

Name: PureData for Analytics

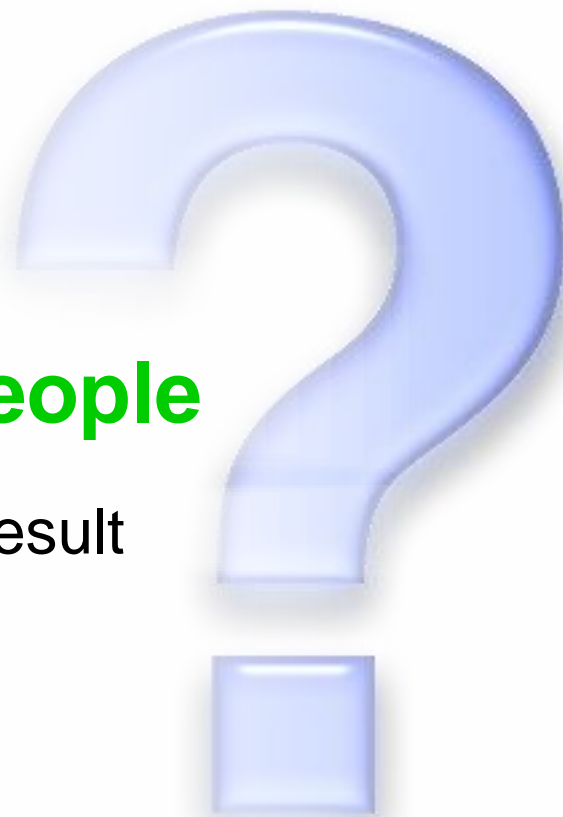
Title: The next generation of Power and Simplicity



Would You Still Use Google

If It Took **3** Days And **7** People

To Get A Search Result



days for a single query

constant tuning

“ **Nearly 70%** of data warehouses experience performance-constrained issues of various types. ”

- Gartner 2010 Magic Quadrant

specialized resources required

months to deploy

Data warehouse or data holding pen?



“ *Many of these ‘large’ Oracle data warehouses are simply holding pens.*

-- Overlooking problems with Oracle's
Exadata
Neil Raden
The Intelligent Enterprise blog

“ *Our existing solution was not keeping up with our growing business demands, nor was it putting us in a position to accommodate new business. We needed to break the cycle of more data, more requirements, more money.*

”

http://intelligent-enterprise.informationweek.com/blog/archives/2009/12/overlooking_pro.html?sessionid=KBTNTOW15M54VQE1GHRSKHWATMY32JVN

IBM Technical **Summit**

Stay ahead.

-- Emory Heisler
VP Global IT Services Wolters Kluwer Health



Customer



Transaction



Business Transaction



Simple Query

Item: 'Shoes'
Cost: '\$34'
Cust: 'James'



Transactional Database

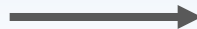


Item	Cost	Cust
Shoes	\$34	James

Business Analyst



Complex Query



Sales & Profit for
Shoes & Belts
Year >= 2005

Data Warehouse



SALES

2010
2009
2008
2007
2006
2005



BI Reports & Dashboards



Meeting Big Data Challenges – Fast and Easy!



PureData

System for Transactions

For apps like E-commerce...

Database cluster services optimized for transactional throughput and scalability

PureData

System for Analytics

For apps like Customer Analysis...

Data warehouse services optimized for high-speed, peta-scale analytics and simplicity

Powered by Netezza technology

PureData

System for Operational Analytics

For apps like Real-time Fraud Detection...

Operational data warehouse services optimized to balance high performance analytics and real-time operational throughput

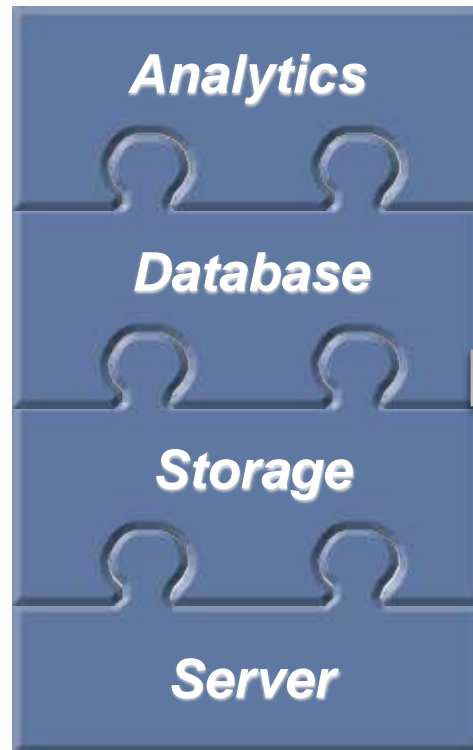
Appliance Simplicity

- Dedicated device
- Optimized for purpose
- Complete solution
- Fast installation
- Very easy operation
- Standard interfaces
- Low cost



Move analytics into the Data Warehouse

- Integrate the server, storage, database and analytics into one optimized package
- Move complex analytics into the database
- Integrated, high performance analytics within the data warehouse



The Simple Appliance for Serious Analytics

Powered by
Netezza Technology

Built-in Expertise

- No indexes or tuning
- Data model agnostic
- Fully parallel, optimized In Database Analytics

Integration by Design

- Server, Storage, Database in one easy to use package
- Automatic parallelization and resource optimization to scale economically
- Enterprise-class security and platform management

Simplified Experience

- Up and running in hours
- Minimal ongoing administration
- Standard interfaces to best of breed Analytics, BI, and data integration tools
- Built-in analytics capabilities allow users to derive insight from data quickly
- Easy connectivity to other Big Data Platform components

New





Transforms the User Experience

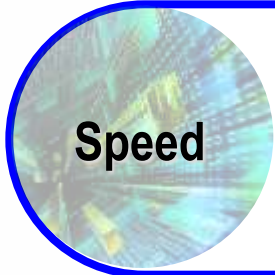
- ✓ Purpose-built analytics engine
- ✓ Integrated database, server and storage
- ✓ Standard interfaces
- ✓ Low total cost of ownership

Speed: 10-100x faster than traditional system

Simplicity: Minimal administration and tuning

Scalability: Peta-scale user data capacity

Smart: High-performance advanced analytics



Speed

Up to **2000X** faster than before
Growing by **30% every month**

"Netezza has allowed us to reduce the complexity of regulatory reporting and processing of exchange data from days down to minutes."



