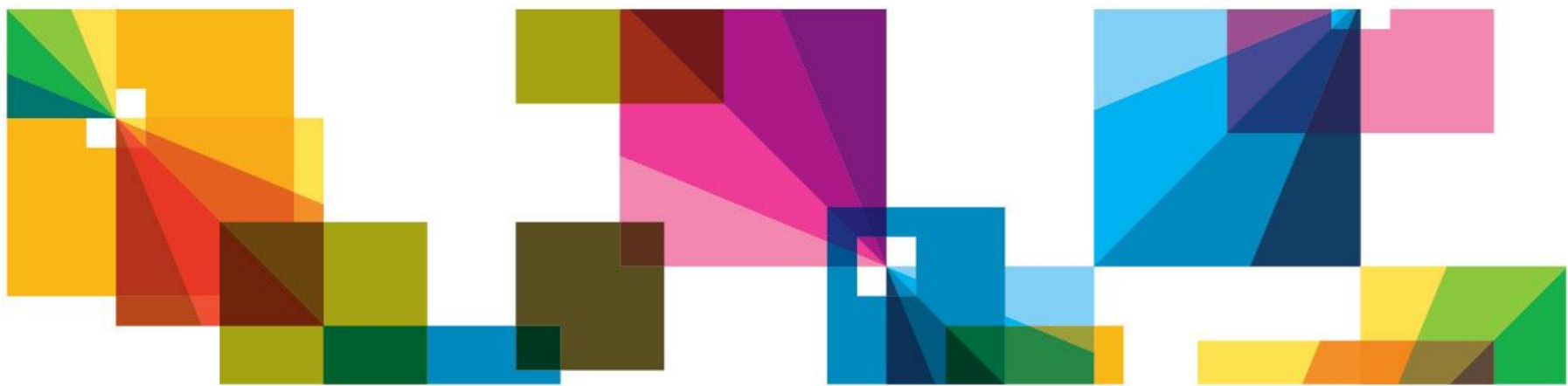


---

**Ernesto Beneduce**

IBM Appliances for Business Analytics



## Agenda

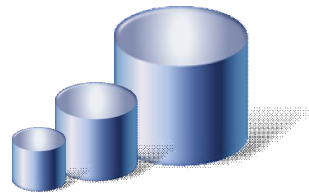
- **Introduction: the rise of appliances**
- **IBM Netezza**
- **IBM DB2 Analytics Accelerator for z/OS**



# DW and BI evolution – the rise of appliances

Driving forces and enablers

Emergence of Business Analytics



**Exploding** data volumes, new workloads

50TB DW are now common    80% unstructured data  
6TB/sec exchanged on the Web



Companies are forced to change their data centers to **address new requirements**

TCO: the appliance-based alternative

*...the use of data warehouse (DW) appliances is surging in the market. Because of their potential value, organizations must investigate DW appliances as an alternative or addition to a custom DW environment.*



*Adapted from a Gartner Group report, 2010*

# IBM Smarter Analytics appliances



## Netezza Analytic Appliances

Rapidly deployed Data Warehouses

Focused set of analytic query capabilities

Rack units optimized for analytics – no customization or tuning needed

Scale & performance thru number of racks

Minimal maintenance

## IBM DB2 Analytic Appliance (powered by Netezza)

Accelerate analytic query response times

Improve price/performance for analytic workloads

Highly secure environment for sensitive data analysis

Transparent to the application

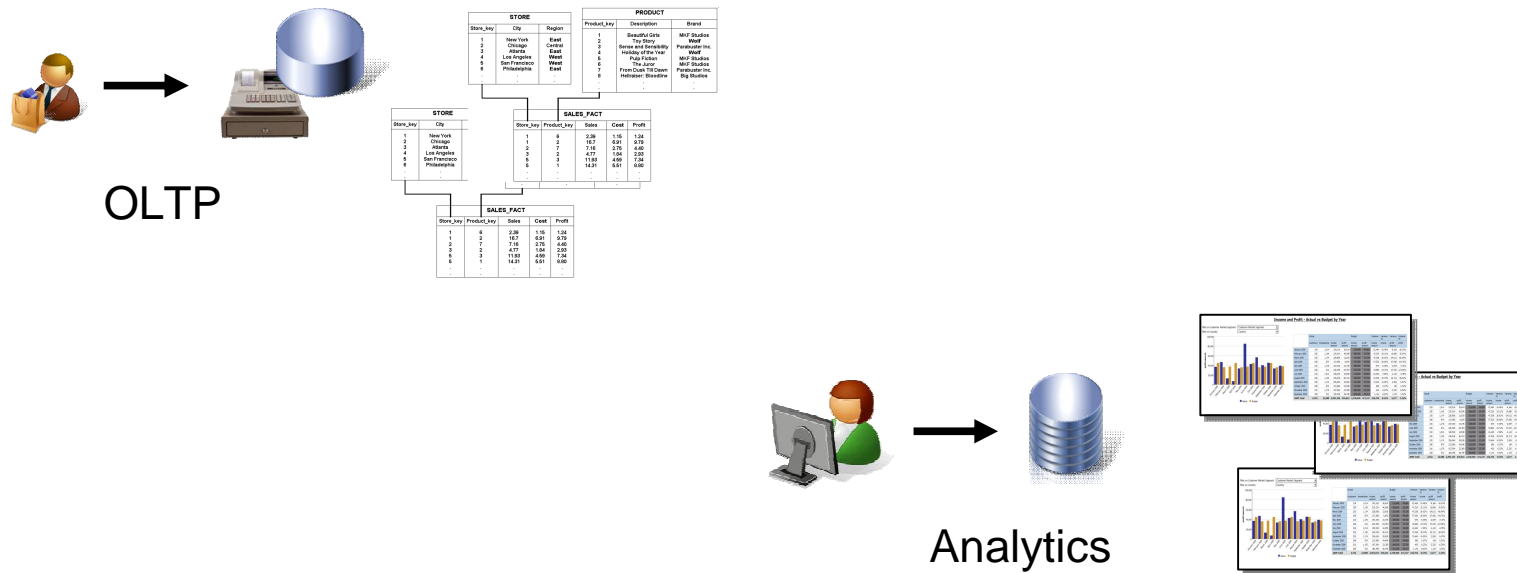
## Why an Appliance ?



