Information Management

IBM

From Small to Internet Scale Analytics, New Approach: Appliance, Fast Insight, Less Complexity.

James Dung

Technical Executive, Information Management

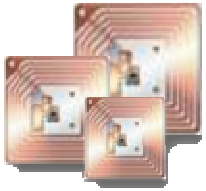
IBM Software Group, AP

November 11, 2011

Information is at the center of a new wave of opportunity

2 Billion Internet users by 2011

1.3 Billion RFID tags in 2005
30 Billion RFID tags by 2010



4.6 Billon Mobile Phones World Wide

Capital market data volumes grew *1,750%* 2003-06

twitter

Twitter process *7 terabytes* of data every day

World Data Centre for Climate
220 Terabytes of Web data
9 Petabytes of additional data

facebook

Facebook process *10 terabytes* of data every day

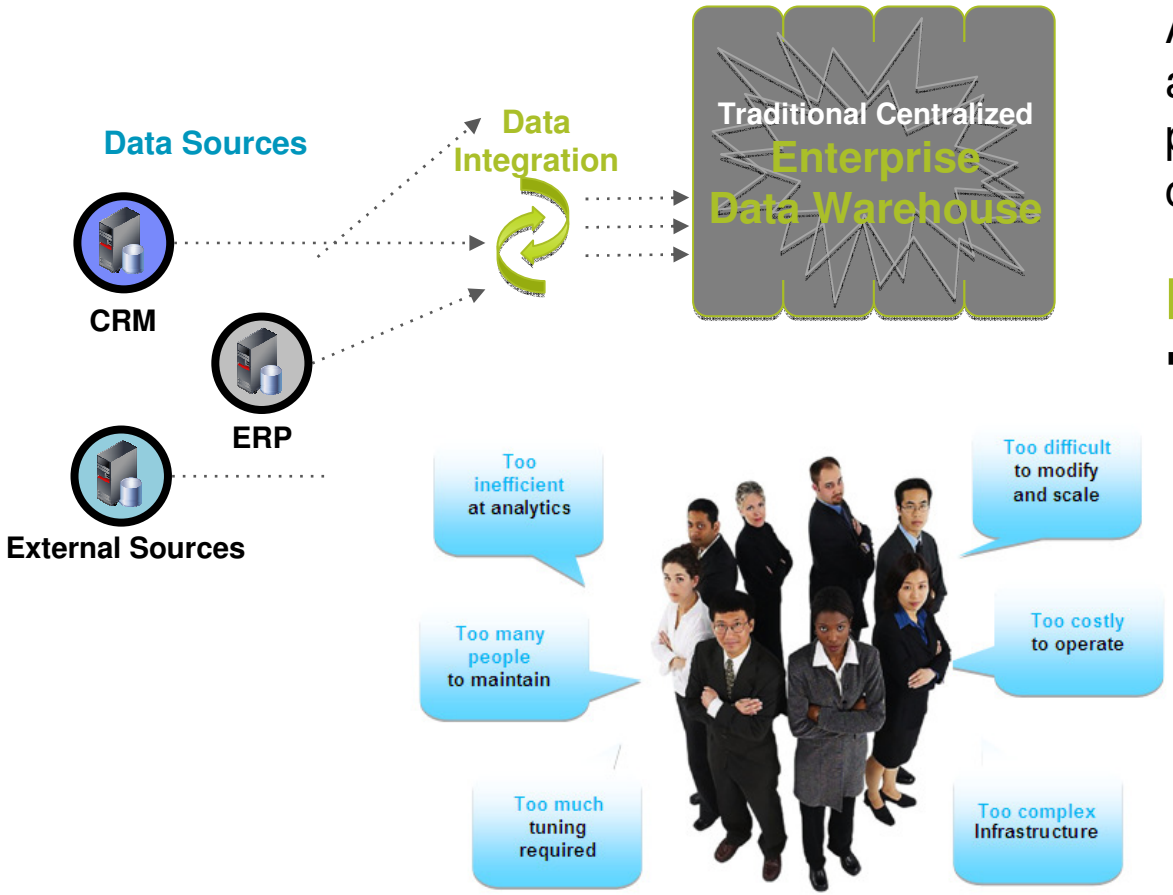
Trends in Data Warehousing and Business Intelligence