Simplicity

Up and running **6 months**
before having any training
200X faster than Oracle system
ROI in less than **3 months**

"Allowing the business users access to the Netezza box was what sold it."

- Steve Taff, Executive Dir. of IT Services



Scalability

1 PB on Netezza
7 years of historical data
100-200% annual data growth

"NYSE ... has replaced an Oracle 10 relational database with a data warehousing appliance from Netezza, allowing it to conduct rapid searches of 650 terabytes of data."

- ComputerWeekly.com



Smart

SUNY Buffalo researchers reduced the time to perform **quintillions of computations** from **27 hours** to **12 minutes**

"Once we had the data on Netezza we were able to do the same analysis and much more complex analysis in minutes. The research draws on medical records, lab results, MRI scans, and patient surveys."

- Dr. Murali Ramanathan, SUNY Buffalo



No indexes and tuning

No storage administration

- No dbspace/tablespace sizing and configuration
- No redo/physical/Logical log sizing and configuration
- No page/block sizing and configuration for tables
- No extent sizing and configuration for tables
- No Temp space allocation and monitoring
- No RAID level decisions for dbspaces
- No logical volume creations of files
- No integration of OS kernel recommendations
- No maintenance of OS recommended patch levels
- No JAD sessions to configure host/network/storage

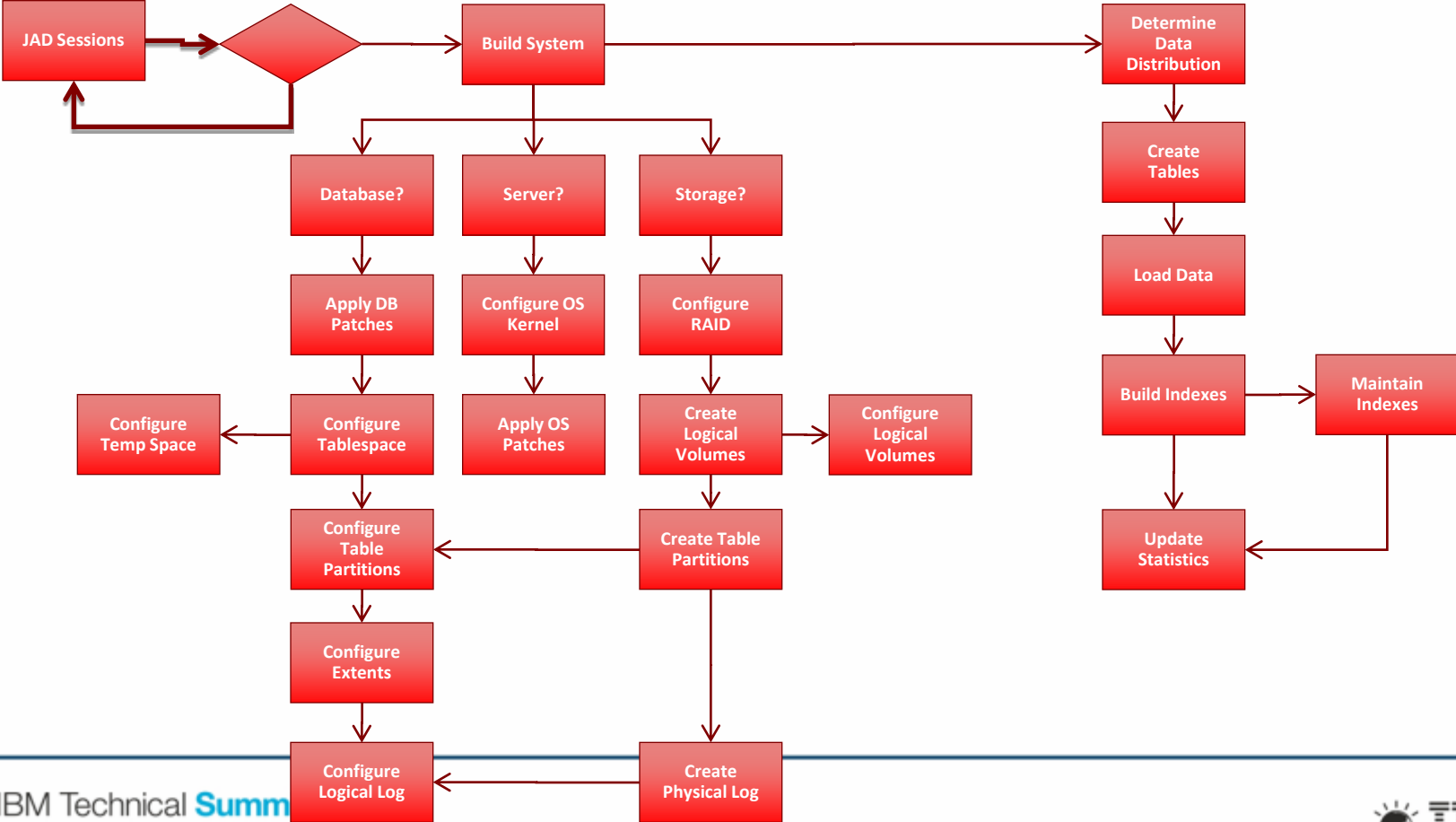
No software installation



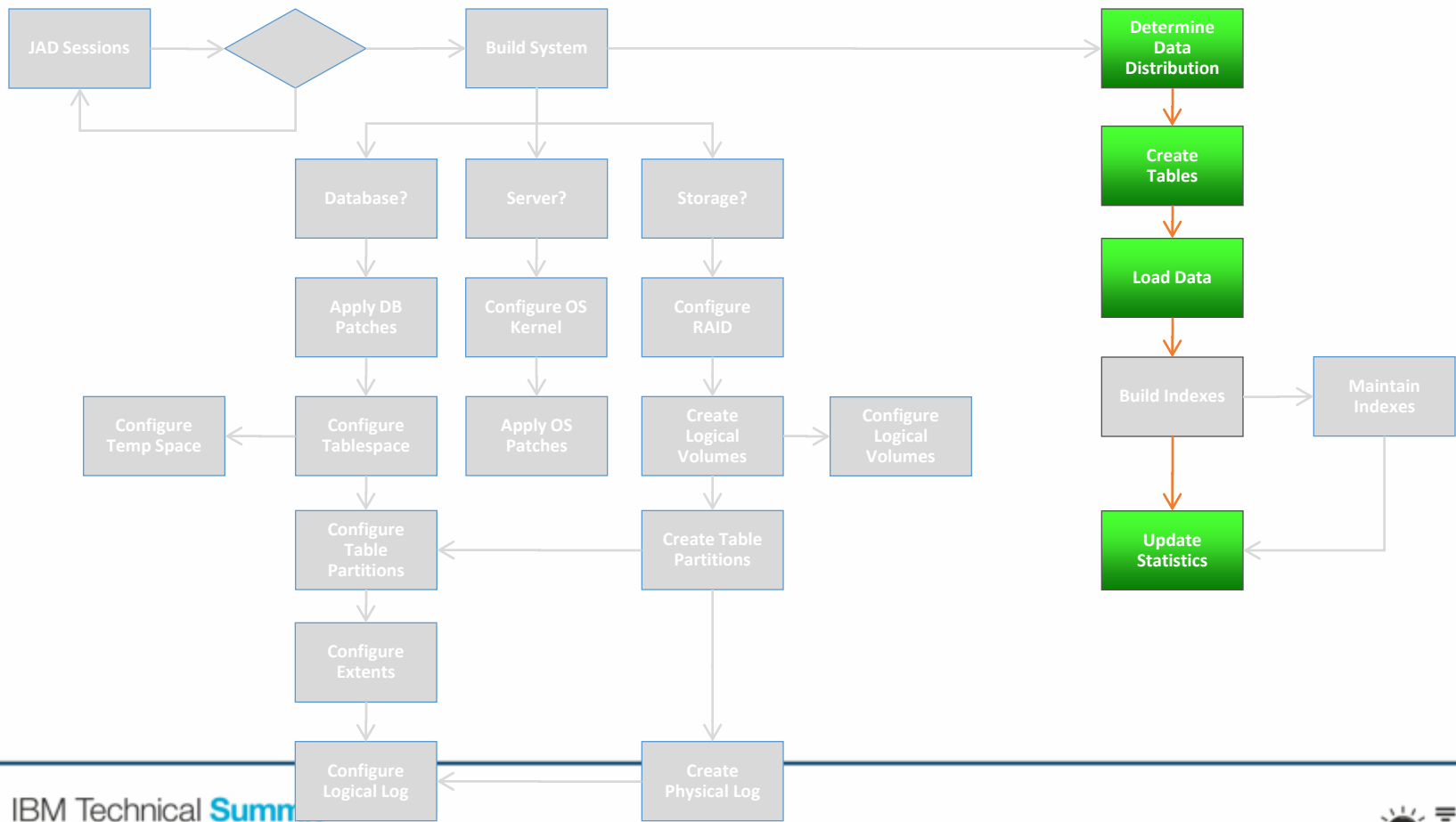
Data experts, not
