Information Management

IBM

IBM Information On Demand Conference 2009

Optimize Your Business for Greater Results with InfoSphere Warehouse

Chi-Chung Hui

Senior I/T Specialist

IBM Software Group, Hong Kong

Agenda

- InfoSphere Family Overview

- InfoSphere Warehouse

  – Modular Design

  – Performance and Scalability

  – No Copy Analytics and Cubing Services

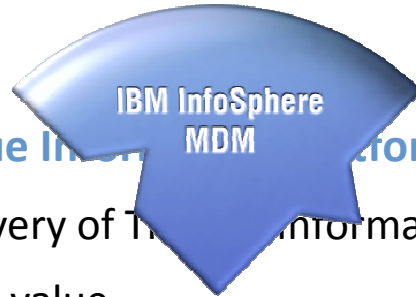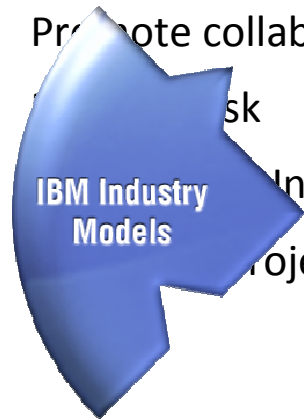 – Deep Compression

  – Workload Management
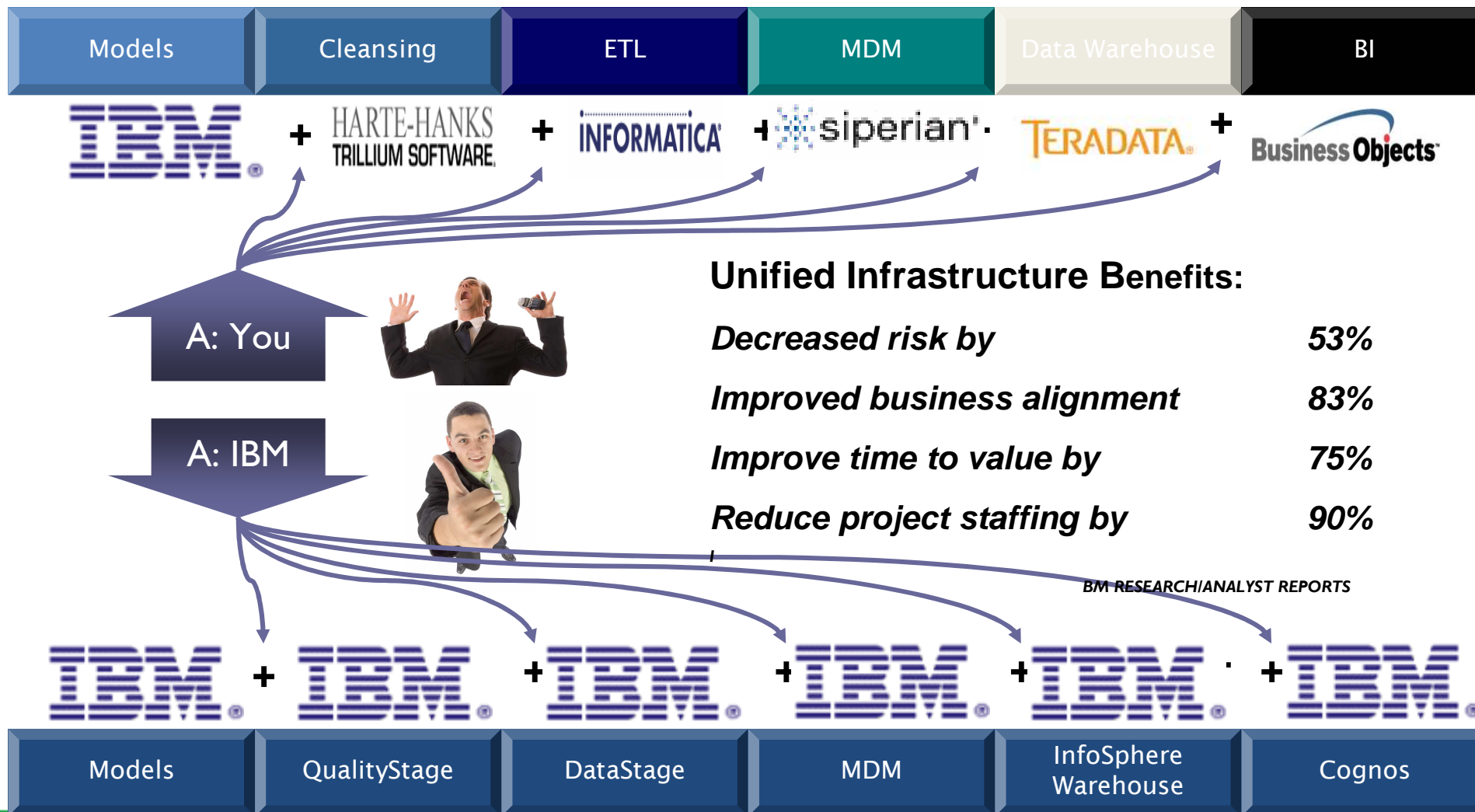
The IBM InfoSphere Vision

IBM InfoSphere MDM

An Industry Unique Information Platform

- Simplify the delivery of Trusted Information

- Accelerate client value

- Promote collaboration

- Reduce risk

- Integrated

- Project to Enterprise

IBM Industry Models

IBM InfoSphere Information Server

Expertise and Experience

Flexibility

IBM InfoSphere Warehouse

Common Metadata

Governance

IBM Positioning in Business Intelligence
Single Integrated Infrastructure Of Best Of Breed Technologies

Models	Cleansing	ETL	MDM	Data Warehouse	BI

IBM + HARTE-HANKS TRILLIUM SOFTWARE. + INFORMATICA + siperian + TERADATA + Business Objects

A: You

A: IBM

Unified Infrastructure Benefits:

Decreased risk by	*53%*
Improved business alignment	*83%*
Improve time to value by	*75%*
Reduce project staffing by	*90%*

BM RESEARCH/ANALYST REPORTS

IBM + IBM + IBM + IBM + IBM · + IBM

Models	QualityStage	DataStage	MDM	InfoSphere Warehouse	Cognos

InfoSphere Warehouse
Full Featured Data Warehouse Platform

Universal Access

Portals & Web	Reporting Solutions	MS Office
Web Services	SQL/MDX	MDX

Advanced Capabilities

No Copy Analytics	Advanced Design & Management	Extreme Performance
OLAP Analytics	Design Studio and Admin Console	Workload Management
Data Mining		Data Compression
Unstructured Analysis	Embedded Data Movement and Transformation	Data Retention

MPP Data Server – DB2 ESE

Platforms

C-Class   D-Class   E-Class   Linux/ Windows   AIX / SUN / HP

What is InfoSphere Warehouse?

The InfoSphere Warehouse is…

1. A **proven building block methodology** that focuses on balance as a key design principle for a warehouse. This methodology has been used by IBM for years.

2. IBM Balanced Warehouse solutions are **rigorously tested and tuned** for performance and reliability.

3. Before a Balanced Warehouse is released, **IBM's laboratories** configure and stress test the systems for maximum reliability and security as well as fine tune them to ensure that the highest levels of query performance are met.

InfoSphere Warehouse is a solution bundle that takes the "guess work" out of configuring a data warehouse

A set of document and best practice exists that help you to build a successful data warehouse!

Incremental Investment: Typical Growth Path

Legend

- Combined Coordinator, Catalog and Single Partition Data BPU
- Coordinator BPU
- Multi Partition Data BPU

