IBM DB2 Analytics Accelerator
Optimizing Your Key Business Decisions

Udo Hertz, Dir. Information Management Development Boeblingen

Information Management

Agenda

- Business and Technology Drivers

- Data Warehousing Solution on zEnterprise
    - IDAA Key Design and Operational Features
    - Powered by Netezza

- IDAA Use Cases and Next Steps

Knowing what happened is no longer adequate.
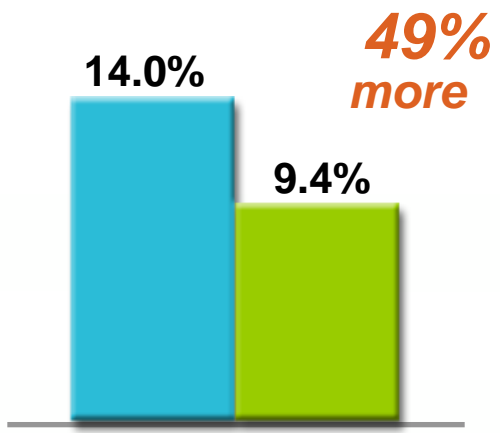
Business leaders need to know
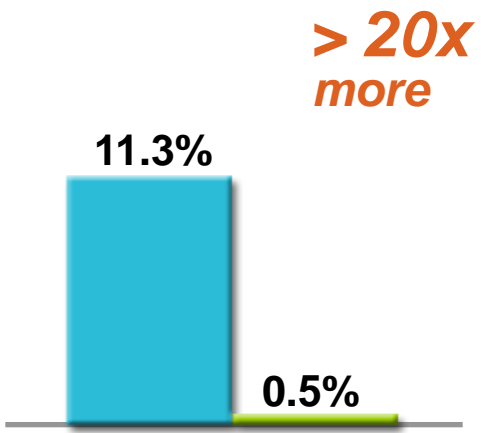
what is happening now,

what is likely to happen next and

what actions they should take.
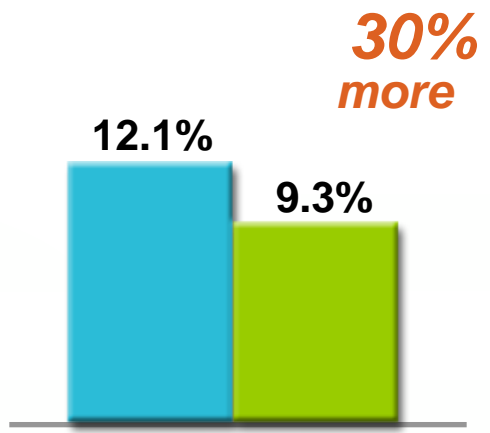
…. *Consistently Outperform*

Revenue Growth	**Profit Growth**	**Return on Invested Capital**
5 Year CAGR (2004-2008)	*5 Year CAGR (2004-2008)*	*5 Year Average (2004-2008)*

49% more

> 20x more
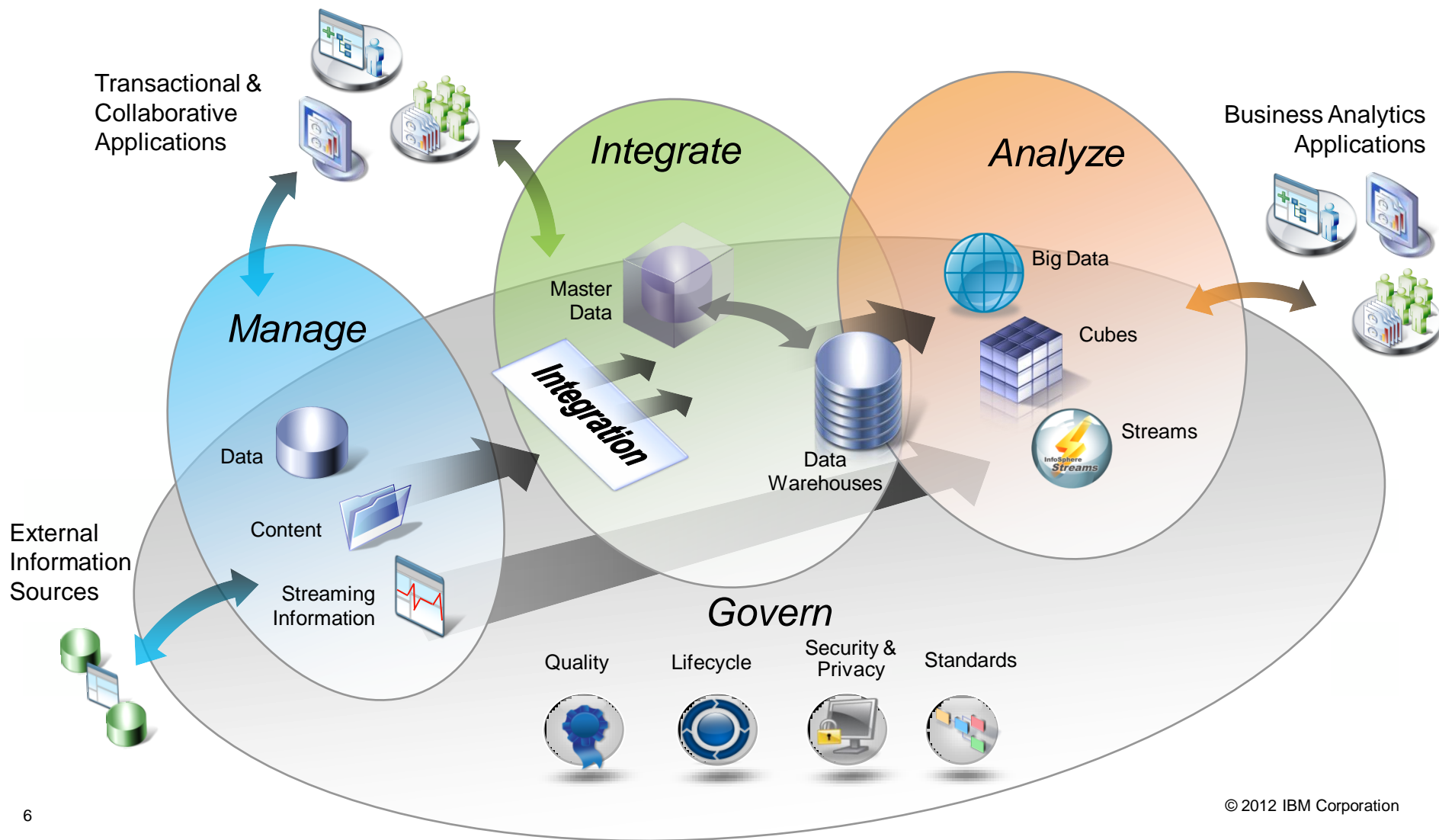
30% more

14.0%
9.4%

11.3%
0.5%

12.1%
9.3%

Finance organizations with business insight
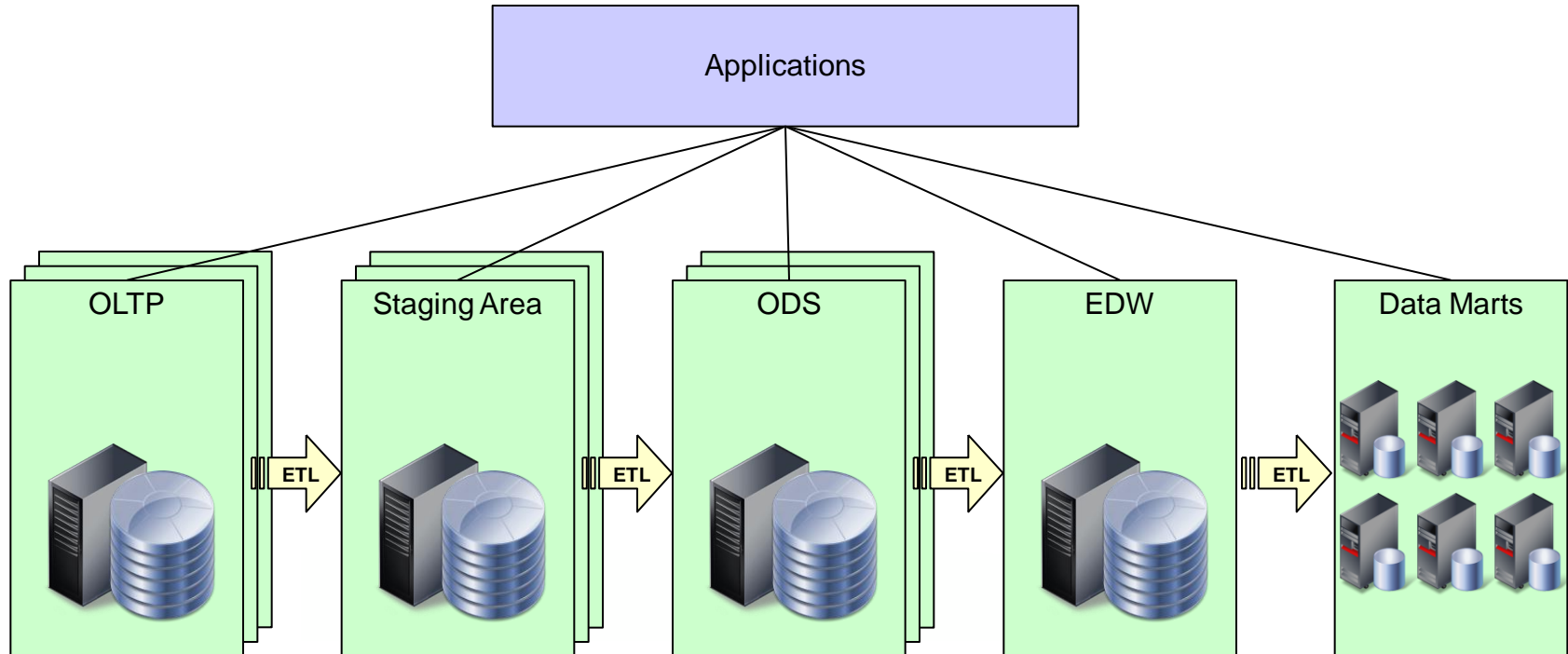
All other enterprises

Revenue Growth: N = 580; Profit: N = 435; ROIC: N = 606
Source: IBM Global Business Services, The Global CFO Study 2010

The market is moving to the strengths of System z

- Enterprises are expanding the role of analytics
  - Better decisions from the right information
  - Informed decisions at the point of contact
  - Consistency of information across organizations
- Which is driving operational characteristics requirements
  - Cost of downtime is escalating
  - The impact of unauthorized intrusion and publishing of private information is overwhelming
  - Stringent Service Level Agreements must be met
- Newer applications demand lower latency of the data
  - Businesses want the most up-to-date information they can get
  - Yesterday's information was good yesterday
- All while focusing on reducing costs/ consolidating
  - Lower costs through reduced complexity
  - Simplified environment with easier administration
  - Lower SW costs
  - Reduced costs through elimination of redundant servers and resources
  - Reduced footprint, environmental, and administrative costs