database experts

“How is PureData for Analytics Simpler?”

Typical Warehouse Implementation Process



Simplicity over Complexity



ORACLE

```
CREATE TABLE "MRDWDDM"."RDWF_DDM_ROOMS_SOLD" ("ID_PROPERTY" NUMBER(5,0) NOT NULL ENABLE, "ID_DATE_STAY" NUMBER(5, 0) NOT NULL ENABLE, "CD_ROOM_POOL" CHAR(4) NOT NULL ENABLE, "CD_RATE_PGM" CHAR(4) NOT
```

ORACLE Indexes

```
NULL  
"CD_ CREATE INDEX "MRDWDDM"."RDWF_DDM_ROOMS_SOLD_IDX1" ON "RDWF_DDM_ROOMS_SOLD"  
NUMB ("ID_PROPERTY" , "ID_DATE_STAY" , "CD_ROOM_POOL" , "CD_RATE_PGM" ,  
NULL "CD_RATE_TYPE" , "CD_MARKET_SEGMENT" ) PCTFREE 10 INITRANS 6 MAXTRANS 255  
STORAGE( FREELISTS 10) TABLESPACE "DDM_DATAMART_INDEX_L" NOLOGGING
```

"ID_ PARALLI

ORACLE Bitmap index

```
NUMB INITRAN  
MAXEXTH CREATE BITMAP INDEX "CRDBO"."SNAPSHOT_MONTH_IDX13" ON  
ENAB DEFAULTS "SNAPSHOT_OPPTY_MONTH_HIST" ("SNAPSHOT_YEAR" ) PCTFREE 10 INITRANS 2  
"CU_ PCTFREQ  
MINEXTH MAXTRANS 255 STORAGE(INITIAL 4194304 NEXT 4194304 MINEXTENTS 2 MAXEXTENTS  
"QY_ MINEXTH  
1 BUFFE 2147483645 PCTINCREASE 0 FREELISTS 1 FREELIST GROUPS 1 BUFFER_POOL  
CHAR PARTIT  
STOR 4194304
```

FREELIS

ORACLE Table Clusters

```
PART "DDM_D  
THAN MAXTRAN  
STOR 100000  
TABLES  
"DDM_ PCTFREE 10 PCTUSED 90 INITRANS 3 MAXTRANS 255 STORAGE(INITIAL  
INITRAN 83886080 NEXT 41943040 MINEXTENTS 1 MAXEXTENTS 1017 PCTINCREASE 0  
LESS MAXEXTH  
STOR DEFAULTS  
PCTFREE 10 INITRANS 6 MAXTRANS 255 STORAGE(INITIAL 4194304 NEXT 4259840  
"DDM_ MINEXTE  
NTS 1 MAXEXTENTS 100000 PCTINCREASE 0 FREELISTS 10 FREELIST GROUPS
```

