**

Source: TDWI Q4 2009 Best Practices Report

Gartner[®]

“Gartner clients increasingly report performance constrained data warehouses during inquiries. Based on these inquiries, we estimate that nearly 70% of data warehouses experience performance-constrained issues of various types.”

Source: Gartner Magic Quadrant for Data Warehouse Database Management Systems, Jan 2010

Biggest BI Challenge - SPEED

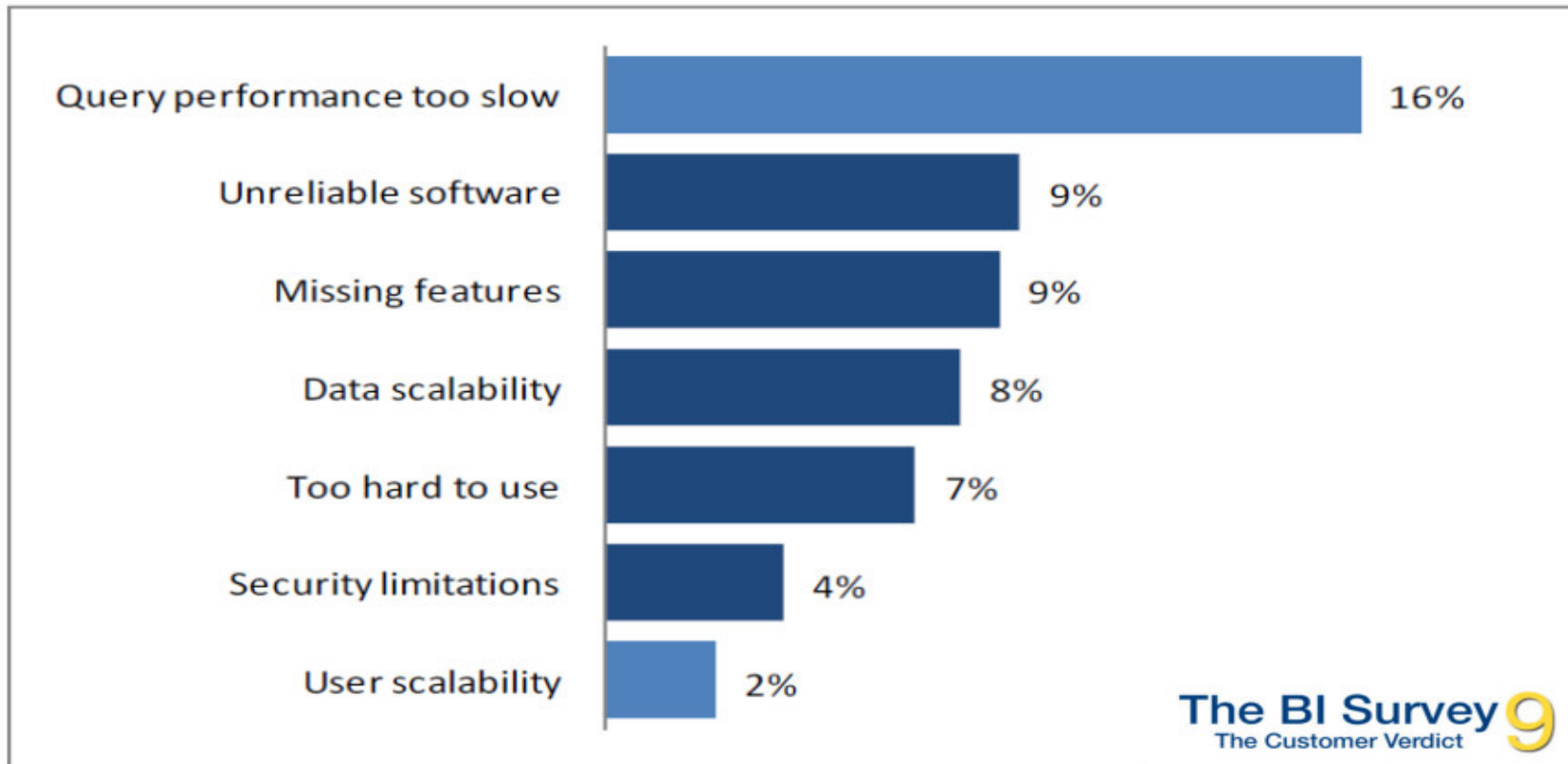


Figure 150: Performance problems (in lighter blue) compared to other product-related problems

Source: 2010 BI Survey 9 – World's largest independent BI Survey 3093 respondents

Cubes and Speed

As cubes grow in size they take longer to load and build

- Processing time might exceed batch window
- Difficulty managing large cubes
- Time required to add new dimensions

On average, how long does it take to add a new source of data to your data warehouse?