- **Dedicated** device
- **Complete** solution
- Standard interfaces
- **Easy** installation
- **Easy** operation
- **Easy** management
- **Easy** support
- **Low cost**

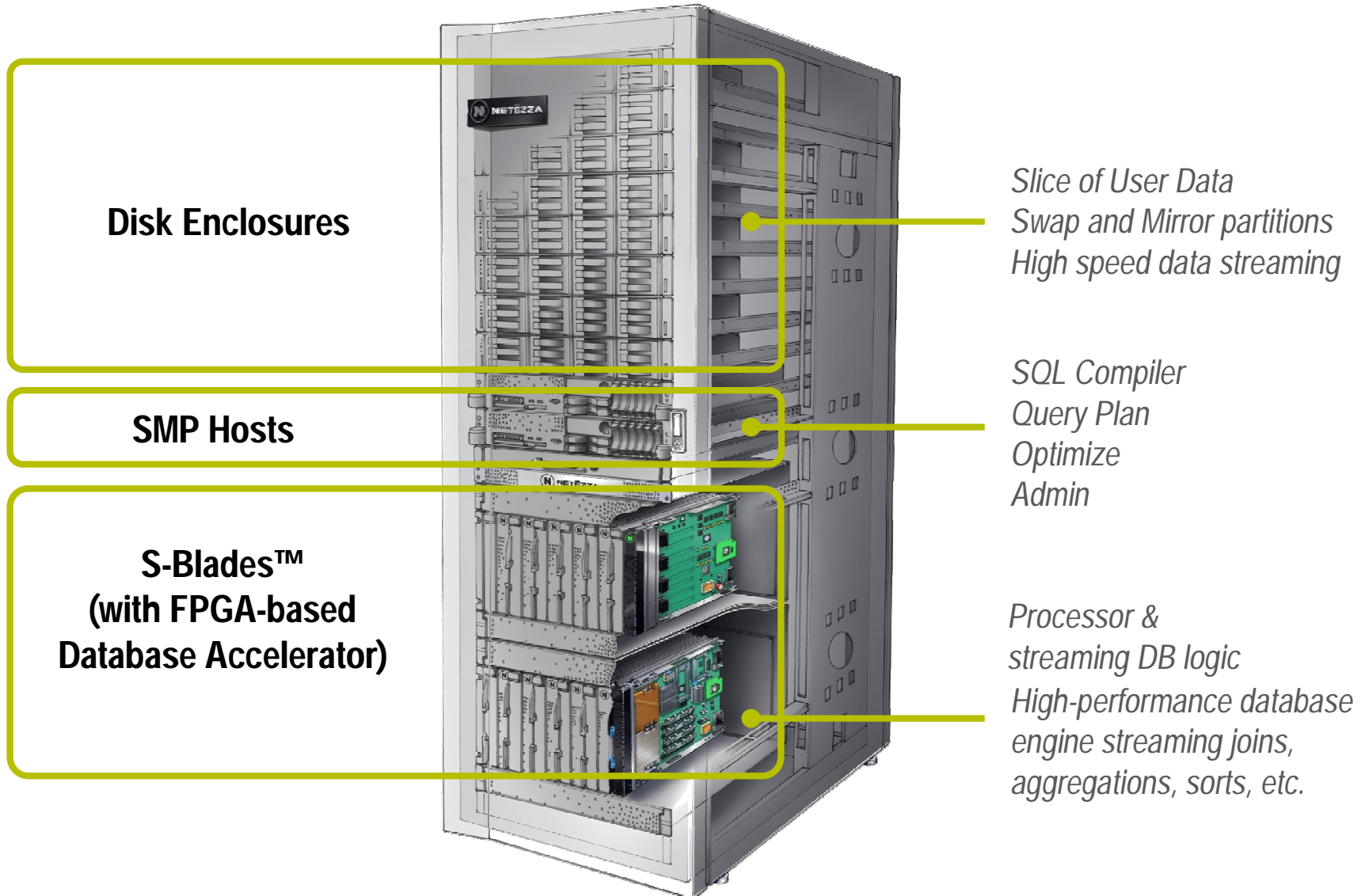
# OLTP and Analytics together ? ...

Other vendors say that appliances must be capable to run both OLTP and analytics workloads together