LESS THAN (1800) PCTFREE 5 PCTUSED 95 INITRANS 4 MAXTRANS 255

STORAGE(INITIAL 16777216 FREELISTS 6 FREELIST GROUPS 1) TABLESPACE

"DDM_ROOMS_SOLD_DATA" NOLOGGING NOCOMPRESS, PARTITION "PART4" VALUES

Netezza

```
CREATE TABLE MRDWDDM.RDWF_DDM_ROOMS_SOLD (  
ID_PROPERTY numeric(5, 0) NOT NULL ,  
ID_DATE_STAY integer NOT NULL ,  
CD_ROOM_POOL CHAR(4) NOT NULL ,  
CD_RATE_PGM CHAR(4) NOT NULL ,  
CD_RATE_TYPE CHAR(1) NOT NULL ,  
CD_MARKET_SEGMENT CHAR(2) NOT NULL ,  
ID_CONFO_NUM_ORIG integer NOT NULL ,  
ID_CONFO_NUM_CUR integer NOT NULL ,  
ID_DATE_CREATE integer NOT NULL ,  
ID_DATE_ARRIVAL integer NOT NULL ,  
ID_DATE_DEPART integer NOT NULL ,  
QY_ROOMS integer NOT NULL ,  
CU_REV_PROJ_NET_LOCAL numeric(21, 3) NOT  
NULL ,  
CU_REV_PROJ_NET_USD numeric(21, 3) NOT NULL  
,  
QY_DAYS_STAY_CUR smallint NOT NULL ,  
CD_BOOK_SOURCE CHAR(1) NOT NULL)
```

distribute on random;

•No indexes

•No Physical Tuning/Admin

•Stripe data randomly, or by Columns

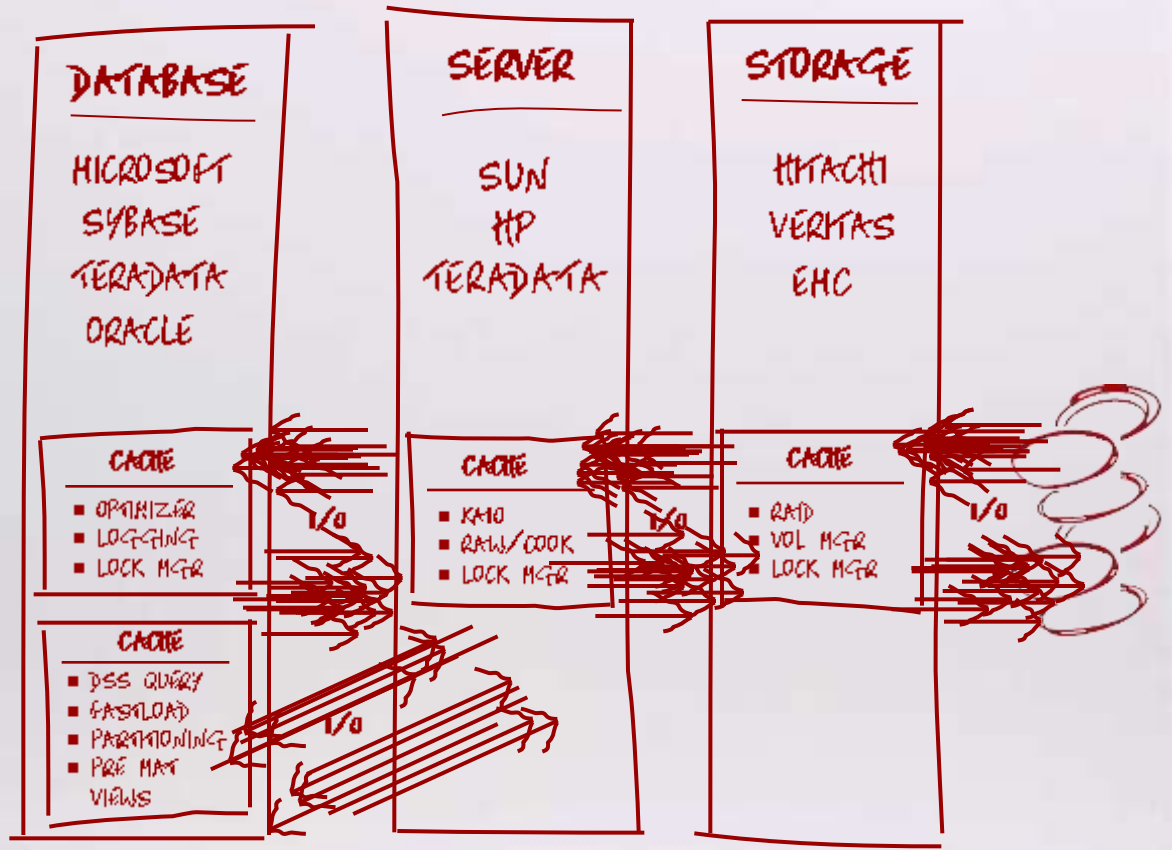
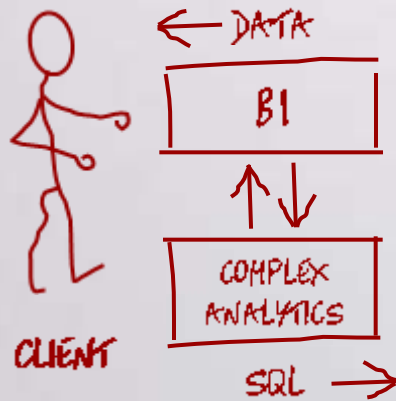
Telecom Call Detail Record FACT (6 billion rows)	Oracle Object Count*	Netezza Object Count
Tables	1	1
Indexes	12	
Table Partitions	47	
Index Partitions	564	
Table Partitions tablespaces	47	
Index Partitions tablespaces	47	
Table Data Files	170	
Index Data Files	122	
TOTAL	1,010	1

