ORACLE®

Oracle Database 11g for Data Warehousing

Hermann Bär - Director Product Management, Data Warehousing

Oracle DW Strategy

- Best Database for BI/DW
  - 30 years of innovation
  - No other database can compare on the breadth and sophistication of Oracle's database features
- Within complete solutions
  - Complete database platform capabilities: ELT and Analytics
  - Complete BI and Performance Management solutions from Oracle
  - Broadest array of third-party technologies and solutions
- On the right hardware infrastructure
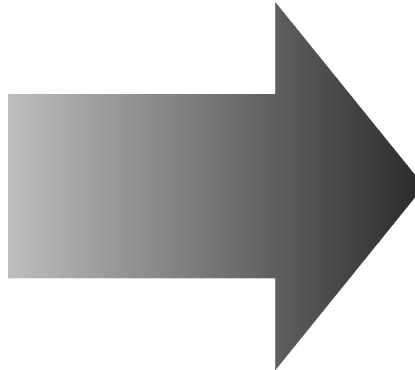
ORACLE®

Complete Solutions

Performance Mgmt: Hyperion

BI Applications: Bus Obj, Cognos

BI Tool: Business Objects, Cognos
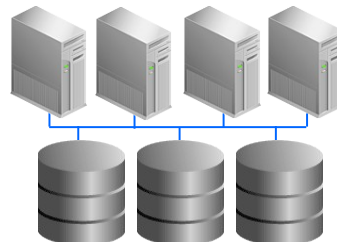
Middleware: IBM, BEA

Database: **ORACLE®**

Performance Mgmt: **ORACLE®**

BI Application: **ORACLE®**

BI Tool: **ORACLE®**

Middleware: **ORACLE®**

Database: **ORACLE®**

Benefits:

- **Integrated stack**

- **Continued best-of-breed**

- **Top-to-bottom performance optimization**

ORACLE®

Oracle Data Warehouse Platform

ORACLE 11g
DATABASE

Analytics

| Multi-Dimensional Calculations | Time Series | Statistics | Data Mining |

Intelligent Query Processing

| Materialized Views | Bitmap Indexes | Partition Elimination | Star Query Optimization |

Scalable Data Management

| Parallel Execution | Partitioning | RAC | Automatic Storage Mgmt | Compression |

Data Integration

| Bulk ETL | Real-Time ETL | Data Quality |

Workload Management

Metadata Management

Security

ORACLE

Oracle10*g* for Data Warehousing
Continuous Innovation

Oracle 7.3

Oracle 8.0

- Partitione
- Partition
- Parallel In
- Parallel In
- Parallel B
- Parallel A
- Parallel C
- Server Ma
- Point-in-T

Oracle8*i*

- Hash and
- Resource
- Progress
- Adaptive
- Server-ba
- Materializ
- Transpor
- Direct Lo
- Functiona
- Partition-
- Security

Oracle9*i*

- List and Ran
- Table Compr
- Bitmap Join
- Self-Tuning
- New Analytic
- Grouping Se
- External Tab
- MERGE
- Multi-Table I
- Proactive Qu
- System Man

Oracle10*g*

- Self-tuning SQL Optimization
- SQL Access Advisor
- Automatic Storage Manager
- Self-tuning Memory
- Change Data Capture
- SQL Models
- SQL Frequent Itemsets
- SQL Partition Outer Joins
- Statistical functions
- and much more ...

Oracle for Data Warehousing
Continuous Innovation

Oracle 7.3

Oracle 8.0

Oracle8*i*

- Hash a...
- Partiti...
- Resou...
- Progre...
- Adapt...
- Server...
- Functi...
- Materi...
- Transp...
- Direct...
- Functi...
- Partiti...
- Securi...

Oracle9*i*

- List and...
- Table Co...
- Bitmap ...
- Self-Tu...
- New Ana...
- Groupin...
- External...
- MERGE ...
- Multi-Ta...
- Proactiv...
- System ...

Oracle1...

- Self-tuning ...
- SQL Access...
- Automatic S...
- Self-tuning ...
- Change Dat...
- SQL Models...
- SQL Freque...
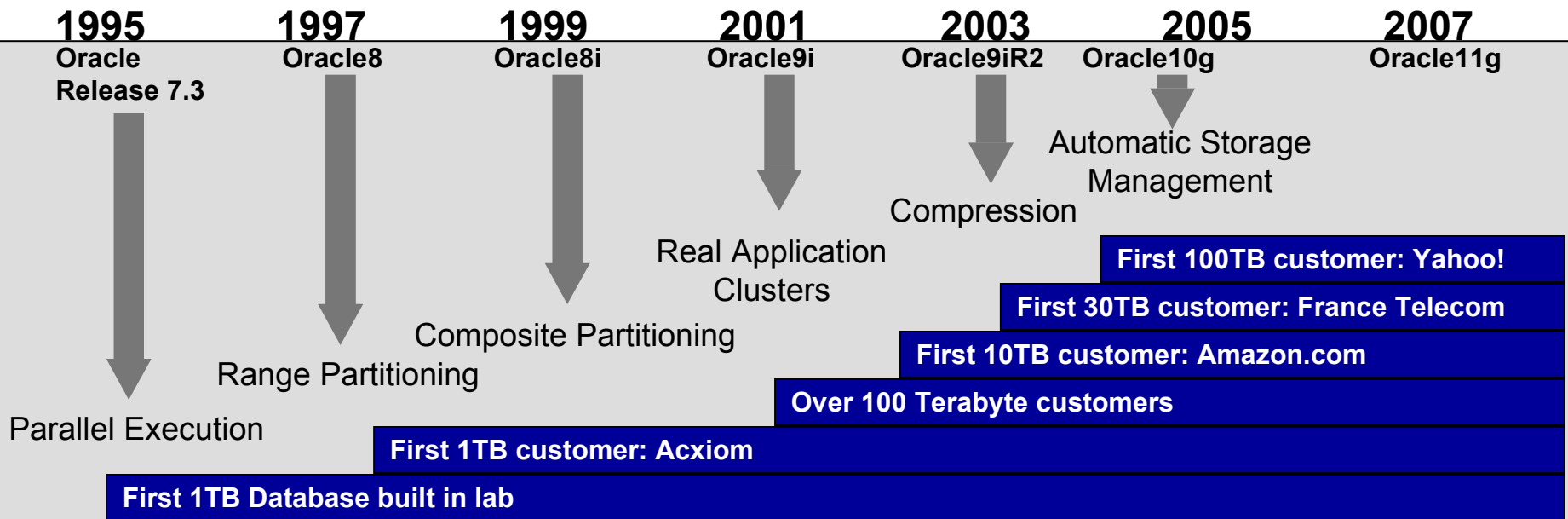- SQL Partitio...
- Statistical fu...
- and much n...

Oracle11*g*

- New composite partitionings
- Virtual column partitioning
- REF partitioning
- Cube-based Materialized Views
- SQL Pivot and Unpivot
- Query Result Cache
- SQL Plan Management
- General Linear Models
- Advanced Compression Option
- OWB included with DB

Continuous R+D Investment in VLDW Continuous Customer Success in VLDW

- Over the past 12+ years, Oracle has steadily introduced major architectural advances for large database support
- Data warehouses have grown exponentially with these new technologies

1995	1997	1999	2001	2003	2005	2007
Oracle Release 7.3	Oracle8	Oracle8i	Oracle9i	Oracle9iR2	Oracle10g	Oracle11g

Automatic Storage Management

Compression

Real Application Clusters

Composite Partitioning

Range Partitioning

Parallel Execution

First 100TB customer: Yahoo!