Delivering trusted information for smarter business decisions across your entire information supply chain

© 2012 IBM Corporation

Traditional Systems Landscape

IBM

Applications

OLTP → ETL → Staging Area → ETL → ODS → ETL → EDW → ETL → Data Marts

Historical reasons:

- Different access patterns
  - → impact on performance
- EDW as the data integration hub
  - → again, impact on performance
- Different life-cycle characteristics
  - → and again, impact on performance
- Different Service Level Agreements (SLA)
  - → Lack of broadly available workload management capabilities
  - → Choice of lower cost-of-acquisition offerings

Negative ramifications:

- Complexity
  - → both in systems management and in applications
- Difficulties in supporting real time analytics
- Inability to match ever more demanding SLA requirements
- High total cost of ownership

Visionary Systems Landscape

```
                          ┌─────────────────────────┐
                          │      Applications       │
                          └─────────────────────────┘
```
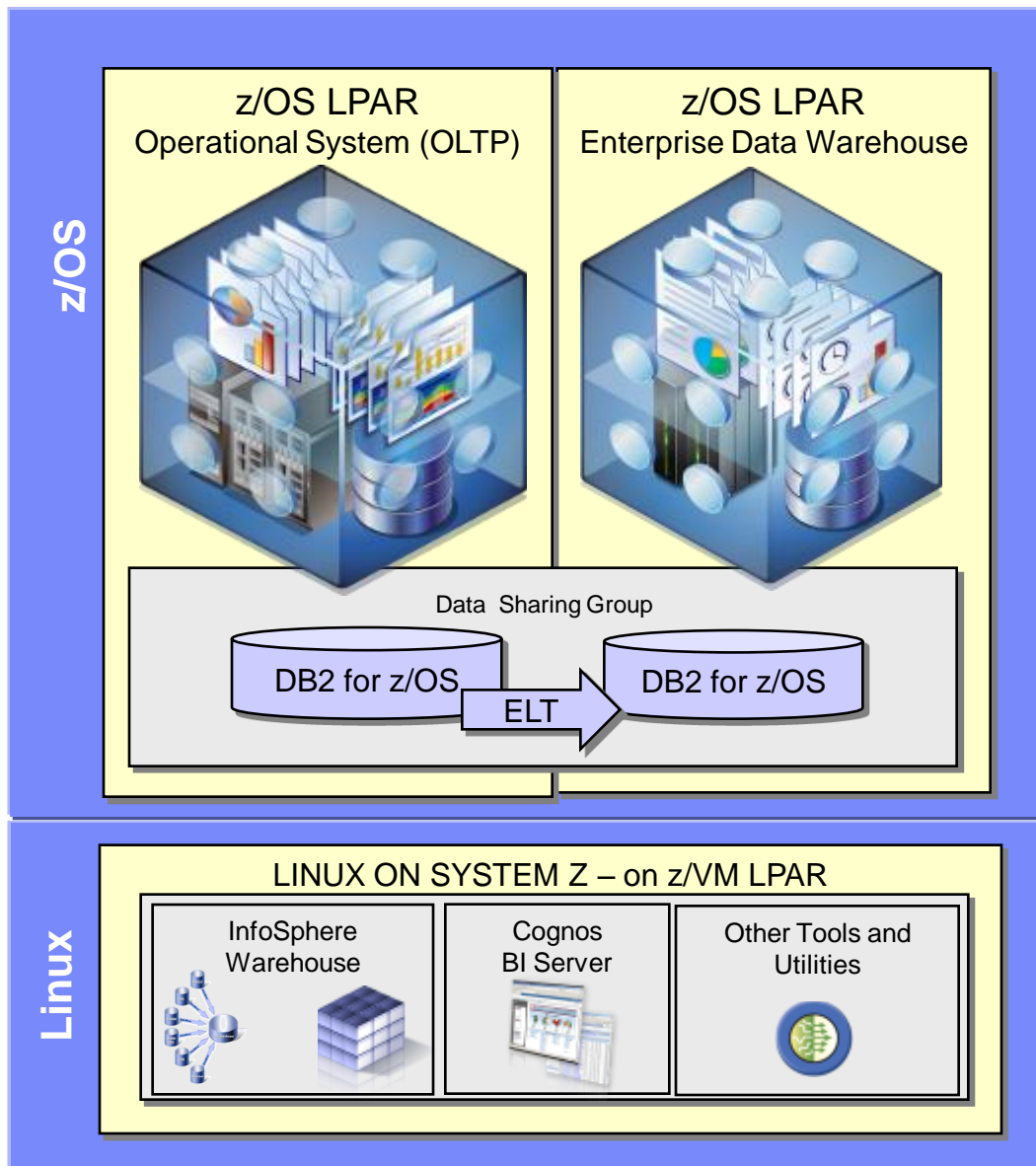
| OLTP | → ELT → | Staging Area | → ELT → | ODS | → ELT → | EDW | → ELT → | Data Marts |

Benefits
- → Consolidating all the components into a single system
- → Uniform access to any data
- → Efficient data movement within the system (ideally, no network)
- → **Opportunity to remove, i.e. consolidate some of the layers**

Challenges
- → Mixed workload management capabilities
- → Ensuring continuous availability, security and reliability
- → Providing seamless scale-up and scale-out
- → Providing universal processing capabilities to deliver best performance for both transactional and analytical workloads **without the need for excessive tuning**

Approaches
- → Columnar stores
- → In-memory databases
- → Hardware acceleration, special purpose processors
- → Appliances

Building on proven technology base
- → System z Data Sharing and Parallel Sysplex technology provides all the needed characteristics except one:
  - Special purpose processing for analytical workloads to minimize the need for manual tuning

Agenda

- Business and Technology Drivers

- Data Warehousing Solution on zEnterprise
  – IDAA Key Design and Operational Features
  – Powered by Netezza

- IDAA Use Cases and Next Steps

A data warehouse solution on a System z foundation

z/OS

z/OS LPAR
Operational System (OLTP)

z/OS LPAR
Enterprise Data Warehouse

Data Sharing Group

DB2 for z/OS → ELT → DB2 for z/OS

Linux

LINUX ON SYSTEM Z – on z/VM LPAR

InfoSphere Warehouse	Cognos BI Server	Other Tools and Utilities

- Minimizes data movement between operational system and data warehouse

- Lowers data latency for time sensitive decisions

- Enables consolidation and simplification of data warehouse and data marts

- Leverages existing high availability, backup, disaster recovery, and security environments

- Provides greater scalability of multidimensional analysis through cubing services (data marts) and DB2 enhancements

DB2 10 for z/OS

IBM

CPU reductions for transactions, queries, and batch

- Out-of-the-box CPU reductions of 5-10% for traditional workloads
- Up to additional 10% CPU savings using new functions or avoiding constraints
- Out-of-the box CPU reductions of up to 20% for new workloads

Scales with less complexity and cost

- 5-10x more concurrent users – up to 20,000 per subsystem
- Significant scale-up capabilities in addition to existing scale-out support
- Consolidate to fewer LPARs and subsystems

Improved operational efficiencies and lower administration cost

- Automatic diagnostics, tuning, and compression

Even better performance

- Elapsed time improvement for small LOBS and Complex Queries

64 bit Evolution Virtual Storage Relief

Temporal Data

Integrated XML Support

Query Processing Enhancements

Business Security & Compliance

Better Productivity

Standardized Cognos Business Intelligence tools

IBM

… delivering information when, where, and how each user needs it

- **Delivers information where, when and how it is needed**
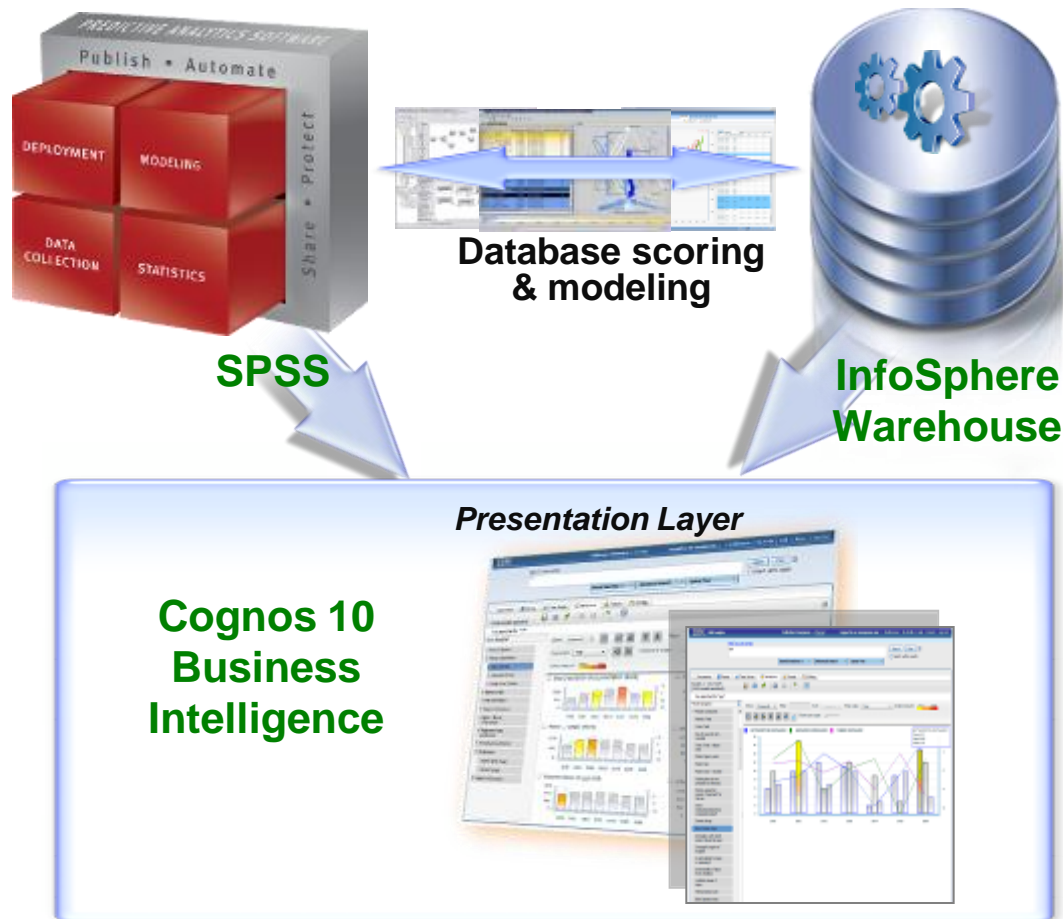 - Self-service reporting and analysis
  - Individualized by user
  - Automated delivery of information in context
  - Author once, consume anywhere