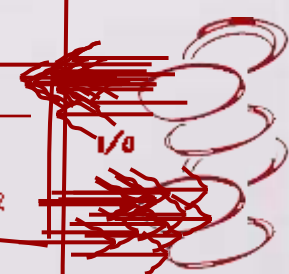
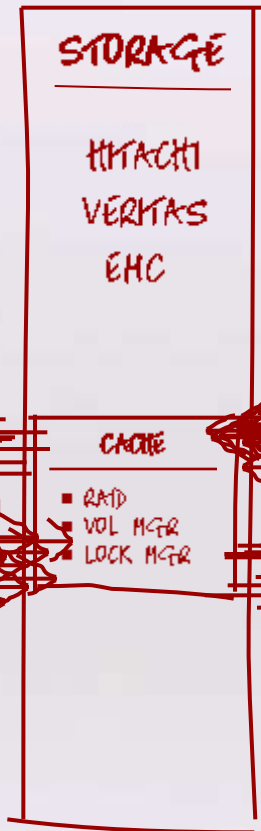
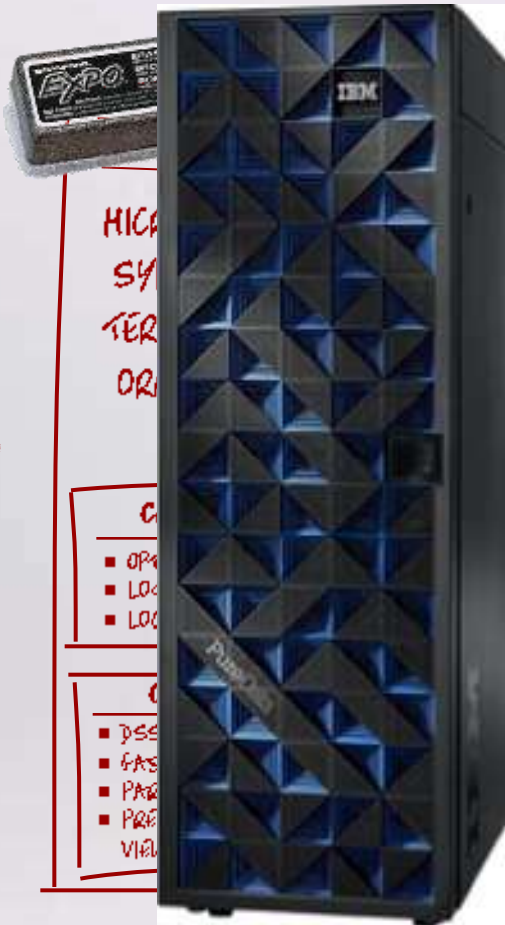
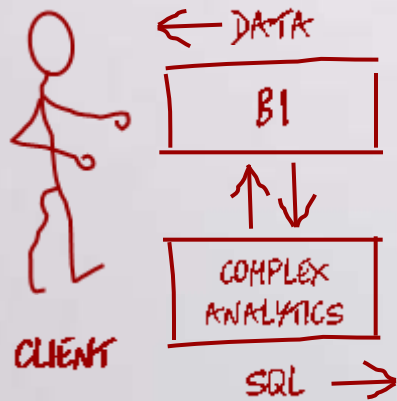
“Look at all the weeks/months worth of effort, DBA design and maintenance that we don't have with Netezza. The appliance claims are true.”

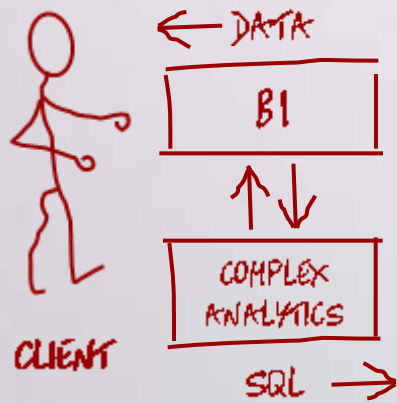
***: Oracle data does not account for ADDITIONAL effort required in configuring and engineering the file system design to accommodate this index management scheme.**



The PureData System for Analytics







LESS IS
MORE



OLE-DB

JDBC

ODBC

SQL



Extract / Load e.g.

INFORMATICA
EXPRESSOR™

Analytics e.g.

COGNOS®

Best in Business Intelligence™

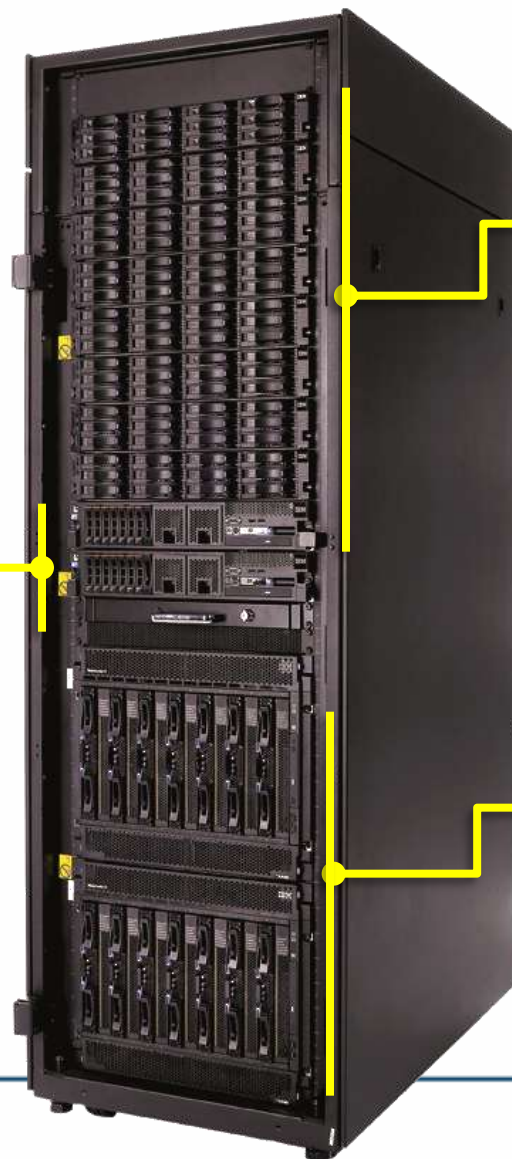
ORACLE®

Optimized Hardware + Software

- Hardware accelerated AMPP
- Purpose-built for high performance analytics
- Requires no tuning

