Emerging Technologies
Shaping the Future of
Data Warehouses & Business Intelligence

Appliances and DW Architecture

John O'Brien
President and Executive Architect
Zukeran Technologies

Zukeran
technologies

1

OBJECTIVES

- To define an "appliance"
- Understand critical components of a DW appliance
- Learn how DW appliances fit within your existing DW architecture

Zukeran
technologies

WHAT IS AN "APPLIANCE?"

General Purpose   Multi–purpose   Purpose specific

AN APPLIANCE IS BORN…

Applications (Operational, BI, DSS)	Purpose
Database	Purpose Specific Appliance
Operating System (Windows, Unix, Linux, etc)	
Computing Resources (CPU, memory)	
Networking Resources	
Storage Resources (Persistence)	

Compatibility

Horizontal versus Vertical Layers
When does it flip?

Zukeran technologies

4

PROVEN APPLIANCES

Google search appliance

Symantec VPN appliance

Linksys Wireless router

TiVo

Zukeran technologies

CHARACTERISTICS OF A GOOD APPLIANCE

- Clear, specific purpose
- Compatible with existing infrastructure
- Ease of installation & use
- Highly cost effective

Zukeran
technologies

MAKING A DW SPECIFIC APPLIANCE?

- Optimize for mixed BI workloads
  - Store and access lots of data
  - Perform analytical queries
  - Support operational report requirements
  - Expand data access to the mainstream DW user
- Combine Server + Optimization Software + Storage
- Ensure ease of use
- Offer low maintenance and TCO
- Integrate with existing infrastructure
  - BI tools with ANSI SQL
  - Ability to load data easily
  - Compatible with standard operating procedures
  - Build in flexibility to support changing requirements

Zukeran
technologies

PRE-CONFIGURED vs. APPLIANCE

- Appliances are not a "pre-configuration" or assembly of tuned products and technologies
- An appliance is built for a specific purpose and can leverage commodity parts
- Appliances require no user serviceable parts

- Test: Does it matter what's inside the appliance?

DW APPLIANCE: HIGHLY COST EFFECTIVE

- Reduce DBA expertise and administration
  - No tablespaces, extents to manage
  - No archive, redo logs to manage
  - No partitioning management
- Reduce need for storage expertise
  - SAN storage architects
  - LUNs, meta volumes
- Relatively low acquisition cost vs. traditional model
  - Server manufacturer annual maintenance
  - Operating System license
  - Storage hardware
- Efficient power consumption

FUTURE DATA WAREHOUSE APPLIANCES?

Will we see these specific appliances in the future?

Do they meet the criteria for success?

- ETL or ELT appliance ?

- BI appliance ?

- EII appliance ?

- OLAP appliance ?

- Audit appliance ?

WHERE DO DW APPLIANCES MAKE SENSE?