- ✓Agile BI
- ✓Self Service BI
- ✓Mobile BI
- ✓Big Data
- ✓Deep Analytics
- ✓Growth of Appliances

Leveraging Data and Information is a business imperative

IBM Software
Information On Demand

The Enterprise Data Warehouse Challenge Today



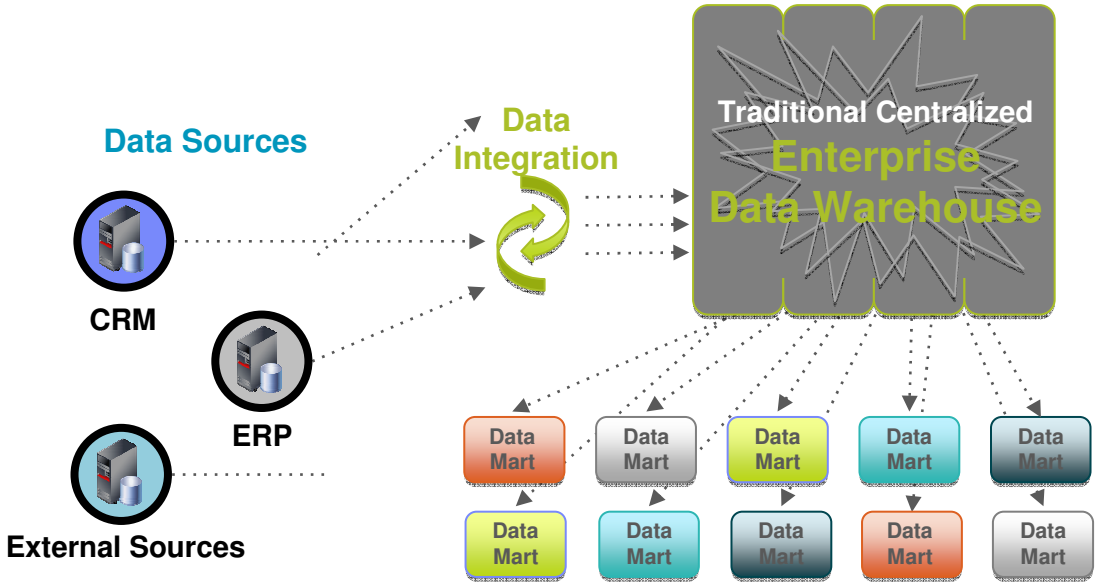
Vision

All enterprise data storage, analytic and operational processing takes place in one central data warehouse

Harsh Reality

- Single EDW cannot handle today's volume, velocity and variety of data and workloads
  Lack of agility, increasing latency
  Business needs not being met

The Enterprise Data Warehouse Challenge
So What is The Reality?