SMP Hosts

- SQL Compiler
- Query Plan
- Optimize
- Admin

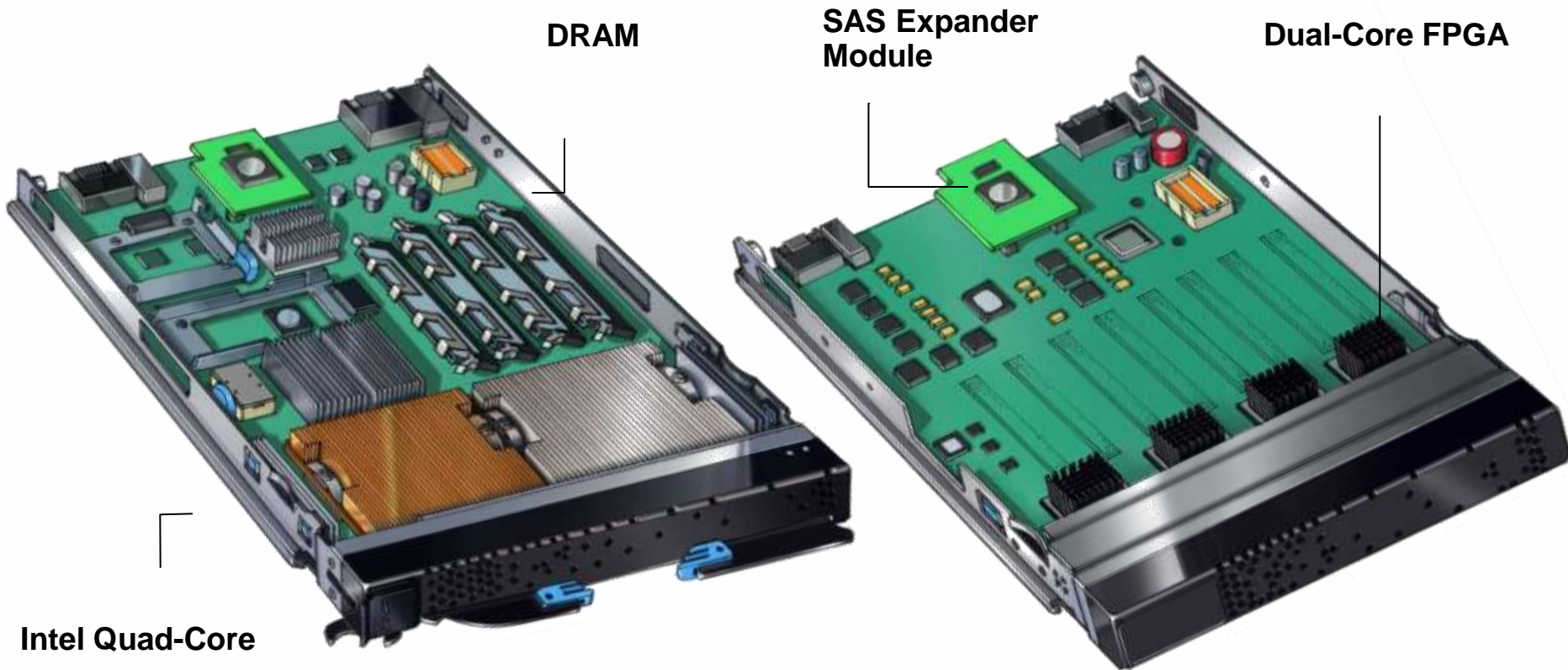


Disk Enclosures

- User data, mirror, swap partitions
- High speed data streaming

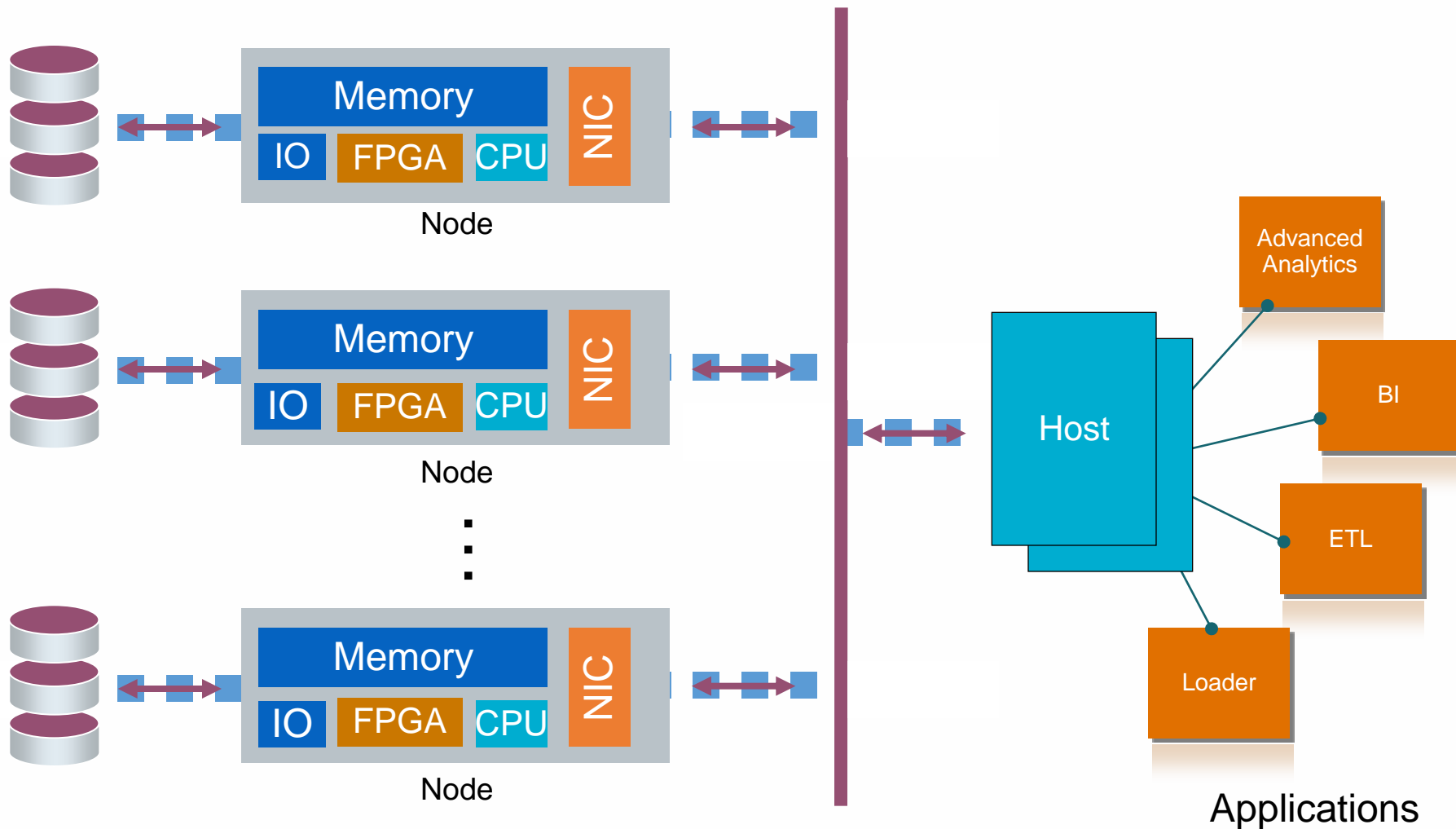
Snippet Blades™

- Hardware-based query acceleration with FPGAs
- Blistering fast results
- Complex analytics executed as the data streams from disk