Average time to add a new data source	
2008	7.0 weeks
2009	8.4 weeks
2010	7.4 weeks

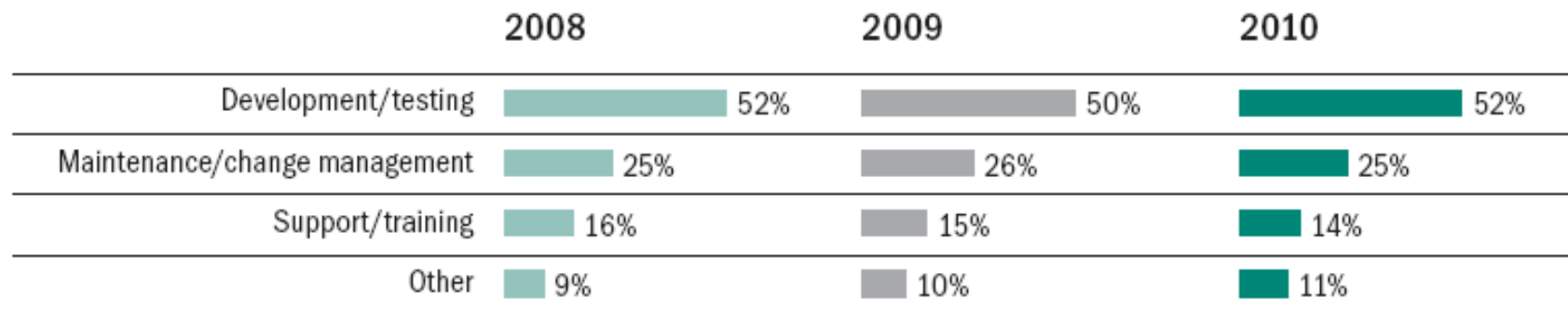
Source: 2010 TDWI Benchmarks

Relational Databases and Speed

Limitations in SQL technology

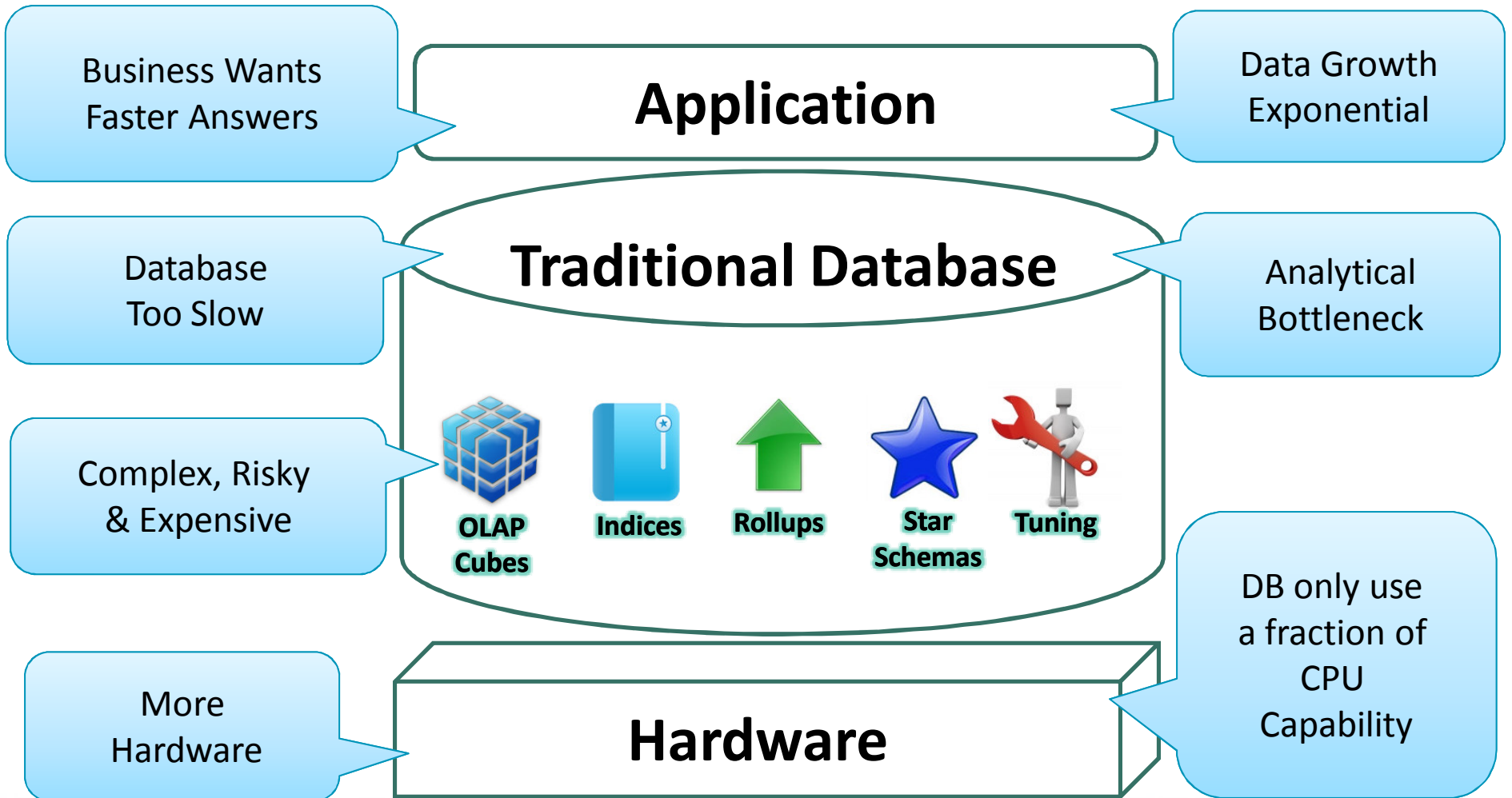
- Adhoc queries too slow
- Indexing/Aggregations cost time & money
- 25% average BI/DW team time used up for maintenance/change management

What percentage of your BI/DW team is allocated to these tasks?



