



Enterprise Healthcare Analytics: Healthcare Data Warehouse Foundation (HDWF)

An Architecture Overview

Michael Schrader

Chief Architect Healthcare Analytics

Oracle Health Sciences GBU

Michael.Schrader@oracle.com



Safe Harbor Statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.



Agenda

- Why BI in Healthcare?
- Methodologies – Inmon, Kimball and Hybrid
- Meta Data
- Performance
 - Partitioning
 - Exadata
 - OLAP
- Single Code Repository



Analytics – at the Core of a Value-Based Health System

- Over the last year and more, the healthcare discussions that have occurred in the USA illustrate the need for healthcare transformation. After all, current estimates by World Health Organization put U.S. health care spending at approximately 16% of GDP.



Analytics – at the Core of a Value-Based Health System

- However, the demand that analytics drive HC transformation is not just revealed in the recent discussions in the USA but around the world. Private and public entities the world over are investigating how to maximize the power of analytics to improve:
 - Patient outcomes
 - Quality of care
 - Operational efficiencies
 - Costs
 - Billing/reimbursements – among others

Analytics are Vital to HC Transformation

1. Find what works

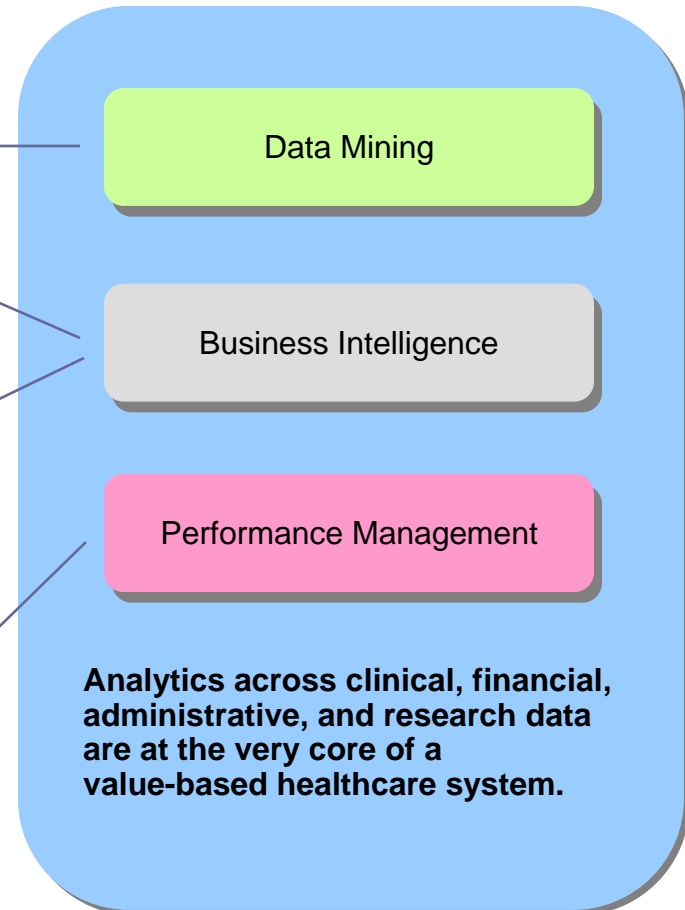
- Has the new FFP transfusion protocol lowered our costs resulting from post-operative infections?
- Will a 0.4% increase in the reimbursement rate from Blue Cross for uncomplicated deliveries restore profitability for this DRG at the individual hospital level and across our enterprise?

2. Do what works

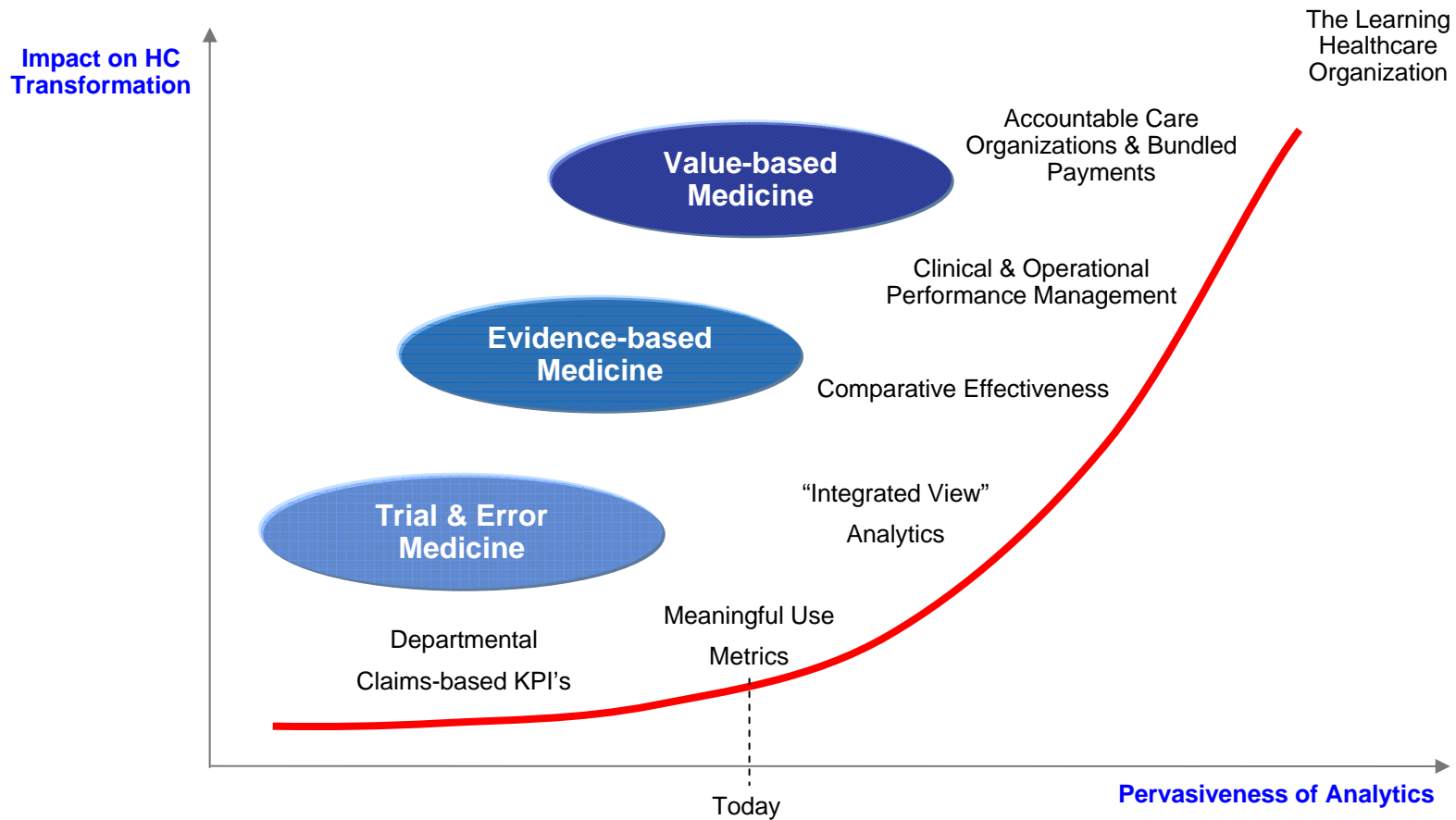
- Do HIV+ patients in our clinics receive anti-retrovirals according to our newly revised evidenced-based guidelines?
- Has the Surgical ICU staff at Cartwright Hospital brought their LoS in line with the enterprise target level without compromising quality metrics?