- **Full range of BI capabilities**
  - Query, reporting, analysis, dashboarding, realtime monitoring

- **Purpose-built SOA platform**
  that fits client environments and scales easily

SPSS Modeler with InfoSphere Warehouse

Database scoring & modeling

SPSS

InfoSphere Warehouse

Presentation Layer

Cognos 10 Business Intelligence

Full breadth of predictive analytics

Data collection, statistics, data mining, predictive modeling, deployment services…

Putting prediction in hands of the business
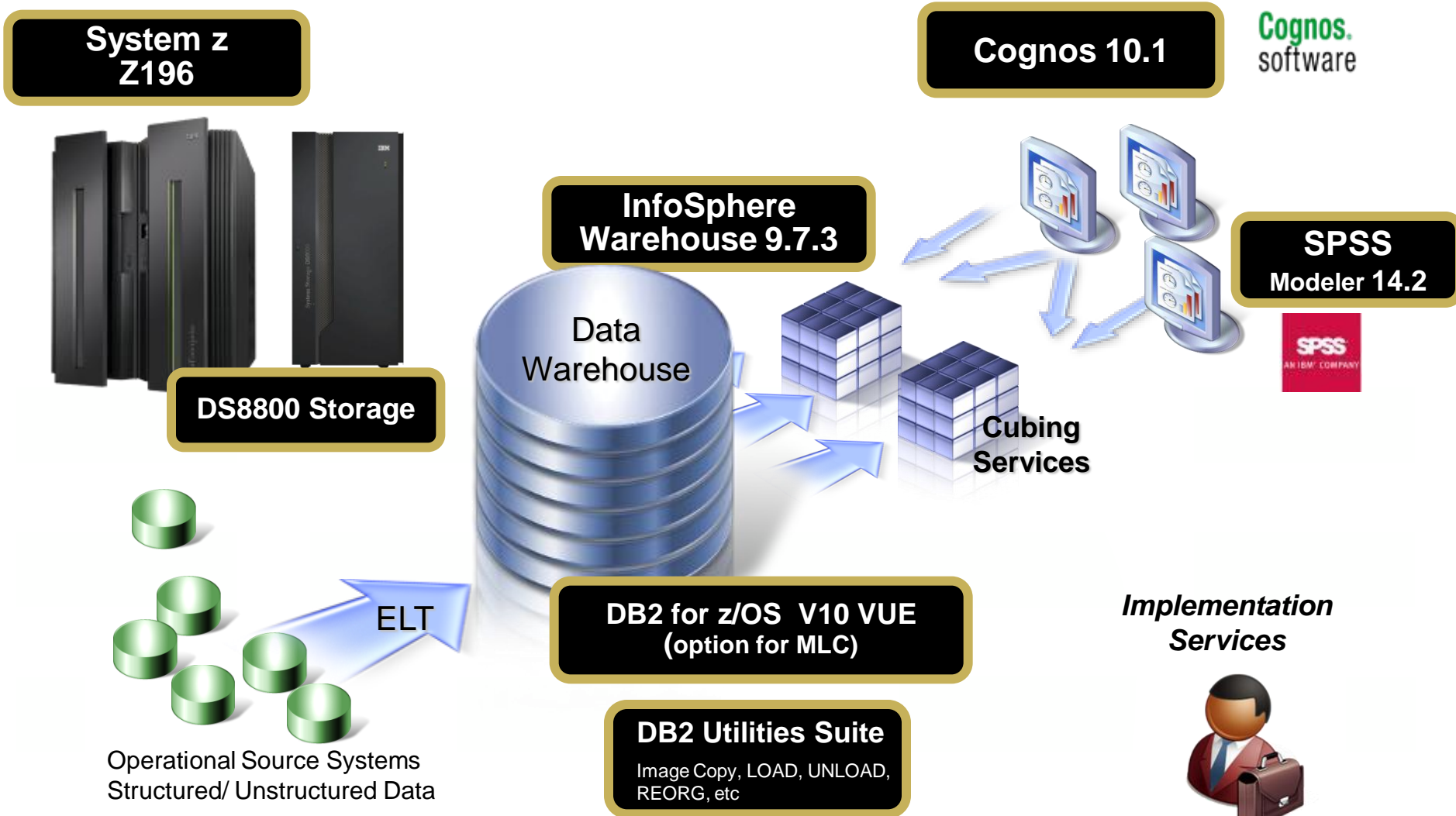
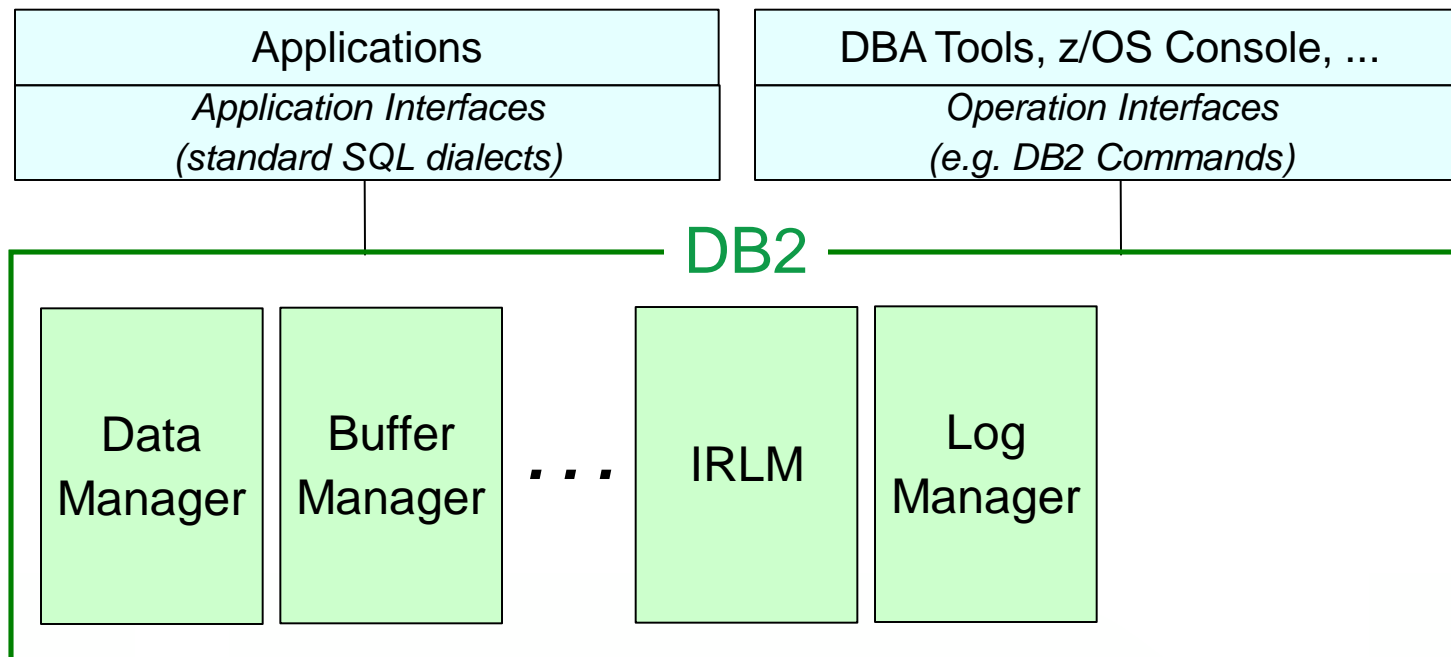Decision Management