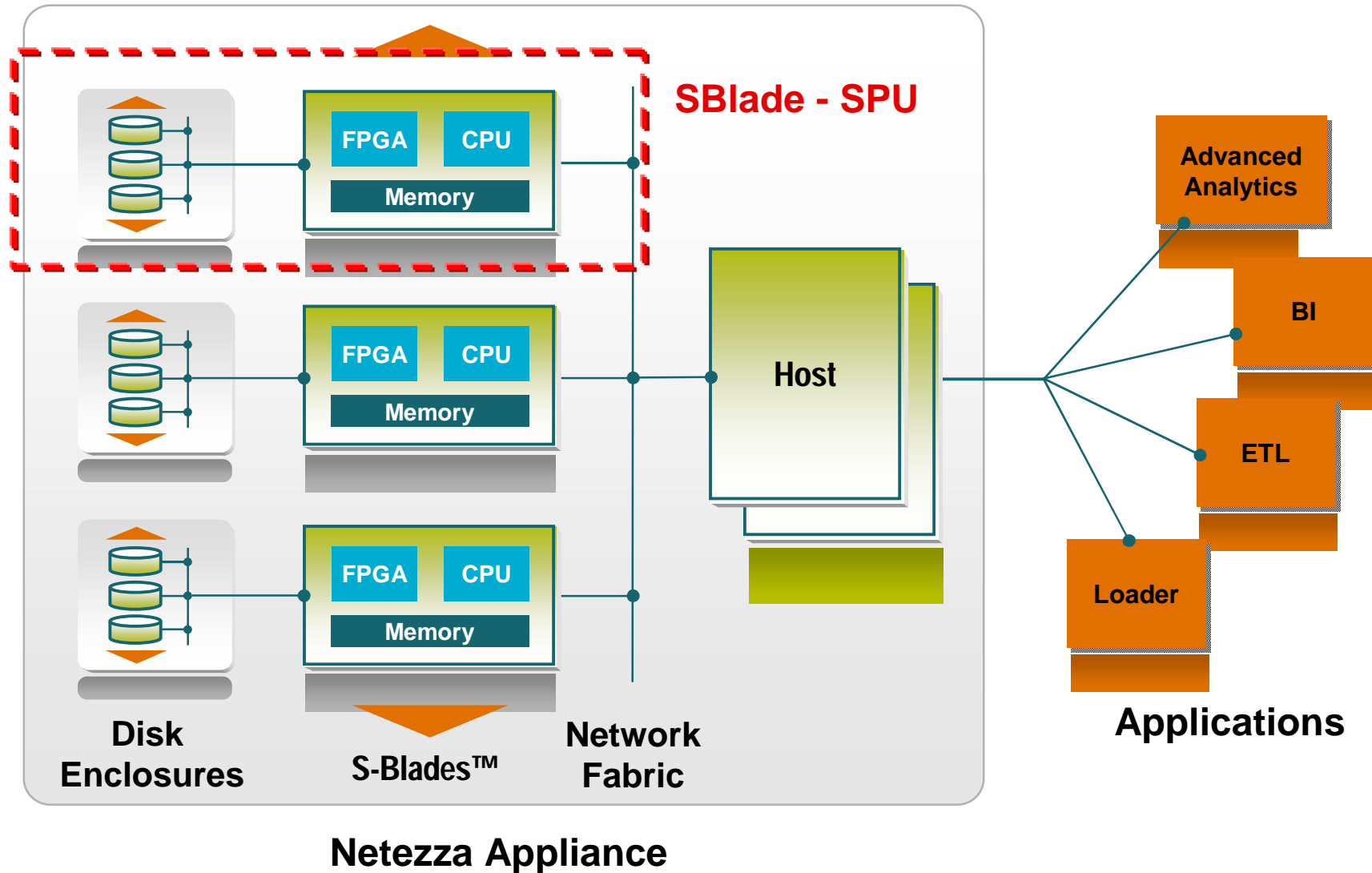


IBM believes that designing and building systems optimized for specific workloads is the best approach