3. Get paid for results

- Meaningful Use
- Avoid “non-reimbursables”
- Pay-for-performance
- Accountable care / bundle payments



Analytics are Vital to HC Transformation



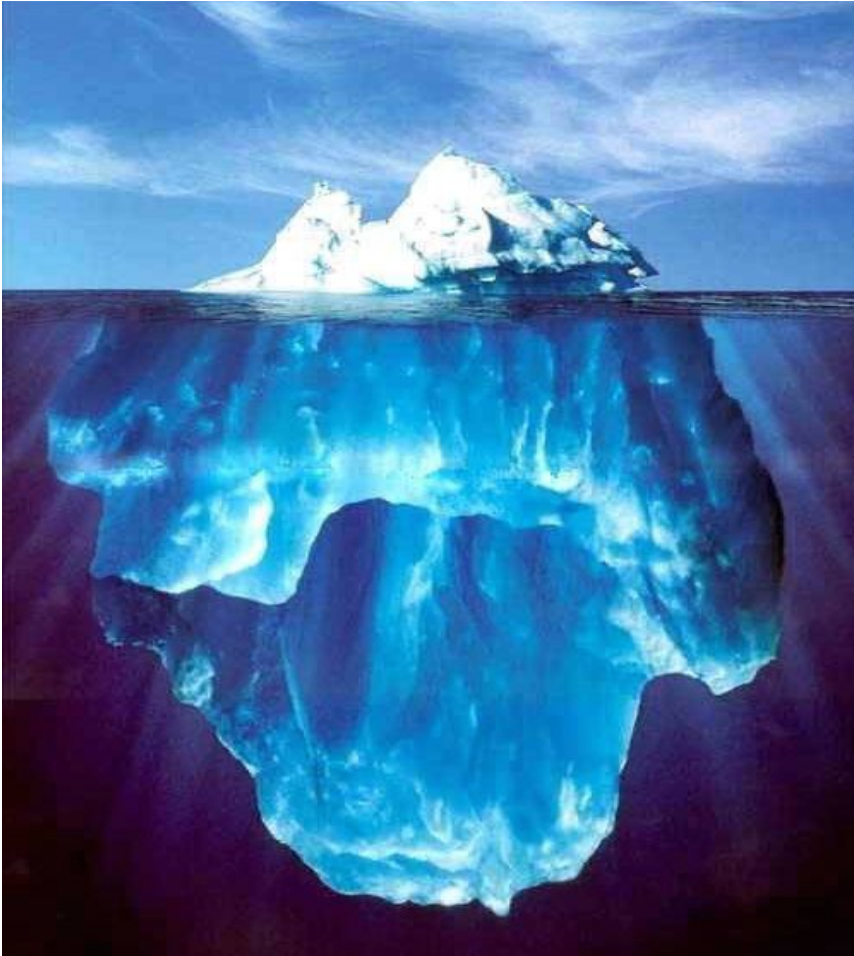
Core Clinical & Operational Systems

Enterprise Data Integration

Analytic Applications

HC Transformation requires much more than just an EMR. It requires integrated clinical, financial, administrative, and research data from across the Provider Enterprise and analytics.