First 30TB customer: France Telecom

First 10TB customer: Amazon.com

Over 100 Terabyte customers

First 1TB customer: Acxiom

First 1TB Database built in lab

ORACLE

Long-standing dichotomy in the DW Market

- "Big brain" Workloads
  - Sophisticated database optimization techniques
    - Advanced Indexing
    - Dimensional query optimizations
    - Materialized views
    - Partition pruning
  - Algorithms and access paths determine performance

- Powerlifting workloads
  - Brute-force query execution
    - Large amounts of hardware
    - Query parallelism, hash partitioning
  - Hardware capabilities determine performance
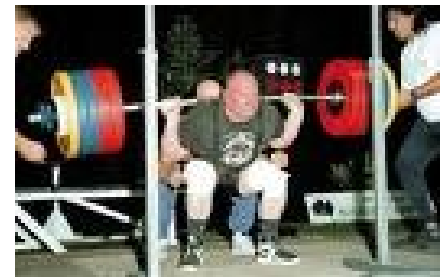
Established Architectural Solutions

Data Mart Layer

Atomic Data Layer
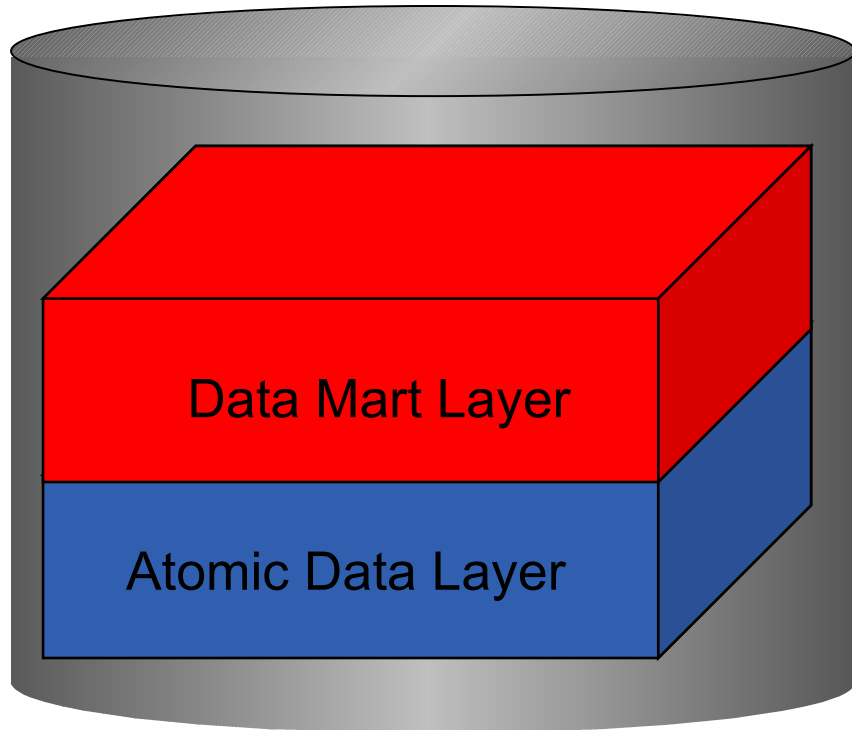
Application-specific performance structures
Summary data / materialized views
Dimensional view of data
Supports specific end-users, tools, and applications

Base data warehouse schema
Atomic-level data, 3nf design
Supports general end-user queries
Data feeds to all dependent systems

Half-Solutions are not the answer

- Enterprise DW solutions must provide both pieces
  - A solution that only provides one part will be limited to simple applications and unable to support growing
  - "If all you have is a hammer …"
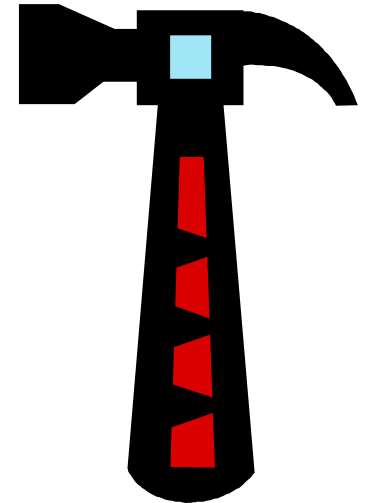
"Big Brain" Strategy

Oracle aggressively continues "Big Brain" strategy

- VLDB
  - Composite Range-Range
  - Composite List-Range
  - Composite List-List
  - Composite List-Hash
  - REF Partitioning
  - Virtual Column Partitioning
  - Compression enhancements
- Performance
  - Query Result Cache
- Manageability
  - Partition Advisor
  - Interval Partitioning
  - SQL Plan Management
  - Automatic SQL Tuning with Self-Learning Capabilities
  - Enhanced Optimizer Statistics Maintenance
  - Multi-Column Optimizer Statistics
  - ASM Fast Resync, Fast VLDB Startup and other enhancements
- SQL
  - SQL Pivot and Unpivot
  - Continuous Query Notification

- OLAP
  - Materialized view refresh and SQL rewrite
  - Continued database integration
    - Cube metadata in the Data Dictionary
    - Fine-grained data security on cubes
  - Simplified application development
    - Fully declarative cube calculations
    - Cost-Based Aggregation
    - Simpler calculation definitions
- Data Mining
  - Simplified development and deployment of models
    - Supermodels: data preparation combined with mining model
    - Additional packaged predictive analytics
    - Integration in database dictionary
  - New algorithms: "General Linear Models"
    - Encapsulates several widely used analytic methods
    - Multivariate linear regression; logistic regression
- ETL
  - OWB Repository installed with Database by default
  - Seibel connector
  - Graphical creation of views, materialized views

Oracle Partitioning:
Ten Years of Development

|	Core functionality	Performance	Manageability
Oracle8	Range partitioning Global range indexes	"Static" partition pruning	Basic maintenance operations: add, drop, exchange
Oracle8i	Hash and composite range-hash partitioning	Partition-wise joins "Dynamic" pruning	Merge operation
Oracle9i	List partitioning		Global index maintenance
Oracle9i R2	Composite range-list partitioning	Fast partition split	
Oracle10g	Global hash indexes		Local Index maintenance
Oracle10g R2	1M partitions per table	"Multi-dimensional" pruning	Fast drop table
Oracle Database 11g	More composite choices REF Partitioning Virtual Column		Interval Partitioning Partition Advisor