Source: 2010 TDWI BI Benchmark Report

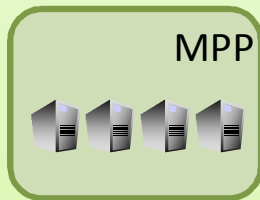
Challenges with Current State of BI



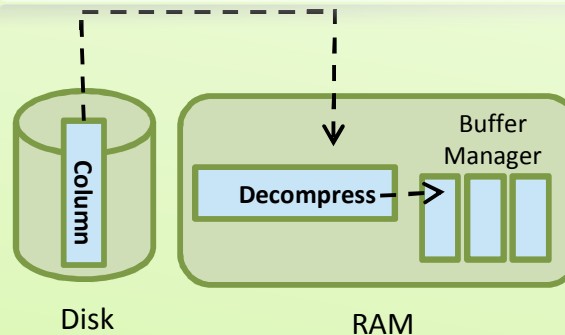
Approaches Used for Achieving Database Performance

Optimizations for parallel processing and minimal data retrieval

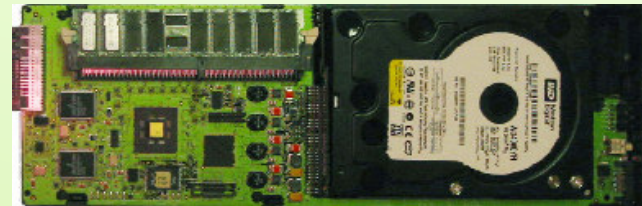
Parallel Processing



Column store with compression



Proprietary hardware



Data Warehouse Appliances

Acceptable performance has been achieved by using more hardware or by intelligently lowering the volume of data to be processed

However, none of these approaches leverages the performance features of today's CPUs i.e. taking the most out of each modern commodity CPU

New Breakthrough VectorWise Analytical Database

**Purpose-Built Analytical
Relational database for BI
and data analysis**

- **Runs blazing
fast/interactive data
analysis**
- **Exploits performance
potential in today's CPUs**
- **Delivers in-memory
performance without being
memory constraint**

VectorWise Sets New Benchmark Record

"Game-changing technology."
Don Feinberg, Gartner Group

***"This is definitely a breakthrough. It delivers
faster results at lower costs."***

Noel Yuhanna, Forrester Research

***"This inevitability puts VectorWise 4 years ahead
of the competition in terms of performance – and it
will remain 4 years ahead until some competitor
finds a way to catch up at a software level. This is
unprecedented."***

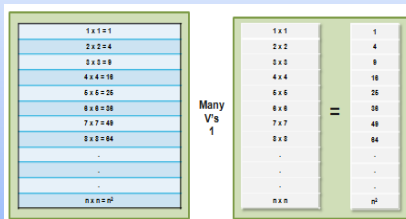
Robin Bloor, The Virtual Circle

VectorWise: On Chip Computing/Columnar Database

Breakthrough technology

Innovations on proven techniques

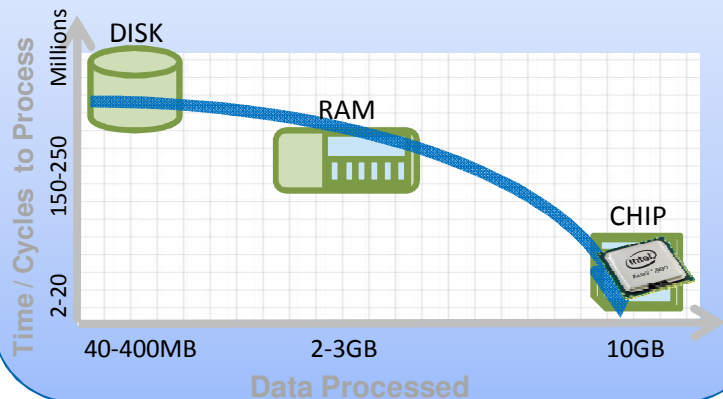
Vector Processing



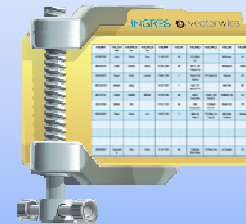
Updateable Column Store

USER NAME	USER NAME	USER NAME	USER NAME	USER NAME	USER NAME	USER NAME	USER NAME	USER NAME	USER NAME
ADMINISTRATOR	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN
ADMINISTRATOR	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN
ADMINISTRATOR	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN
ADMINISTRATOR	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN
ADMINISTRATOR	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN
ADMINISTRATOR	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN
ADMINISTRATOR	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN
ADMINISTRATOR	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN
ADMINISTRATOR	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN	ADMIN

On Chip Computing

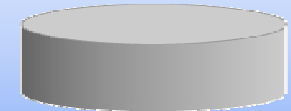


Automatic Compression



Automatic Storage Indexes

Minimize IO



Parallel Processing

VectorWise Technology

- **Vector processing**
 - Exploits super-scalar features using SIMD capabilities of today's CPUs
- **Optimizes memory hierarchy**
 - Maximizes use of CPU cache
 - Fewer requests to RAM and disk
- **Data Compression/De-Compression**
 - Optimized compression enabling very fast de-compression for overall performance enhancement
 - Vectorized de-compression
 - Automatic compression through ultra-efficient algorithms
- **Automatic Indexing**
 - System generated Storage Indexes
 - Easy identification of candidate data blocks for queries
- **Integration**
 - Standard SQL and interfaces
 - Common BI/Data Integration tools