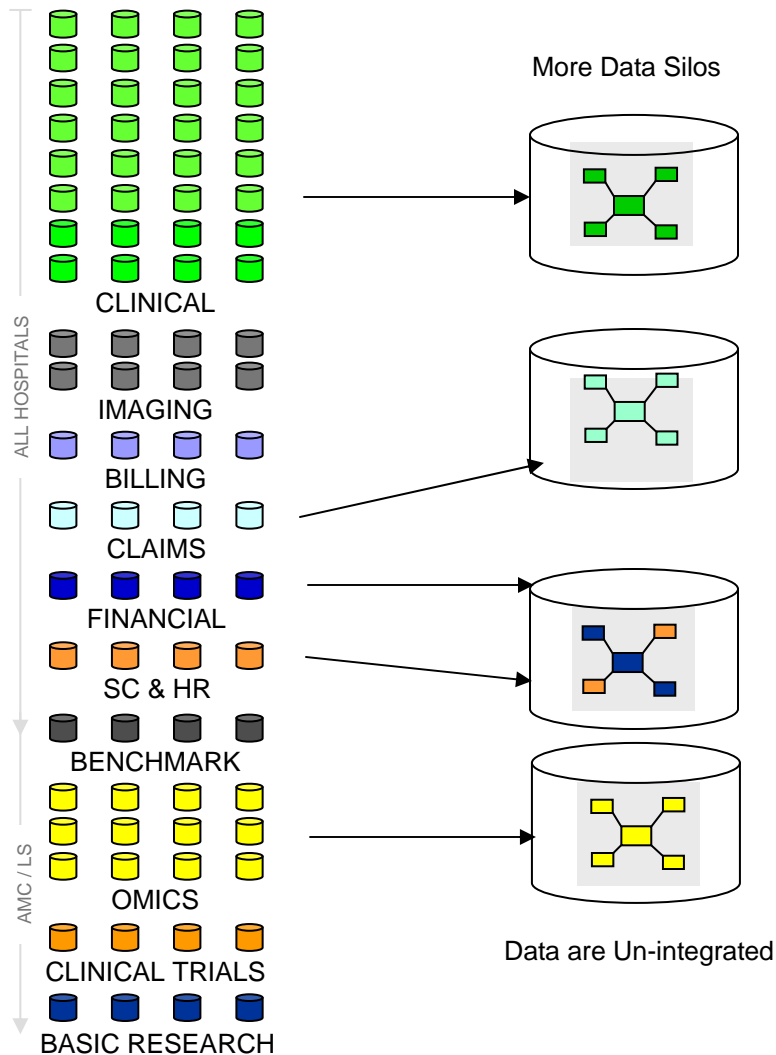
Analytics: More Than Meets the Eye



▶ Data analysis and presentation is the FUN PART.

▶ Data acquisition from myriad complex clinical, financial, administrative, and research source systems and the attendant cleansing, integration, and warehousing of these data is the HARD PART.

“Silo Analytics” Are Not the Answer



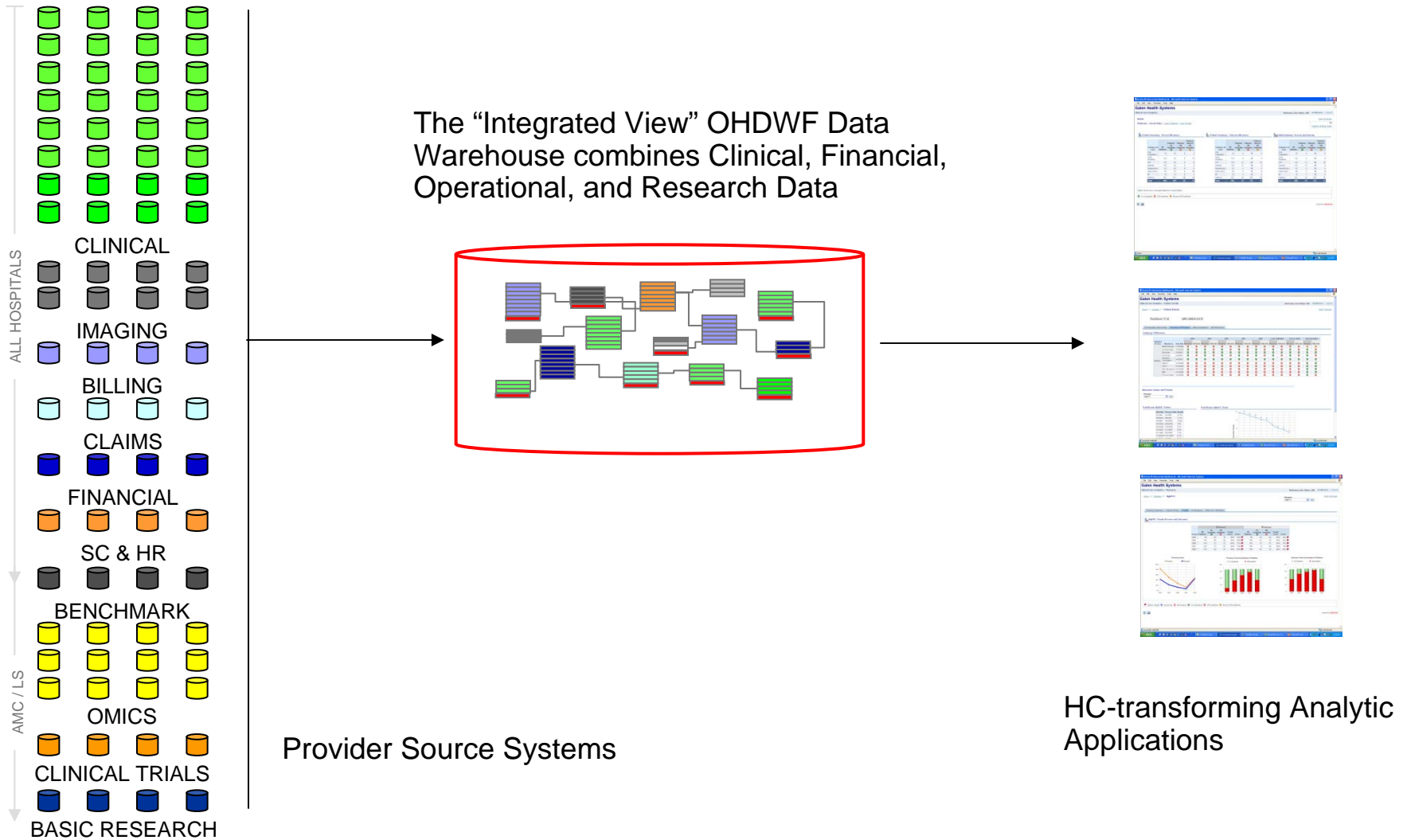
Silo Analytic Solutions cannot answer questions that span silos like:

“How is our nurse overtime policy affecting our ICU quality and patient satisfaction?”

“Do our cardiac care reimbursements reflect our improved quality measures?”

“How has our “Just In Time” ventilator policy affected the profitability of our ICUs?”

Transformational Analytics Require An Integrated View of the Provider Enterprise





What is HDWF?