Reaction

Lines of business resort to ad hoc solutions creating sprawl

Result

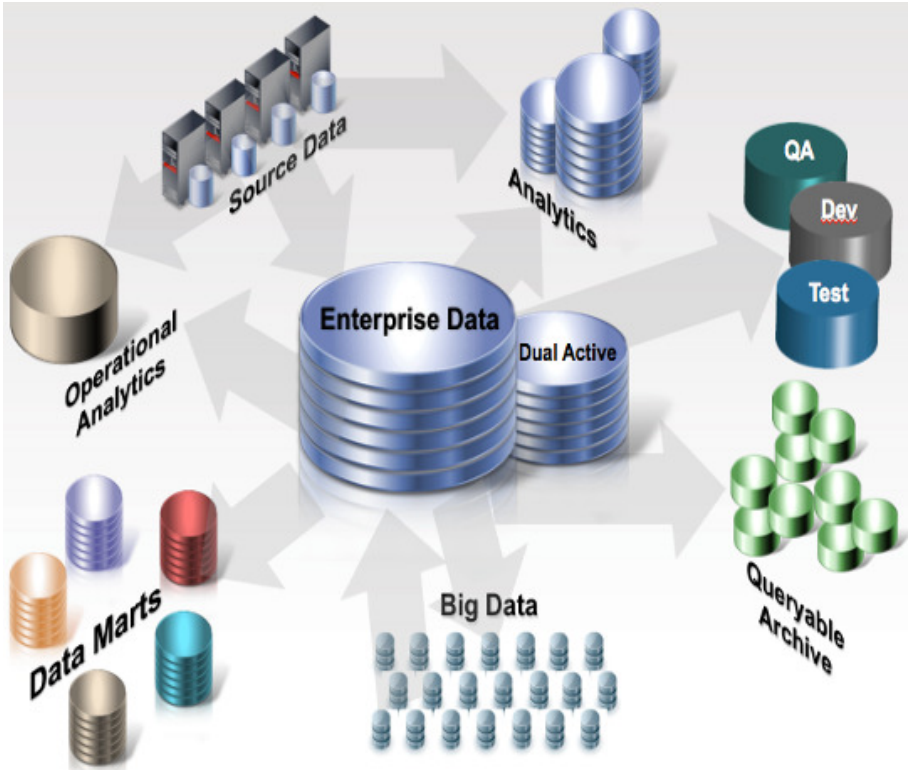
- Data silos limit enterprise-wide analytics and visibility
- Lack of true governance
- Increased strain on EDW, shorter lifespan
- Inability to scale
- Escalating complexity, inefficiency and cost

Smart Consolidation for Smarter Warehousing

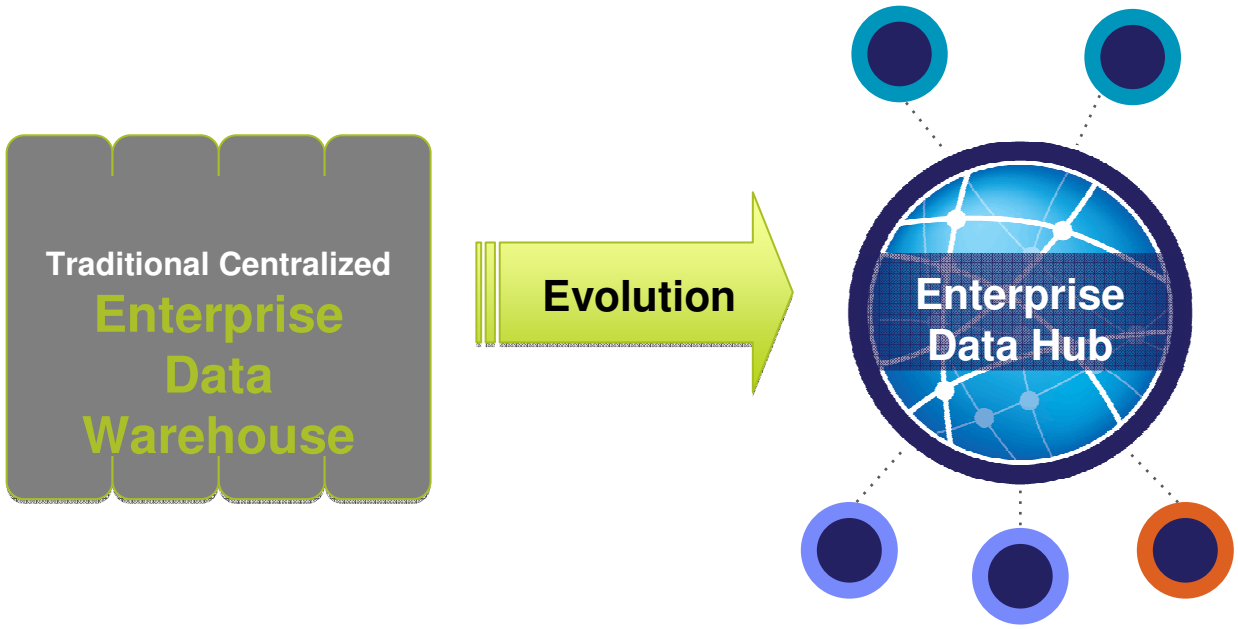
Break Free from Complexity

Issues with the Traditional Approach

- **Multiple workloads and types of queries**
- **Complications in performance**
- **Lack of agility**
- **Lack of responsiveness to the business**
- **Unprepared to handle Big Data**