Modern CPU Instruction Capabilities

▪ SIMD

- Traditional CPU processing: Single Instruction, Single Data (SISD)
- Modern CPU processing capabilities: Single Instruction, Multiple Data (SIMD)

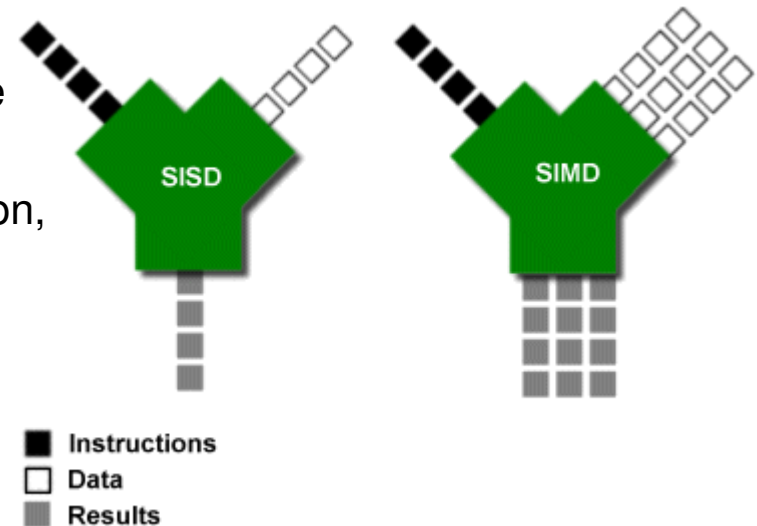
▪ Out-of-order execution

▪ Chip multi-threading

▪ Large L2/L3 caches

▪ Streaming SIMD Extensions for efficient SIMD processing

▪ Hardware accelerated String Processing



Vector Processing

Traditional Scalar Processing

One operation performed on one element at a time

Large overheads

$1 \times 1 = 1$
$2 \times 2 = 4$
$3 \times 3 = 9$
$4 \times 4 = 16$
$5 \times 5 = 25$
$6 \times 6 = 36$
$7 \times 7 = 49$
$8 \times 8 = 64$
.
.
.
$n \times n = n^2$

Many V's
1

Vector Processing

1×1	=	1
2×2		4
3×3		9
4×4		16
5×5		25
6×6		36
7×7		49
8×8		64
.		.
.		.
.		.
$n \times n$		n^2