# Netezza architecture (1/2)

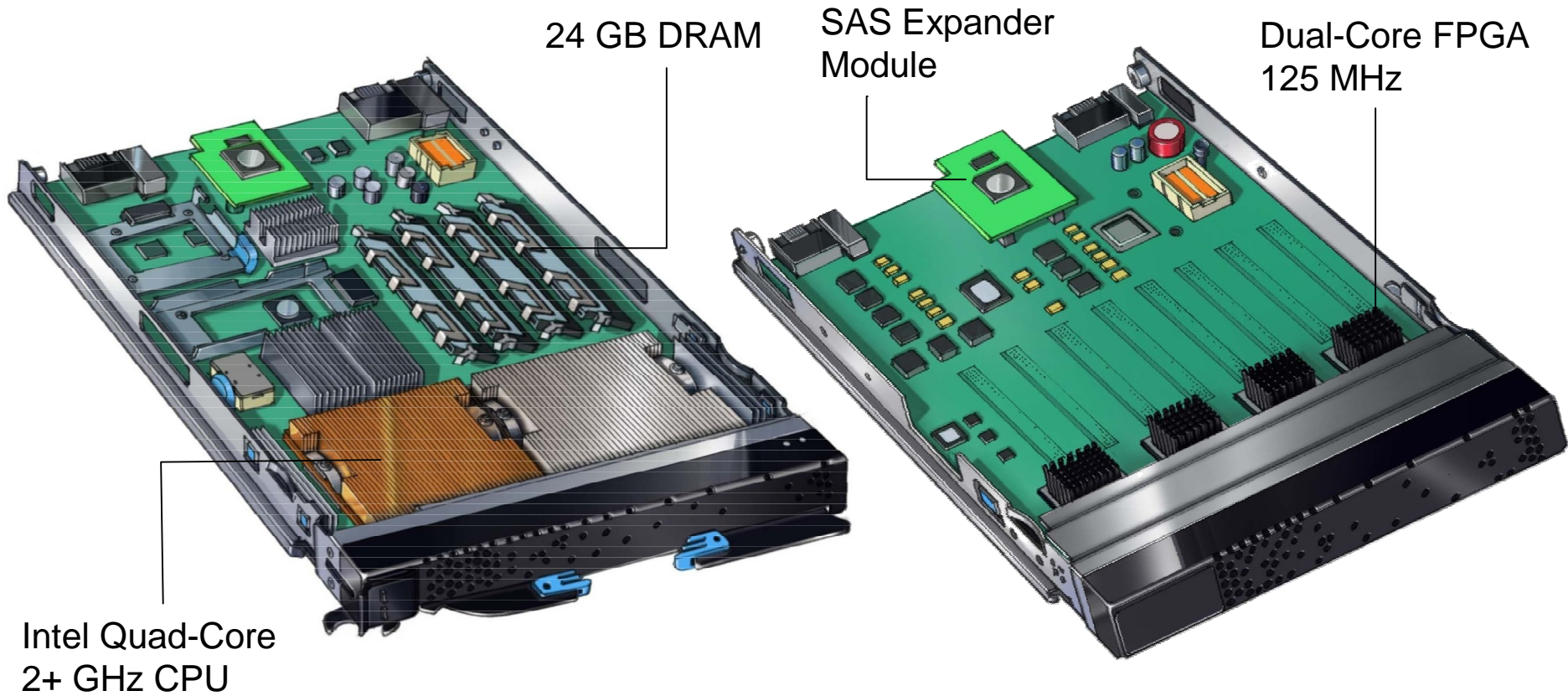


# Netezza Architecture (2/2)





# S-blades

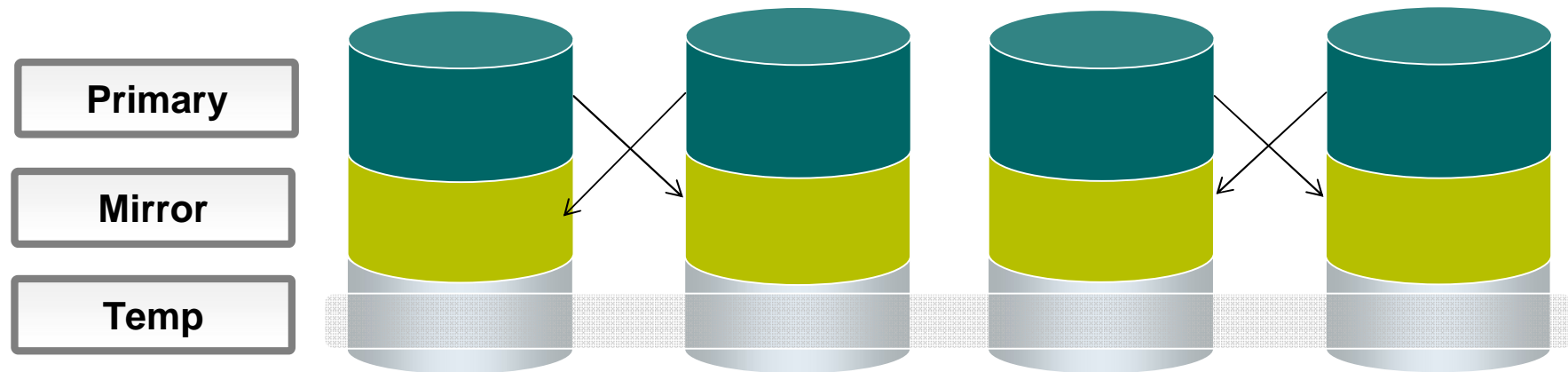


IBM BladeCenter Server

Netezza DB Accelerator

**1 S-Blade = 8 Snippet Processors**

## Disk Mirroring and Failover



- All user data and temp space mirrored
- Disk failures transparent to queries and transactions
- Failed drives automatically regenerated
- Bad sectors automatically rewritten or relocated

# IBM Netezza Data Warehouse Appliance Family



IBM Netezza 100  
(Skimmer)

Development & Test System



IBM Netezza 1000  
(TwinFin)