Driving better business outcomes
- Attract and retain profitable customers
- Detect and prevent fraud
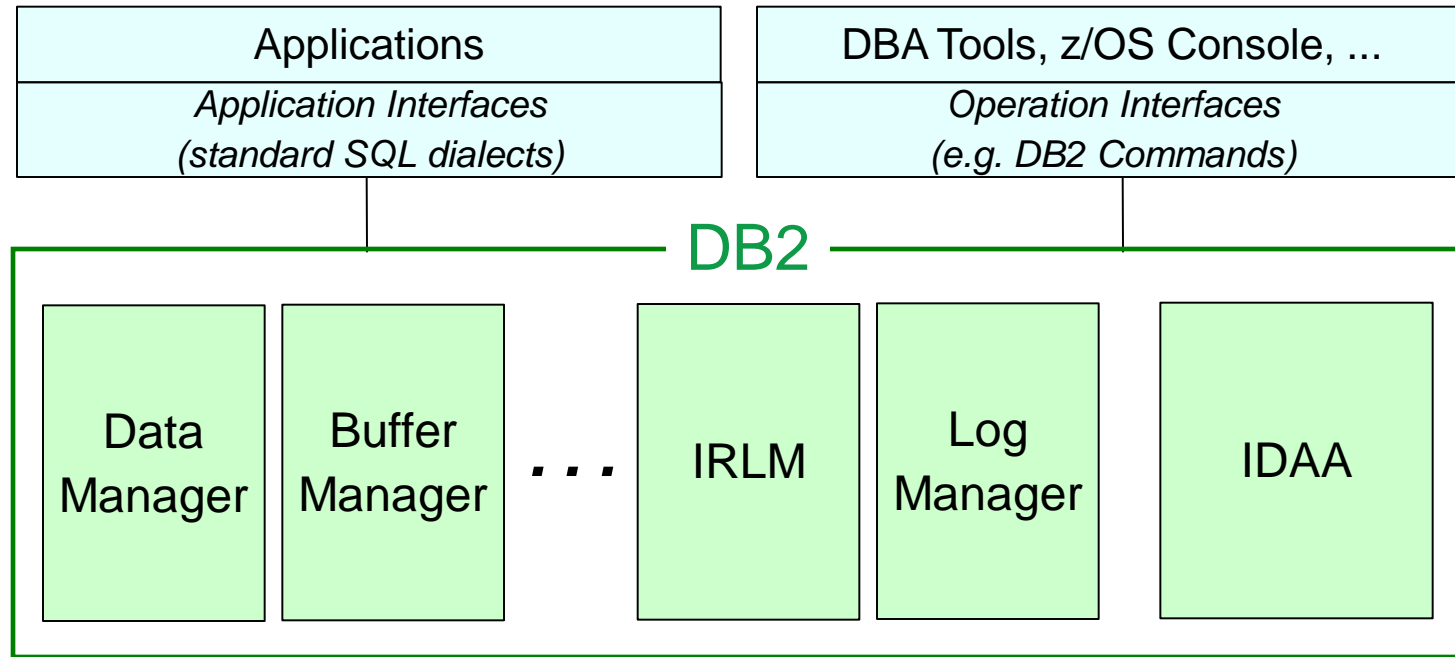- Improve resource allocation
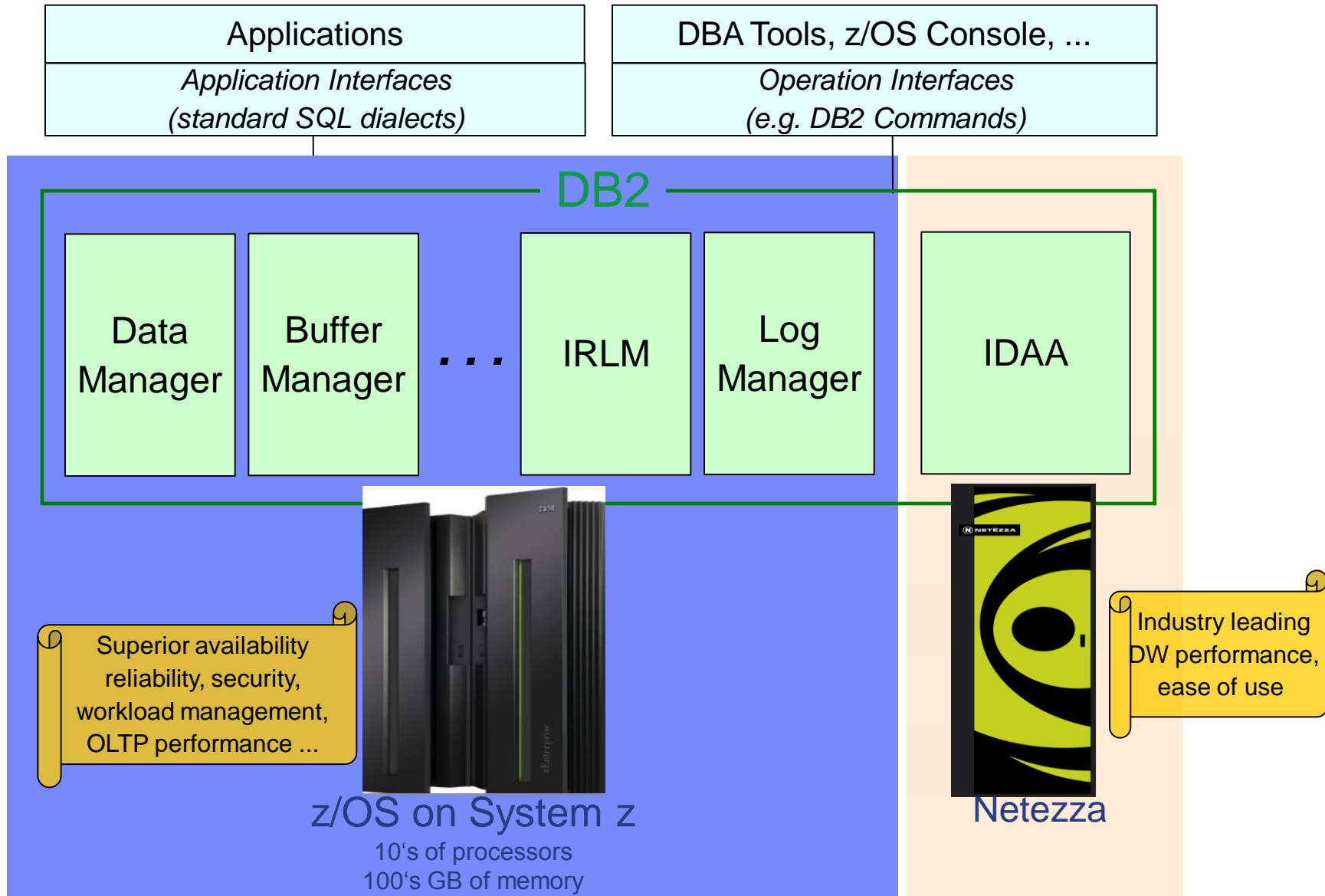
IBM Smart Analytics System 9700
High Value Data Warehousing – Standard Configuration

System z Z196

Cognos 10.1

Cognos. software

InfoSphere Warehouse 9.7.3

DS8800 Storage

SPSS Modeler 14.2

SPSS
AN IBM COMPANY

Data Warehouse

Cubing Services

ELT

Operational Source Systems
Structured/ Unstructured Data

DB2 for z/OS V10 VUE (option for MLC)

DB2 Utilities Suite

Image Copy, LOAD, UNLOAD, REORG, etc

Implementation Services

DB2 Components

Applications	DBA Tools, z/OS Console, ...
Application Interfaces (standard SQL dialects)	*Operation Interfaces (e.g. DB2 Commands)*

DB2

Data Manager	Buffer Manager	. . .	IRLM	Log Manager

IBM DB2 Analytics Accelerator as a Virtual DB2 Component

Applications	DBA Tools, z/OS Console, ...
Application Interfaces (standard SQL dialects)	*Operation Interfaces (e.g. DB2 Commands)*

DB2

Data Manager	Buffer Manager	. . .	IRLM	Log Manager	IDAA

Deep DB2 Integration

Applications	DBA Tools, z/OS Console, ...
Application Interfaces (standard SQL dialects)	*Operation Interfaces (e.g. DB2 Commands)*

DB2

Data Manager	Buffer Manager	. . .	IRLM	Log Manager	IDAA

Superior availability reliability, security, workload management, OLTP performance ...

Industry leading DW performance, ease of use

z/OS on System z
10's of processors
100's GB of memory

Netezza

Query Execution Process Flow

Application

Application Interface

Optimizer

DB2 for z/OS

Query execution run-time for queries that cannot be or should not be off-loaded to IDAA

IDAA DRDA Requestor

Heartbeat

SMP Host

SPU
CPU FPGA
Memory

SPU
CPU FPGA
Memory

SPU
CPU FPGA
Memory

SPU
CPU FPGA
Memory
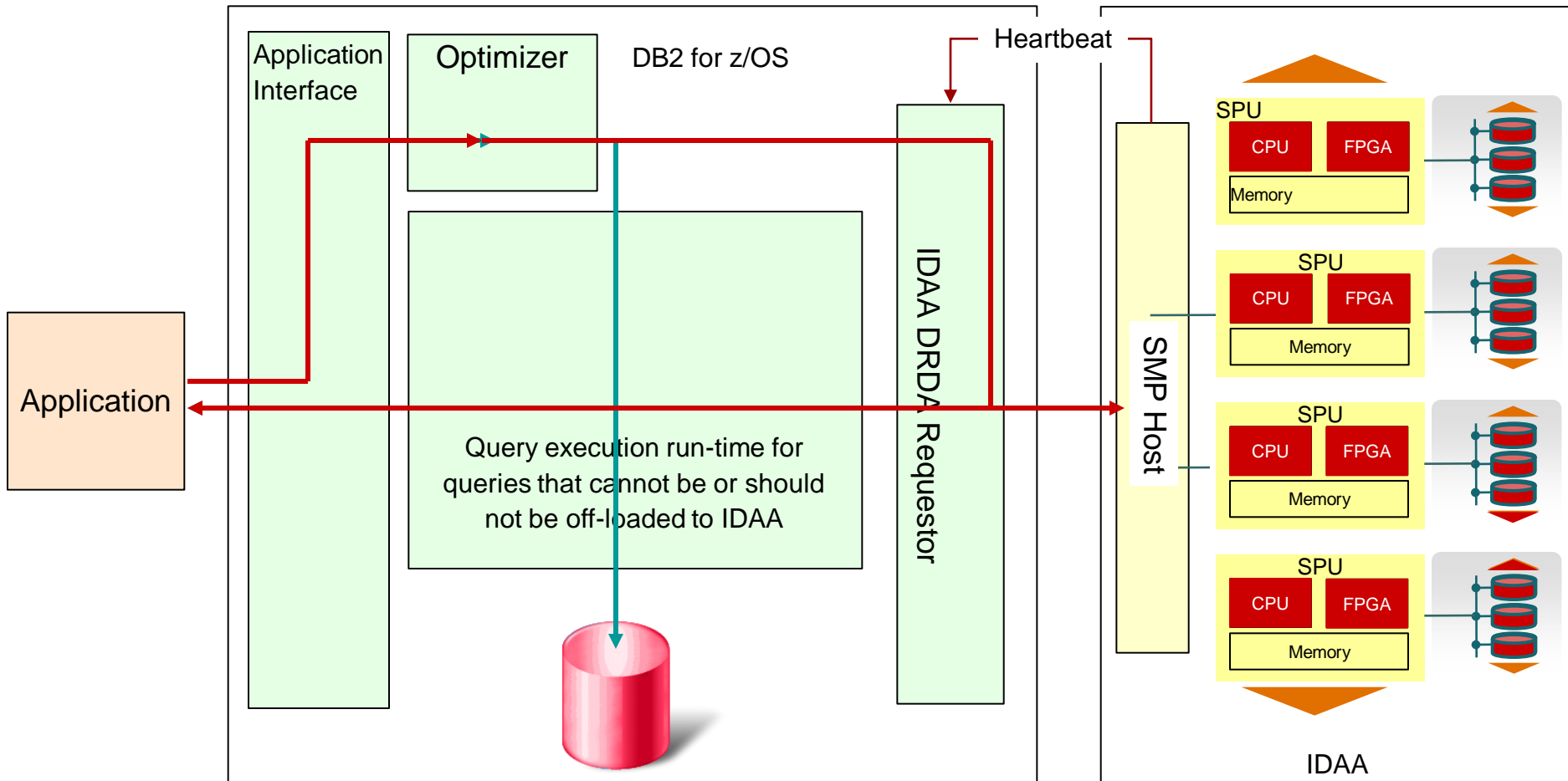
IDAA

→ Queries executed without IDAA

→ Queries executed with IDAA

→ Heartbeat (IDAA availability and performance indicators)

Fast Time to Value

- IBM DB2 Analytics Accelerator (Netezza 1000-12)
  - ➔ Production ready - 1 person, 2 days

- Table Acceleration Setup … **2 Hours**
  - DB2 "Add Accelerator"
  - Choose a Table for "Acceleration"
  - Load the Table (DB2 copy to Netezza)
  - Knowledge Transfer
  - Query Comparisons

- Initial Load Performance …
  - ➔ 400 GB "Loaded" in 29 Min
    570 million rows (Loads of 800GB to 1.3TB/Hr)

- Actual Query Acceleration … **1908x faster**
  - ➔ 2 Hours 39 Minutes to 5 Seconds