The Corporate Information Factory and the Web Environment

by Bill Inmon and Claudia Imhoff
Copyright ©2001, all rights reserved
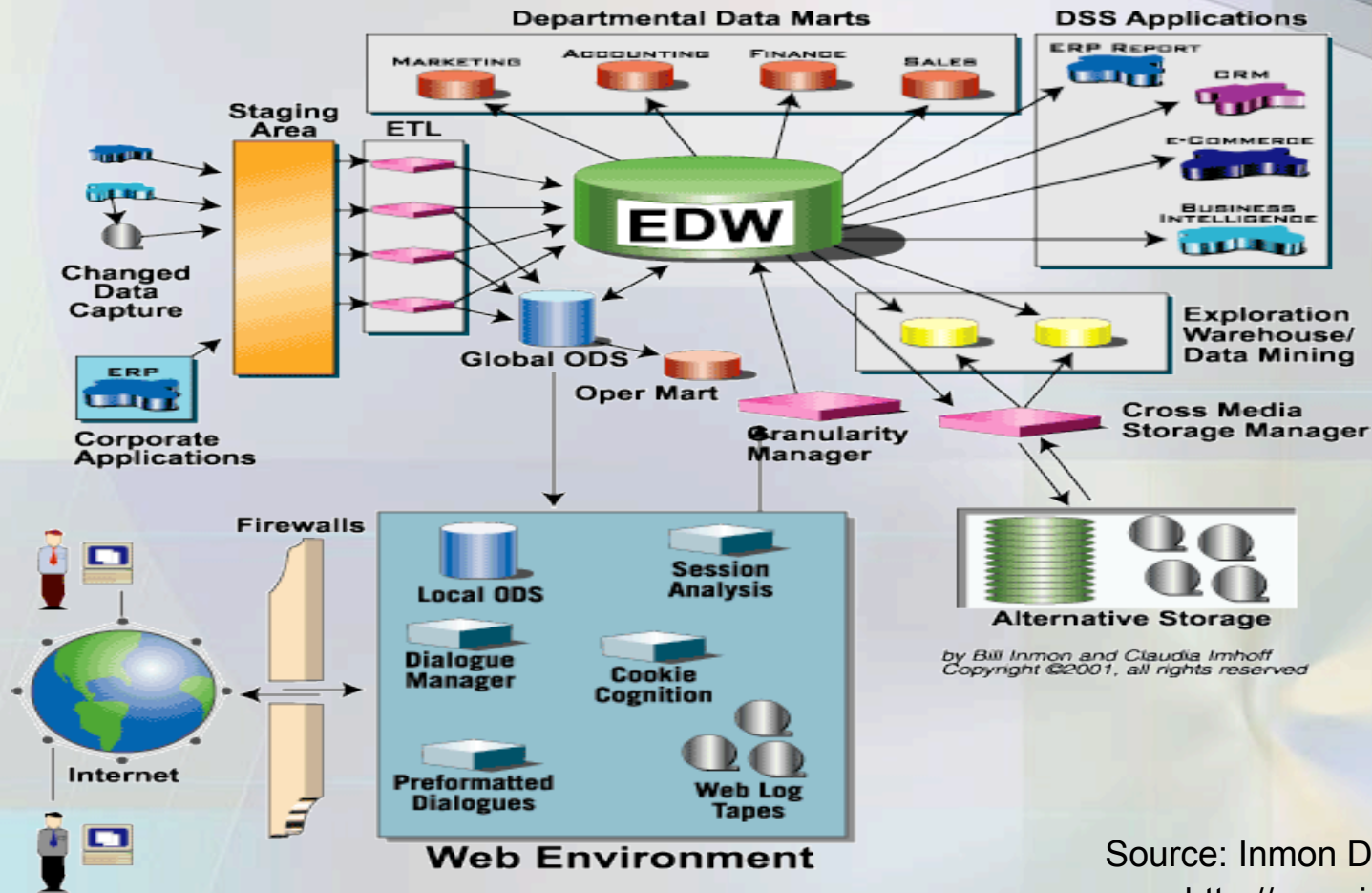
Source: Inmon Data Systems
http://www.inmoncif.com

DW APPLIANCE: HIGH PERFORMANCE

- Massive Parallelism inherent to the architecture
- Improved I/O throughput rates with effective I/O

Ethernet
ODBC
JDBC

Host Database

CPU
CPU
CPU
CPU
CPU

Backplane

Zukeran
technologies
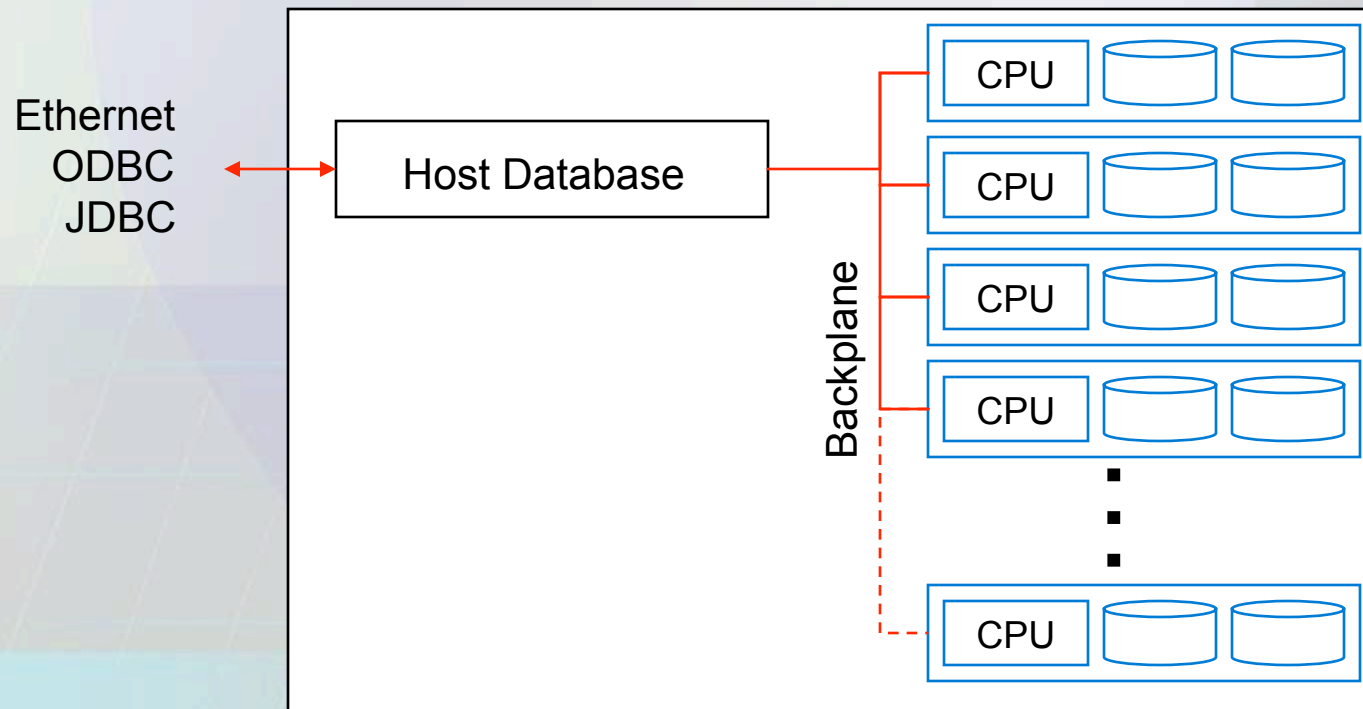
DW APPLIANCE: LARGE SCALE DATABASES