IBM BladeCenter Server

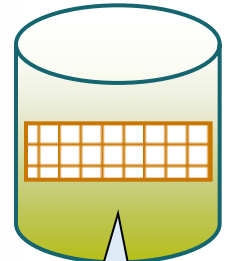
Netezza DB Accelerator



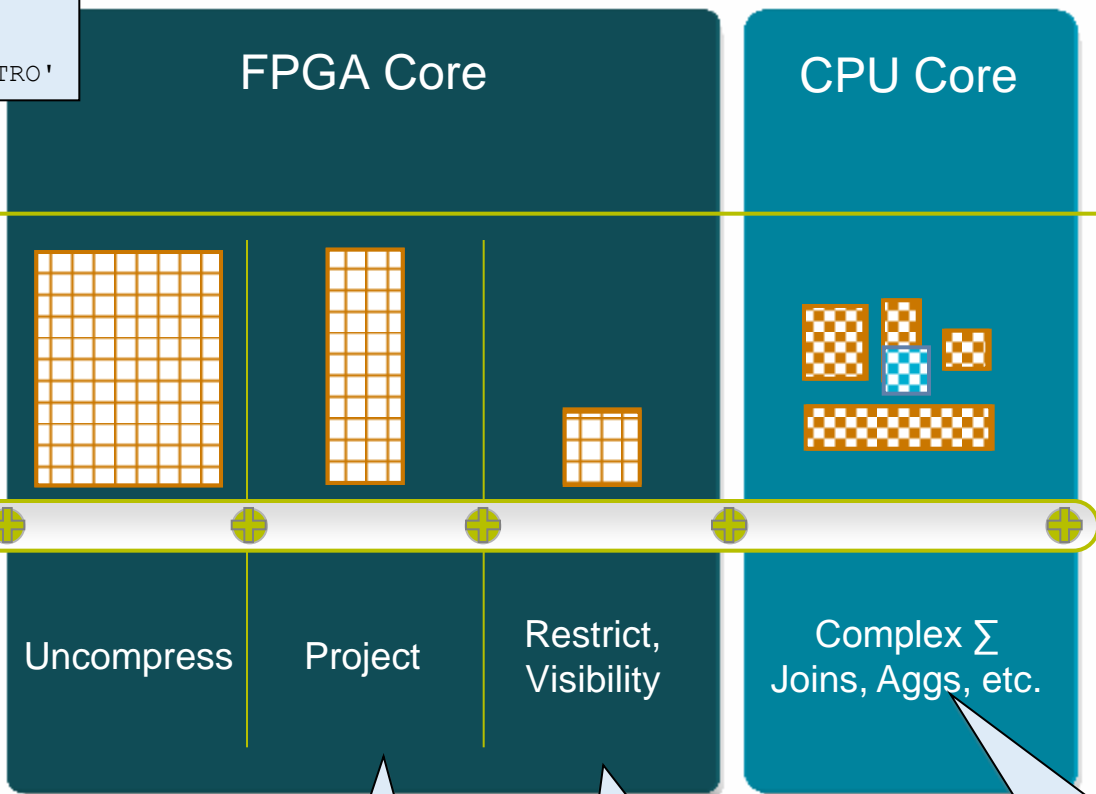
```

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

```



Slice of table
MTHLY_RX_TERR_DATA
(compressed)



```

select DISTRICT,
       PRODUCTGRP,
       sum(NRX)

```

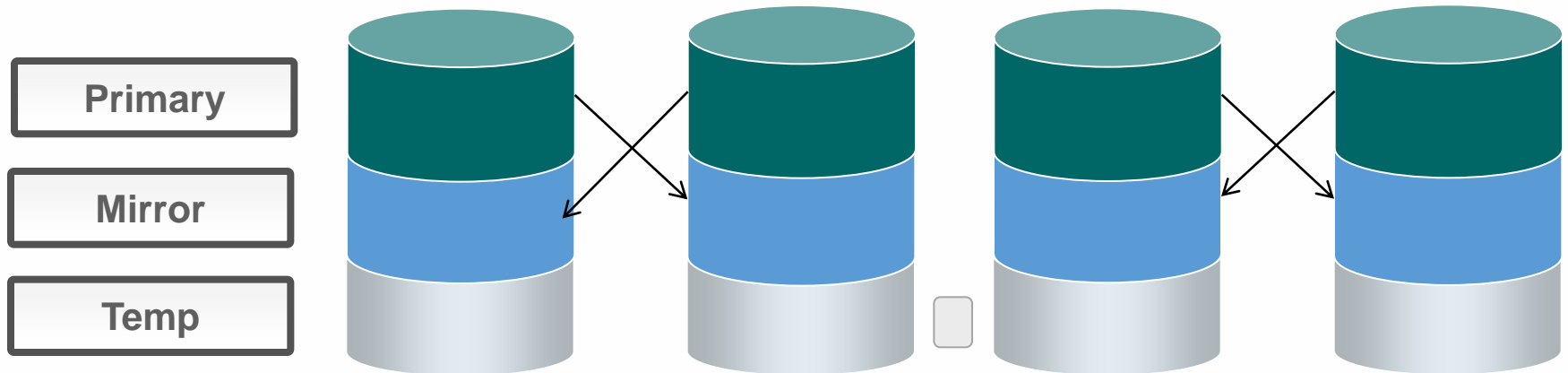
```

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