- CPU Utilization Reduction
  - ➔ 35% to ~0%

Actual customer results, October 2011

Query	Total Rows Reviewed	Total Rows Returned		DB2 Only			DB2 with IDAA			Times Faster
				Hours	Sec(s)		Hours	Sec(s)		
Query 1	2,813,571	853,320		2:39	9,540		0.0	5		1,908
Query 2	2,813,571	585,780		2:16	8,220		0.0	5		1,644
Query 3	8,260,214	274		1:16	4,560		0.0	6		760
Query 4	2,813,571	601,197		1:08	4,080		0.0	5		816
Query 5	3,422,765	508		0:57	4,080		0.0	70		58
Query 6	4,290,648	165		0:53	3,180		0.0	6		530
Query 7	361,521	58,236		0:51	3,120		0.0	4		780
Query 8	3,425.29	724		0:44	2,640		0.0	2		1,320
Query 9	4,130,107	137		0:42	2,520		0.1	193		13

Queries run faster

- **Save CPU resources**
- **People time**
- **Business opportunities**

Actual customer results, October 2011

DB2 Analytics Accelerator: "we had this up and running in days with queries that ran over 1000 times faster"

DB2 Analytics Accelerator: "we expect ROI in less than 4 months"

Accelerating decisions to the speed of business

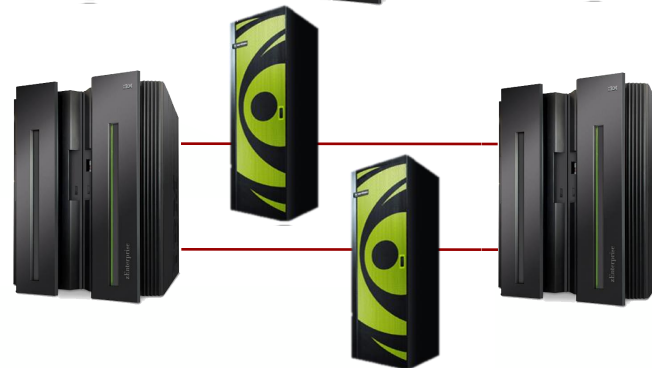# Live Demo: IDAA – BI Day with mixed workload on zEnterprise

Users doing complex reports

Users doing intermediate reports

Users doing simple reports

Simulated Cognos Report users

Initiated from Demo Web Front-end

Each report executes one or more queries

Complex Report

Intermediate Report

Intermediate Report

Simple Report

Simple Report

Web-Interface I
- Define Query-Mix

Parallel connections

Web-Interface II
- Enable/Disable Accelerator
- Display accelerator status
- Monitor query elapsed times

DB2 Analytics Accelerator (IDAA)

Monitoring RMF, Omegamon, …

Real World Business Analytics Workload:
Concurrent/multi-user Cognos Reporting workload

Demo Web Front-end

Hybrid Data Server

Demo focuses on hybrid benefits provided with IDAA as single database for mixed queries (transactional/complex)

- *Reduced complexity for Administration and Application Programming*

- *Throughput and response times*

Transactional Queries

Complex Queries

DB2 using DS8000

Demo Database size about 500 GB
Uses a typical star schema data warehouse, includes 80 tables

Connectivity Options

Multiple DB2 systems can connect to a single IDAA

A single DB2 system can connect to multiple IDAAs

Multiple DB2 systems can connect to multiple IDAAs

Full flexibility for DB2 systems:

- residing in the same LPAR
- residing in different LPARs
- residing in different CECs
- being independent (non-data sharing)
- belonging to the same data sharing group
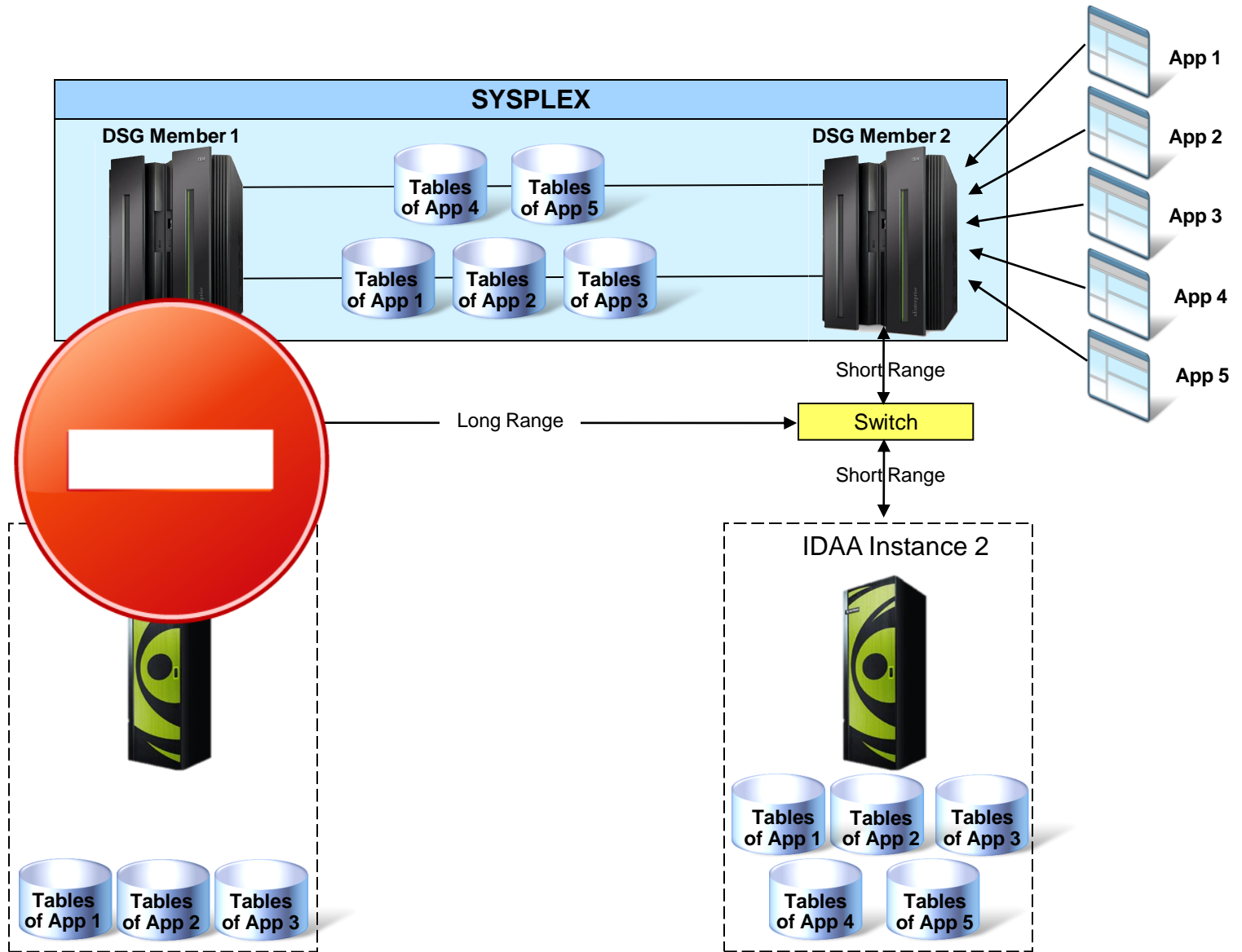- belonging to different data sharing groups
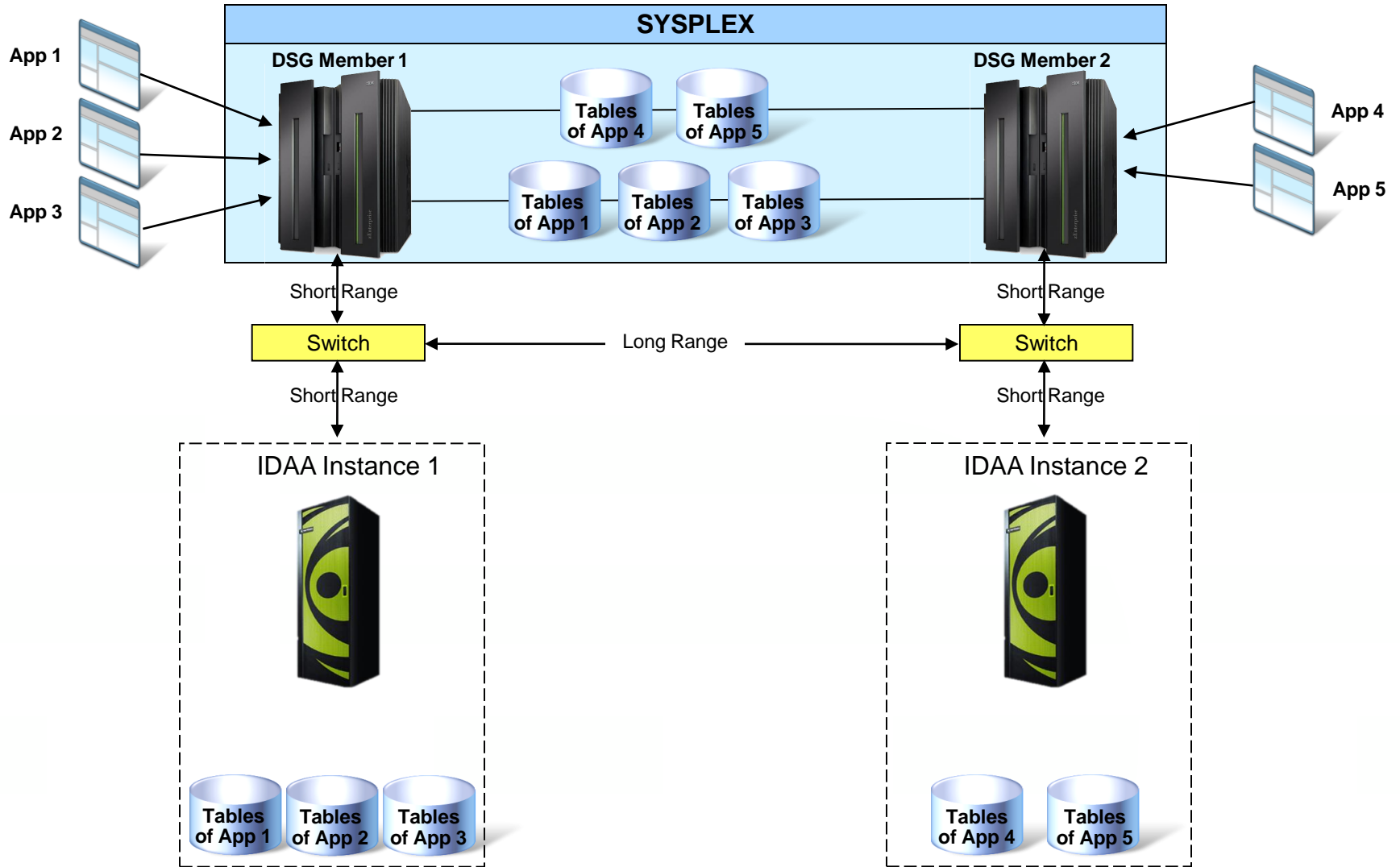
Better utilization of IDAA resources
Scalability
High availability