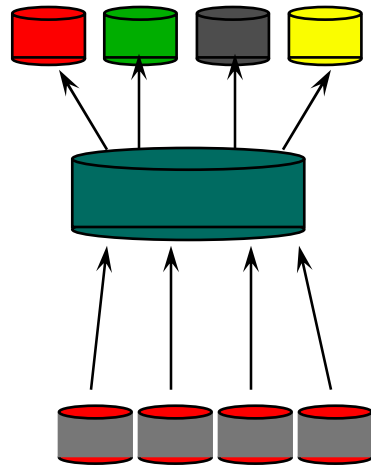
- **The HDWF provides an integrated view of enterprise-wide clinical and operational healthcare data that is optimized for healthcare business intelligence. This model consists of a logical and physical data model along with detailed user documentation.**
- **By leveraging Oracle's strong healthcare domain expertise, HDWF provides an industry standard-compliant foundation schema that is modern, relevant, topical, and addresses the needs of most healthcare segments. This foundation provides an integrated base for business information with fully defined entities, attributes, and relationships. You can take advantage of pre-built and pre-tested solution sets designed by industry experts that deliver relevant insights, are actionable, and aimed at improving both top-line and bottom-line results. HDWF, combined with Oracle technology, provides all of the components required for a complete and extendable Healthcare Data Warehouse framework that is designed to reduce complex and costly integration requirements and total cost of ownership.**
- **The solution takes advantage of Oracle's scalability and reliability, using Oracle's familiar optimization, parallelism and performance engineering within the database.**



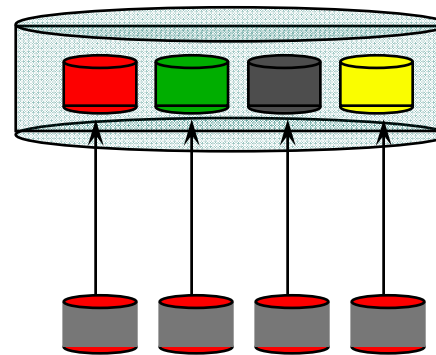
Architectural Discussion

- Methodologies – Inmon, Kimball and Hybrid
- Meta Data
- Performance
 - Partitioning
 - Exadata
 - OLAP
- Single Code Repository

Data Warehouse Architectures



**Enterprise Data
Warehouse**
Bill Inmon



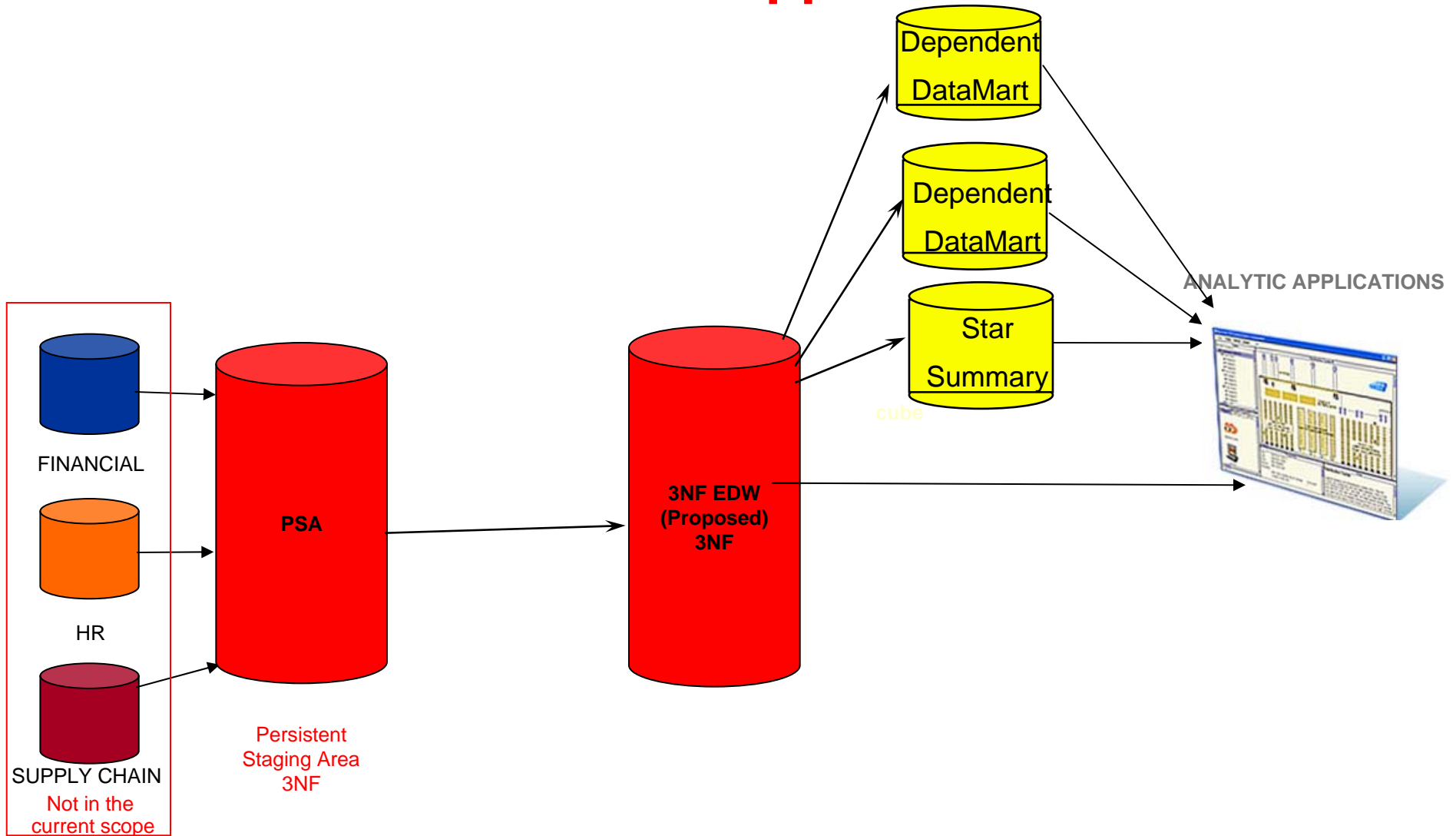
**Bus Architecture
Data Mart**
Ralph Kimball