Admin BCU 1

Admin BCU 2

MEM

CPU CPU

MEM

CPU CPU

BPU
DB2
MEM
CPU
Data BCU 1

BPU
DB2
MEM
CPU
Data BCU 2

BPU
DB2
MEM
CPU
Data BCU 3

BPU
DB2
MEM
CPU
Data BCU 4

BPU
DB2
MEM
CPU
Data BCU 5

BPU
DB2
MEM
CPU
Data BCU 6

BPU
DB2
MEM
CPU
Data BCU n
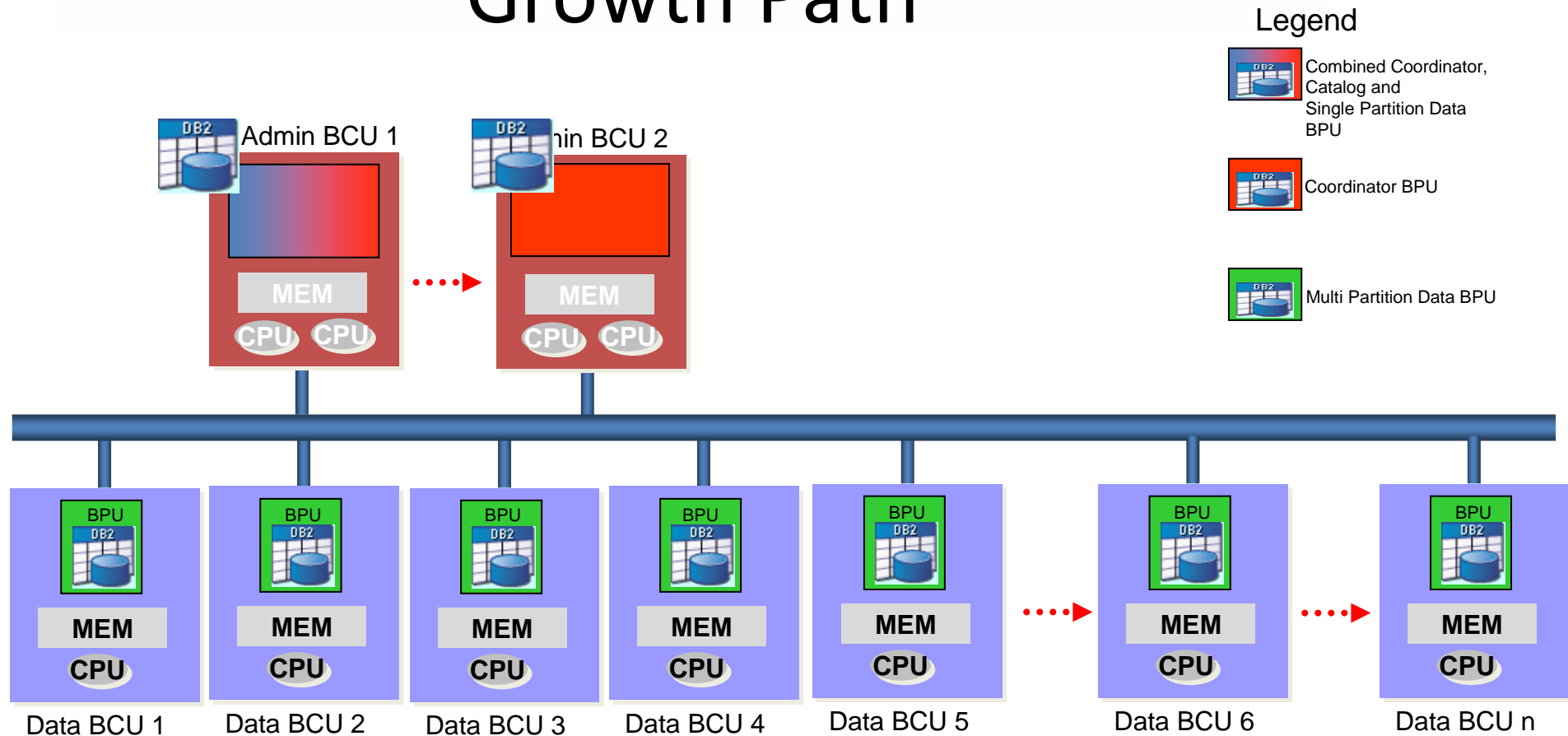
IBM InfoSphere Balanced Warehouse
Pre-built, Pre-packaged Warehouse Appliance Offering

- **Reduces Risk**
  - Standard Pre-tested Configuration
  - Ongoing Supported Package
  - Predictable performance

- **Reduces your Time to Value**
  - Create Table Ready
  - Faster to Deploy
  - Built for Performance
  - Easy to Integrate into any environment

- **Reduces Warehouse Cost of Ownership**
  - Automated Management
  - Single Product Support
  - System Built for Performance

- **Configured for your Business Requirements**
  - Multiple configurations
  - Modularly expandable
  - Available in different sizes

E-Class

Modular for flexible Scale out
Highly Scalable EDWs

E7100 for large scale and complex workloads

D-Class

Modular for flexible Scale out
Advanced Departmental Data Marts and
Growing Data Warehouses

D5100 for great performance and high availability

C-Class

Business Partner / IBM

Easy to Deploy Warehousing
applications and reporting tools

C4000 – IBM/BP - up to 4TB[1]
C3000 – BP - up to 1.5TB[1]
C1000 - BP - up to 500GB[1]

| <1 | 3 | 10 | 50 | 150+ |

[1] Raw Active User Data

8

Features for Data Warehousing Solutions
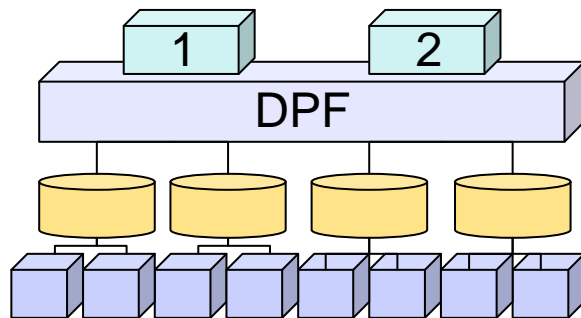
- **Distribute** data across nodes (**DPF**)
- Range partition the rows in table partitions per node (**Range Partitioning**)
- Cluster the data by multidimensional column values (**MDC**)
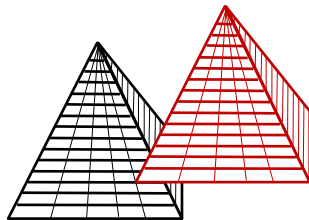
- Precalculate summary / aggregate data by group by's into **MQTs**.
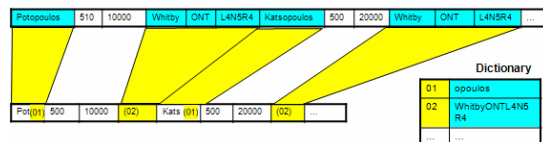
- Reduce storage and increase performance with advanced **Compression**.
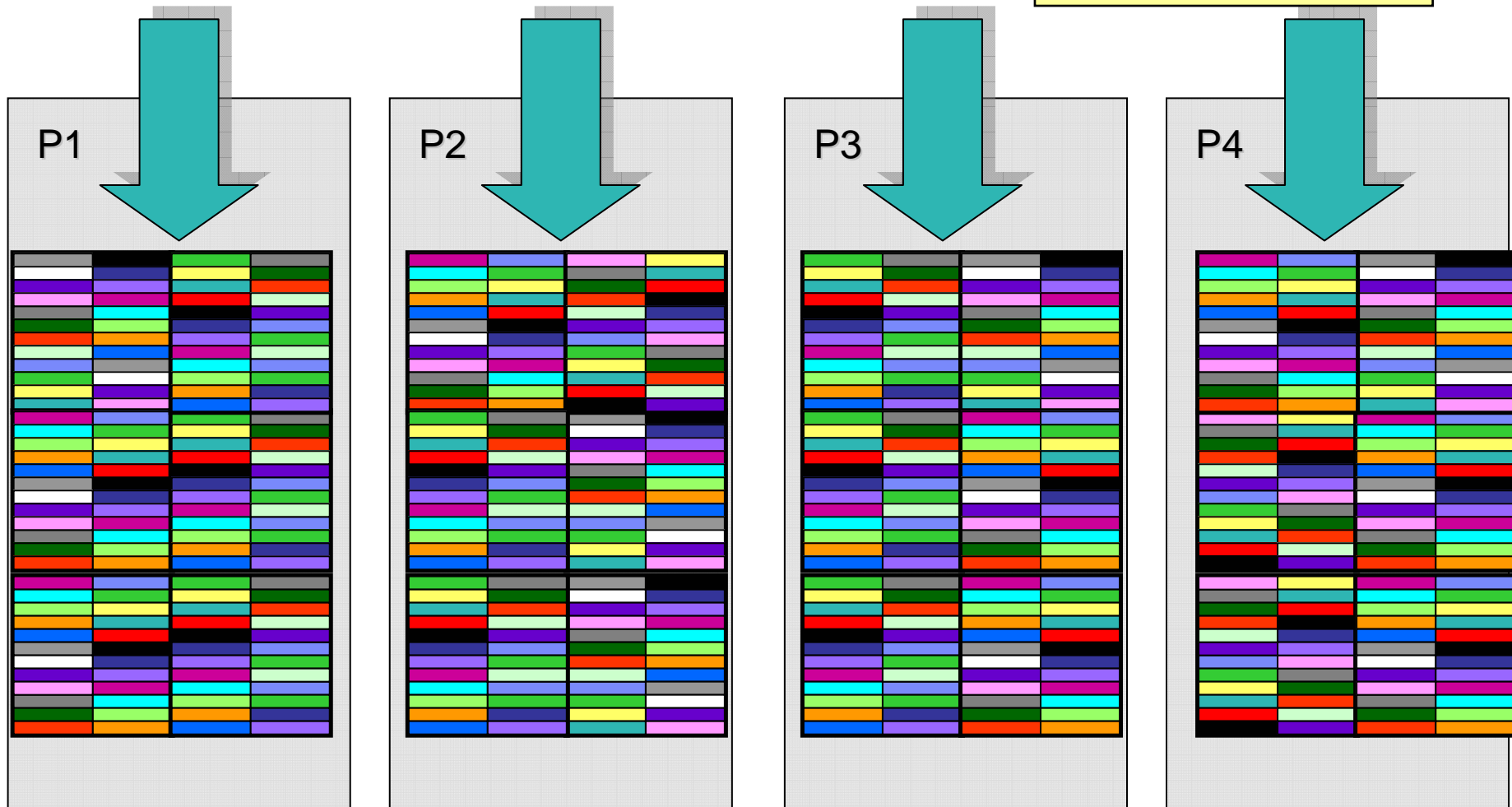
No Partitioning