Agenda
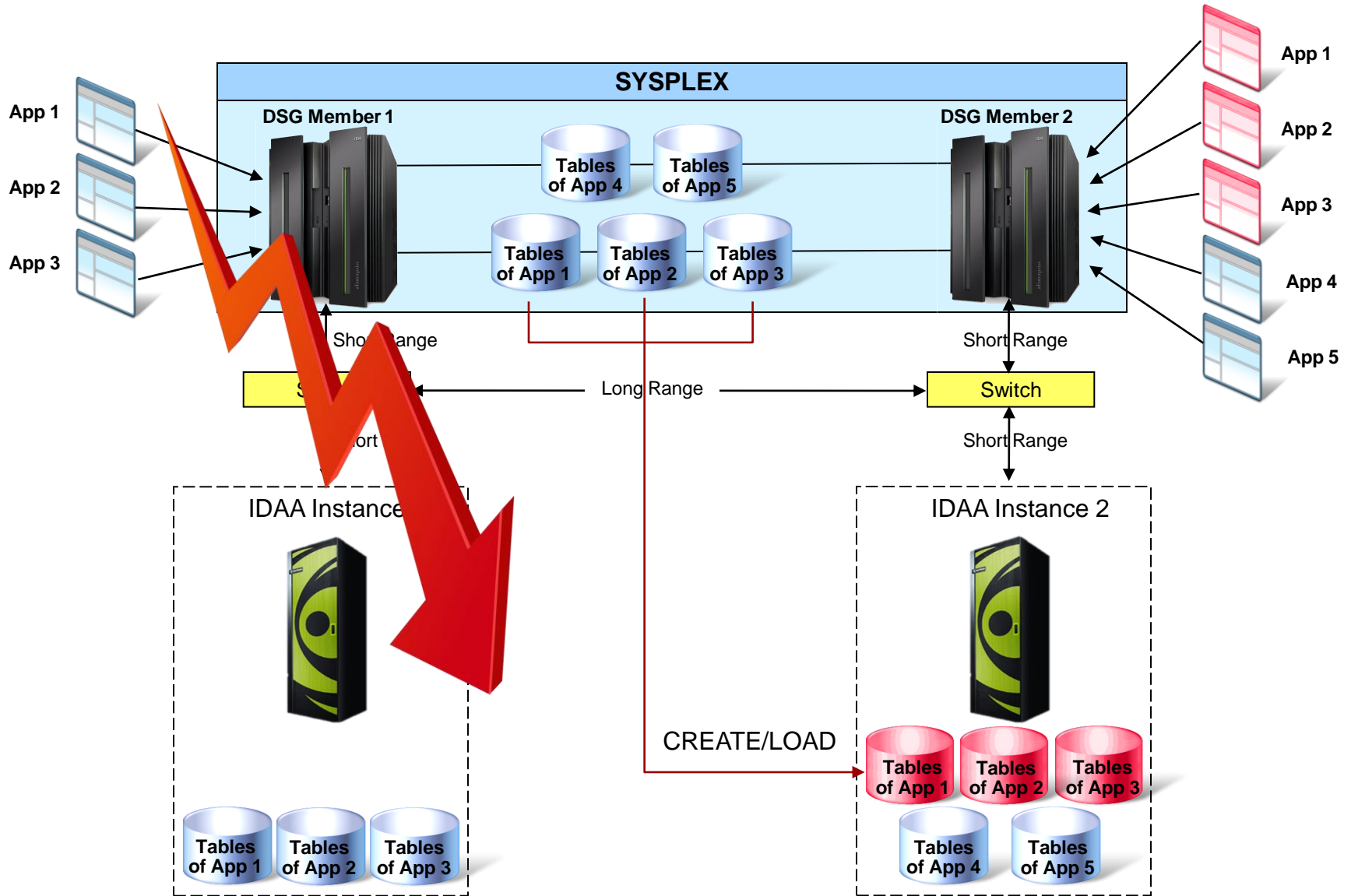
- Business and Technology Drivers

- Data Warehousing Solution on zEnterprise
  - IDAA Key Design and Operational Features
  - Powered by Netezza

- IDAA Use Cases and Next Steps

IDAA V2 Powered by IBM Netezza 1000 Appliance

Disk Enclosures

SMP Hosts

Snippet Blades™
(S-Blades, SPUs)
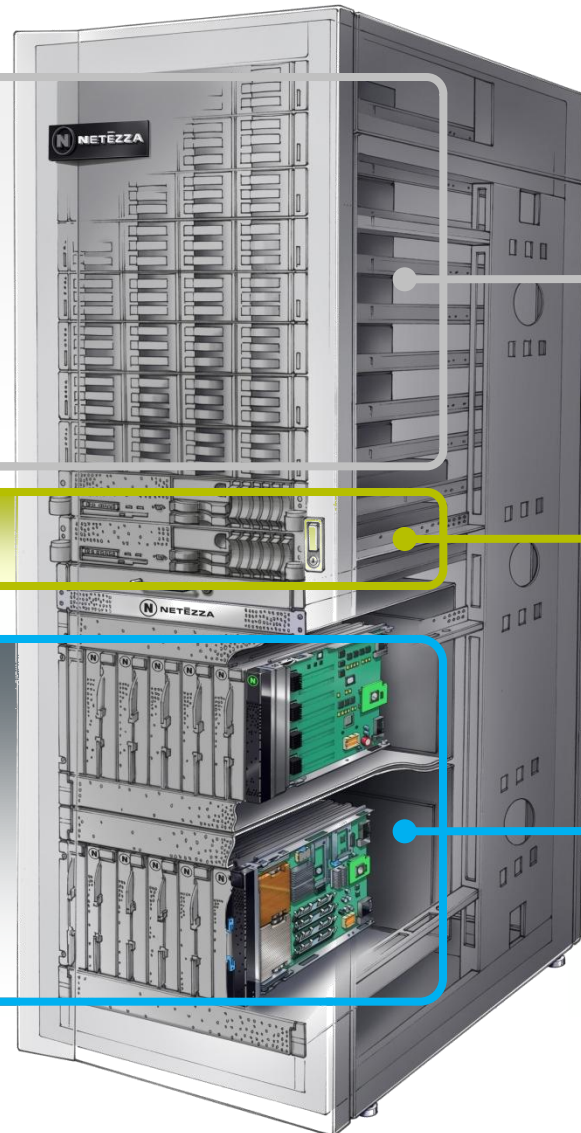
Slice of User Data
Swap and Mirror partitions
High speed data streaming
High compression rate
EXP3000 JBOD Enclosures
12 x 3.5" 1TB, 7200RPM, SAS (3Gb/s)
max 116MB/s (200-500MB/s compressed data)
e.g. in model 1000-12:
8 enclosures → 96 HDDs
32TB uncompressed user data (→ 128TB)

IDAA Server
SQL Compiler, Query Plan, Optimize
Administration
2 front/end hosts, IBM 3650M3 or 3850X5
clustered active-passive
2 Nehalem-EP Quad-core 2.4GHz per host

Processor &
streaming DB logic
High-performance database
engine streaming joins,
aggregations, sorts, etc.
e.g. in model 1000-12: 12 back/end SPUs
(more details on following charts)
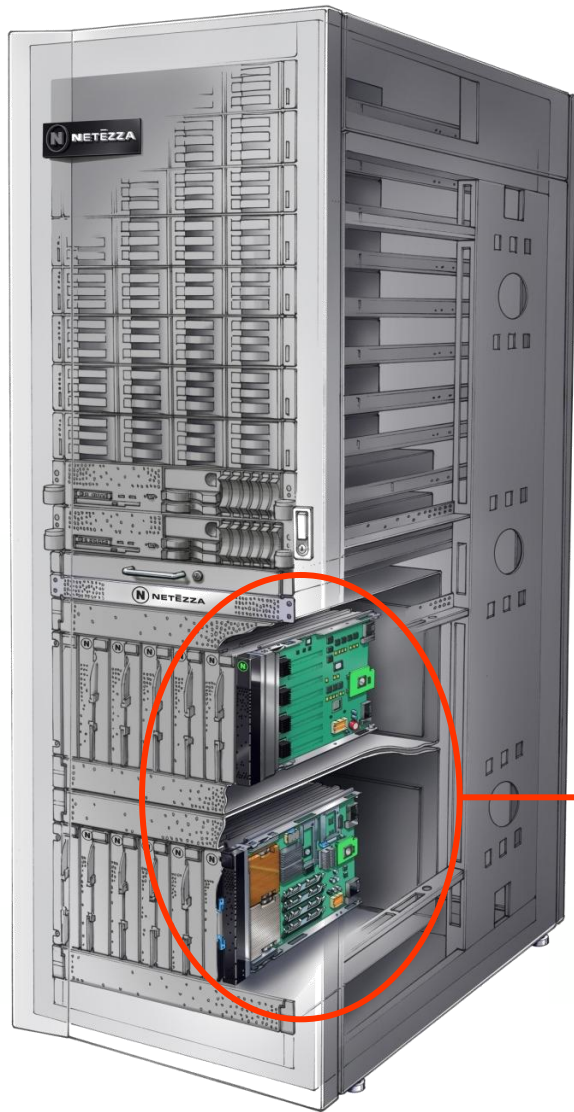
# Applying Data Stream Processing to DB2 Queries

… the key to the speed

```
select DISTRICT,
       PRODUCTGRP,
       sum(NRX)
from    MTHLY_RX_TERR_DATA
where   MONTH = '20091201'
and     MARKET = 509123
and     SPECIALTY = 'GASTRO'
```


DB2 for z/OS

CPU Core

FPGA Core

Complex ∑ Joins, Aggs, etc.