- How do you effortlessly scale beyond 2TB as your data warehouse matures?

- Federated architecture
- Rapid expansion and data redistribution
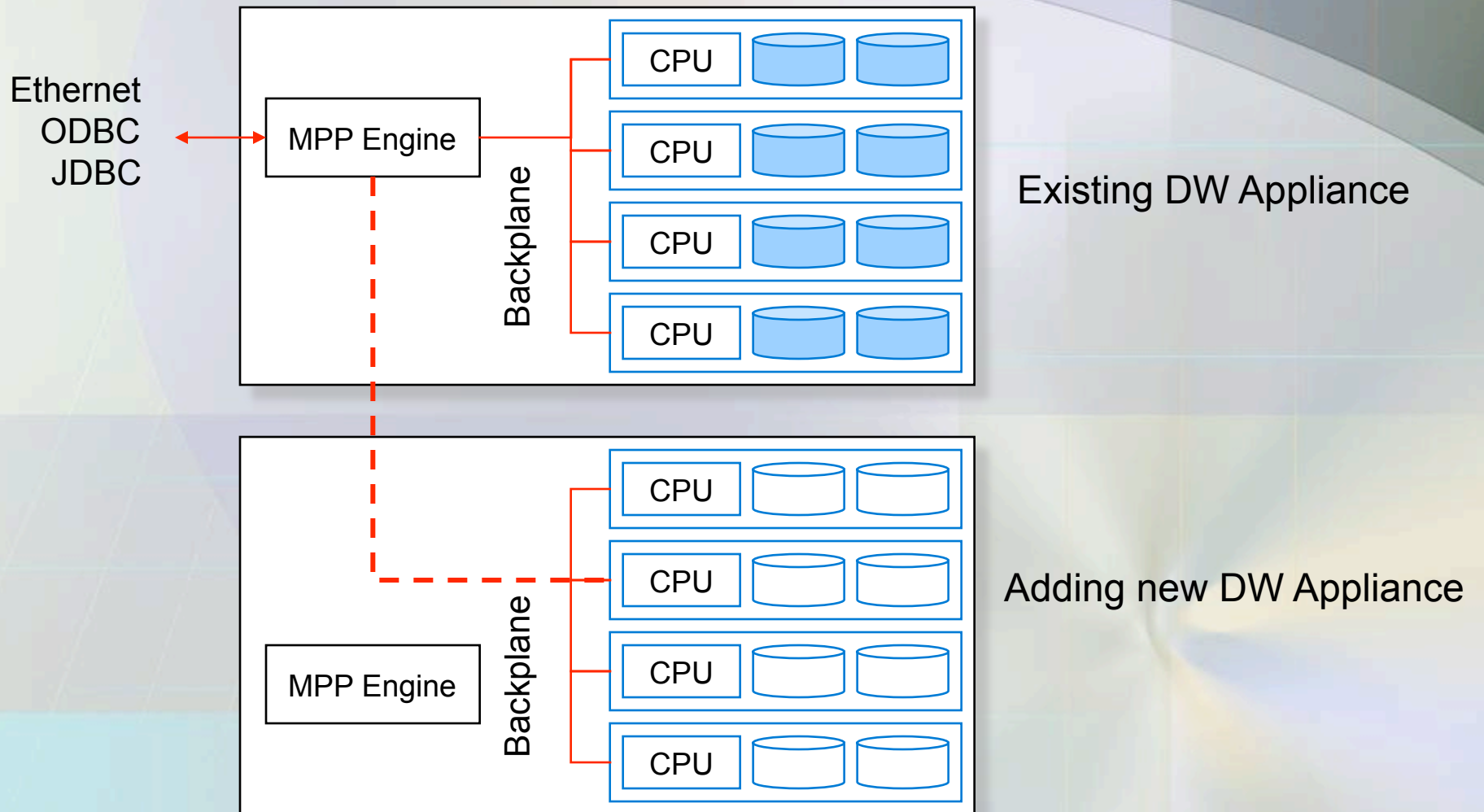
Zukeran
technologies