```

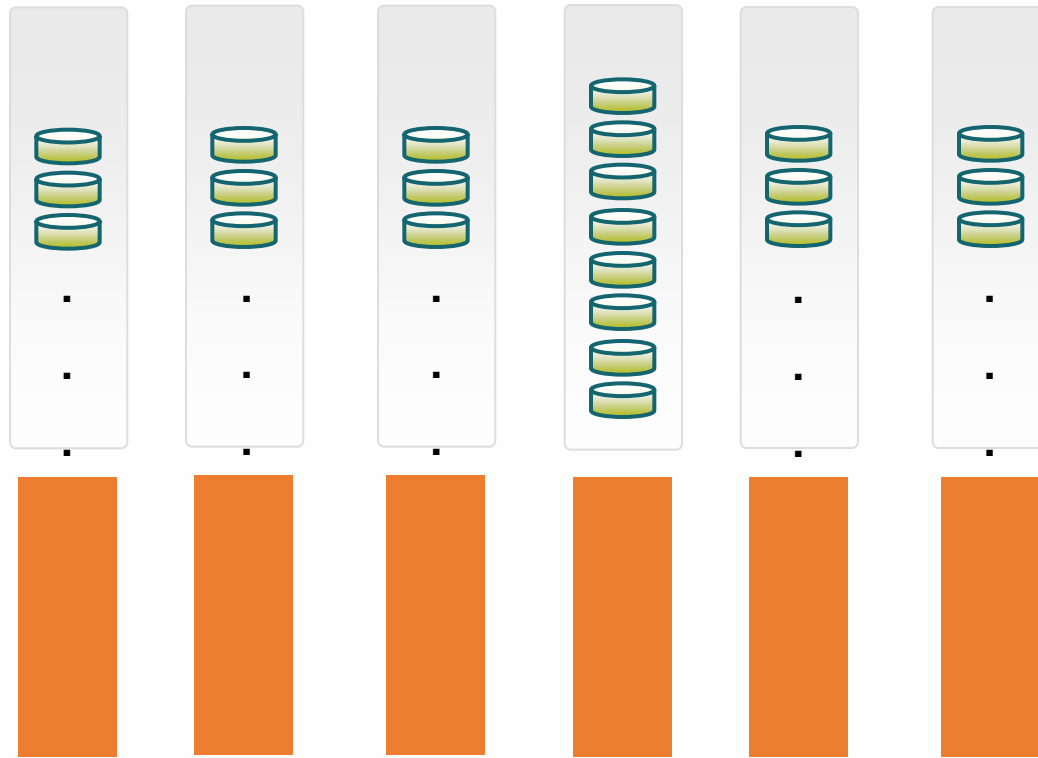
sum(NRX)

Disk Mirroring and Failover



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

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



- 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

- 14 PureData for Analytics S-Blades™
- 2 Intel Quad-Core 2+ GHz CPUs
- 4 Dual-Engine 125 MHz FPGAs
- 24 GB DDR2 RAM
- Linux 64-bit Kernel

Scales from
¼ Rack to 10 Racks

32 TB to 1.2 PB of
User Data

- User Data Capacity: 128 TB**
- Data Scan Speed: 145 TB/hr**
- Load Speed (per system): 5+ TB/hr

- Power Requirements: 7.6 kW
- Cooling Requirements: 7.8 kW

IBM Technical Summit **** 4X compression assumed**

IBM PureData and Netezza Data Warehouse Appliance Family



IBM Netezza 100

Development & Test System

1 TB to 10 TB



**IBM PureData System
for Analytics**

Data Warehouse
High-Performance Analytics

1 TB to 1.5 PB



...

1/4 Rack 1/2 Rack 1 Rack 2 Racks 3 Racks 10 Racks

S-Blades 4 7 14 28 42 140

Snippet Processors 32 56 112 224 336 1120

Capacity Raw (TB) 8 16 32 64 96 320

Compressed (TB) 32 64 128 256 384 1280

Netezza In-Database Analytics

Transformations

Mathematical

Geospatial

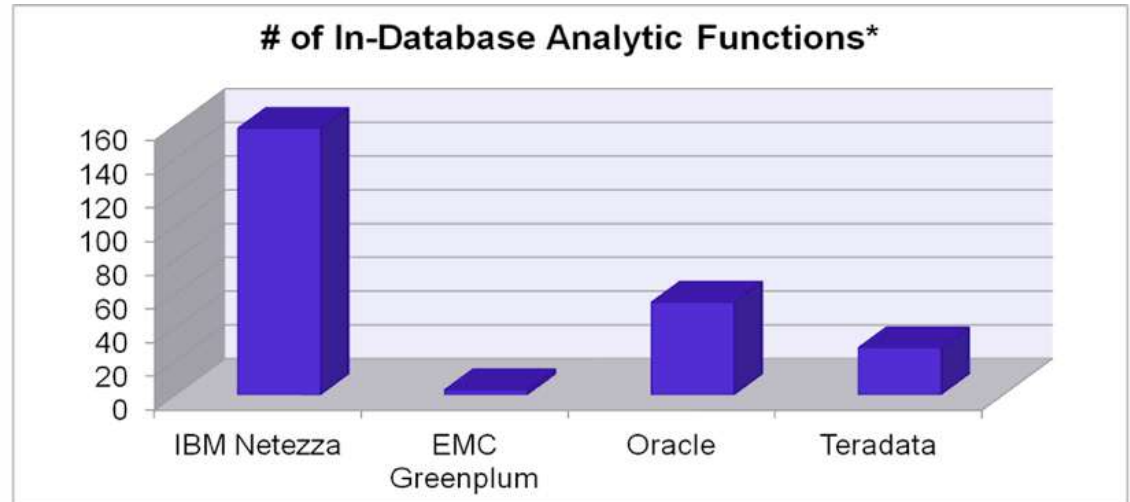
Predictive

Statistics

Time Series

Data Mining

The MOST in-Database Analytic Functions



**Approximate number of analytic functions based on competitors' documentation
Does not include matrix or spatial functions. With spatial and matrix, INZA has over 200 functions.*

- ✓ No data movement
- ✓ Analyze deep and wide data
- ✓ High performance, parallel computation

Analytic Applications

BI / Reporting | Exploration / Visualization | Functional App | Industry App | Predictive Analytics | Content Analytics

IBM Big Data Platform

Visualization & Discovery

Application Development

Systems Management



Accelerators

Hadoop System



Stream Computing



Data Warehouse



Information Integration & Governance

*PureData
System for Analytics*



Disclaimer

The information contained in this presentation is provided for informational purposes only.

While efforts were made to verify the completeness and accuracy of the information contained in this presentation, it is provided “As Is”, without warranty of any kind, express or implied.

In addition, this information is based on IBM’s current product plans and strategy, which are subject to change by IBM without notice.

IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this presentation or any other documentation.

Nothing contained in this presentation is intended to, or shall have the effect of:

Creating any warranty or representation from IBM (or its affiliates or its or their supplies and/or licensors); or

Altering the terms and conditions of the applicable license agreement governing the use of IBM software.