Restrict Visibility

Project

Decompress

S-Blade Table Cache

```
sum(NRX)
```

```
where MONTH = '20091201'
and   MARKET = 509123
and   SPECIALTY = 'GASTRO'
```

```
select DISTRICT,
       PRODUCTGRP,
       sum(NRX)
```

```
Slice of table
MTHLY_RX_TERR_DATA
    (compressed)
```
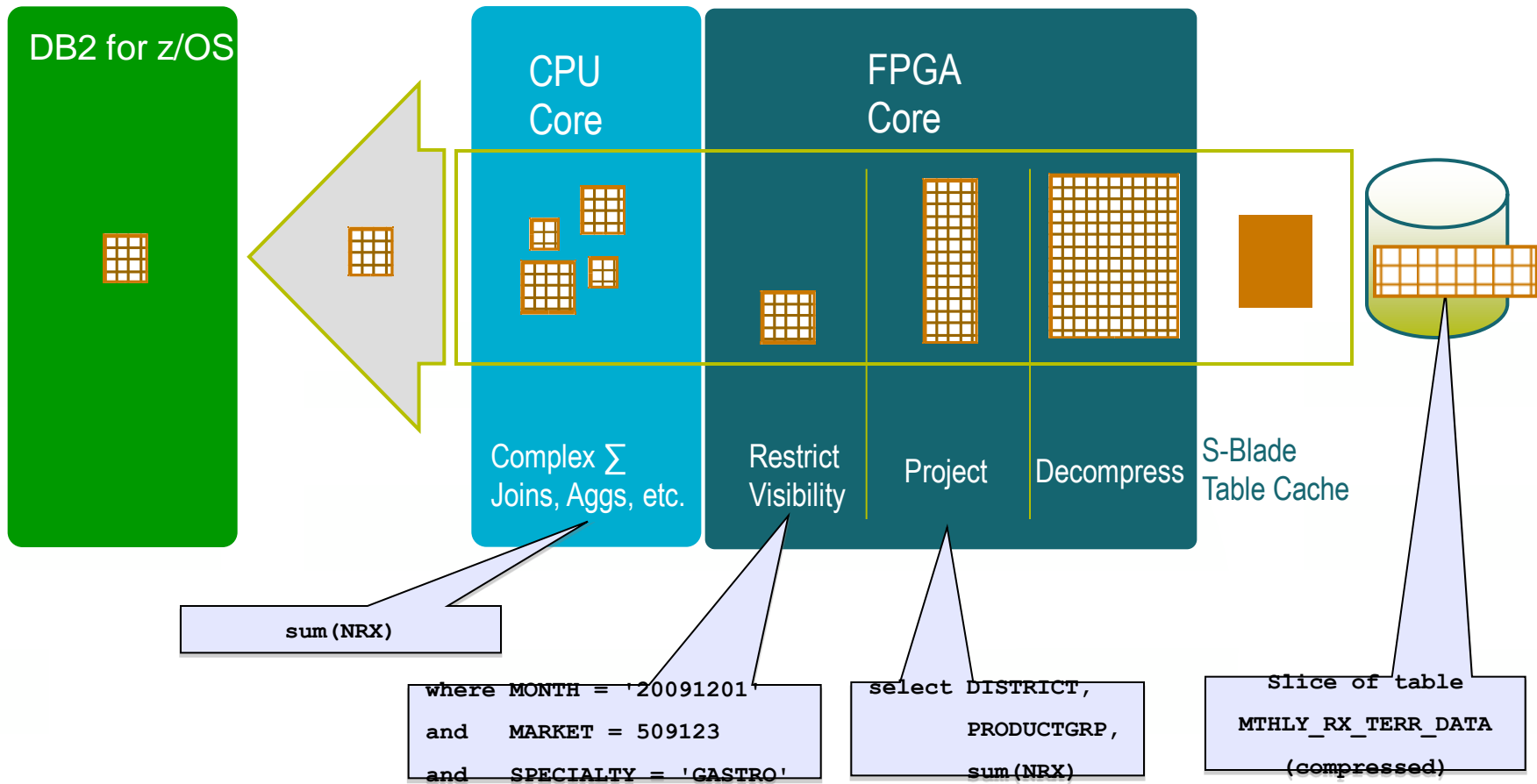
IBM Netezza 1000 Appliance Models
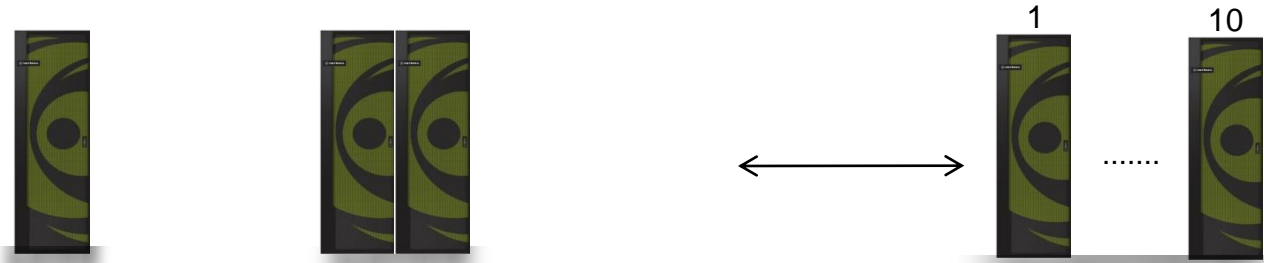
Model	1000-3	1000-6	1000-12	1000-24	1000-36	1000-48	1000-72	1000-96	1000-120
Cabinets	1/4	1/2	1	2	3	4	6	8	10
Processing Units	24	48	96	192	288	384	576	768	960
Capacity (TB)	8	16	32	64	96	128	192	256	320
Effective Capacity (TB)*	32	64	128	256	384	512	768	1024	1280

Capacity = User data space
Effective Capacity = User data space with compression (4x compression assumed)
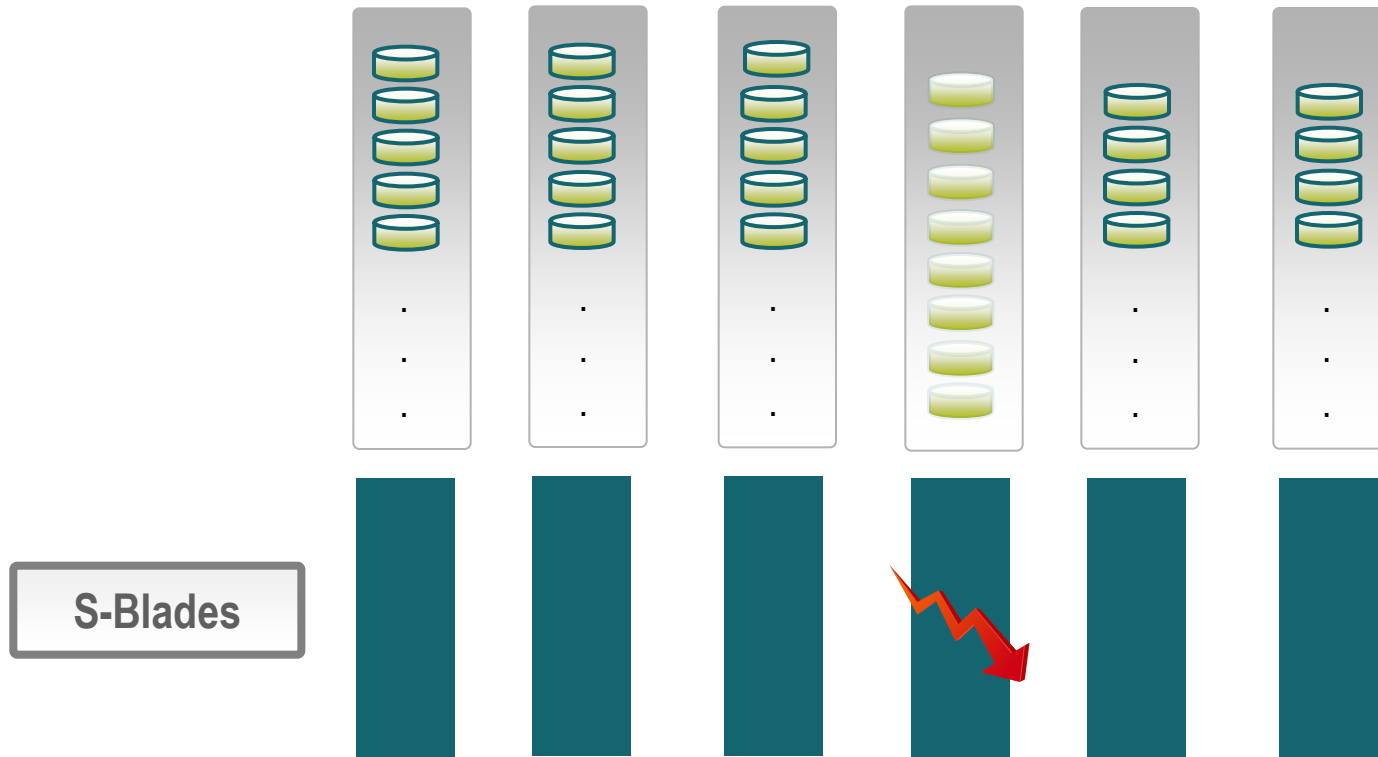
Shielding Against Disk Failures
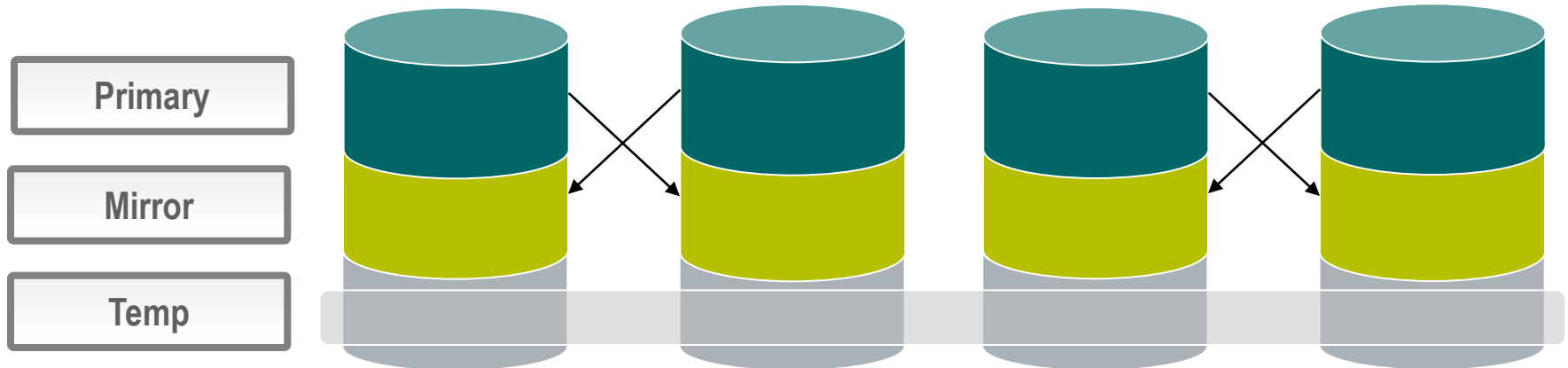
- All user data and temp space mirrored