Speedup = 1
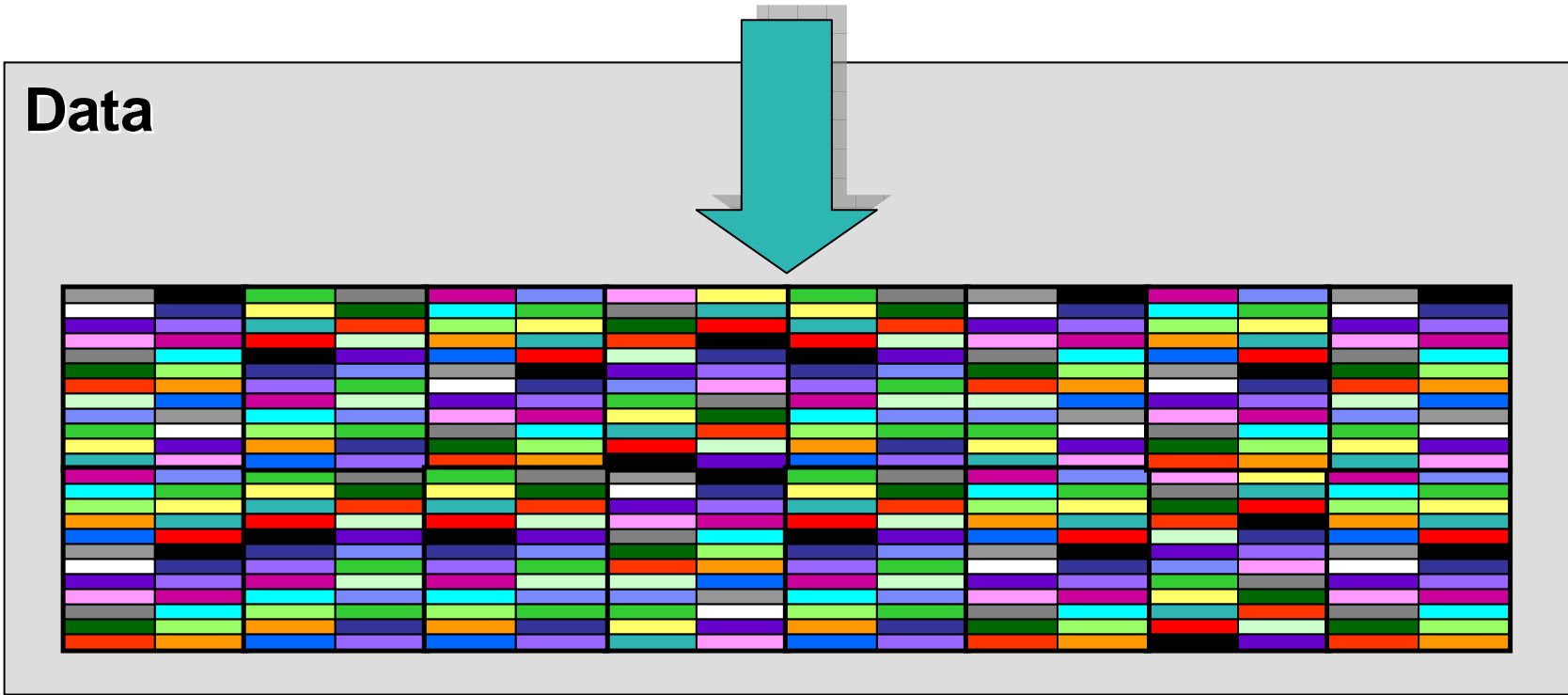
Data
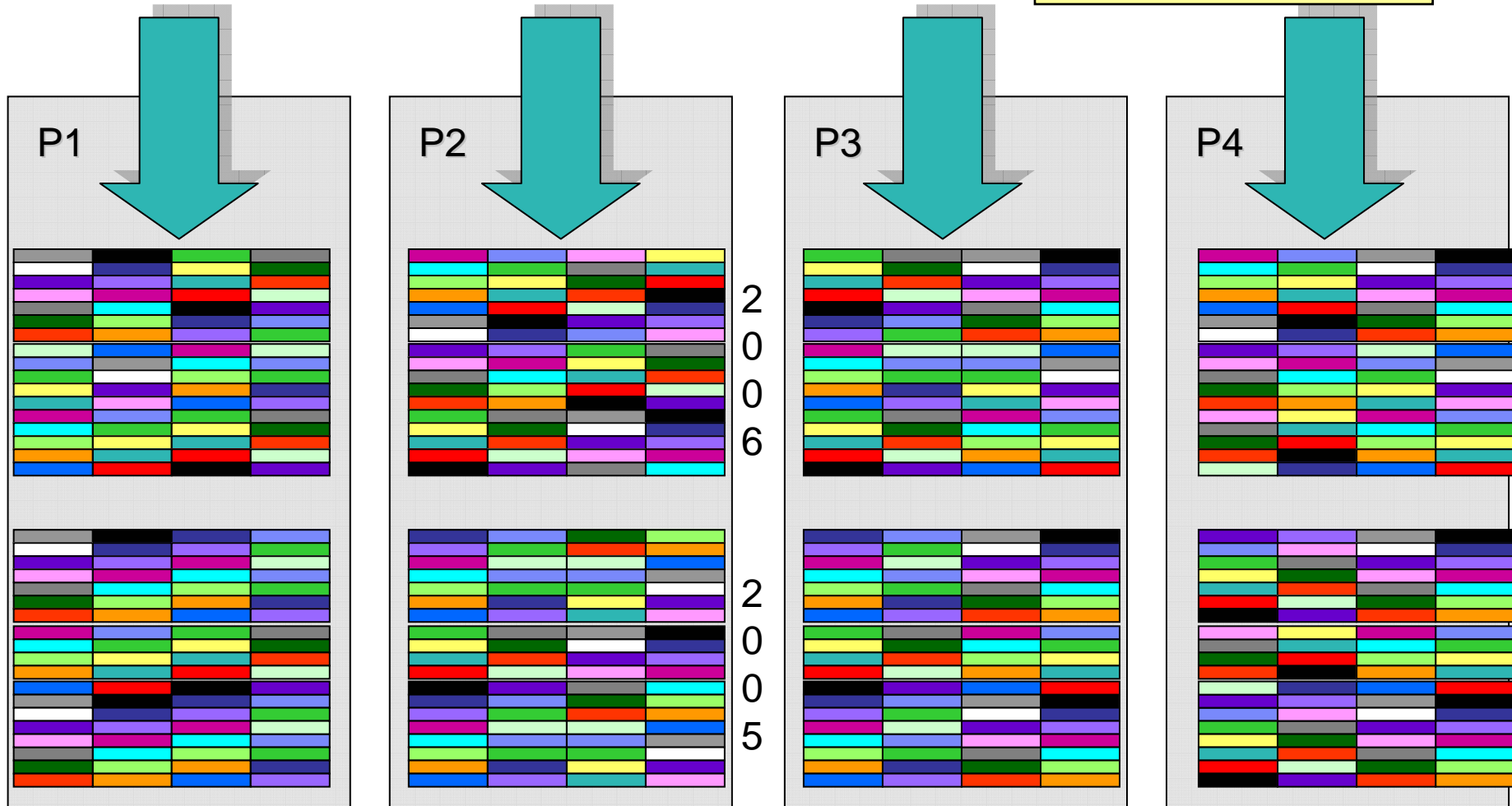
Divide Data and Put in Multiple Servers
Divide & Conquer Parallelism

Speedup = 4

Add Range Partition: Eliminate Unused Partitions
When analyze 2006 data only…

Speedup = 8

P1 P2 P3 P4

2006

2005

Further Group Data by Types (Assumed 6 Types)
When analyze 2006 purple data only...

Speedup = 48!
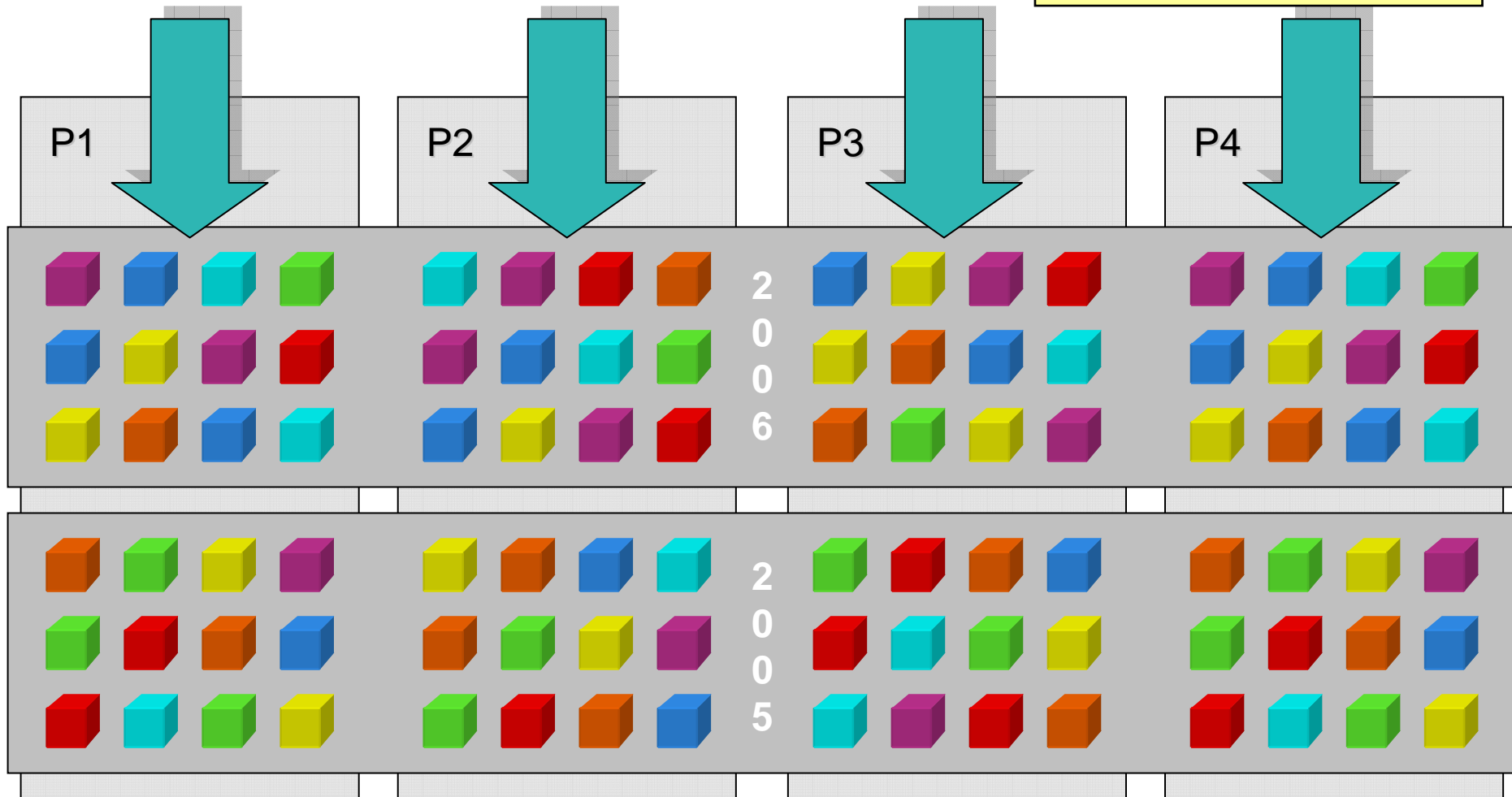
Before any Index is Built At All