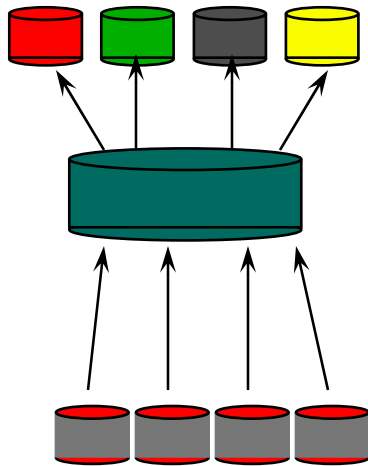
Methodologies - Bill Inmon

- Father of Data Warehousing – One central repository of all enterprise data.
- Corporate Information Factory with Caudia Imhoff
- “top down”

Oracle Healthcare Intelligence Architecture – Inmon Approach



Enterprise Data Warehouse - Inmon



Advantages

- ◆ Single version of the Truth
- ◆ Common definition of data
- ◆ Robust data transformation
- ◆ 3rd Normal Data

Issues

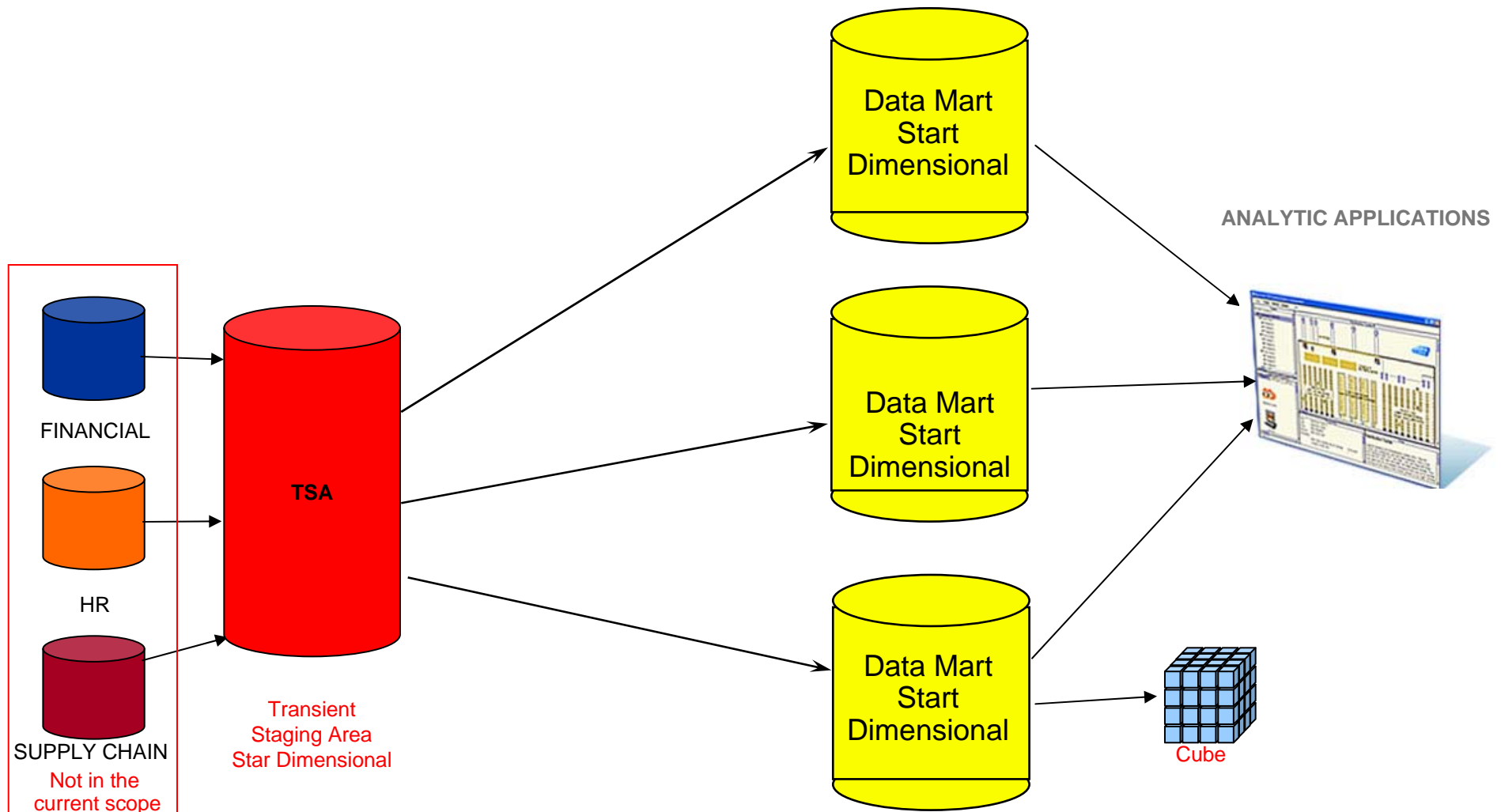
- ◆ Very Time Intensive and Expensive
- ◆ Complex ETL processes
- ◆ Drilling Down from Data Marts
- ◆ Data, Data and more Data



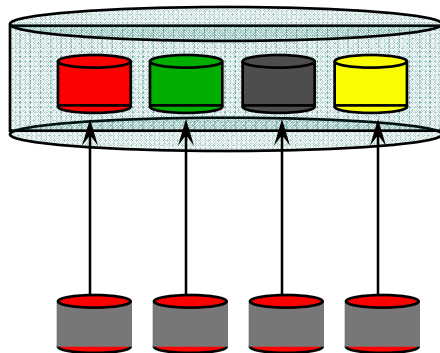
Methodologies – Ralph Kimball

- **Dimensional Model – Star Schema**
- **Data Warehouse Bus Architecture**
 - **Suite of dimensions that all data marts conform to set of Standardized facts**
 - **Central Meta Data repository defining the DWB**
 - **Collection of Data Marts become the EDW**
- **“Bottom Up”**