One operation performed on a set of data at a time

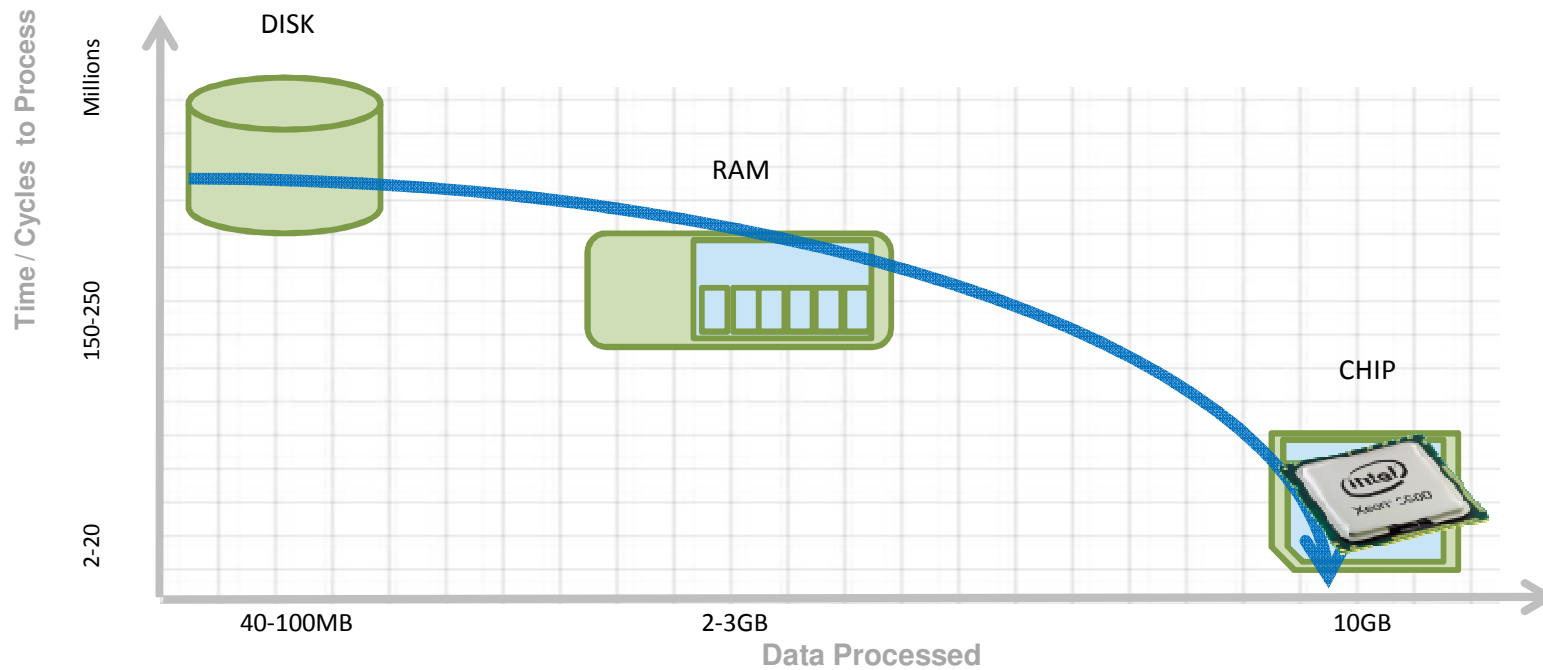
No overheads

Process even 1.5GB per second

Processing in Chip Cache

GB/s	Measure of Throughput
Cycles	Amount of CPU time required to process data

Using CPU cache is far more faster & efficient



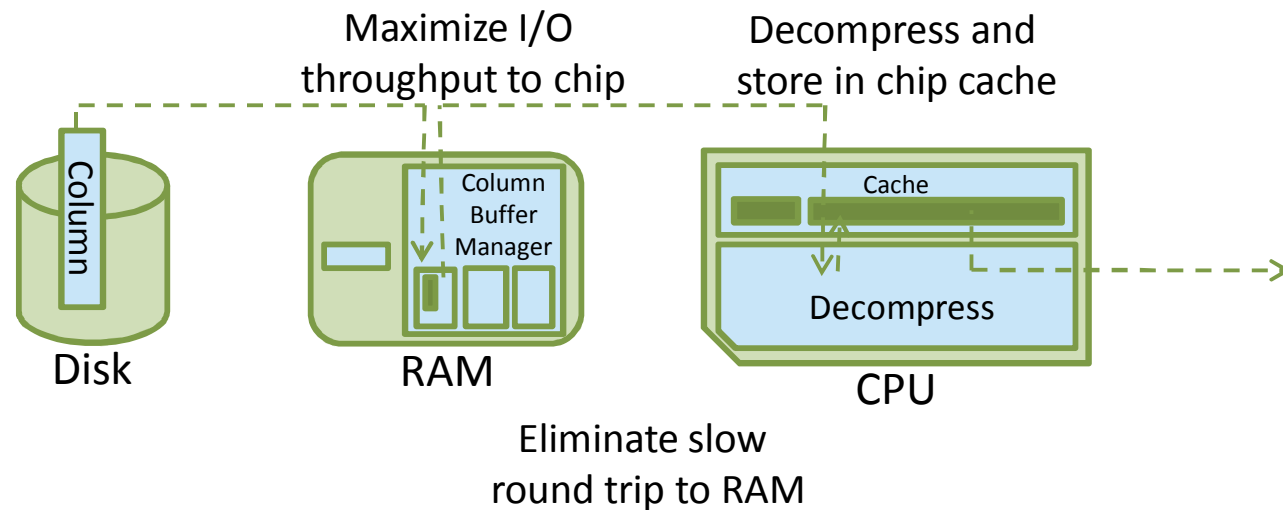
Updateable Column Store

- Only access relevant data
- Enable incremental updates efficiently
 - Traditionally a weakness for column-based stores

Cust_Num	Cust_surname	Cust_first_name	Cust_mid_name	Cust_DOB	Cust_Sex	Cust_Addr_1	Cust_Addr_2	Cust_City	Cust_State
46328927956	Jones	Steven	Sean	17-JAN-1971	M	333 StKilda Rd		Melbourne	Vic
98679975745	Smith	Leonard	Patrick	04-APR-1964	M	Unit 12, 147 Trafalgar Sqr		Birmingham	London
52634346735	Rogers	Cindy	Carmine	11-MAR-1980	F	Belmont Rail Service	421 Station St	Belmont	CA
346737347347	Andrews	Jenny		14-SEP-1977	F	Apt1, 117 West 42 nd St		New York	NY
88673477347	Cooper	Sheldon	Michael	30-JUN-1980	M	Ingres Corporation	Level 2, 426 Argello St	Redwood City	CA
34673447568	Kollwitz	Rolf		22-DEC-1975	M	IBM Headquarters	123 Mount View Crs	Atlantic City	PN
99554443044	Wong	Penny	Lee	13-NOV-1981	F	Ming On Tower 1	1777 Moa Tzu Tung Rd	Ming Now Province	Shanghi


Optimized Compression & Fast De-Compression

- **Column-based compression with multiple algorithms**
 - Automatically determined by VectorWise
- **Vectorised decompression**
 - Only for data processing in CPU cache





Storage Index

- **Always automatically created**
 - **Automatically maintained**
 - **Stores min/max value per data block**
 - **Enables database to efficiently identify candidate data blocks**
- 

VectorWise Features

Performance

- 10x-75x faster for BI, analytics & reporting
- In-memory performance without memory restraints
- Near real-time updatable database
- Delivers results in seconds not minutes
minutes not hours