Single Version with Near Real-time Warehouse

Giving customers the right information to enable bigger sales

Customer Self Service

Customer Service/ Care

Optimizing the business with right-time information

Online Sales

Marketing/Sales Interactions

Focus is on having the right information at the right time

Financial Trading

Financial Planning

Making the right decision at the right time

Understanding the choices and planning with accurate information

Fraud Avoidance

Regulatory Compliance

Supply Chain Optimization

InfoSphere Warehouse and No Copy Analytics

- The No Copy approach to delivering business analytics has many benefits
  - Enables low latency, right-time Analytics
  - Increased Flexibility and adaptability to information changes
  - Reduced development and Management Effort
  - Flexibility to add and change Analytical Applications cost effectively

- Leading to outstanding business results across the spectrum, including
  - Business trend detection and analysis
  - Fraud Detection
  - Customer Churn Prediction
  - Event Based Decision Support

- Overall Greater Return On Investment and Lower Cost of Ownership!

How does No-Copy Analytics work?
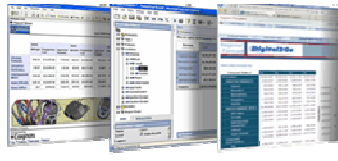
COGNOS AN IBM COMPANY — IBM

DataQuant

Microsoft Office

3rd Party

Universal Access
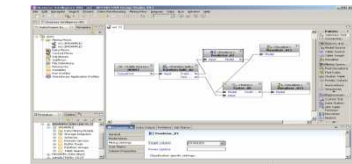
Data Mining
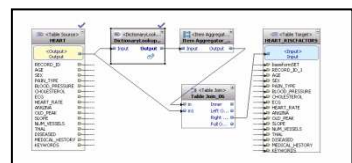
Unstructured Text

Annotation

InfoSphere Warehouse
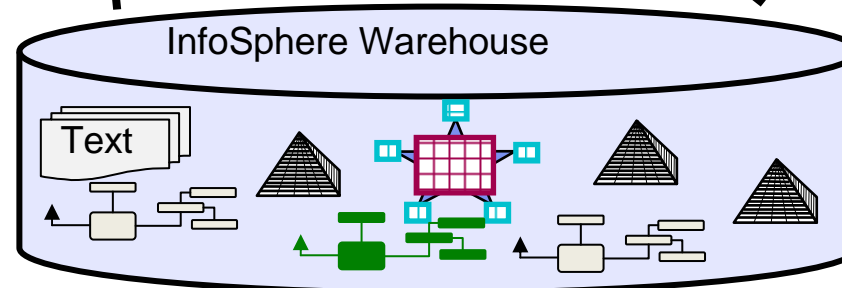
Text
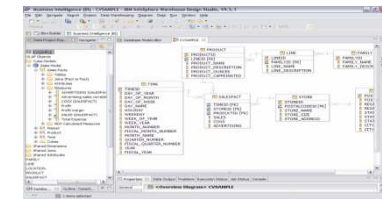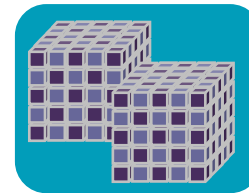
Cubing

- No data extraction
- Data always (only) in the data warehouse

- Full (universal) access to all analytical information
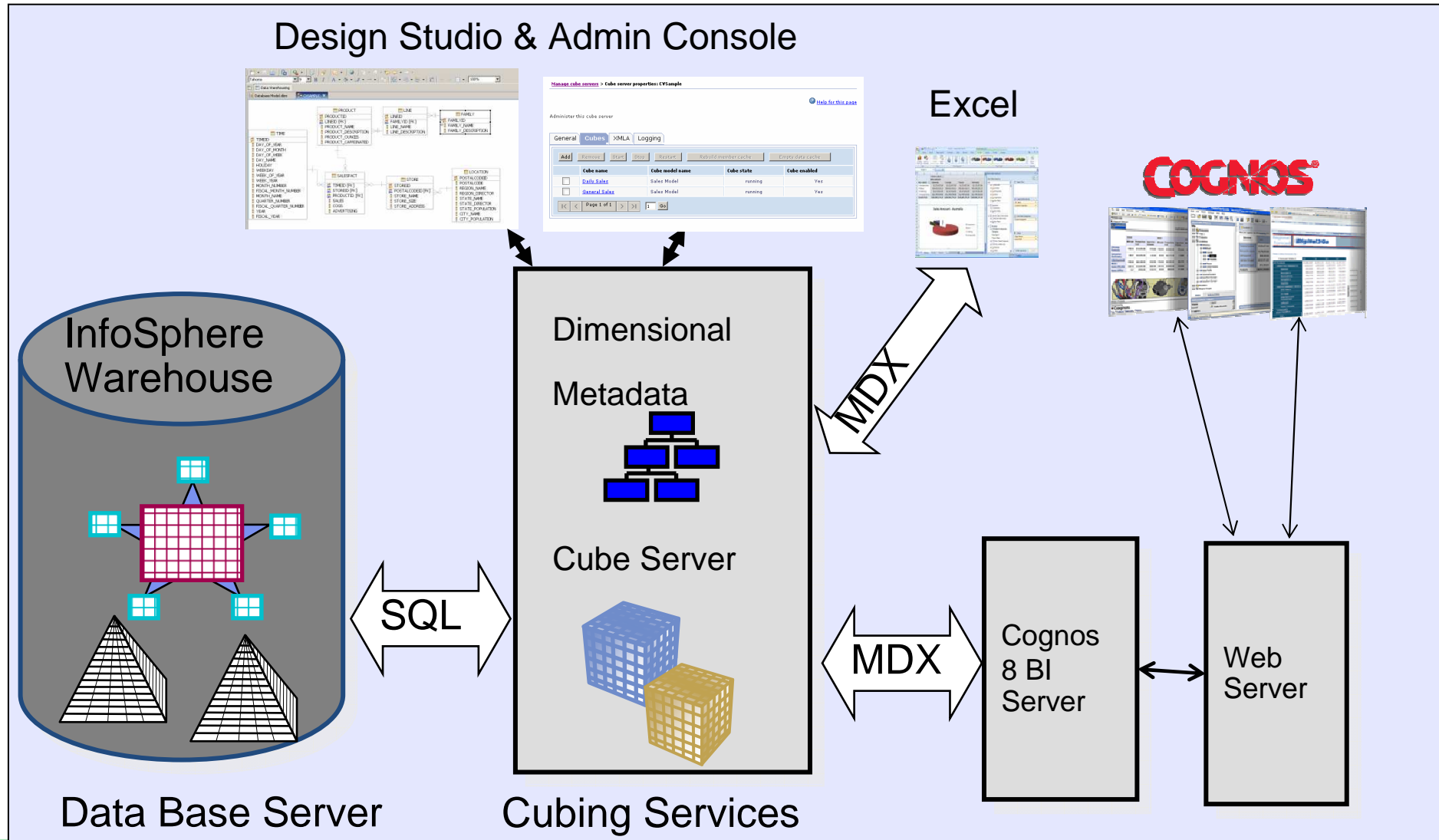- Low latency response to new/changed data