Data Warehouse  
High-Performance Analytics



IBM Netezza High Capacity  
Appliance

Queryable Archiving  
Backup/DR

# Netezza 1000 – Scalability

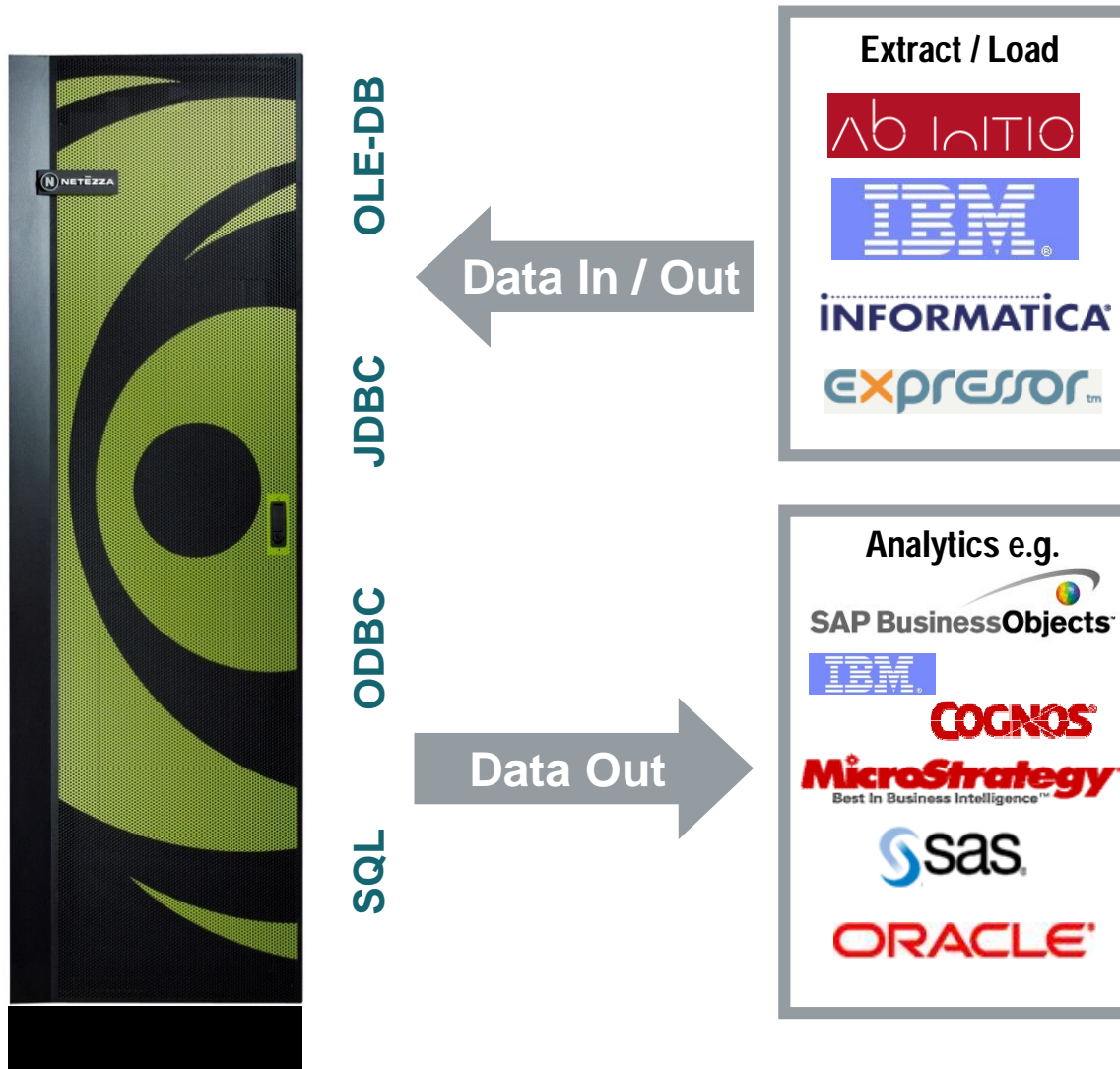
1.....10



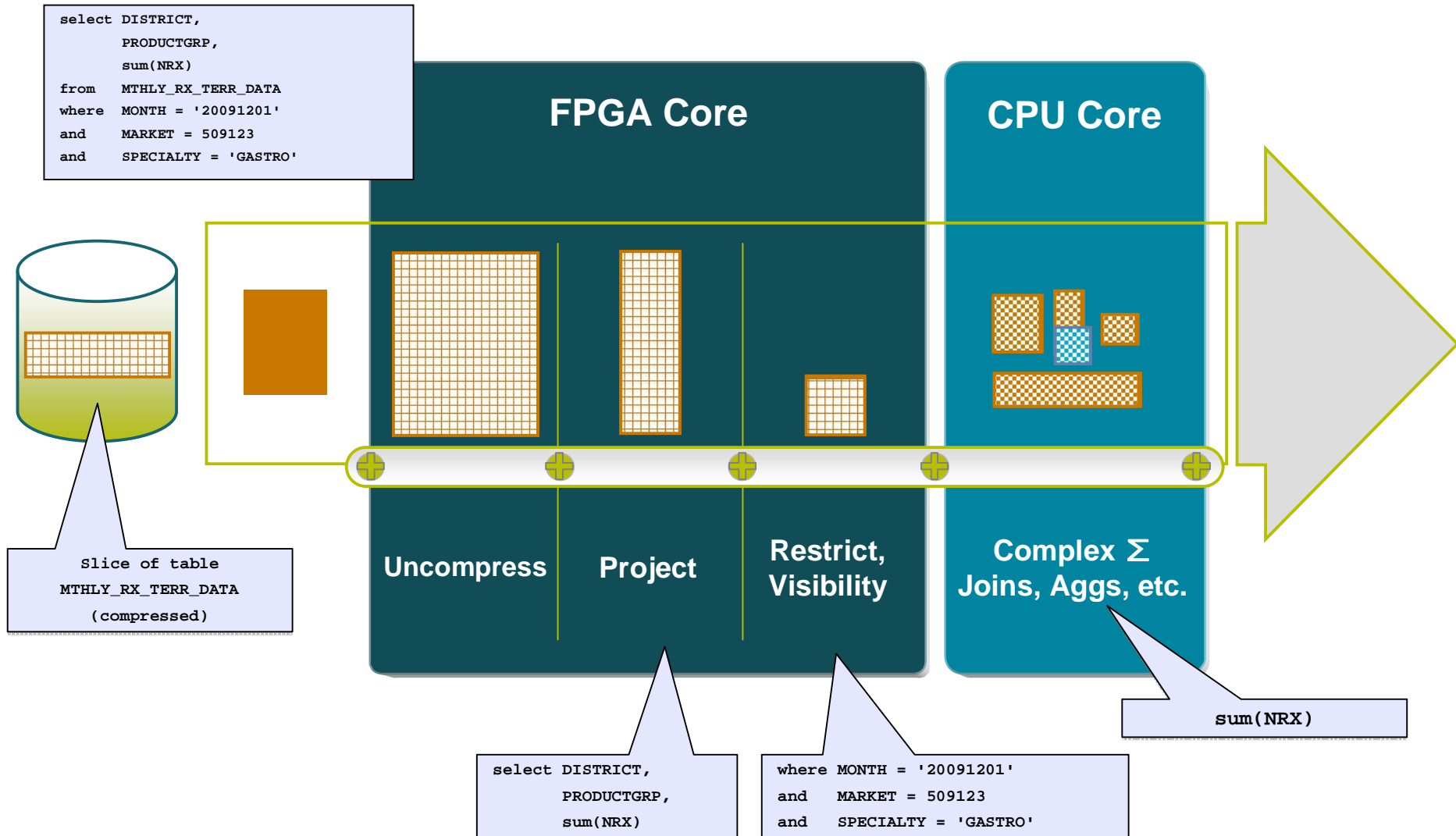
	TF3	TF6	TF12	TF24	TF48	TF120
Snippet Processors	24	48	96	192	384	960
Capacity (TB)	8	16	32	64	128	320
Compression (TB)	16	32	64	128	256	640