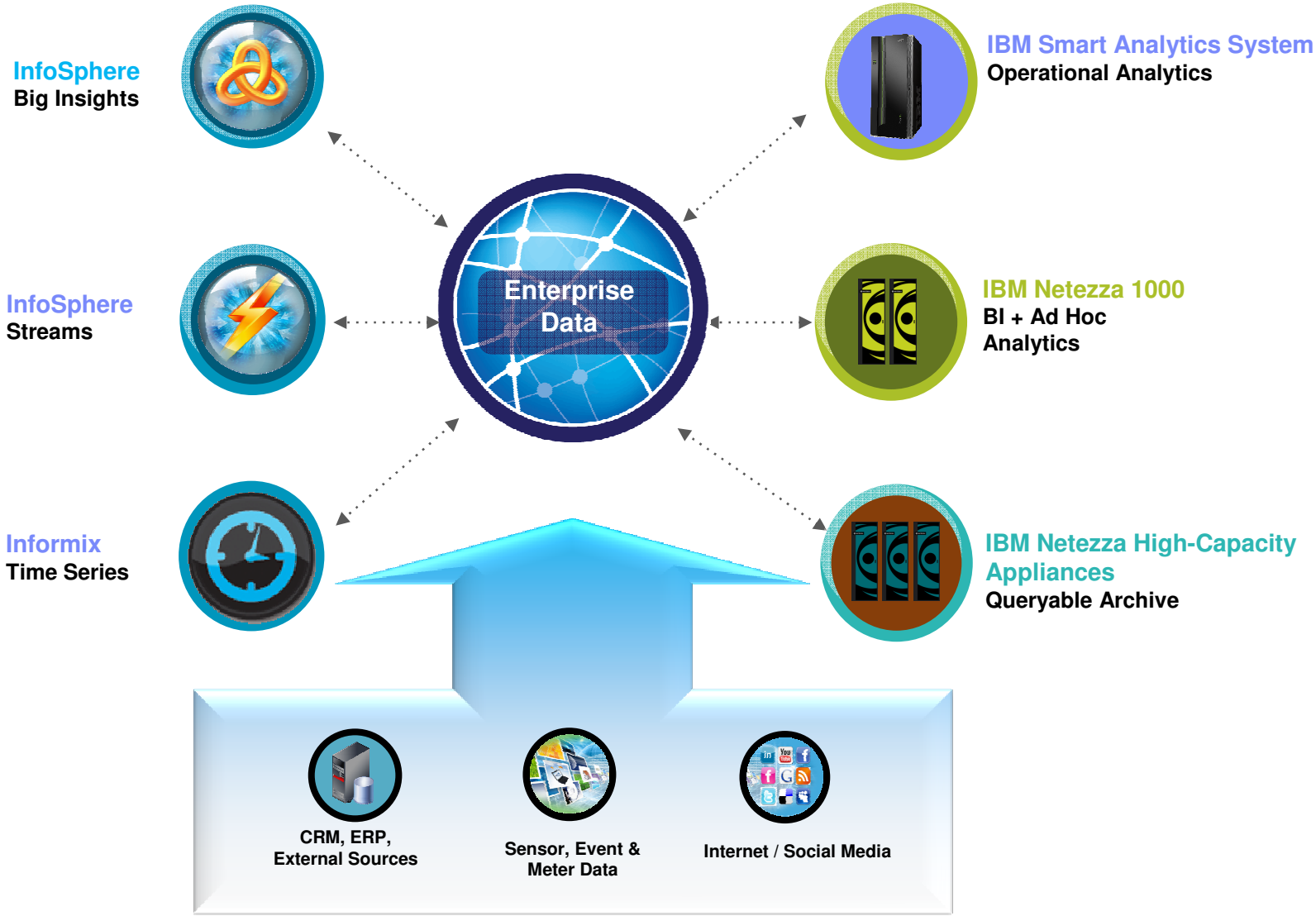
The New Concept of "Smart EDW Consolidation" Evolving to a Logical Data Warehouse



How is this different?