InfoSphere Warehouse Cubing Services

A business-centric OLAP abstraction of the data warehouse information

COGNOS AN IBM COMPANY | **IBM** | Office | 3rd Party

Universal Access

Cubing Services

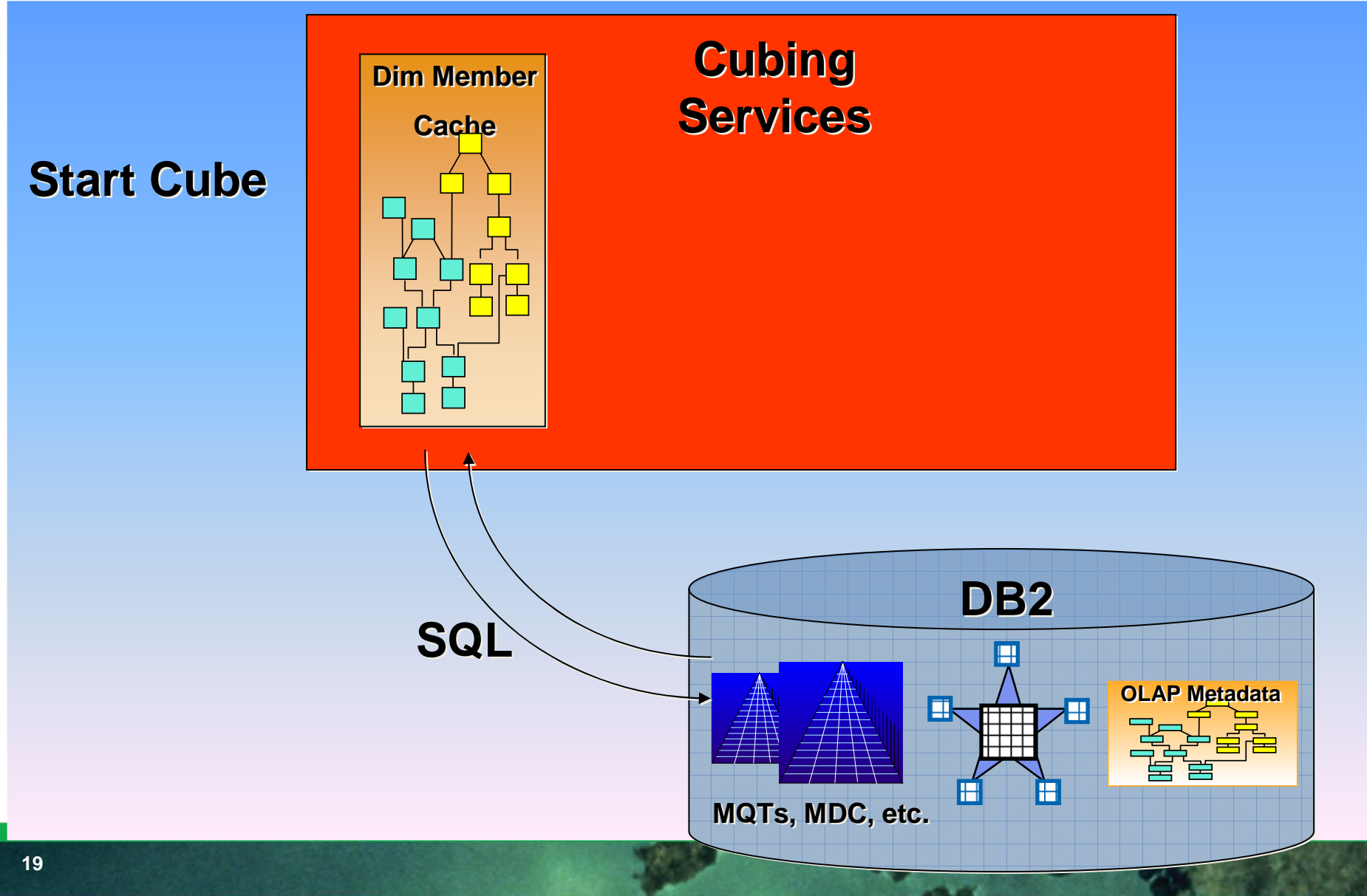**InfoSphere Warehouse**
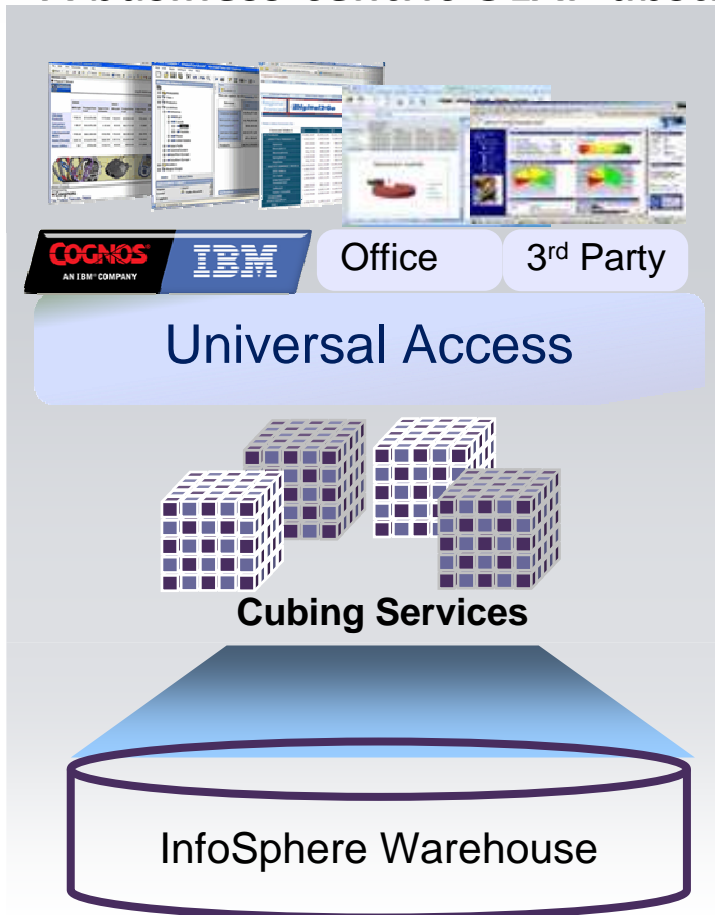
- Benefits
  - Scalable low-latency OLAP
  - Very large dimensions over very large data sets
  - Optimized for the InfoSphere Warehouse
  - Integrated design and maintenance
  - No copy analytics with universal heterogeneous BI tool access
  - Support MDX, ODBO, XMLA

- Business Value
  - Enriched reporting and decision support for more members of the organization – greater business impact from information assets
  - Accelerated delivery of high performance business intelligence solutions across a spectrum of business processes and industries.

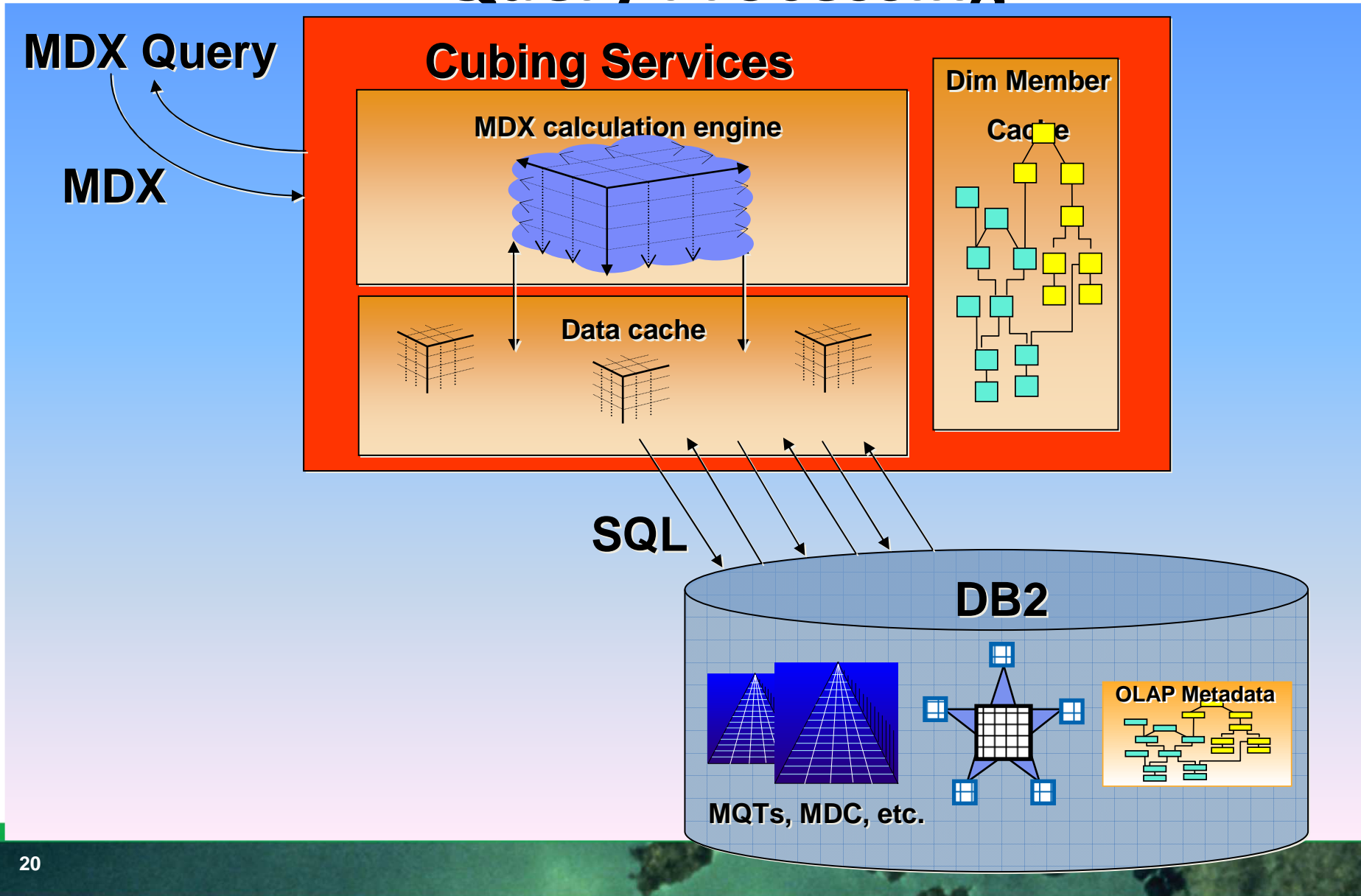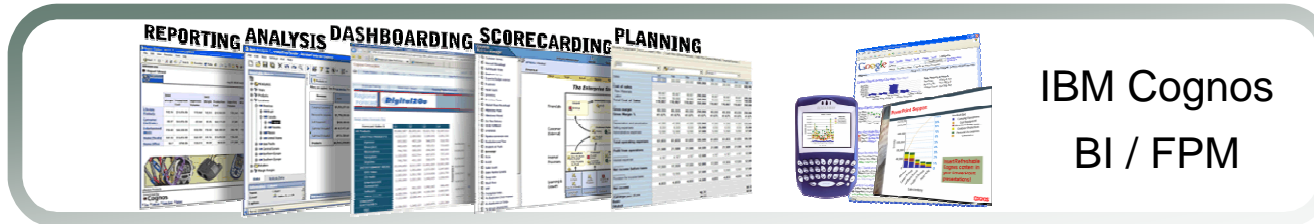# Cubing Services Architecture with Cognos

Design Studio & Admin Console

Excel

InfoSphere Warehouse

Dimensional

Metadata

Cube Server

SQL

MDX

MDX

Cognos 8 BI Server

Web Server

Data Base Server

Cubing Services

Cube Server in Action – Startup

Start Cube

Cubing Services

Dim Member Cache

SQL

DB2

MQTs, MDC, etc.