Capacity	=	User Data space
Compression	=	Effective User Data Space

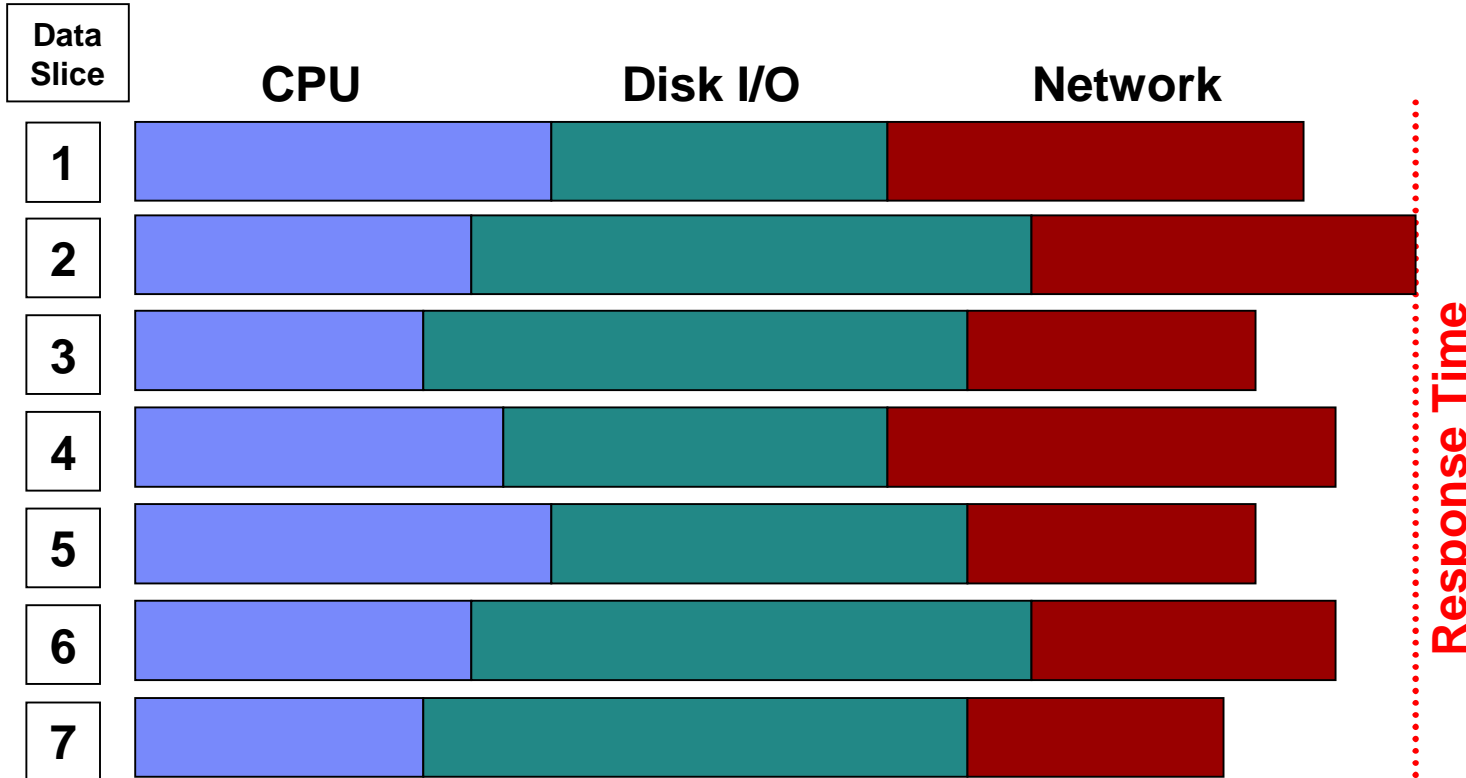
# Netezza – Standard Interfaces



# How S-Blades process queries



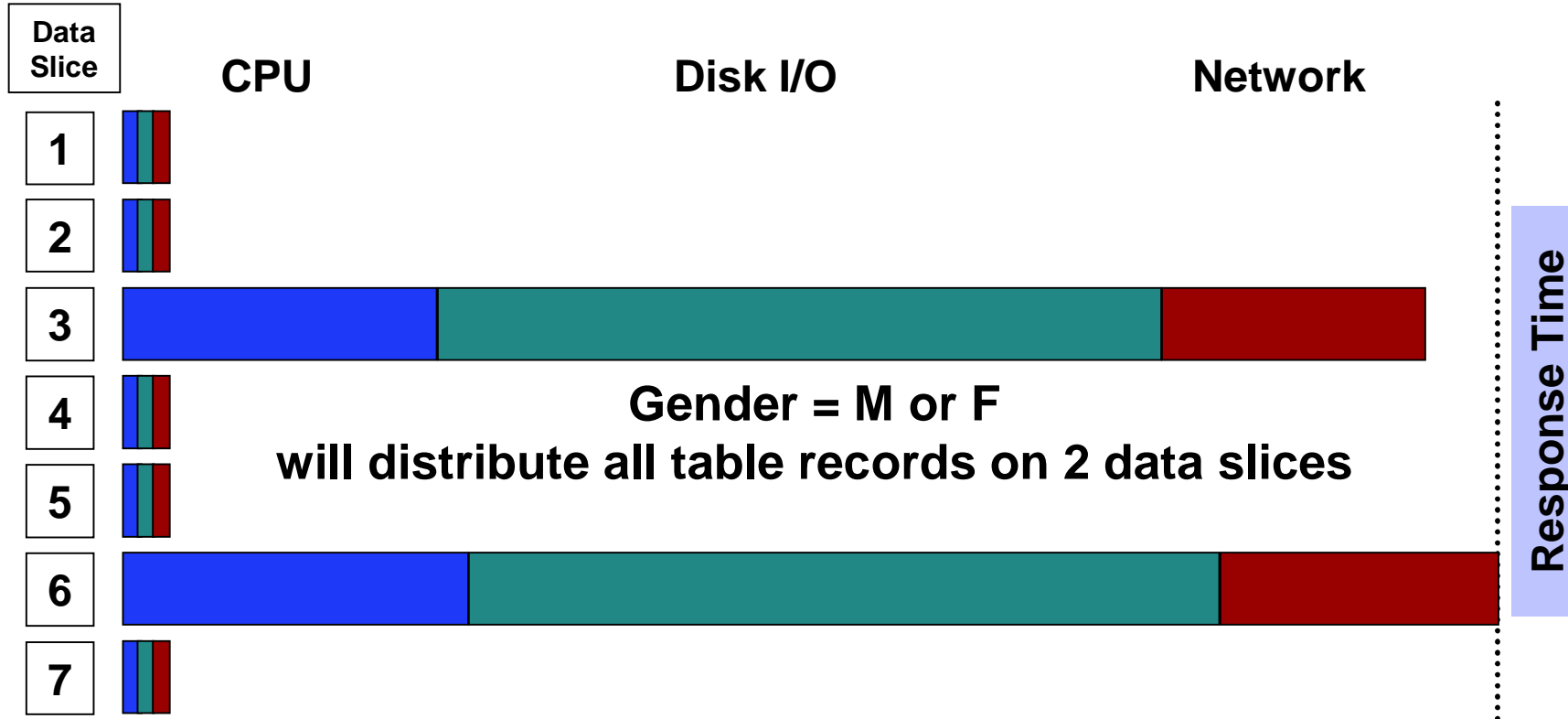
# Distributions and Performance



Response time is affected by the completion time for all of the data slices in the MPP array.