Oracle Healthcare Intelligence Architecture – Kimball Approach



Kimball Approach



Advantages

- ♦ Single version of the Truth
- ♦ Common definition of data
- ♦ Fast Incremental Delivery
- ♦ Cross-functional data availability
- ♦ De-centralized, intuitive data navigation functionality
- ♦ Balanced workload capacity for data growth and increased users

Issues

- ♦ Challenging to get agreement for Conform Dimension
- ♦ Summary across marts difficult
- ♦ Data Integration – More ETL



Methodologies – Hybrid

- The first half is Inmon – Data Integration
- The second half is Kimball – Data Access
- With HOLAP cubes



Metadata Discussion

- ***Metadata Types***
- Metadata is very important for the management and operation of a data. In today's environment, there exist legal requirements that enforce the importance of metadata. Metadata can be classified into four areas:
 - Business Metadata;
 - Operational Metadata;
 - Technical Metadata and;
 - Process Metadata.



Metadata Discussion

- ***Business Metadata***
- Business metadata describes the business meaning of data. It includes business definitions of the objects and metrics, hierarchies, business rules, and aggregation rules.
- ***Operational Metadata***
- Operational metadata stores information about who accessed what and when. This information is not only important for legal requirements but for the design of the data warehouse itself. For example, we can identify that a particular data mart is not being utilized. Should we eliminate the data mart? Should we be providing better education for the end users.



Metadata Discussion

- ***Technical Metadata***
- Technical Metadata describes the data structures and formats such as table types, data types, indexes and partitioning method. Also, it describes the location of the data elements. With technical metadata, version control of database structures is possible. Services?
- ***Process Metadata***
- Process Metadata describes the data input process. It includes data cleansing rules, source target maps, transformation rules, validation rules and integration rules. Also, it includes the history of data input such as the time of the data loads.

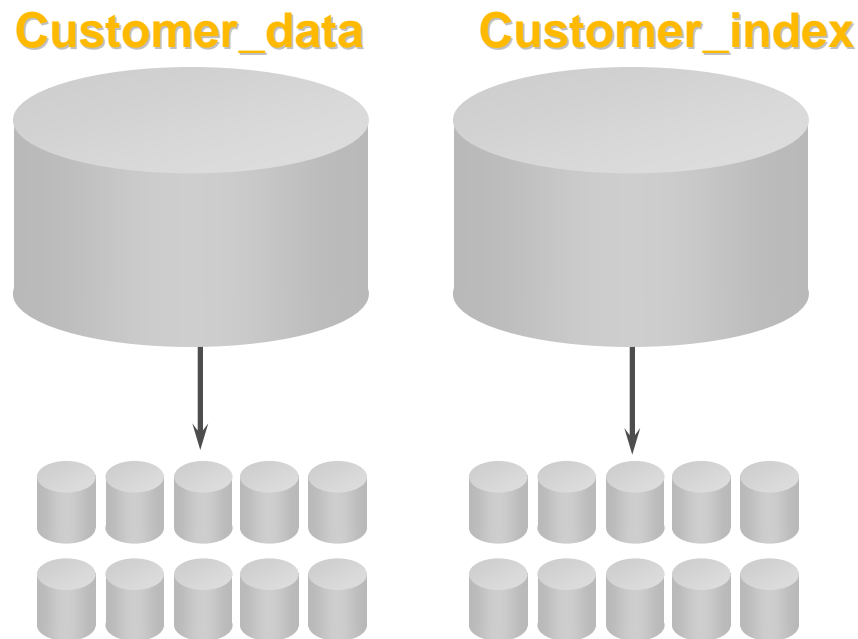


Performance

- **Performance**
 - Partitioning
 - Exadata
 - OLAP

What Are Partitioned Tables and Indexes?

- What Are Partitioned Tables and Indexes? The table and index is stored in many physical partitions.
- Partitioning allows tables to growth to many, many Tbs.





Partitioning is Key

- Reducing Downtime for Backup & Recovery. Backup and restore can be done on individual partitions
- Reducing time for loading new data
- Reducing I/O bottlenecks for Very Large Databases (VLDBs)
- Support for Parallel Queries - This is very important for data warehousing applications where tables can be several hundred gigabytes. The optimizer eliminates partitions that do not need to be scanned
- Partition transparency - When I insert a new row into a partitioned table, Oracle automatically selects the correct partition to store it in
- High Availability - Partitions that may be unavailable do not affect queries or DML operations on other partitions within that table



11g Partitioning New Features – Interval Partitioning

- Range partitioning allows you to create partitions based on ranges of the values of the partition key column. Here is an example of the range partitioned table:
- create table sales
(sales_id number,
sales_dt date)
partition by range (sales_dt
(partition p0701 values less than (to_date('2010-02-01','yyyy-mm-dd')),
partition p0702 values less than (to_date('2010-03-01','yyyy-mm-dd')));
Here you have defined partitions for January 2010 and February 2010 only, so what happens if a record is inserted into the table that has the sales_dt in March 2010?
- The insert will fail with the following error:
- ORA-14400: inserted partition key does not map to any partition Obviously you need to add a partition for March 2010 before you can insert a record. But this is often easier said than done. Often you can't afford to create a lot of partitions beforehand and too few of them may result in this error.



11g Partitioning New Features – Interval Partitioning

- Wouldn't it be better if Oracle somehow automatically sensed the need for new partitions and then created them? Oracle Database 11g does, with a feature called Interval Partitioning. Here, you don't define partitions and their boundaries but merely an interval that defines each partition's boundaries. Here is the same example in interval partitioning:
- ```
create table sales(
 sales_id number,
 sales_dt date)
 partition by range (sales_dt)interval (numtointerval(1,'MONTH'))
(partition p0701 values less than (to_date('2010-02-01','yyyy-mm-dd')));
```
- Note the clause: interval followed by the interval. Here you have instructed Oracle to create intervals of one month each. You have also created the initial partition named p1001, for the January 2010 data. Now suppose you insert a record with June 2010 data:
- ```
SQL> insert into sales values (1,'01-jun-010'); 1 row created.
```



11g Partitioning New Features – Interval Partitioning

- Now suppose you enter a value lower than highest value, such as May 1, 2010. It should ideally have its own partition, as your partition interval is a month.
- SQL> insert into sales6 values (1,'01-may-10'); 1 row created.
- The partition was created by splitting the SYS_P41 partitions (for June). Thus, Oracle automatically creates and maintains the partitions when you define an interval partitioning scheme.