- *Enterprise Data Warehouse evolves from "doing it all"*
  - Analytical work is offloaded and optimized
  - Cold data is offloaded to maintain costs
  - Operational BI is offloaded and optimized
  - Big Data augments the capability
- *The Logical Data Warehouse Components*
  - Enterprise Data – either virtual or physical
  - Analytical systems (operational and deep analytics)
  - Data Lifecycle Management
  - Governance

The New Concept of "Logical EDW" Building Blocks

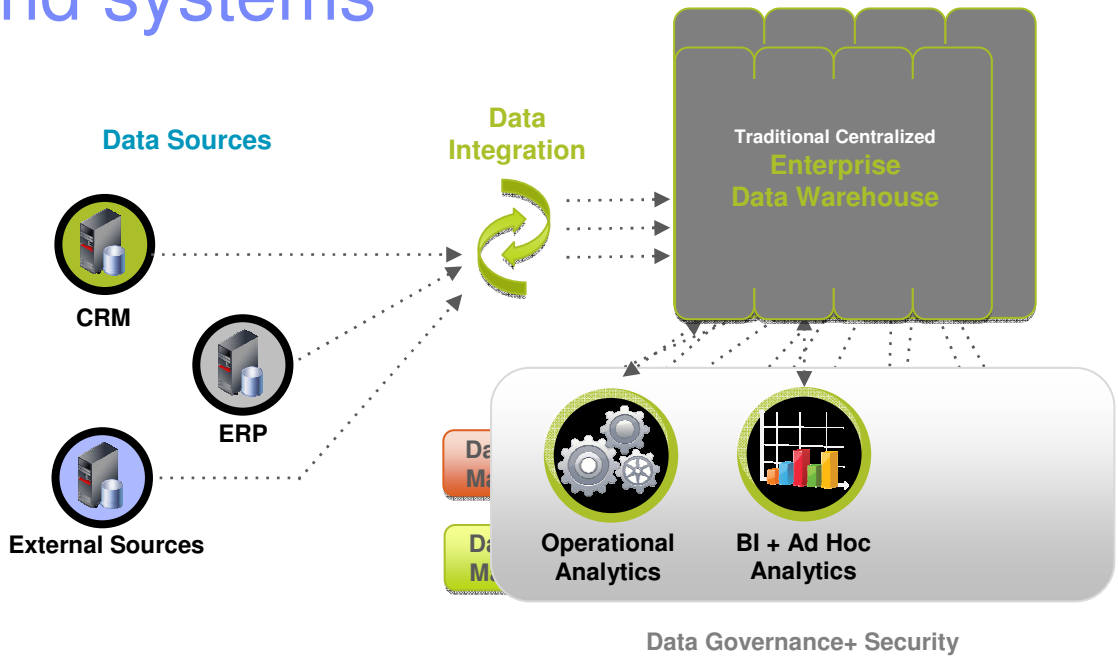


Components or Entry Points in an LDW

Workload Type	Description
ETL/ELT/Data Integration	**Data staging, bulk and trickle-feed data loading, ETL, ELT.**
Data Governance	**Master data management (MDM), changed data capture (CDC), data quality (DQ), etc.**
Operational Intelligence	**Low latency real-time query and Operational BI support; BI reporting and dashboard updating.**
Complex Event Processing	**Real-time event processing for data compliance, data security, fraud detection, etc.**
Analytics/Advanced Analytics	**Light-to-moderate or heavy decision support, data mining, complex in-database analytics.**
Line of Business Marts/Warehouses	**Data warehouse appliances for specific LoB applications—retail analytics, ERP, etc.**
Big Data Processing	**InfoSphere Big Insights(Hadoop) grid to analyze massive unstructured data sets.**
Real-Time Analytics (Big Data)	**InfoSphere Streams system for high-volume stream capture and analysis.**
Time Series Data Processing	**Informix TimeSeries for optimized storage and processing of time series / time interval data**
Queryable Archiving	**High-capacity federated storage for data to which future or intermittent access is required.**
Backup/Recovery	**High-capacity, write-only systems for non-queryable archiving and/or disaster recovery.**
Exploration Sandbox	**Replicated data for use in data exploration, non-production analytics.**
Test/Dev/Prototyping	**Non-production systems for application development, prototyping, and testing.**
Short-Request/Transactional	**OLTP or other short-request query activity.**