Partitioning in Oracle Database 11g
Virtual Column-Based Partitioning

ORDERS

```
ORDER_ID    ORDER_DATE   CUSTOMER_ID...          REGION AS (SUBSTR(ORDER_ID,6,2))
----------  -----------  ----------- --          ------
9834-US-14  12-JAN-2007        65920                  US
8300-EU-97  14-FEB-2007        39654                  EU
3886-EU-02  16-JAN-2007         4529                  EU
2566-US-94  19-JAN-2007        15327                  US
3699-US-63  02-FEB-2007        18733                  US
```
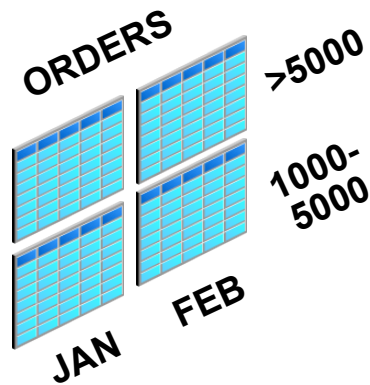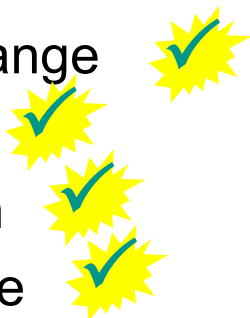
- `REGION` requires **no storage**
- Partition by `ORDER_DATE`, `REGION`
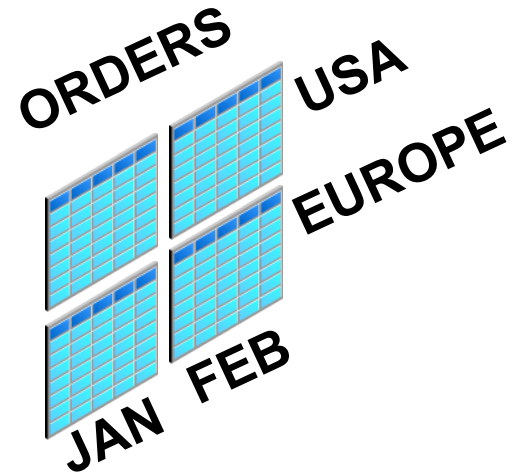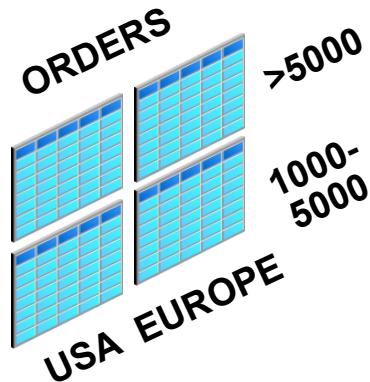
Partitioning in Oracle Database 11g
Complete Composite Partitioning

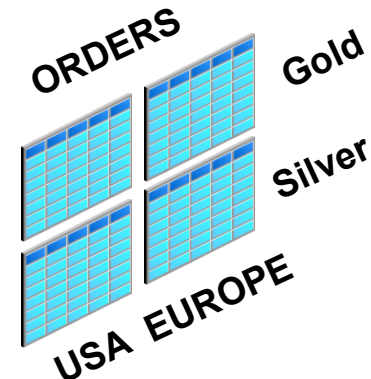- Range – range ✓
- List – list ✓
- List – hash ✓
- List – range ✓

RANGE-RANGE
Order Date by Order Value

LIST-RANGE
Region by Order Value

LIST-LIST
Region by Customer Type

Advanced Compression

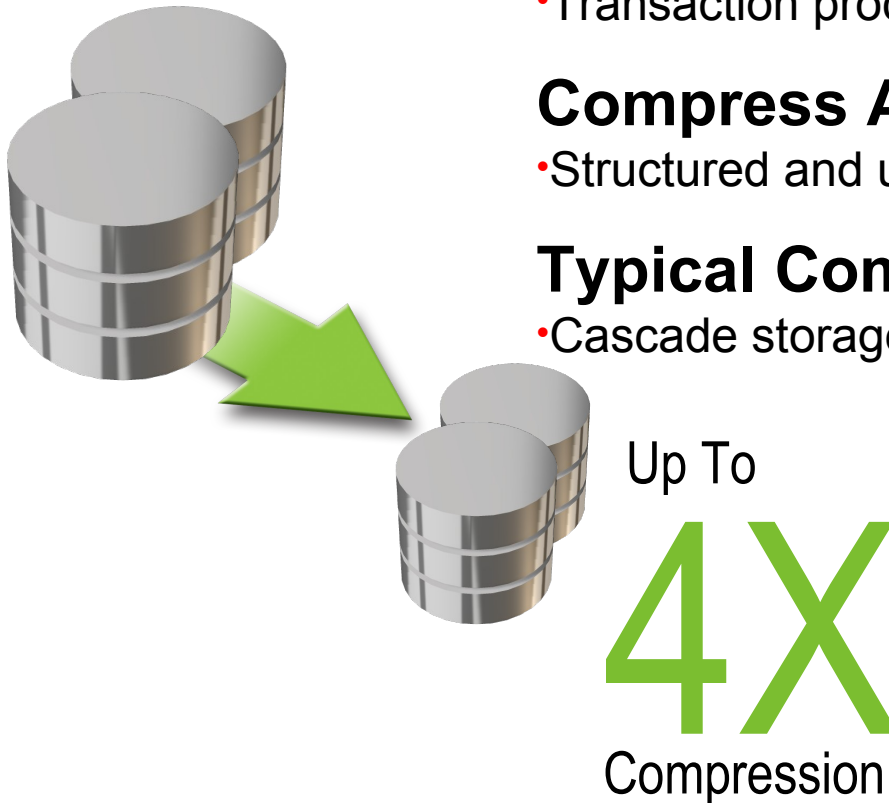## Compress Large Application Tables
•Transaction processing, data warehousing

## Compress All Data Types
•Structured and unstructured data types

Typical Compression of 2-4X
•Cascade storage savings throughout data center

Up To

4X

Compression

Database Result Cache

- Automatically Caches results of queries, query blocks, or pl/sql function calls
  - Cache is shared across statements and sessions on server
  - Significant speed up for read-only / read-mostly data
  - Full consistency and proper semantics
    - Cache refreshed when any underlying table updated
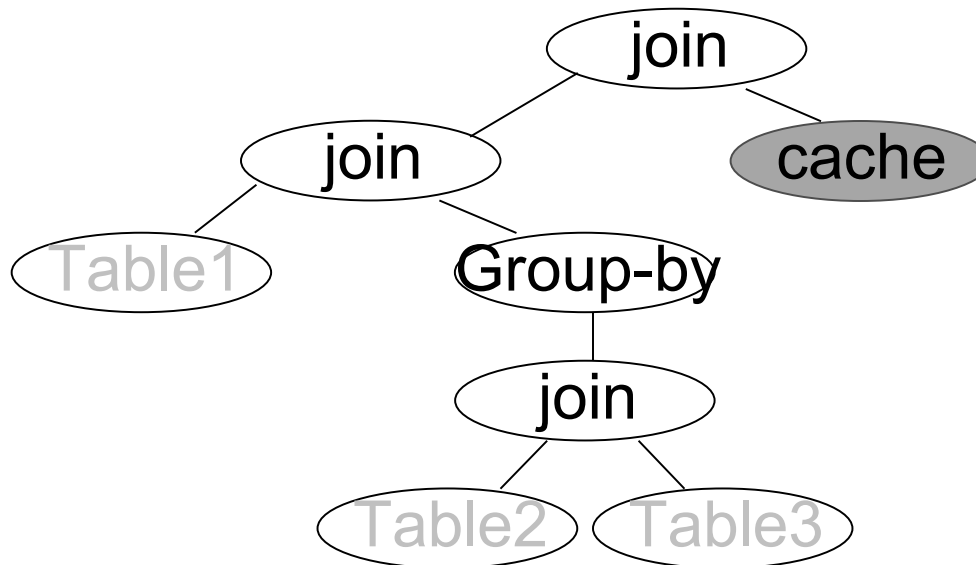
Q2: Uses result transparently