- Disk failures transparent to queries and transactions

- Failed drives automatically regenerated

- Bad sectors automatically rewritten or relocated

Shielding Against S-Blade™ Failures
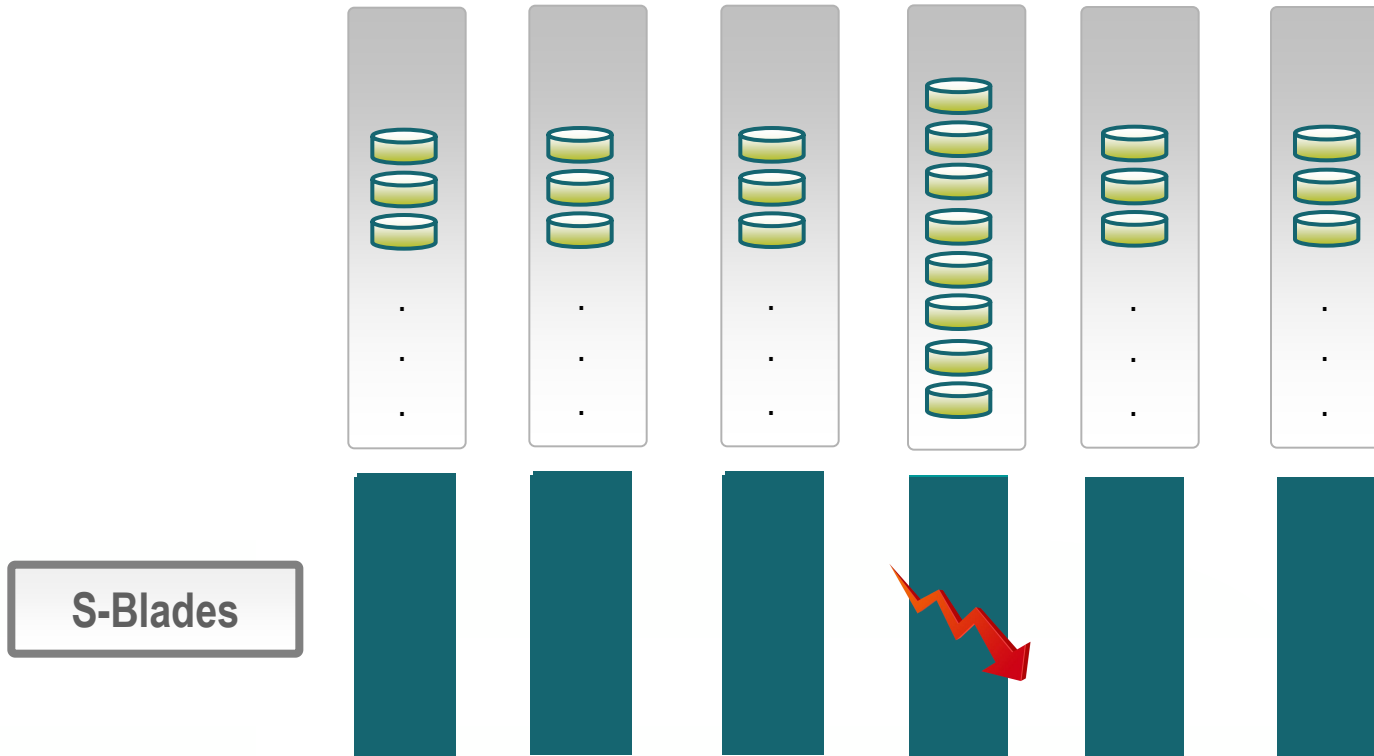
S-Blades

- S-Blade failure is automatically detected

Shielding Against S-Blade™ Failures

S-Blades

- Drives automatically reassigned to active S-Blades within a chassis
- Read-only queries (that have not returned data yet) automatically restarted
- Transactions and loads interrupted
- Loads automatically restarted from last successful checkpoint

DB2 Analytics Accelerator

Accelerating decisions to the speed of business

Blending System z and Netezza technologies to deliver unparalleled, mixed workload performance for complex analytic business needs.
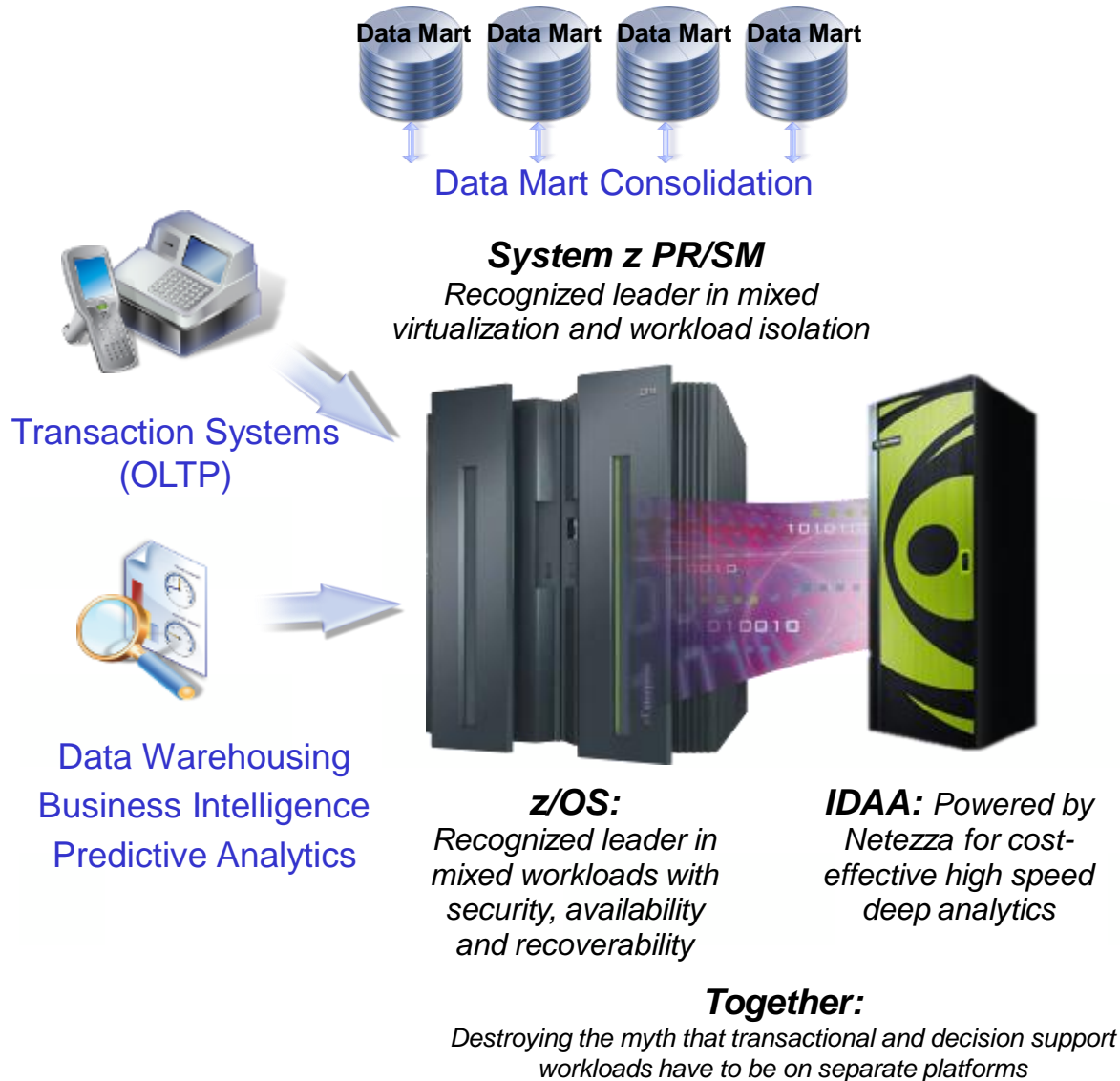

Get more insight from your data

- Fast, predictable response times for "right-time" analysis

- Accelerate analytic query response times

- Improve price/performance for analytic workloads

- Minimize the need to create data marts for performance

- Highly secure environment for sensitive data analysis

- Transparent to the application

Agenda

- Business and Technology Drivers

- Data Warehousing Solution on zEnterprise
  - IDAA Key Design and Operational Features
 - Powered by Netezza

- **IDAA Use Cases and Next Steps**

zEnterprise with IDAA - The Ultimate Data Server Platform

IBM

Data Mart **Data Mart** **Data Mart** **Data Mart**

Data Mart Consolidation

System z PR/SM
Recognized leader in mixed virtualization and workload isolation

Transaction Systems (OLTP)

Data Warehousing
Business Intelligence
Predictive Analytics

z/OS:
Recognized leader in mixed workloads with security, availability and recoverability

IDAA: *Powered by Netezza for cost-effective high speed deep analytics*

Bringing it all together

- *Better Business Response*

- *Reduced Costs*

- *More Available*

- *More Secure*

- *Reduced Data Movement*

- *Reduced Data Latency*

- *Reduced Complexity*

- *Reduced Resources*

Together:
Destroying the myth that transactional and decision support workloads have to be on separate platforms

IBM

- On average, **70% of the data** that feeds data warehousing and business analytics solutions **originates on the System z platform**
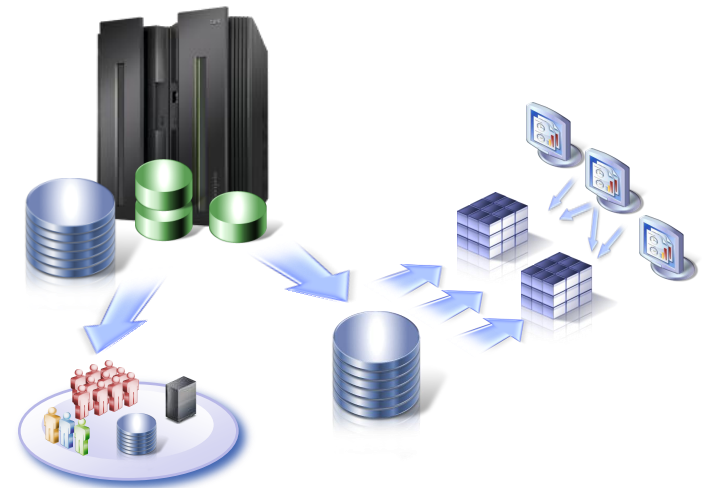
How is that data being analyzed today?	*IDAA use case*
Analyzed on distributed EDW	**Consolidation of EDW**
Analyzed on a sprawl of Data Marts	**Consolidation of Data Marts**
Analyzed on distributed EDW	**Offload data and analytics from the EDW** to optimized DB2 z with IDAA
Not being analyzed yet (new workload)	**More Insights** • new BI applications • new operational BI
With long running queries (>1min) in DWH or Operational Data Stores on DB2 for z/OS.	**Modernization**

- Partner with IBM to assess best IDAA entry point for your needs:
    - jointly assess your Data Warehouse & BI architecture and current challenges
    - define IDAA architecture
    - identify distributed workload for IDAA
    - define use case for new project
    - conduct Workload Assessment of long running queries on DB2 for z/OS
    - define PoC and next steps

- Contact the Data Warehousing on System z Center of Excellence at dwhz@de.ibm.com