SCALABILITY: FEDERATED ARCHITECTURE

- Logical areas of the data warehouse architecture move to new appliances
  - Data marts, ODS, staging
- Subject areas or conformed dimensions stay together unless the capability to perform database joins across platforms exist
  - EII tool, remote tables
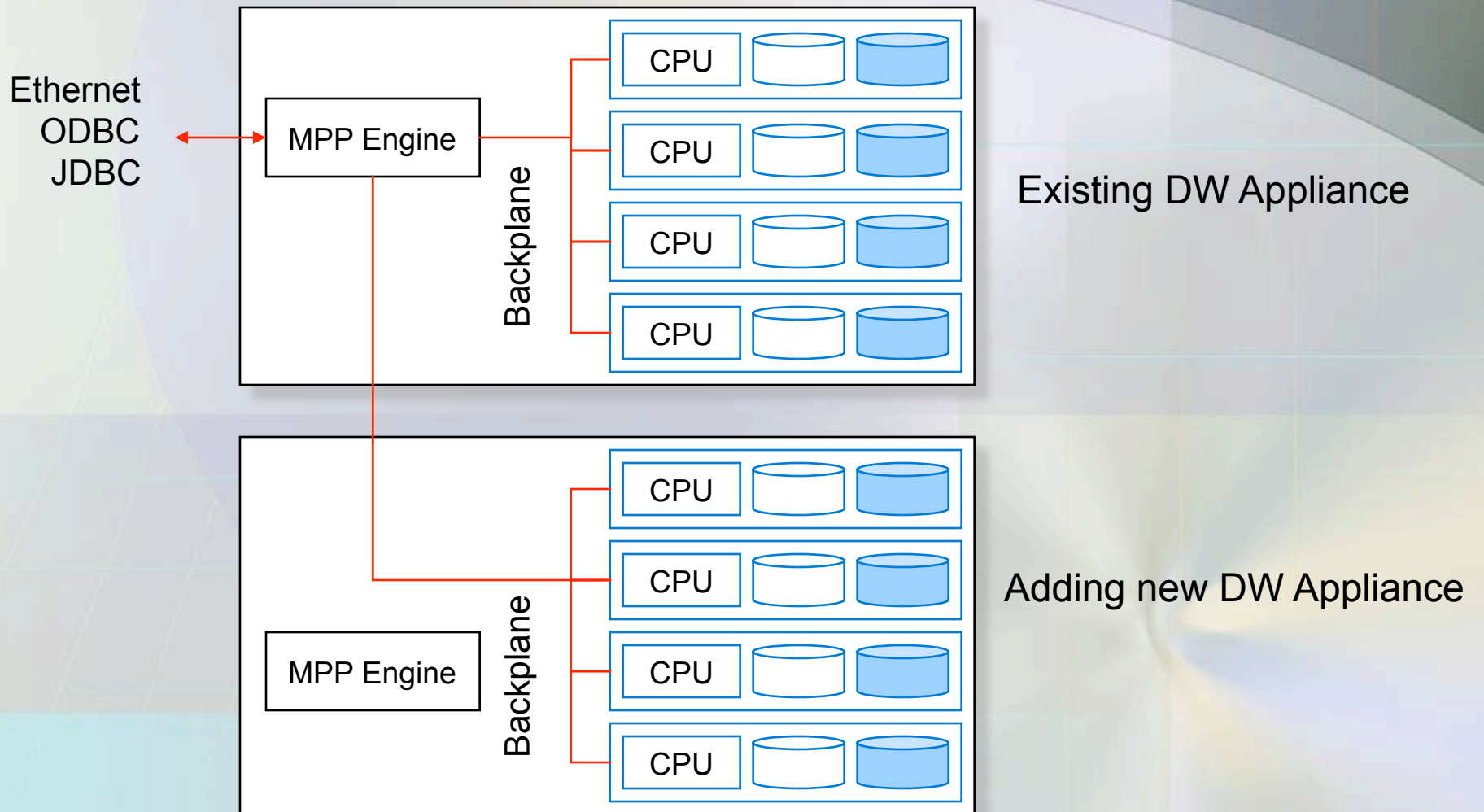- Logical groups such as North America DW, Euro DW or corporate entities