SQL Query Result Cache

- **Retail customer data (~50 GB)**
- **Concurrent users submitting queries randomly**
  - **executive dashboard application with 12 heavy analytical queries**
- **Cache results *only* at in-line view level**
  - **12 queries run in random, different order with 4 queries benefiting from the cache**

Legend:
- ■ No cache
- ■ Cache
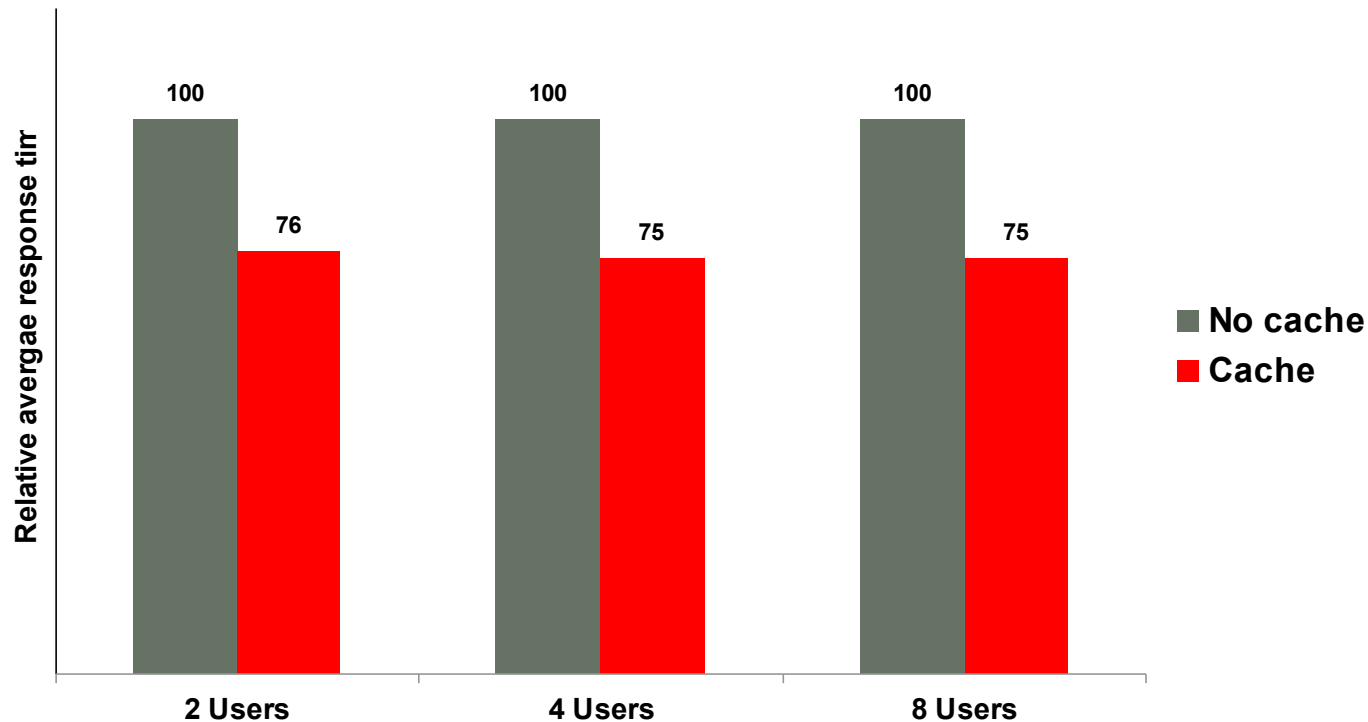
Transparent "Big Brain" Features

- Materialized Views
  - Transparent rewrites of expensive queries
    - Including rewrites on remote objects
  - Incremental automatic refresh
- Bitmap Indexes
  - Optimal storage
  - Ideal for star or star look-a-like schemas
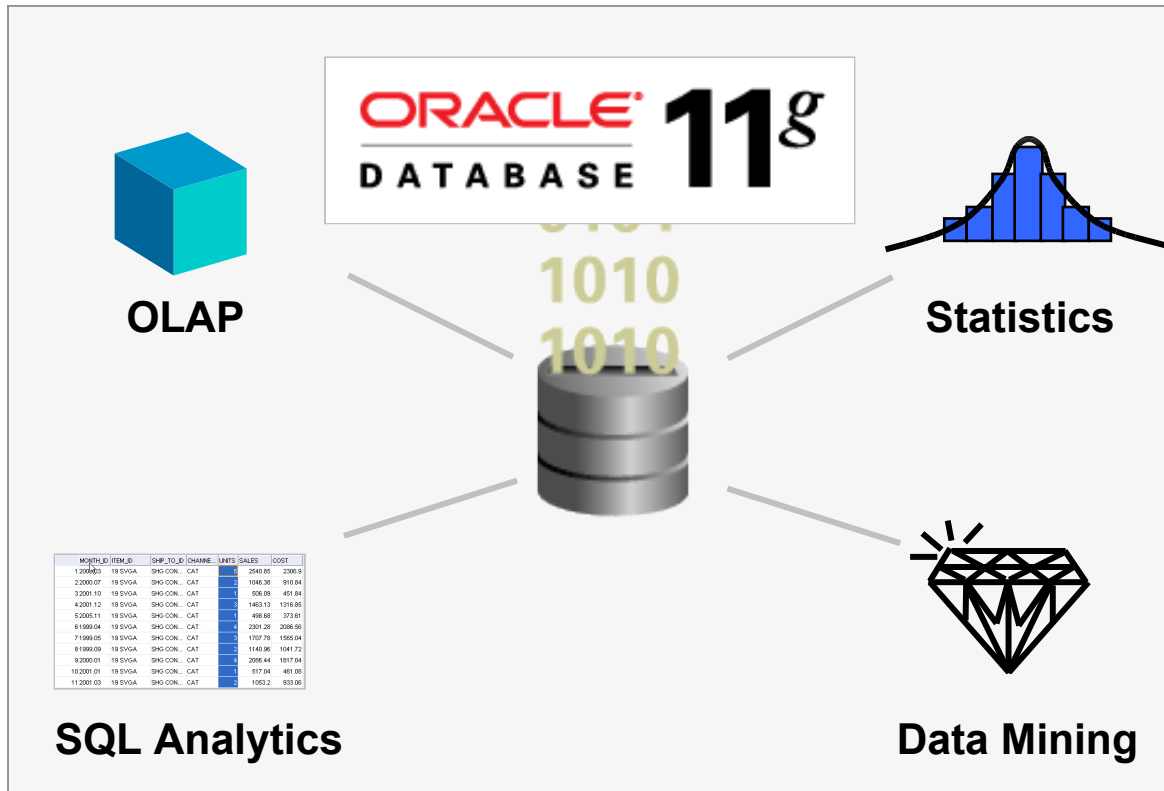- SQL Access Advisor – based on workload
  - Materialized view advice
  - Index advice
  - Partition advice
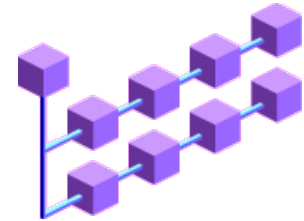
Integrated Analytics