OLAP Metadata

Cube Server in Action – Query Processing

MDX Query

MDX

Cubing Services

MDX calculation engine

Data cache

Dim Member Cache

SQL

DB2

MQTs, MDC, etc.

OLAP Metadata

Cubing Services and BI Tools

REPORTING ANALYSIS DASHBOARDING SCORECARDING PLANNING

IBM Cognos

BI / FPM

IBM® COGNOS® 8 Framework Manager

Framework Manager allows you to model metadata and publish packages.

IBM® COGNOS® 8

Projects

Create a new project
Open a project...

Recent Projects

Name
IoD Demo Proje
CS Project 1

My Content

My home

IBM Cognos content

Administration

My Actions

Query my data

Analyze my business

Create professional reports

Manage my events

Cubing Services

Deep Compression Using a Compression Dictionary

- Repeating patterns within the data (and just within each row) is the key to good compression. Text data tends to compress well because of reoccurring strings as well as data with lots of repeating characters, leading or trailing blanks

Name	Dept	Salary	City	State	ZipCode
Fred	500	10000	Plano	TX	24355
John	500	20000	Plano	TX	24355

Fred	500	10000	Plano	TX	24355	John	500	20000	Plano	TX	24355	…

Fred	(01)	10000	(02)	John	(01)	20000	(02)	…

Dictionary

01	Dept 500
02	Plano, TX, 24355
…	…
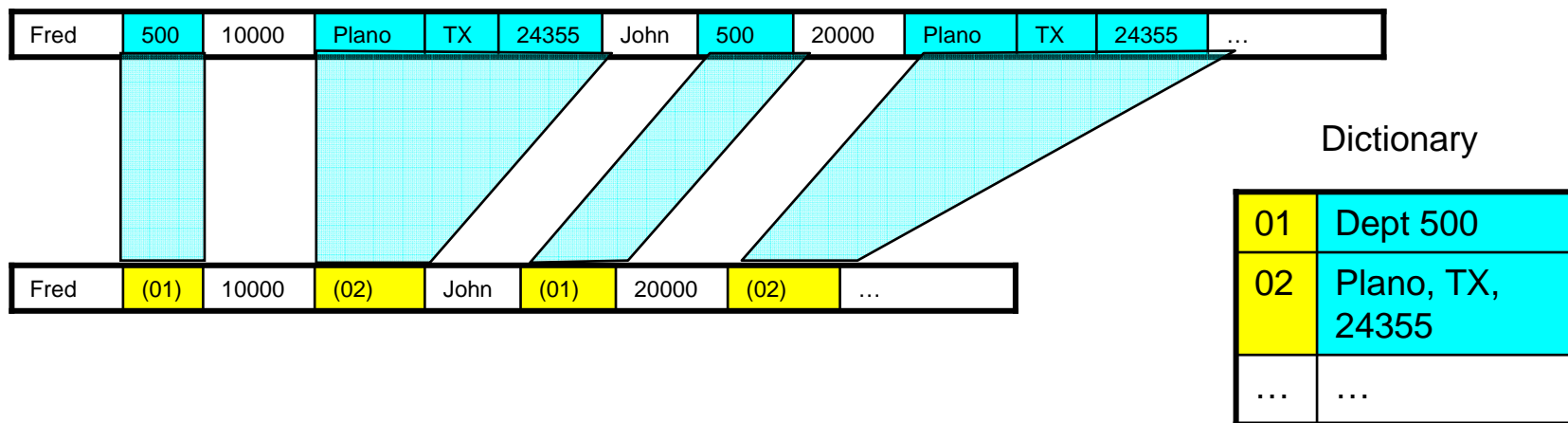
Row Compression

Uncompressed Row

SOCKS BLUE DALLAS TEXAS

Compressed Row

x'01C'
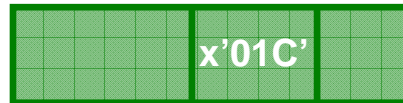
Common sequences of
consecutive bytes in row
replaced with 12 bit symbol

Data page with
uncompressed rows
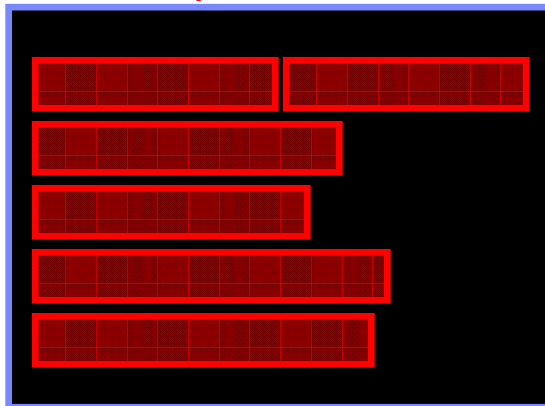
Data page with
compressed rows
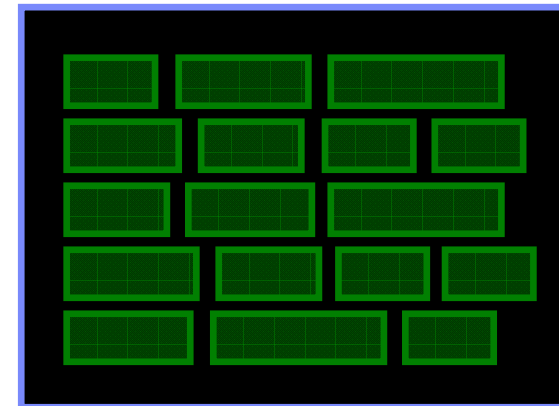
- Less storage

- Over 70% off in storage saving in real cases!

- Much faster, for example, table scan
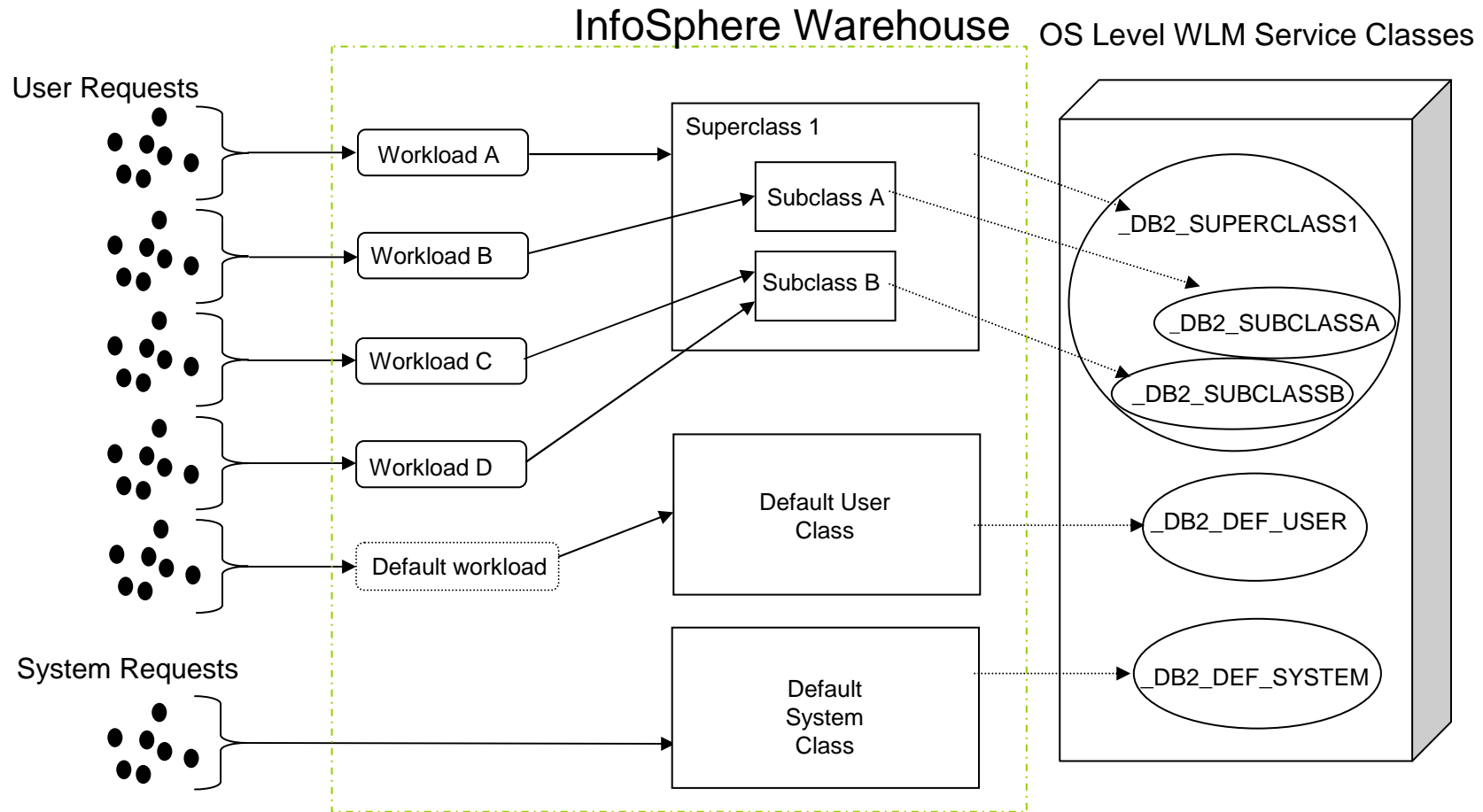
Even Faster Performance with Compression

Overall Workload Performance:
Uncompressed Data versus Compressed Data

Faster!