Smart Consolidation 1:

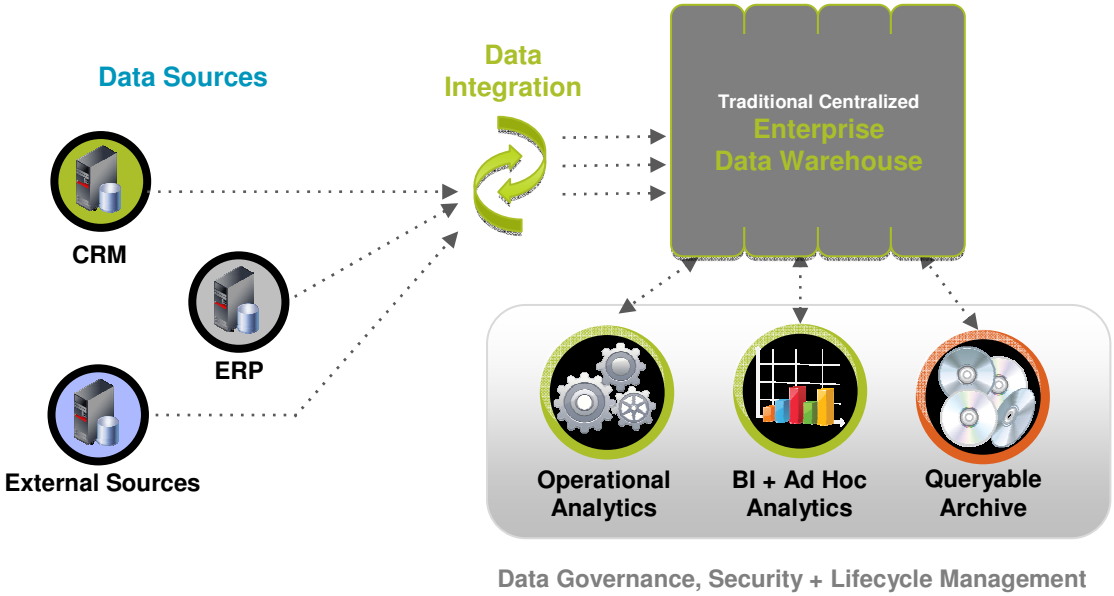
Consolidate infrastructure with purpose-built appliances and systems



- Reduce data mart sprawl
- Offload analytics from the EDW to appliances optimized for performance
- True data governance