Exadata

“Exadata is the fastest growing product in Oracle’s history”

- Oracle President, Charles Phillips

“The Exadata sales pipeline is fast approaching the \$1 billion mark”

- Oracle CEO, Larry Ellison

Exadata Architecture

Scaleable Grid of industry standard servers for Compute and Storage

- Eliminates long-standing tradeoff between Scalability, Availability, Cost

Database Grid

- 8 compute servers (1U)



- 64 Intel cores
- 576 GB RAM

InfiniBand Network

- 3 36-port 40Gb/s switches
- Unified Net- servers & storage
- 324 FC Ports equivalent



Storage Grid

- 14 storage servers (2U)



- 112 Intel cores in storage

- 100 TB SAS disk, or
336 TB SATA disk

- 5 TB PCI Flash

- Data mirrored across
storage servers

Exadata V2 Goals



- ***Ideal Database Platform***
 - Best Machine for **Data Warehousing**
 - Best Machine for **OLTP**
 - Best Machine for **Database Consolidation**
- **Unique** Architecture Makes it
 - Fastest, Lowest Cost

Scales to 8 Racks by Just Adding Cables



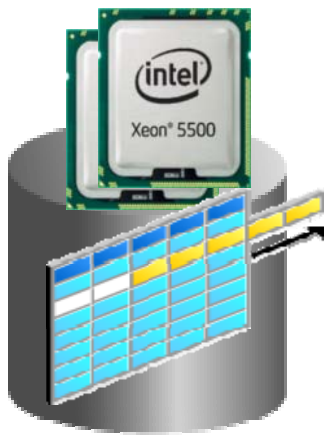
8 Rack Exadata

- Total Compute Cores 1408
- TB of Disk 2640
- InfiniBand Ports 604

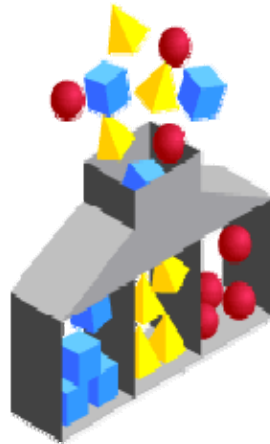
Comparable To

- 20** IBM Power 595s
- 13** Rack EMC Symmetrix
- 4300** FC Ports

Keys to Speed and Cost Advantage



Exadata Intelligent
Storage Grid



Exadata Hybrid
Columnar Compression

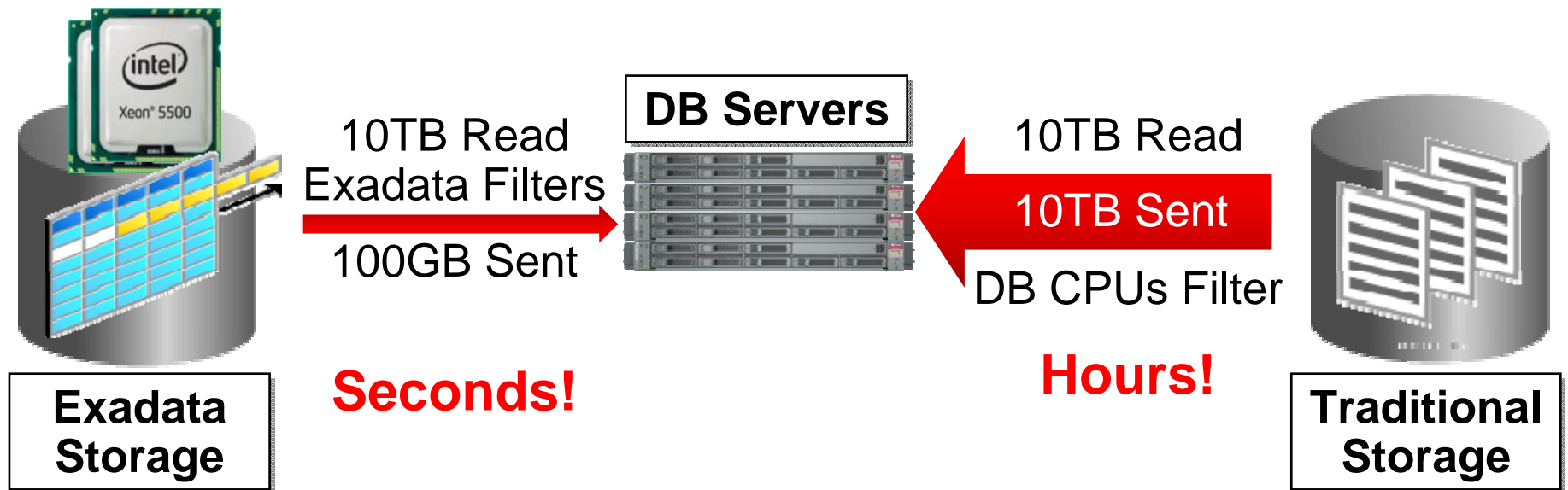


Exadata Smart
Flash Cache

Exadata Intelligent Storage Grid

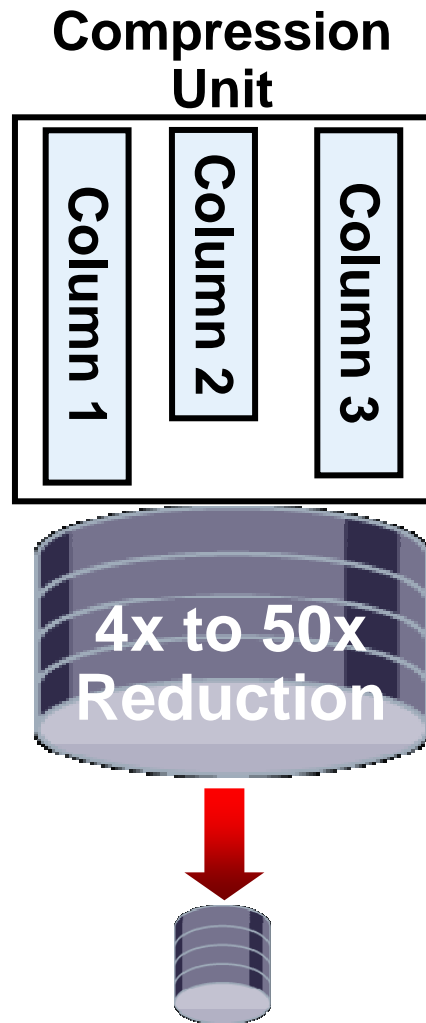
Most Scalable Data Processing

- Data Intensive processing runs in Exadata Storage Grid
 - Filter rows and columns as data streams from disks (112 Intel Cores)
 - Scale-out storage removes bottlenecks
- Example: How much product X sold in month Y



Exadata Hybrid Columnar Compression

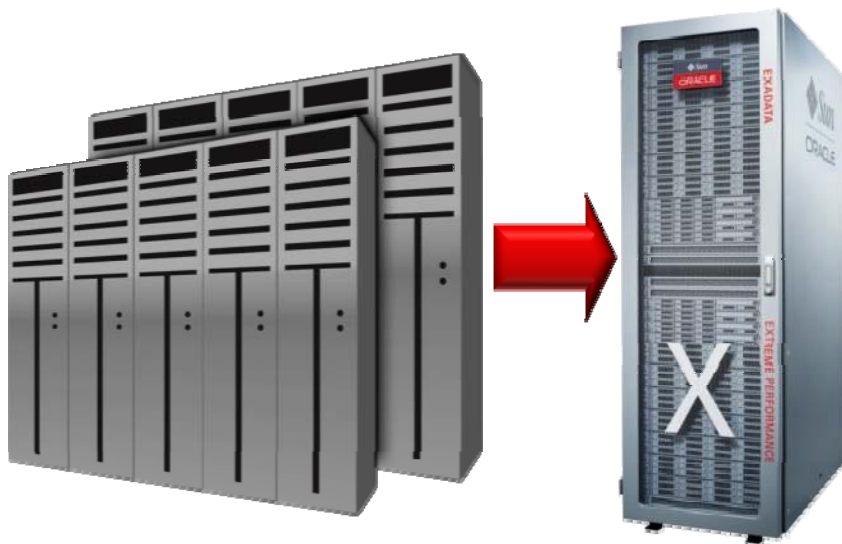
How it works



- Tables are organized into sets of a few thousand rows
 - Compression Units (CUs)