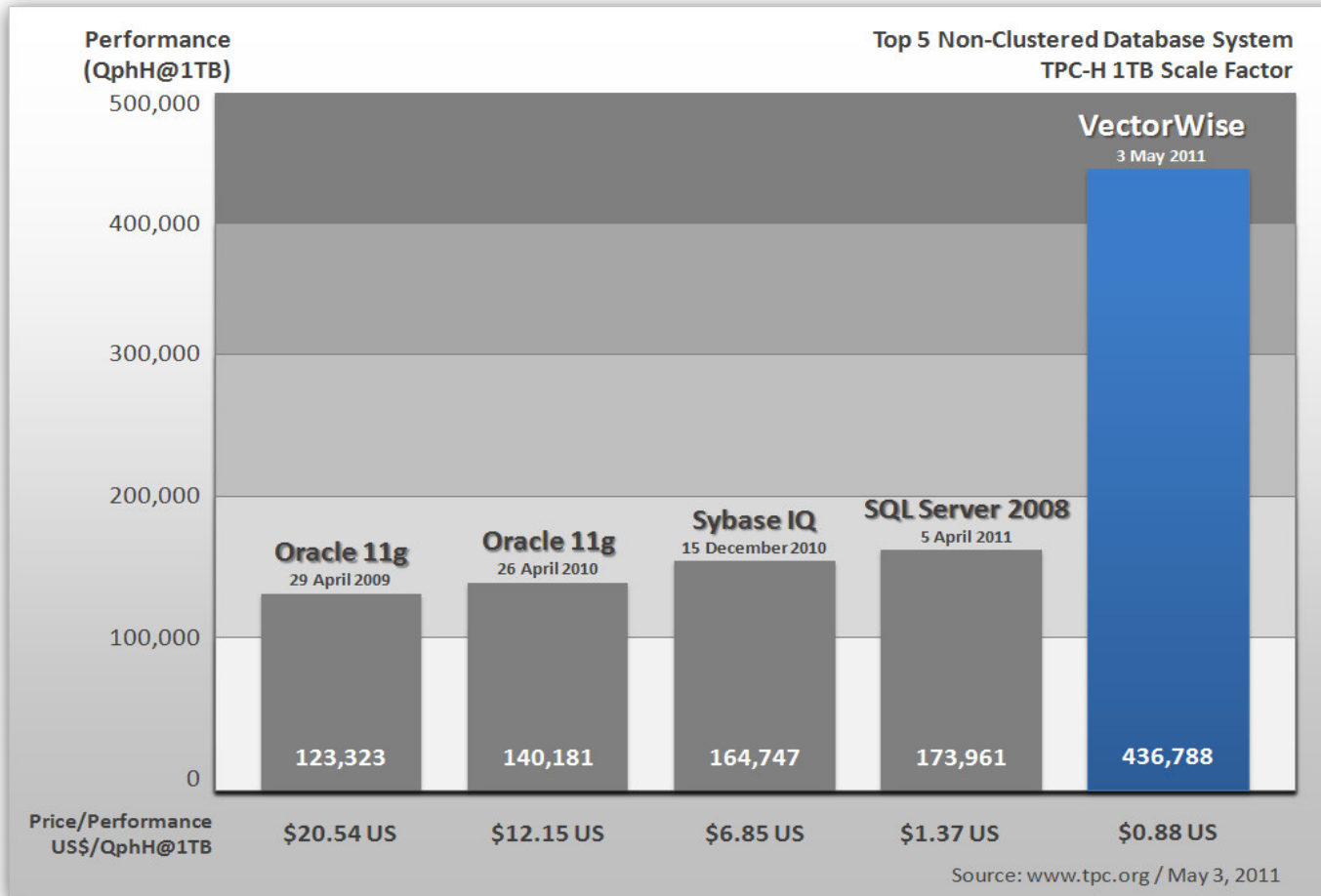
Usage & Integration

- Uses ANSI standard queries & SQL statements
- Eliminate/reduce Cubes, aggregate tables, roll ups, indexes....
- Self indexing & self tuning database
- Deliver BI projects faster with lower cost & risk

TCO

- Maximize utilization of CPUs in low cost commodity hardware
- Handle tens of terabytes scale data with a single server
- Requires commodity hardware
- Does not require MPP