- Bring the analytics to the data
- Leverage core database infrastructure

ORACLE®

Native Support for Pivot and Unpivot

```
SALESREP      Q1     Q2     Q3     Q4
---------- ----- ----- ----- -----
       100   230    240    260    300
       101   200    220    250    260
       102   260    280    265    310
```

```
SALESREP QU    REVENUE
---------- -- ----------
       100 Q1        230
       100 Q2        240
       100 Q3        260
       100 Q4        300
       101 Q1        200
       101 Q2        220
       101 Q3        250
       101 Q4        260
       102 Q1        260
       102 Q2        280
       102 Q3        265
       102 Q4        310
```

Native Support for Pivot and Unpivot

QUARTERLY_SALES

```
SALESREP       Q1      Q2      Q3      Q4
---------- ----- ----- ----- -----
100        230     240     260     300
101        200     220     250     260
102        260     280     265     310
```

```
SALESREP   QU      REVENUE
---------- --  ----------
100 Q1            230
100 Q2            240
100 Q3            260
100 Q4            300
101 Q1            200
101 Q2            220
101 Q3            250
101 Q4            260
102 Q1            260
102 Q2            280
102 Q3            265
102 Q4            310
```

```
select * from quarterly_sales
unpivot include nulls
(revenue for quarter in (q1,q2,q3,q4))
order by salesrep, quarter ;
```

Native Support for Pivot and Unpivot
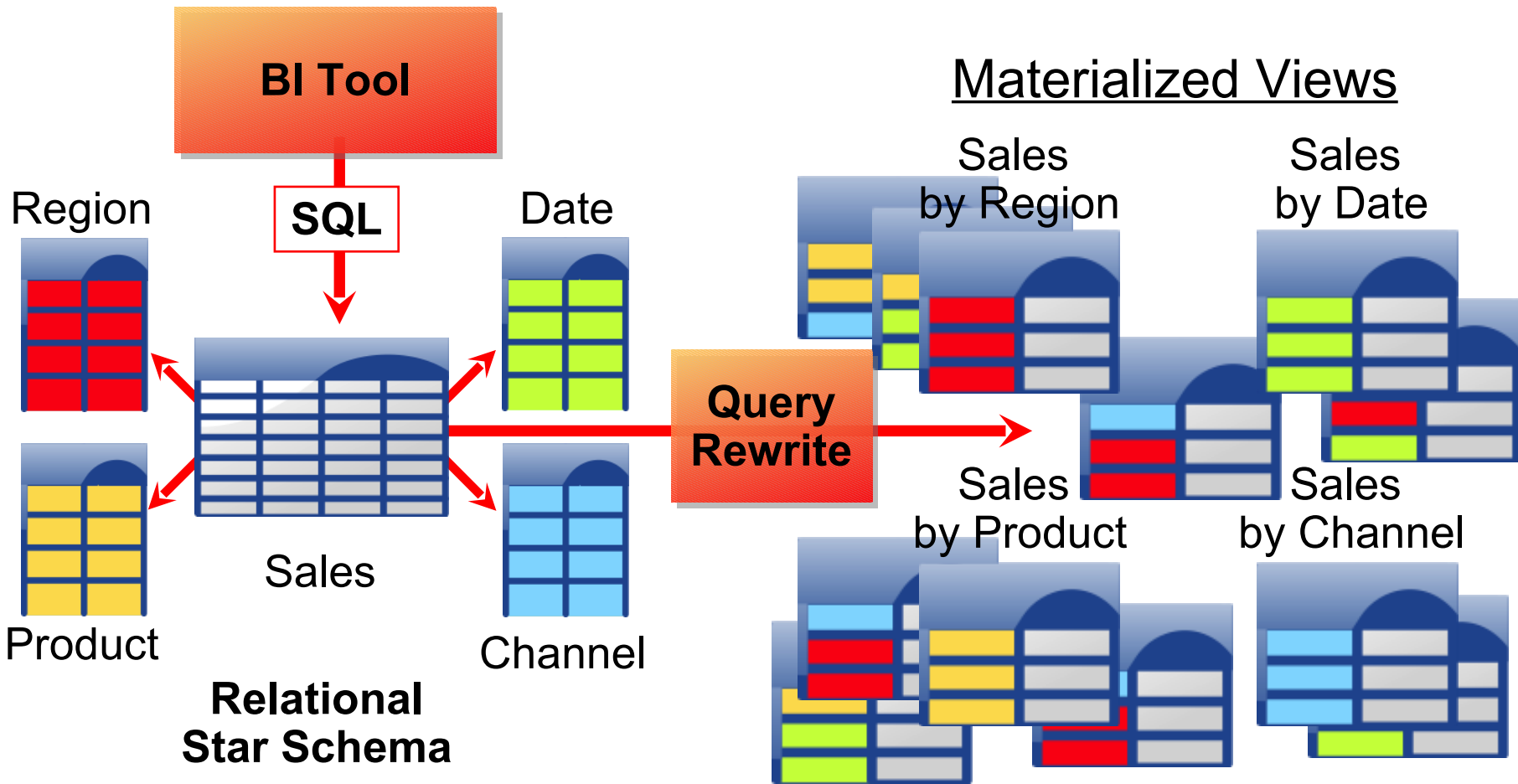
SALES_BY_QUARTER

```
              SALESREP QU    REVENUE
             ---------- -- ----------
              100 Q1          230
              100 Q2          240
              100 Q3          160
              100 Q4           90
              100 Q3          100
              100 Q4          140
              100 Q4           70
              101 Q1          200
              101 Q2          220
              101 Q3          250
              101 Q4          260
              102 Q1          260
```

```
SALESREP    'Q1'   'Q2'   'Q3'   'Q4'
---------- ----- ----- ----- -----
100         230    240    260    300
101         200    220    250    260
102         260    280    265    310
```

```
select * from sales_by_quarter
pivot (sum(revenue)
for quarter in ('Q1','Q2','Q3','Q4'))
order by salesrep ;
```
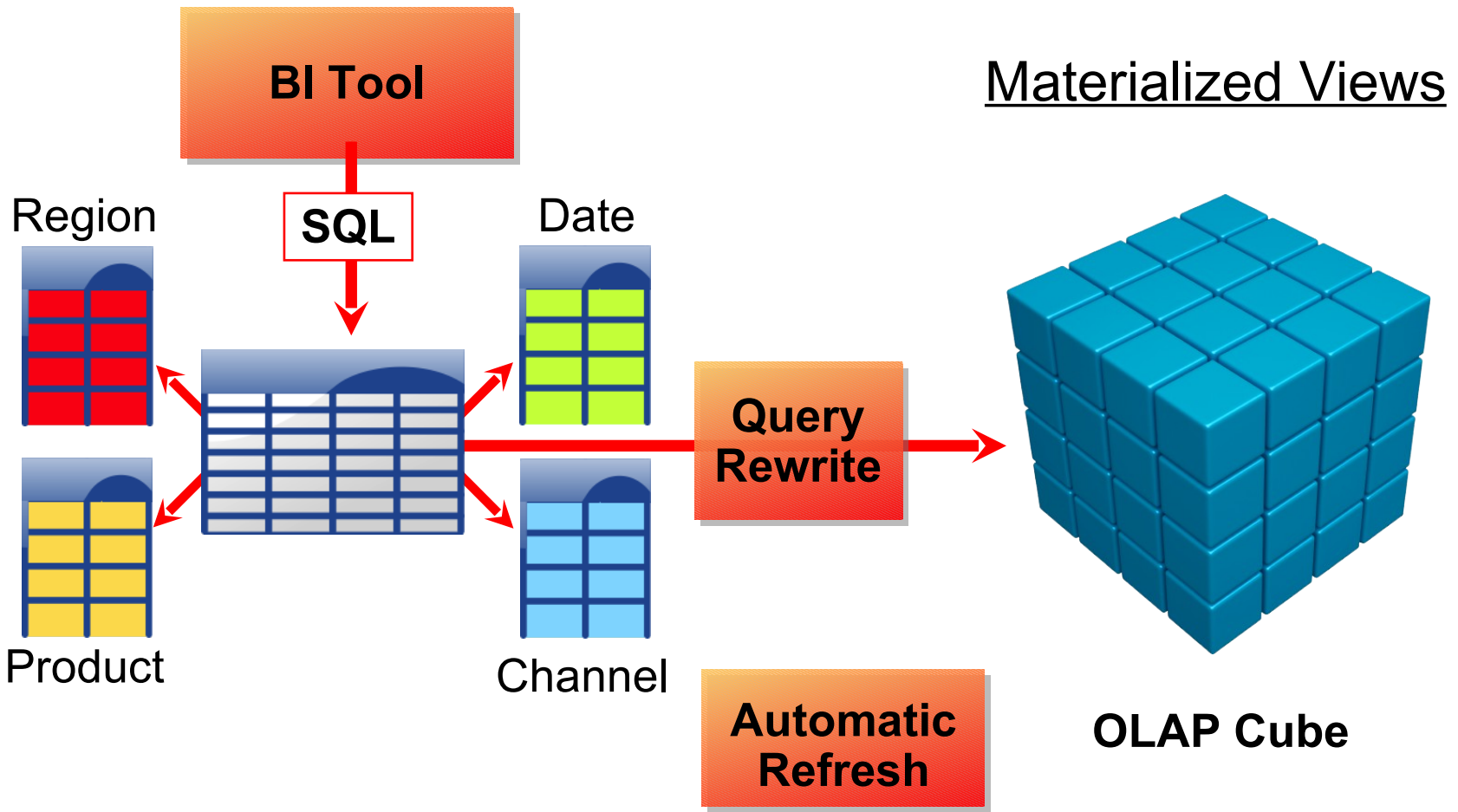
Business Intelligence Analysis
Typical Architecture Today