100%	
90%	
80%	
70%	
60%	
50%	
40%	
30%	
20%	
10%	
0%	

Uncompressed Data   Compressed Data
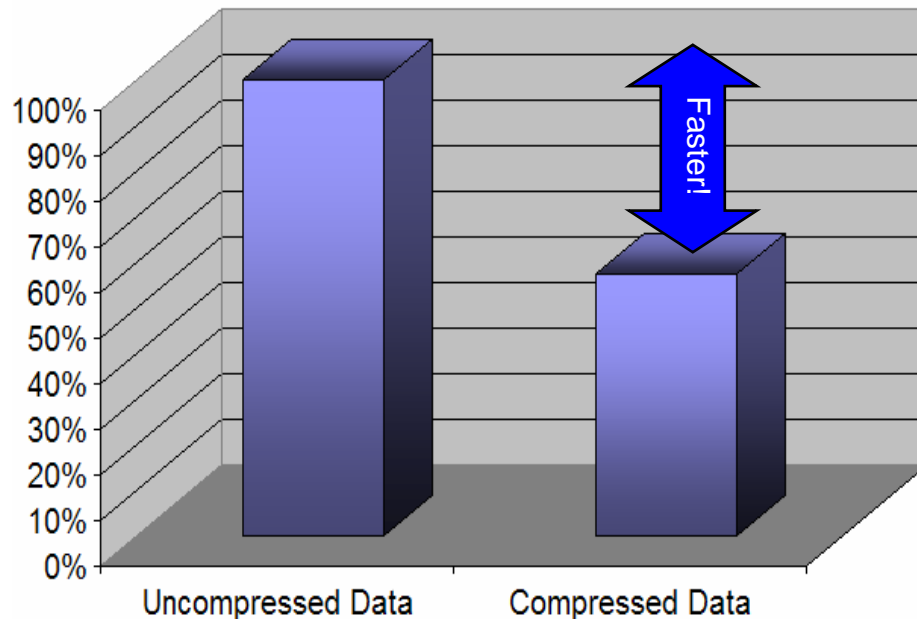
"With the new compression technology in DB2 Viper, we realized an **80 Percent improvement in space savings** for our most critical tables in our data warehouse. We were even more pleased with this technology when we found that Viper's compression capability helped us process queries to the database an average of **40 Percent faster** than before. We're looking forward to seeing the same results with our Operational Data Store and OLTP systems." Donny Ledbetter, Senior DBA
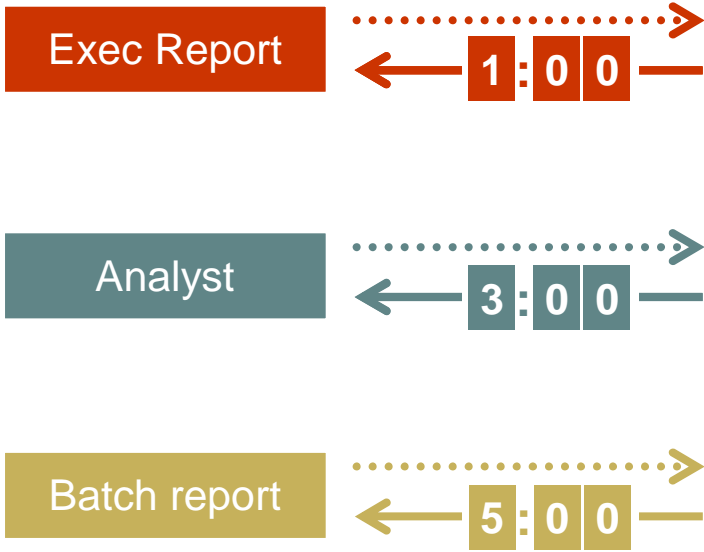
AutoZone
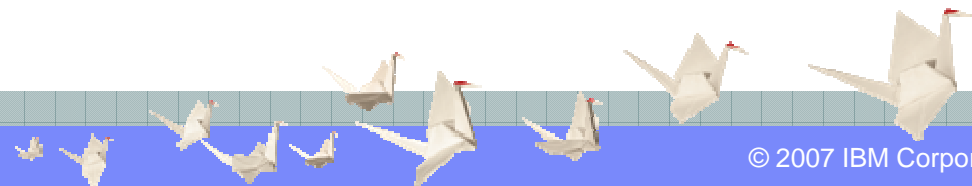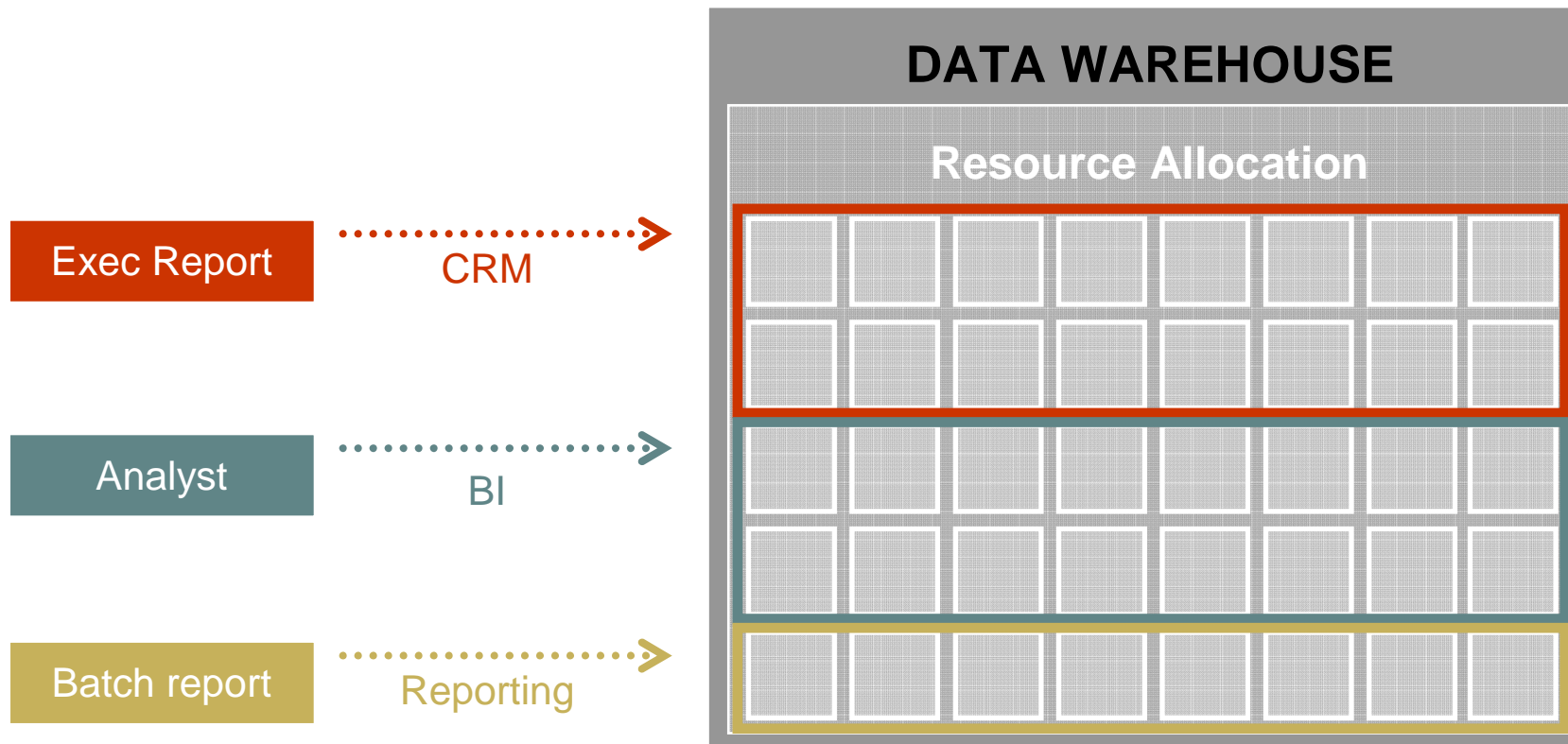
Workload Management Guarantees Service Level Agreement

InfoSphere Warehouse   OS Level WLM Service Classes

User Requests

Workload A → Superclass 1

Workload B → Subclass A

Workload C

Subclass B

Workload D

Default workload → Default User Class

System Requests

Default System Class

_DB2_SUPERCLASS1

_DB2_SUBCLASSA

_DB2_SUBCLASSB

_DB2_DEF_USER

_DB2_DEF_SYSTEM

Why does it matter?
# Consider a typical warehouse…

Unacceptable Response Time!