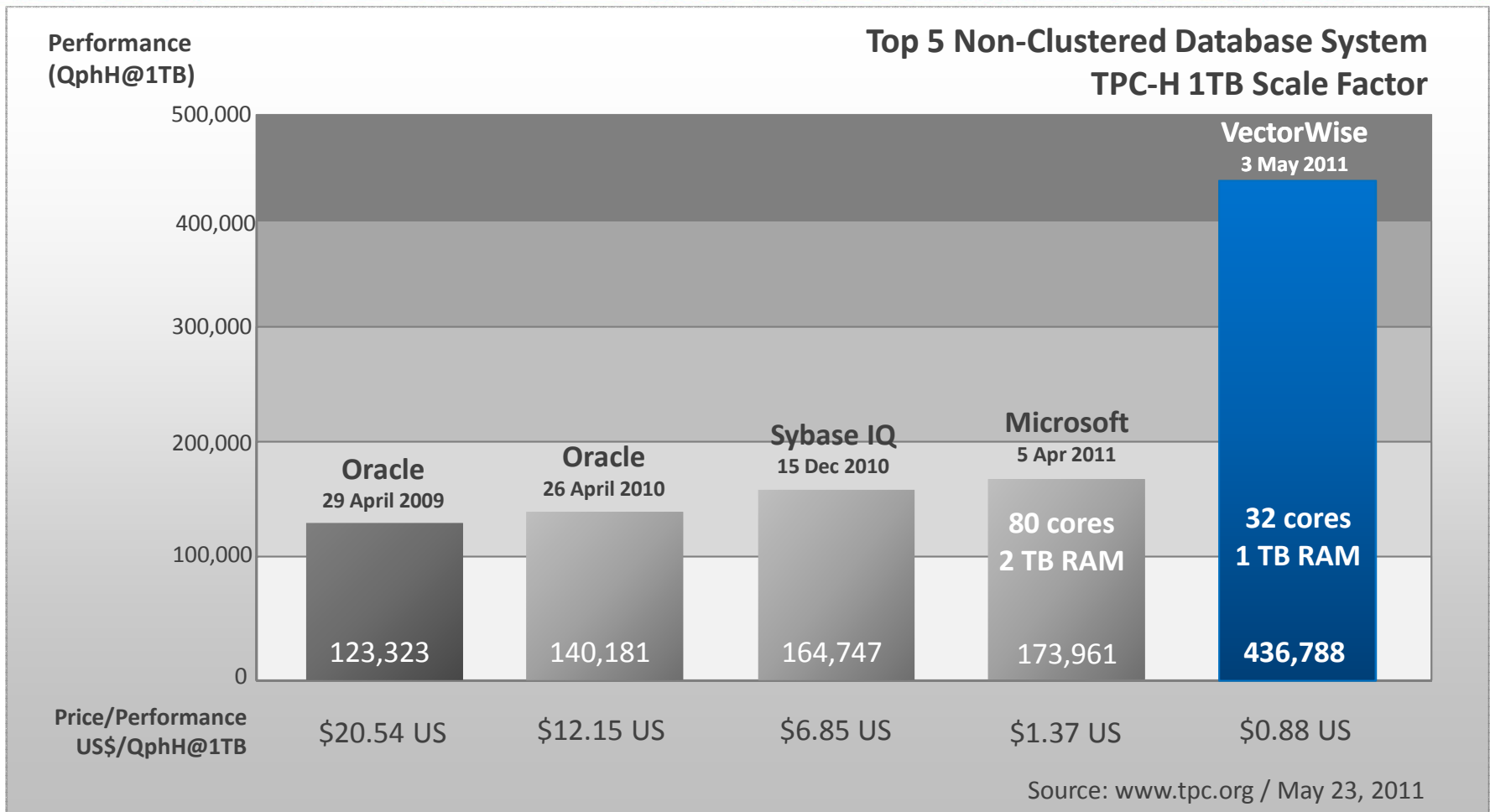
VectorWise TPC-H Performance Benchmark



TPC, TPC Benchmark, TPC-H, QppH, QthH and QphH are trademarks of the Transaction Processing Performance Council (TPC)

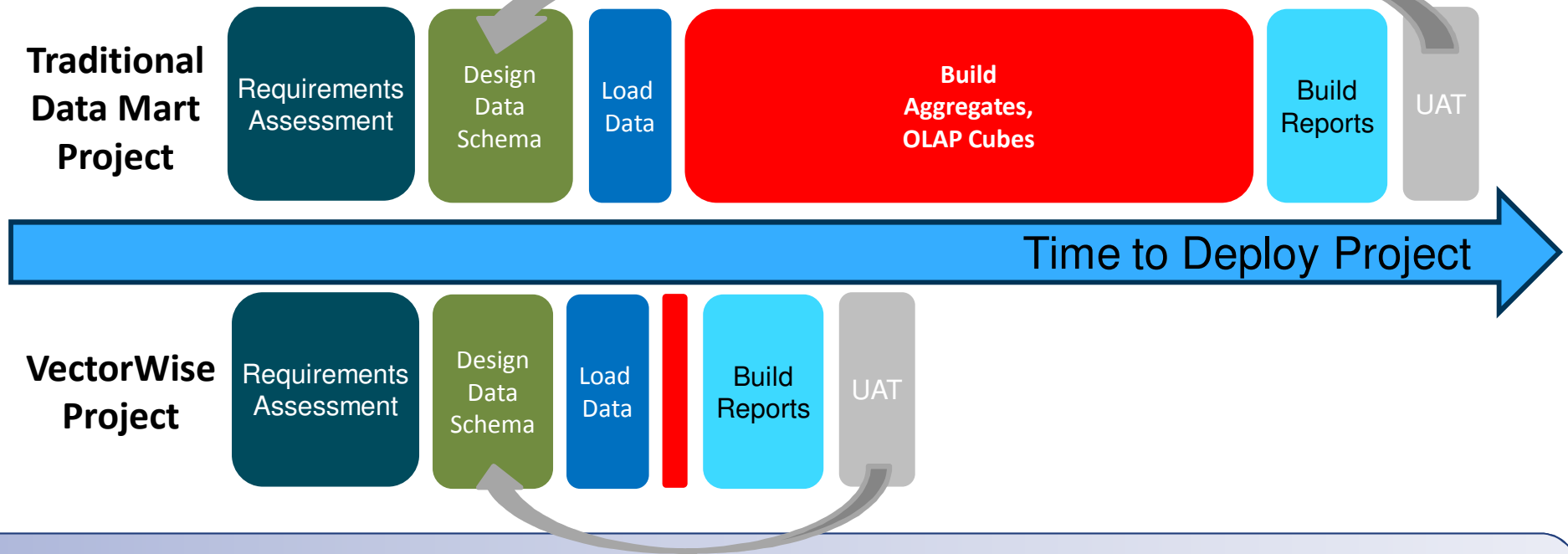
TPC Benchmark Results – 1TB

<1/2 Hardware, >2.5x Performance



TPC, TPC Benchmark, TPC-H, QppH, QthH and QphH are trademarks of the Transaction Processing Performance Council (TPC)