BI Tool

SQL

Region

Product

Sales

Date

Channel

Relational Star Schema

Query Rewrite

Materialized Views

Sales by Region

Sales by Date

Sales by Product

Sales by Channel

New in Oracle Database 11g
Cube-Organized Materialized Views

Materialized Views

BI Tool

SQL

Region

Date

Product

Channel

Query Rewrite

Automatic Refresh

OLAP Cube

Oracle Warehouse Builder
Packaging

Enterprise ETL Option

- Performance
- Productivity
- Reusability
- Metadata Management

Data Quality Option

- Data Profiling
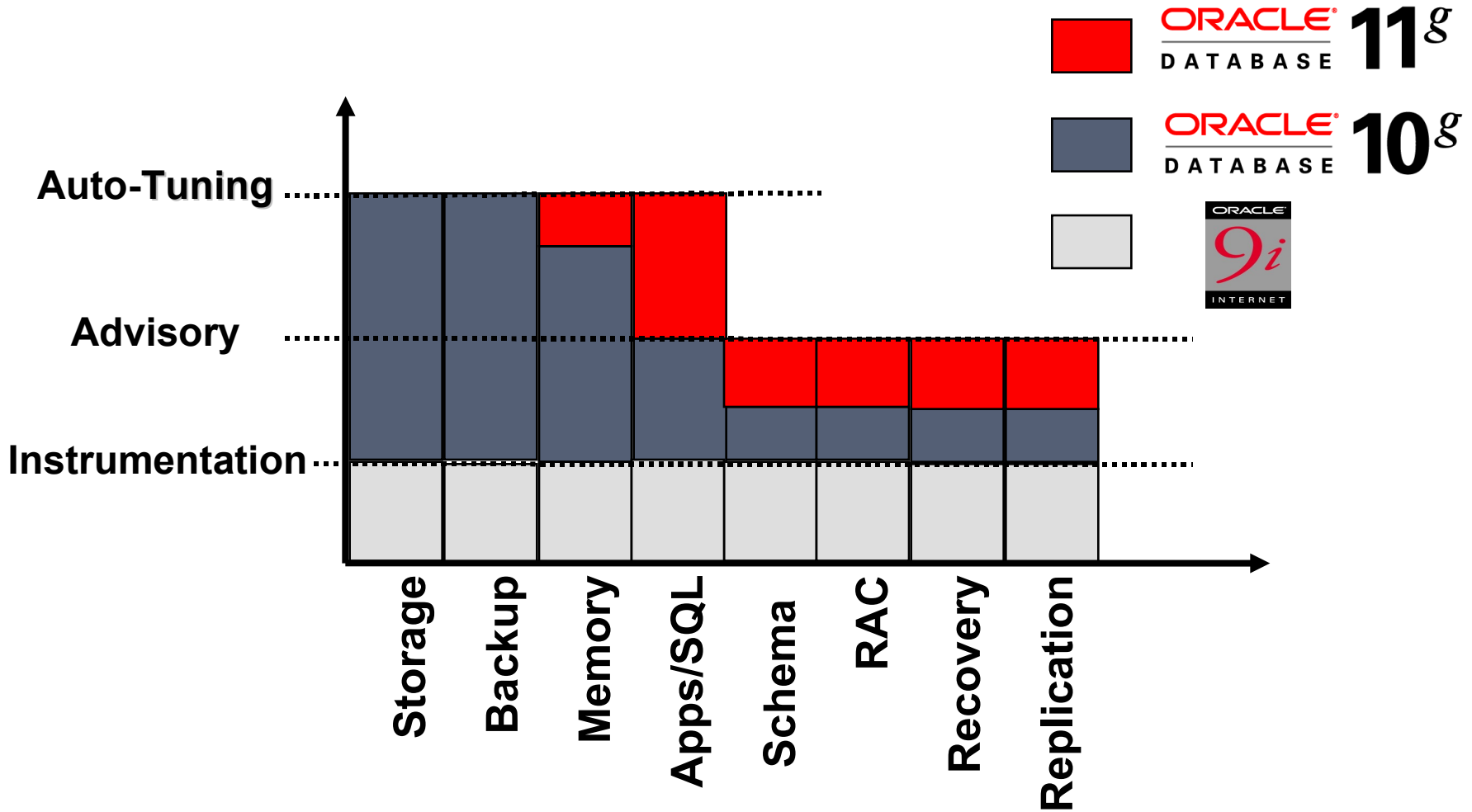- Anomaly Detection
- Business Rules
- Audit

ERP/CRM Connectors

- Oracle EBS
- PeopleSoft
- Siebel
- SAP

Core Features
(No extra cost with database SE/SE1/EE)