DATA WAREHOUSE

Resource Usage

Exec Report

1 : 0 0

Analyst

3 : 0 0

Batch report

5 : 0 0
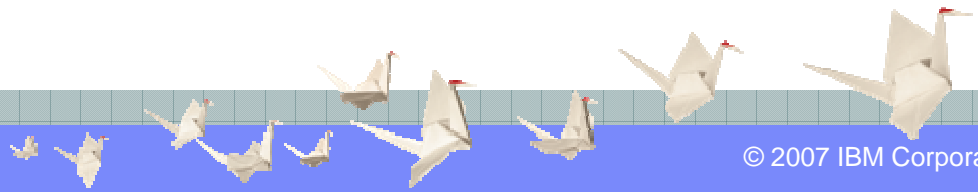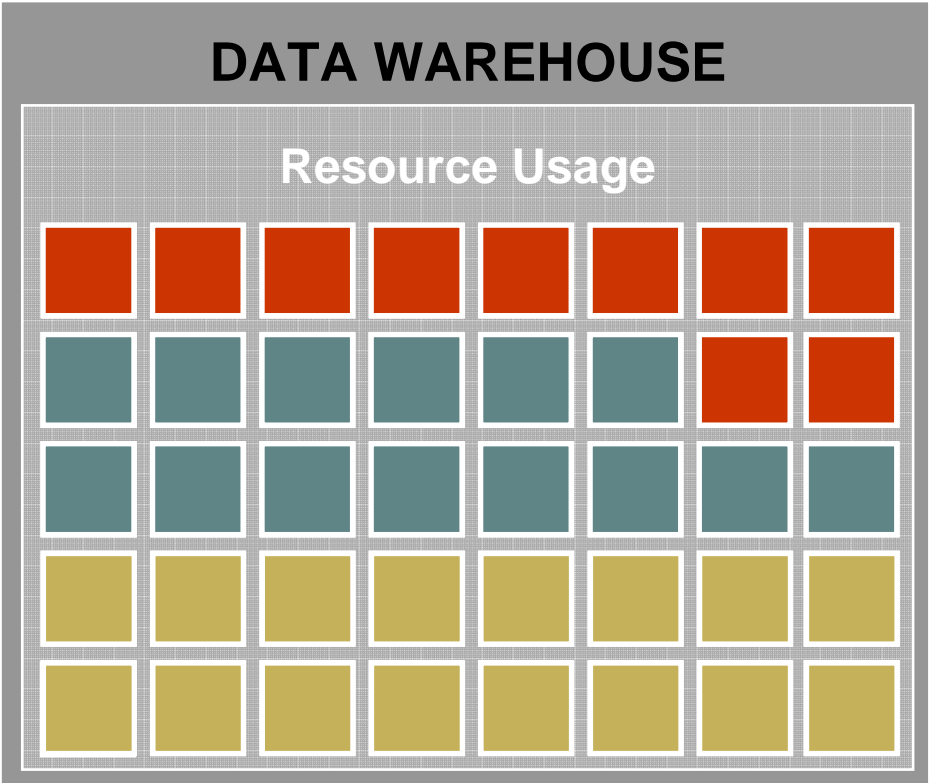
Extreme workload management warehouse lets you align resources according to business priorities …

DATA WAREHOUSE

Resource Allocation

Exec Report — — — — — > CRM

Analyst — — — — — > BI

Batch report — — — — — > Reporting

… to ensure operational applications get the response time required—regardless of the number and types of requests

DATA WAREHOUSE

Resource Usage

Exec Report → 0:05

Analyst → 2:00

Batch report → 8:00
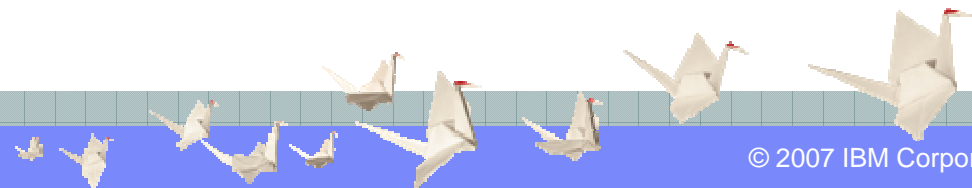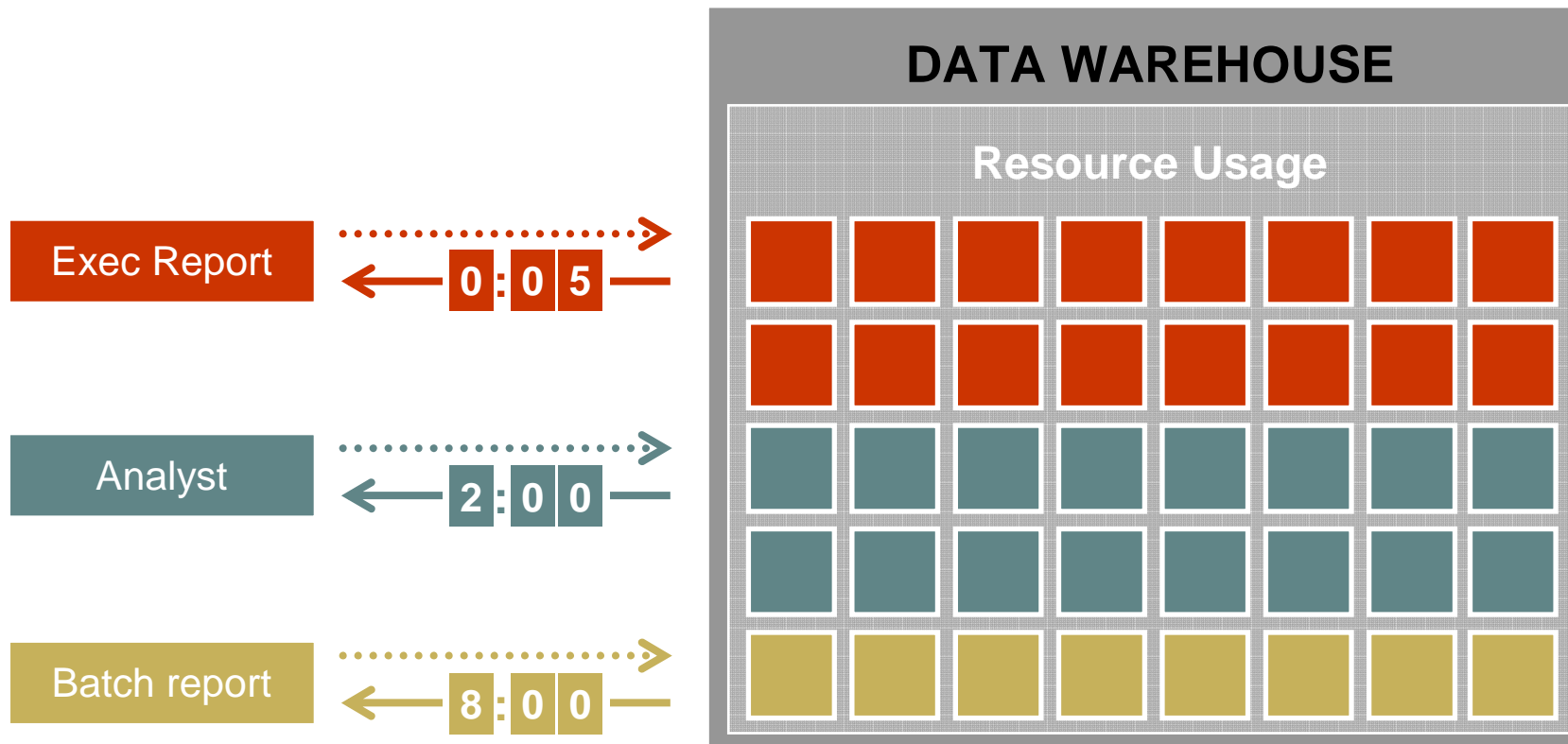
Unexpected Workload is put on Queue

ID	Query	started	released	completed	wait	exec	total
5238	Q3	17:03:34	17:03:34	17:11:09	00:00	07:35	07:35
5239	Q1	17:03:34	17:03:34	17:06:35	00:00	03:01	03:01
5240	Q1	17:03:34	17:03:34	17:06:43	00:00	03:09	03:09
5241	Q3	17:03:35	17:03:35	17:11:09	00:00	07:34	07:34
5242	Q3	17:03:35	17:03:35	17:11:09	00:00	07:34	07:34
5243	Q1	17:03:35	17:03:35	17:06:21	00:00	02:46	02:46
5244	Q3	17:03:35	17:03:35	17:11:09	00:00	07:34	07:34
5245	Q1	17:03:35	17:03:35	17:06:36	00:00	03:01	03:01
5246	Q1	17:03:36	17:05:30	17:07:15	01:54	01:45	03:39
5247	Q3	17:03:36	17:06:14	17:13:28	02:38	07:14	09:52
5248	Q3	17:03:36	17:06:15	17:13:28	02:39	07:13	09:52
5249	Q3	17:03:36	17:06:21	17:13:13	02:45	06:52	09:37
5250	Q1	17:03:36	17:06:26	17:08:15	02:50	01:49	04:39
5251	Q1	17:03:36	17:06:33	17:08:11	02:57	01:38	04:35
5252	Q3	17:03:36	17:06:33	17:13:26	02:57	06:53	09:50
5253	Q3	17:03:36	17:06:35	17:13:27	02:59	06:52	09:51
5254	Q1	17:03:36	17:06:36	17:08:29	03:00	01:53	04:53
5255	Q3	17:03:36	17:06:43	17:13:22	03:07	06:39	09:46
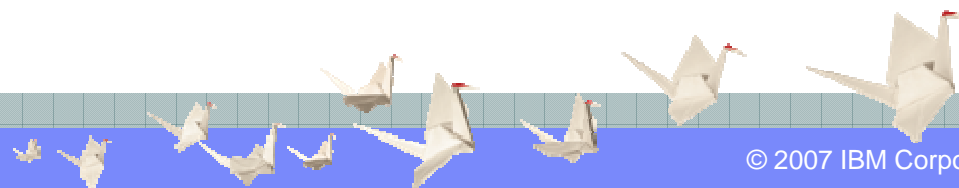
Research Showed Delayed Extra Jobs Improves Response Time and Performance (IEEE Concurrency Journal @ 1999)

Summary: Smart Enterprise Information Management Architecture

MPP & High Speed

Core/Sources
- Cards
- Treasury
- Loans
- Deposits
- External
- Other Sources

Data Warehouse Design

Data Warehouse Model

Services Data Model

Business Solution Templates

Enterprise Data Warehouse
- Staging Area
- Data Warehouse
- Real-time Data
- Archiving Area

Master Data Management
- SIF
- MDM Repository
- Business Services

Enterprise Data Integration

OLAP

Business Views

ESB

Operational Consumers
- Core
- CRM

Information Lifecycle Management & Administration

Metadata Management & Metadata Repository

Business Applications
- Customers
- Risk
- Products

Data Analysis & Reporting
- Optimization
- Predictive Modelling
- Mgt Reporting

31

Information Management

IBM

IBM Information On Demand Conference 2009

Thank You!