VectorWise BI Tuning & Complexity



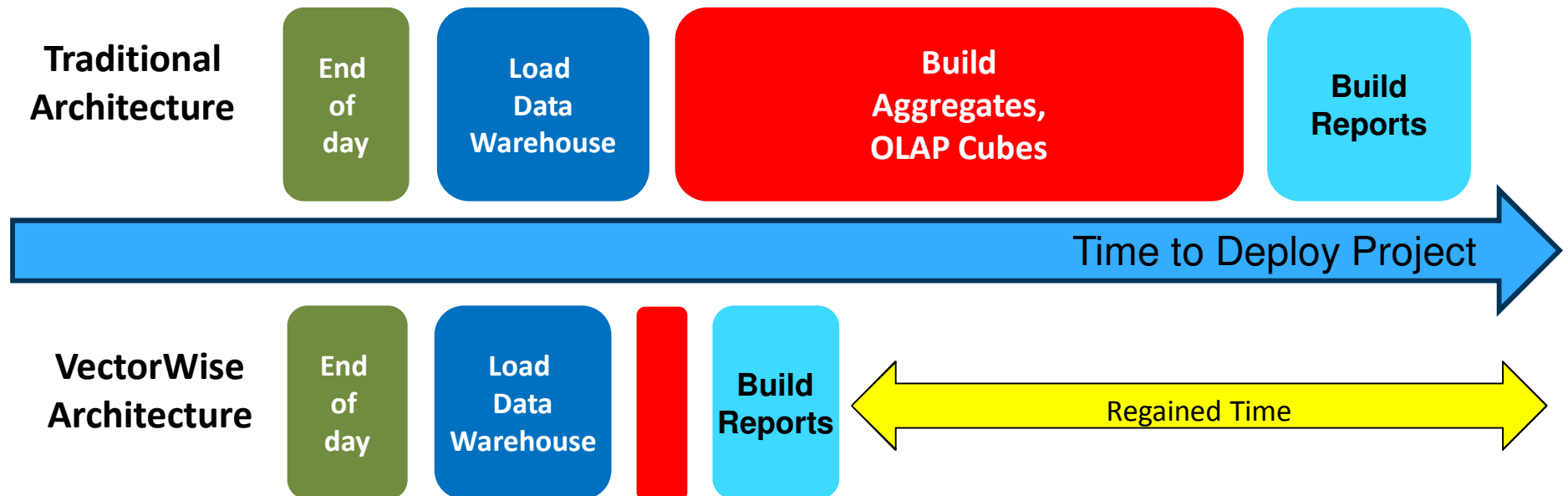
2010 TDWI BI Benchmark Report

*Average time to build a complex report or dashboard
(20 dimensions, 12 measures, and 6 user access roles)*

2008	6.7 weeks
2009	6.3 weeks
2010	6.6 weeks

VectorWise BI Tuning & Complexity

Fast Processing Everyday!!!



VectorWise TPC-H Price/Performance Benchmark

Price/Performance Benchmarks TPC-H 100GB Scale Factor

Non-Clustered Results - Price / Composite queries per hour (QphH)

Previous Benchmark 0.51 US\$/QphH

VectorWise 0.38 US\$/QphH

25.5% better price/performance than previous-benchmark record holder

Source: www.tpc.org / February 15, 2011

TCO

- Spend less on infrastructure
- Spend less on BI tuning

Analytical Databases - Illustrative Use Cases

Telcos/VAS	Retail	FSI	Web 2.0
Store & analyze CDR, VAS downloads & other subscriber/network data for: <ul style="list-style-type: none"> - Revenue assurance - Price optimization - Customer loyalty/churn - Marketing effectiveness - Service level effectiveness - Network performance 	Store & analyze data for: <ul style="list-style-type: none"> - Customer loyalty - Buying behavior - Marketing effectiveness - SKU level analysis 	Store & analyze transaction, market & customer data for: <ul style="list-style-type: none"> - Risk management & compliance - Quantitative analysis of financial models - Claims data analysis - Fraud detection - Credit rating - Marketing effectiveness 	Store & analyze data for: <ul style="list-style-type: none"> - Weblog data - Online behavior - Buying behavior - Marketing effectiveness
Healthcare & Biotech	Transportation	Manufacturing	Government
Store & analyze data for: <ul style="list-style-type: none"> - Patient data records - Clinical data analysis - Drug discovery & development analysis 	Store & analyze data for: <ul style="list-style-type: none"> - Passenger traffic data - Customer behavior - Customer loyalty - Marketing effectiveness 	Store & analyze data for: <ul style="list-style-type: none"> - Supply chain - Product quality - Strategic procurement 	Store & analyze data for: <ul style="list-style-type: none"> - Fraud detection - Cyber security - Immigration control

VectorWise – Illustrative Real-world Use Cases

- **Financial Services**
 - Hedge fund - Risk management in position analysis
 - Bank - Risk management & compliance reporting, Interactive BI/Analytics Platform
- **Telecom**
 - 3G operator - CDR analysis for better customer insight and cross/upsell
 - BSS solutions provider - Telecom analytics
- **Web 2.0/Social Media**
 - Social media portal – Analyzing user traffic analysis for better targeted advertising
 - Freight exchange – Customer behavior analytics
- **Retail**
 - Data aggregator – Customer and infomercial analytics
 - Solution provider – Retail analytics
- **Energy**
 - Services provider to Utilities – Cloud-based smart metering solution
- **Govt.**
 - Tax authority – Tax compliance analysis

More Information



www.ingres.com/products/vectorwise



VectorWise LinkedIn User Group

SkyInsight
VectorWise Cloud Data Analytics



QUESTIONS