ORACLE®
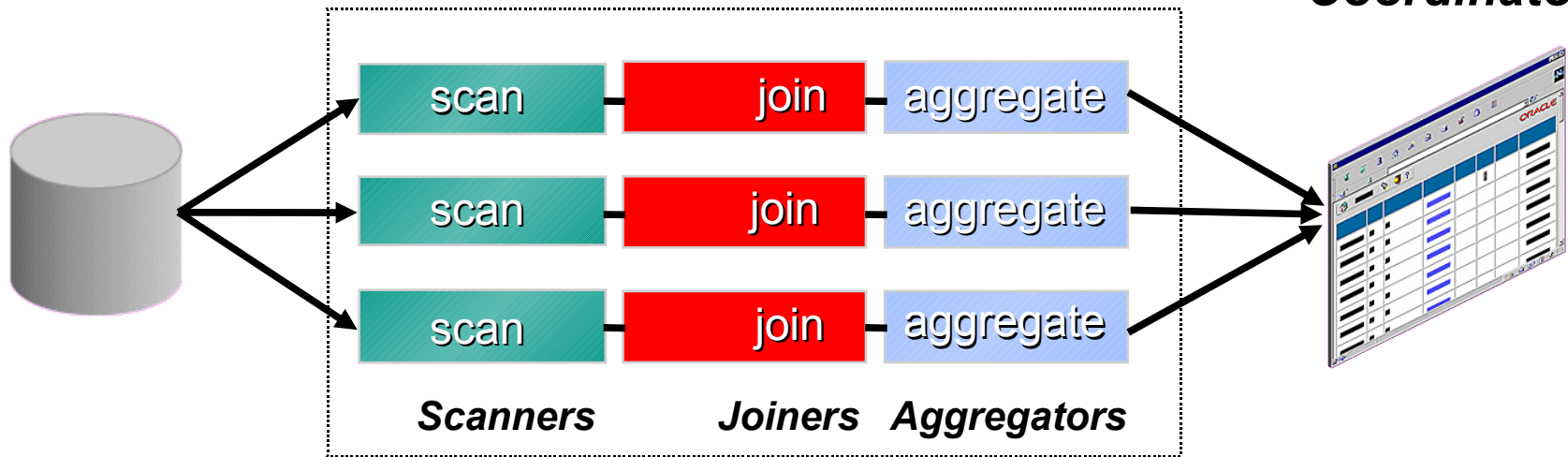
Self Managing Database

"Powerlifting" Strategy

Parallel Execution

```
select c.cust_last_name
, sum(s.amount_sold)
from customers c, sales s
where c.cust_id = s.cust_id
group by c.cust_last_name ;
```
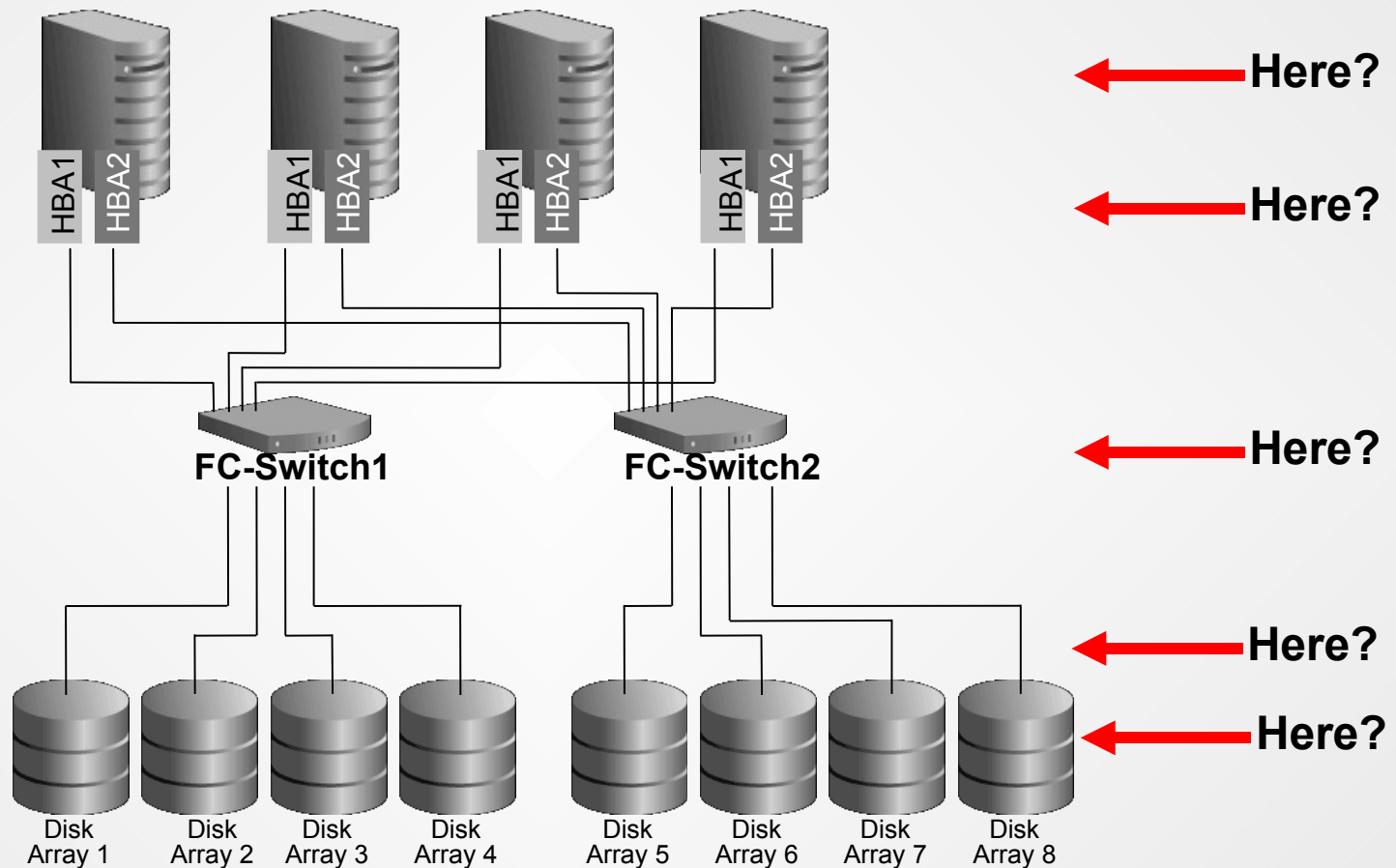
Data on Disk   **Parallel Servers**   **Coordinator**