Consolidate and Optimize Analytic Systems for a Quick Business Win

Smart Consolidation 2:
Optimize infrastructure through Lifecycle Management

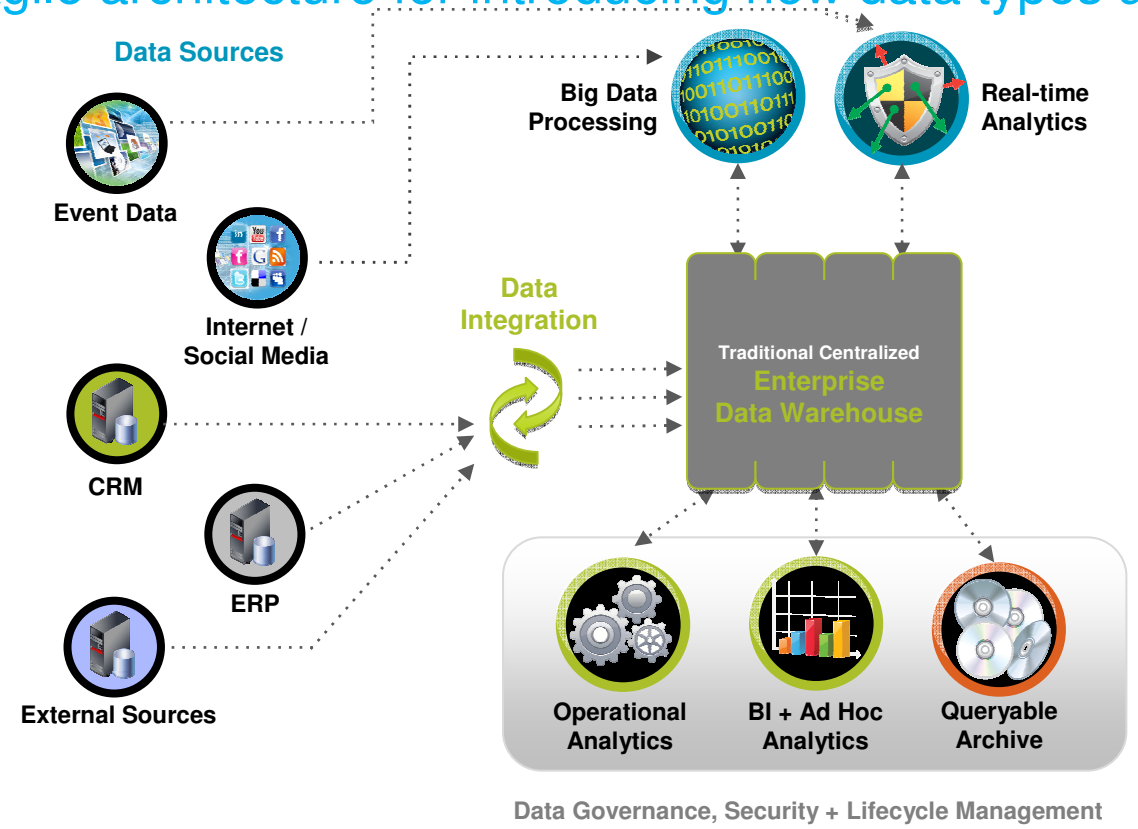


- Queryable Archive for cost-effective querying and analytics on massive data
- Increase accessibility of information
- Lower cost of ownership
- Reduce stress on EDW

Introduce Lifecycle Management to Lower Cost of Ownership

Smart Consolidation 3:

Agile architecture for introducing new data types and analytic models

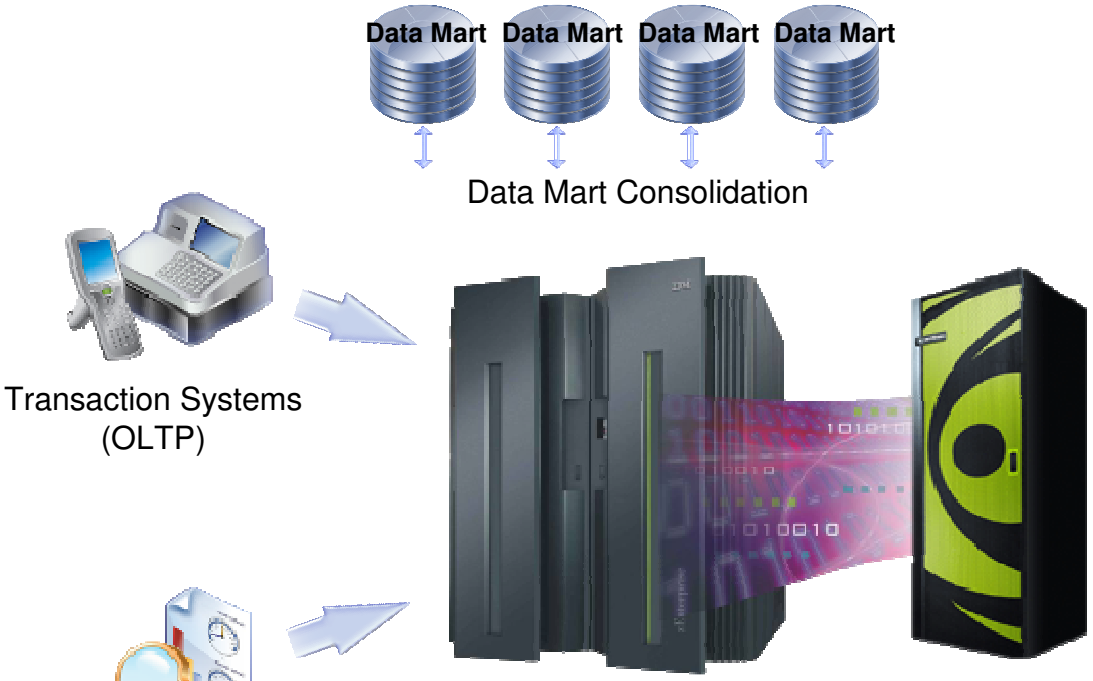


- Extend data warehouse by adding big data and real-time analytic processing
- Analytics on new data types
- Optimized support for different types of compute