Data evenly distributed across partitions represent the **KEY** factor for achieving best performances

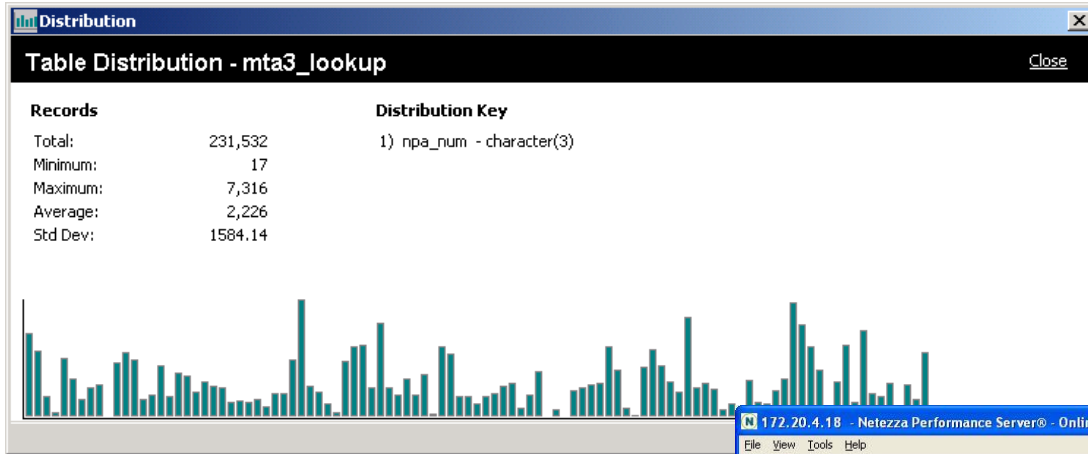
# Hash Distributions and data skewing



**Select a distribution key with unique values and high cardinality**

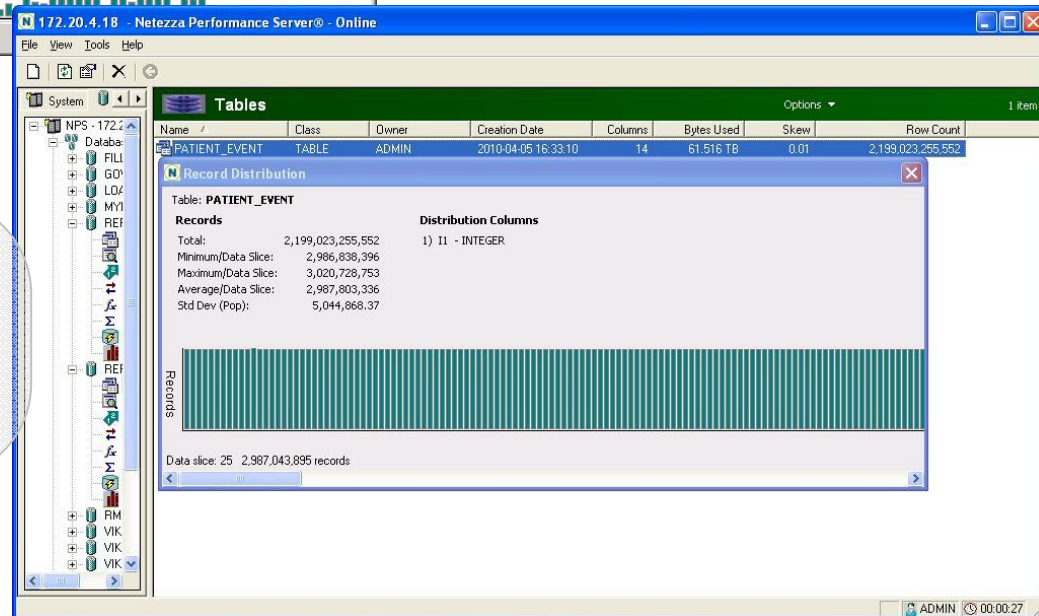


# Distribution monitoring with NZAdmin



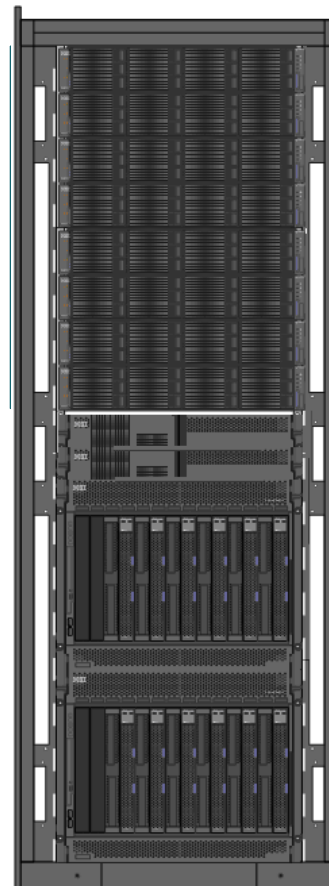
Bad !

Good !



# SW IOT Netezza Centro di Eccellenza

- **IBM Netezza 1000-12 a Segrate**
- Dedicato a Clienti e Business Partners: ambiente completo di hardware e software
- Esecuzione di Demo e Proof of Concepts



- 8 Disk Enclosures
- 96 1TB SAS Drives (4 hot spares)
- RAID 1 Mirroring

- 2 Hosts (Active-Passive):
- 2 Quad-Core Intel 2.6 GHz CPUs
- 7x146 GB SAS Drives
- Red Hat Linux 5 64-bit

- 12 Netezza S-Blades™:
- 2 Intel Quad-Core 2+ GHz CPUs
- 4 Dual-Engine 125 MHz FPGAs
- 16 GB DDR2 RAM
- Linux 64-bit Kernel