scan	join	aggregate
scan	join	aggregate
scan	join	aggregate

Scanners   *Joiners*   *Aggregators*

ORACLE

Typical Customer DW Platform

- **Where is the performance bottleneck?**

Only Balanced Configurations Drive Optimized Performance

An Unbalanced Configuration

100%
Possible Efficiency

Database CPUs Memory Actuators LUNs Disks Raid

< 50%
Achieved Efficiency

A Balanced Configuration

100%
Possible Efficiency

Database CPUs Memory Actuators LUNs Disks Raid

100%
Achieved Efficiency
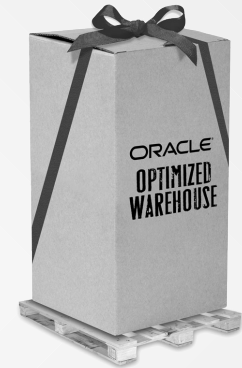
ORACLE®

Full Range of Options

Custom Solutions

Reference Configurations

Optimized Warehouse

Flexibility for the most demanding data warehouse

Documented best-practice configurations for data warehousing

Scalable systems pre-installed and pre-configured: create-table ready
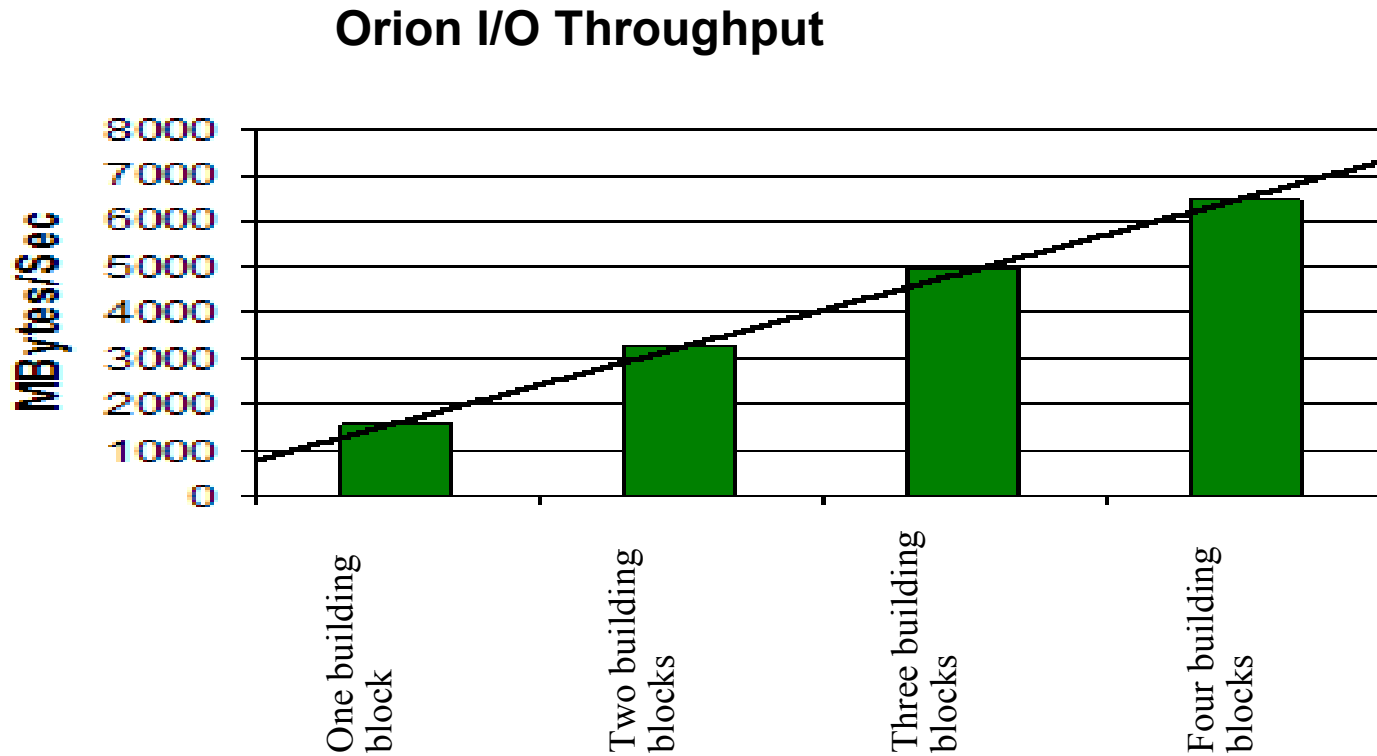
Pre-configured, Pre-installed, Validated

Complete Flexibility

ORACLE®

Building Block Scale-Out

- Add power in a balanced fashion

ORACLE®

Provide linear hardware scalability

Orion I/O Throughput

Reduce Risk & Accelerate Deployment

Custom Configuration

Reference Configuration

Oracle Optimized Warehouse

Time to implement

Months.....

Weeks.....

Custom Configuration
Testing and Validation
Installation and Configuration
Acquisition of Components
Pre-Implementation System Sizing

Reference Configuration
Testing and Validation
Installation and Configuration
Acquisition of Components

Faster Deployment Lower Risk Increased Flexibility

Oracle Optimized Warehouse
Oracle Optimized Warehouse

< 1 Week to implement

ORACLE

Oracle Optimized Warehouse Availability

Scalability   Limited   High   Very High

Oracle DW Strategy

- Best Database for BI/DW
  - 30 years of innovation
  - No other database can compare on the breadth and sophistication of Oracle's database features
- Within complete solutions
  - Complete database platform capabilities: ELT and Analytics
  - Complete BI and Performance Management solutions from Oracle
  - Broadest array of third-party technologies and solutions
- On the right hardware infrastructure

#1 for Data Warehousing
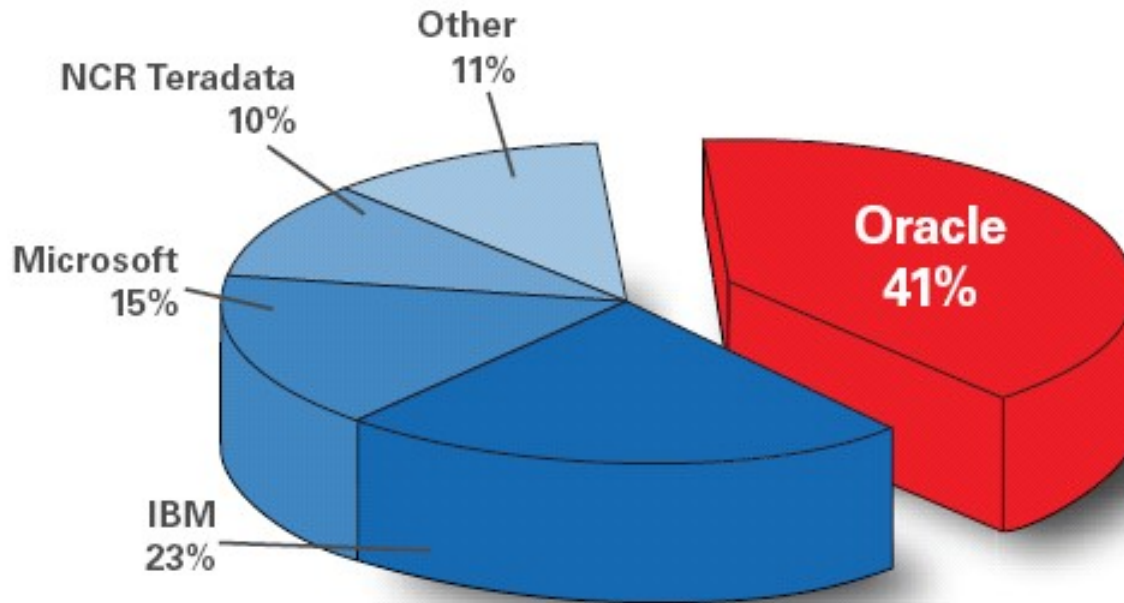
**Worldwide Data Warehouse Management
Market Share, 2006**

ORACLE

Q & A

ORACLE