SCALABILITY: DATA REDISTRIBUTION

- Adding capacity may cause data to be redistributed

    – Buckets

    – SLA Driven

    – Redistribution

Zukeran
technologies

SCALABILITY: DATA REDISTRIBUTION

Ethernet
ODBC
JDBC

MPP Engine

Backplane

CPU

CPU

CPU

CPU

Existing DW Appliance

MPP Engine

Backplane

CPU

CPU

CPU

CPU

Adding new DW Appliance

Zukeran
technologies

MAINTAINING PERFORMANCE WITH SCALABILITY

Ethernet
ODBC
JDBC

MPP Engine

Backplane

CPU

CPU

CPU

CPU

Existing DW Appliance

MPP Engine

Backplane

CPU

CPU

CPU

CPU

Adding new DW Appliance

Zukeran
technologies

17

DW APPLIANCE: BACKUP AND RECOVERY

- Built-in RAID mirroring and striping

- The challenge is that large systems take massive amount of computing and network resources to backup changes in a high volume environment

- Recommended to compress and save load files since loads are faster than recoveries

- In a fully mirrored environment, chances of needing to restore from a backup are low

- Veritas or other standard APIs are available backups

Zukeran technologies

DW APPLIANCE: FAST LOAD/UNLOAD

- Minimum 500GB per hour load rates

- Near physical limitations disk I/O and Ethernet connection to appliance

- Utilizes high performance ODBC drivers and special loader utilities

Zukeran technologies

IMPACT OF DATA WAREHOUSE APPLIANCES

Zukeran
technologies

WHAT A FAST DATABASE CAN DO FOR YOU

- **Before ETL tools**
  - ETL was hand coded programs
  - ETL was code in database procedural languages

- **ETL tools offered**
  - Faster development with 3GL
  - Better performance than database code & SQL

- **High Performance data warehouse appliance**
  - Faster queries on large data sets
  - → High Performance SQL

Zukeran
technologies

ELT – A PARADIGM SHIFT

- Transformations typically are performed after extraction and before loading limiting the database workload which is thought to be query intense for data warehouses.

- High performance data warehouse appliances are being used to transform the data inside the high performance database and the results stored in the appliance or in other databases
  - Some companies point to ETL license savings as part of the ROI for appliances
  - Sunopsis is a tool being used in these cases

Zukeran technologies

ROLAP vs. MOLAP

- Multidimensional database engines were created to make up for the performance deficiencies of relational OLTP databases at that time.

- MOLAP cubes are known for their high performance interactive capabilities, complex calculations and dimensional and hierarchical analysis but are challenged in:
  - Real-time environments
  - Large datasets
  - High degree of dimensional and hierarchical branches

- MOLAP cubes are also a duplication a data in another database format, the multidimensional database.

- ROLAP leverages on the underlying database for performance characteristics

- ROLAP metrics can be atomic level, pre-calculated, pre-rolled up depending on what works best

Zukeran technologies

FOCUS ON BUSINESS VALUE OF DATA

- ## Data Architecture and Modeling
  - High performance databases shouldn't make up for poor data analysis and modeling efforts

- ## Good Requirements, Analysis and Reports
  - Make sure that the wrong answer doesn't just come back faster…

Zukeran technologies

WHAT THE BUSINESS WANTS FROM DW

- ## On-Demand
  - New products and projects

- ## Lower Cost
  - Overall TCO and long term maintainability

- ## Flexibility
  - Simpler infrastructure to build and go
  - Able to update as the business needs change

- ## New capabilities
  - Scalability
  - Ability to store more data with less cost

Questions and Discussion

Zukeran
technologies