# IBM DB2 Analytics Accelerator

## *Marrying the best of IBM System Z and Netezza*

**IBM  
System z**



**Mixed Workload System**

**IBM  
Netezza**



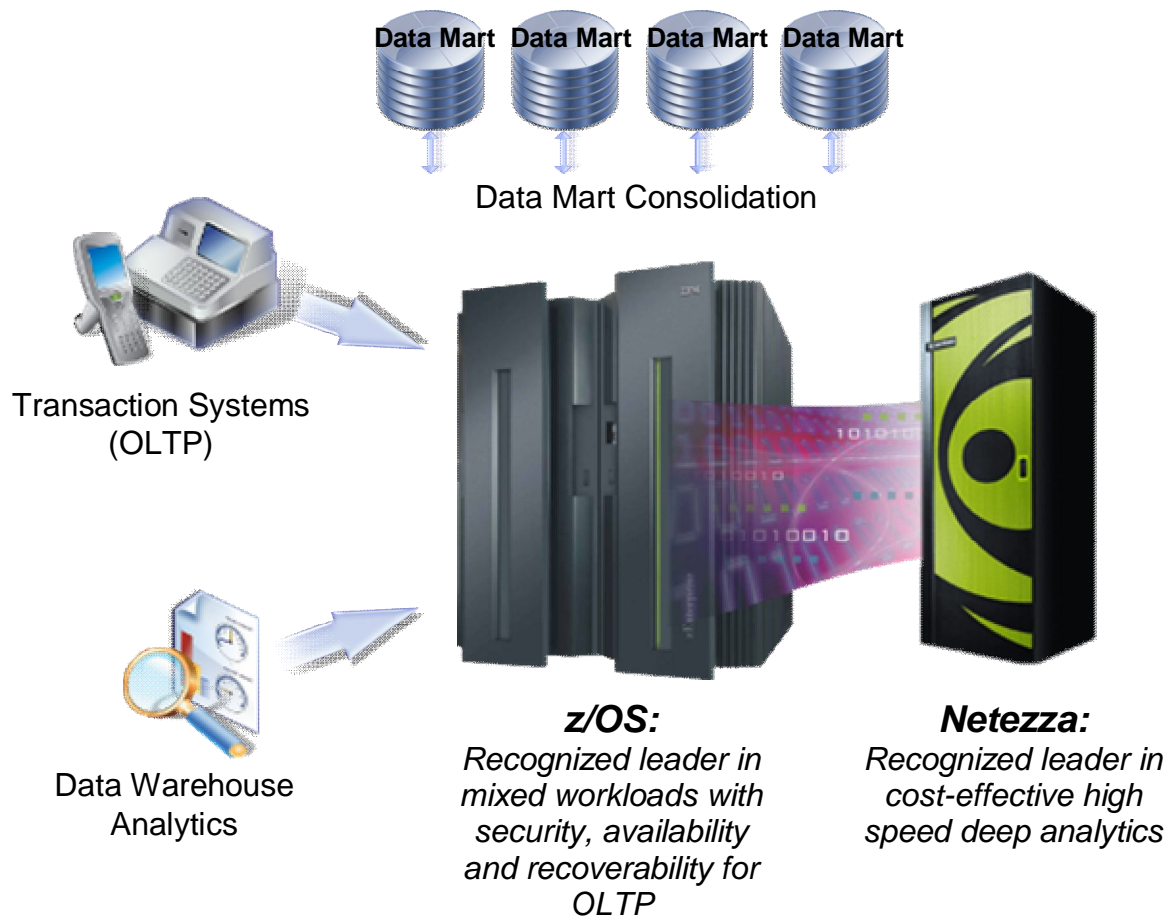
**Focused Appliance**

*Capitalizing on the strengths of both platforms while driving to the most cost effective, centralized solution - destroying the myth that transaction and decision systems had to be on separate platforms*

**Very diverse workload**

**Very focused workload**

# Combining the best transaction system with the best analytics system



## Best in OLTP

*Industry recognized leader in mission critical transaction systems*

## Best in Data Warehouse

*Proven appliance leader in high speed analytic systems*

## Best in Consolidation

*Unprecedented mixed workload flexibility and virtualization providing the most options for cost effective consolidation*

# It's all about speed !

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

- **US Customer** “...we had this up and running in days with queries that ran over 1000 times faster”
- **Europe Customer** “...we expect ROI in less than 4 months”

*Accelerating decisions to the speed of business*

# IDAA adoption assessment - Quick Workload Test

## Customer

- Collecting information from dynamic statement cache, supported by step-by-step instruction and REXX script (small effort for customer)
- Uploading compressed file (up to some MB) to IBM FTP server

## IBM / Center of Excellence

- Importing data into local database
- Quick analysis based on known DB2 Analytics Accelerator capabilities

**IBM Smart Analytics Optimizer -- Workload**  
Center of Excellence, Datawarehouse on System z, IBM Research & Development

**Query Summary**

	Total	With potential	Uncertain	W/o potential
Queries	23	11 (48%)	5 (22%)	7 (30%)
Query Blocks	23	11 (48%)	5 (22%)	7 (30%)
Elapsed Time	144801.47	106821.61 (74%)	8150.21 (6%)	29829.66 (21%)
CPU Time	21300.25	11420.12 (54%)	1453.14 (7%)	8426.98 (40%)

Queries	23	100%
... no eligible blocks	7	30%
... not read only	0	0%
... with very large dim.	1	4%
... with quant. pred.	0	0%

Leaf Query Blocks	23	100%
... with aggregations	23	100%
... with UDFs	0	0%
... with unsupported functions	6	26%
... with unsupported join types	0	0%
... with very large dimensions	1	4%

Start trace time: Apr 2, 2010 9:41 AM  
End explain time: Apr 2, 2010 4:37 PM  
Min stmt cached: Apr 2, 2010 9:44 AM  
Max stmt cached: Apr 2, 2010 4:37 PM

Disclaimer: Information provided in this document is for information purposes only and does not guarantee characteristics nor imply supported features of IBM products. V20100331  
Schema: IGBURG -- Date: 4/12/10 -- Page: 1

**Report for a first assessment:**

- Acceleration potential for
  - Queries
  - Estimated time
  - CP cost

Query blocks with acceleration potential  
Query blocks with uncertain potential  
Query blocks without acceleration potential

Elapsed time with acceleration potential  
Elapsed time with uncertain potential  
Elapsed time without acceleration potential