- Within CU, data is organized by column, then compressed
 - Column organization brings similar values close together, enhancing compression
- Useful for data that is bulk loaded and queried
 - Update activity is light
- Compared to best conventional algorithms – Gzip, Bzip2
 - Typically 2X the compression, 10X the performance
- Exadata servers offload filtering, projection, etc. for scans on compressed data
 - Indexed accesses return compressed blocks to database so buffer cache benefits from compression

Exadata Smart Flash Cache

Extreme Performance OLTP

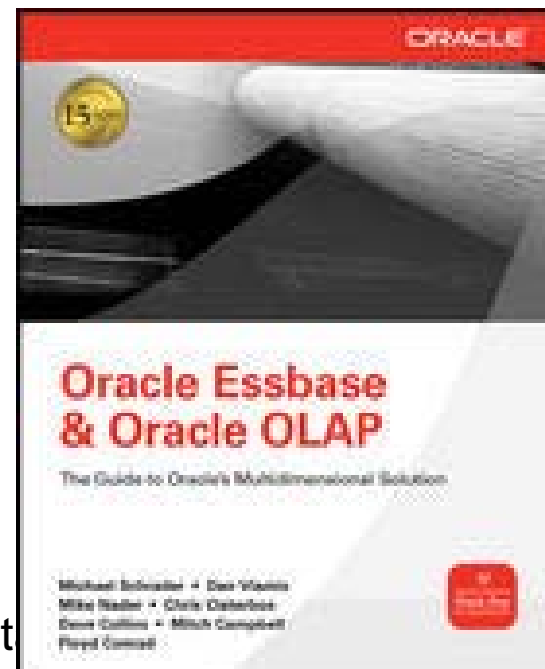


**5X More I/Os than
1000 Disk Enterprise
Storage Array**

- Exadata has **5 TB** of flash
- **56 Flash PCI cards avoid disk controller bottlenecks**
- **Intelligently manages flash**
 - Smart Flash Cache holds hot data
 - Avoids large scan wipe-outs of cache
- **Gives speed of flash, cost of disk**
- Exadata flash cache achieves:
 - Over **1 million IO/sec from SQL** (8K)
 - Sub-millisecond response times

Performance OLAP -Oracle Essbase & Oracle OLAP: A Guide to Oracle's Multidimensional Solution

- Published by Oracle Press – Sept 2009
 - Michael Schrader
 - Dan Vlamis
 - Michael Nader
 - Chris Clasterbos
 - David Collins
 - Floyd Conrad
 - Mitchell Campbell
- Cover both Oracle Essbase and Oracle OLAP Data
- 500 Pages
- See <http://mhprofessional.com/product.php?isbn=0071621822> for details



Slide 38

JK1

Place the names in alphabetical order.

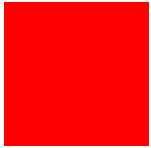
Kopcke, 10/15/2008



Code Repository

- **Code Repository**

- The Code Repository is used to contain all coded (and non-coded) lookup values used within the HDWF. The Code Repository hosts customer-defined/developed terminologies as well as standardized terminologies provided by commercial, national and international standards bodies (e.g. ICD-9-CM, SNOMED, NDC, etc.). Code System versioning is supported.



ORACLE®