Extend the Warehouse with Big Data

DB2 Analytics Accelerator for zOS

Combining the best transaction system with the best analytics system



z/OS:

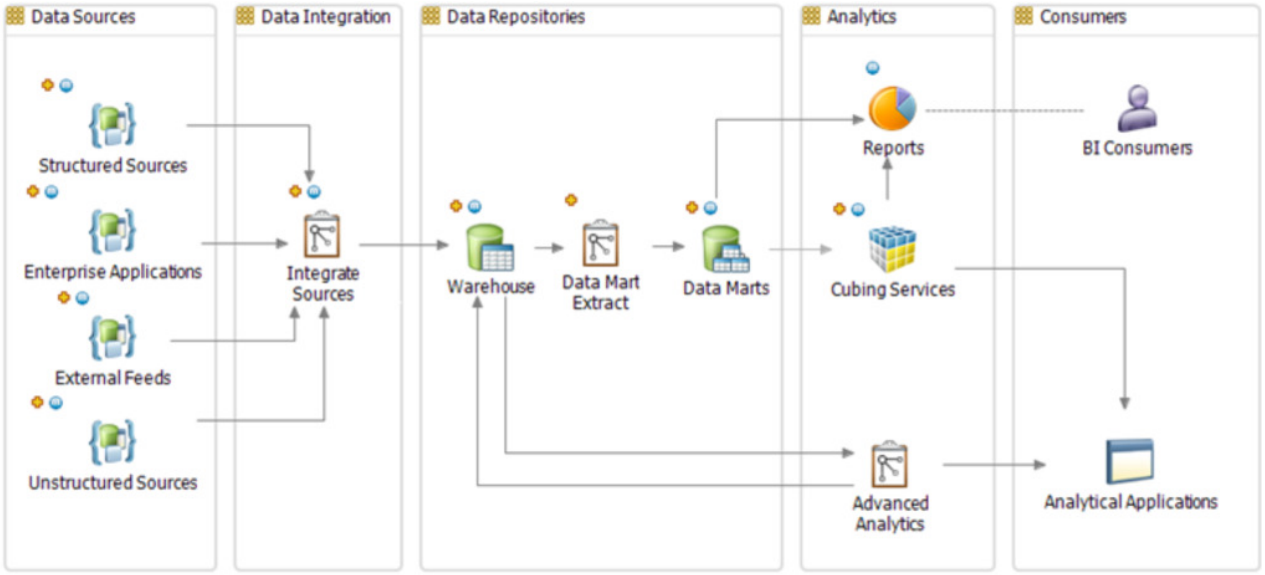
Recognized leader in mixed workloads with security, availability and recoverability for OLTP

Netezza:

Recognized leader in cost-effective high speed deep analytics

Integrate all Components Using Policy Based Management Simplified Tooling

Vision • Execution • Completion



- Achieving the right mix of simplicity and flexibility through a common design perspective (InfoSphere Blueprint)
- Common end-to-end project vision, considering
- Guided implementation
- Control and insight of the information roadmap and its evolution through a collaborative project lifecycle

IBM's Logical Data Warehouse Architecture

Business Benefits: Smarter Warehousing & Analytics

- **Business agility, performance and time-to-value**
- **Scales and adapts to changing requirements**
- **Balances consolidated, purpose-built approach**
- **Maintains governance while controlling costs**
- **Simplifies delivery and operations**
- **Flexible – multiple entry points**

Smart Consolidation for Smarter Warehousing
White Paper: http://bit.ly/oFuh71
Web:http://thinking.netezza.com/

How does Netezza Achieve Appliance Simplicity

How a purpose-built appliance for high-speed data warehousing can enable blazingly fast analytics (The True Appliances)

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

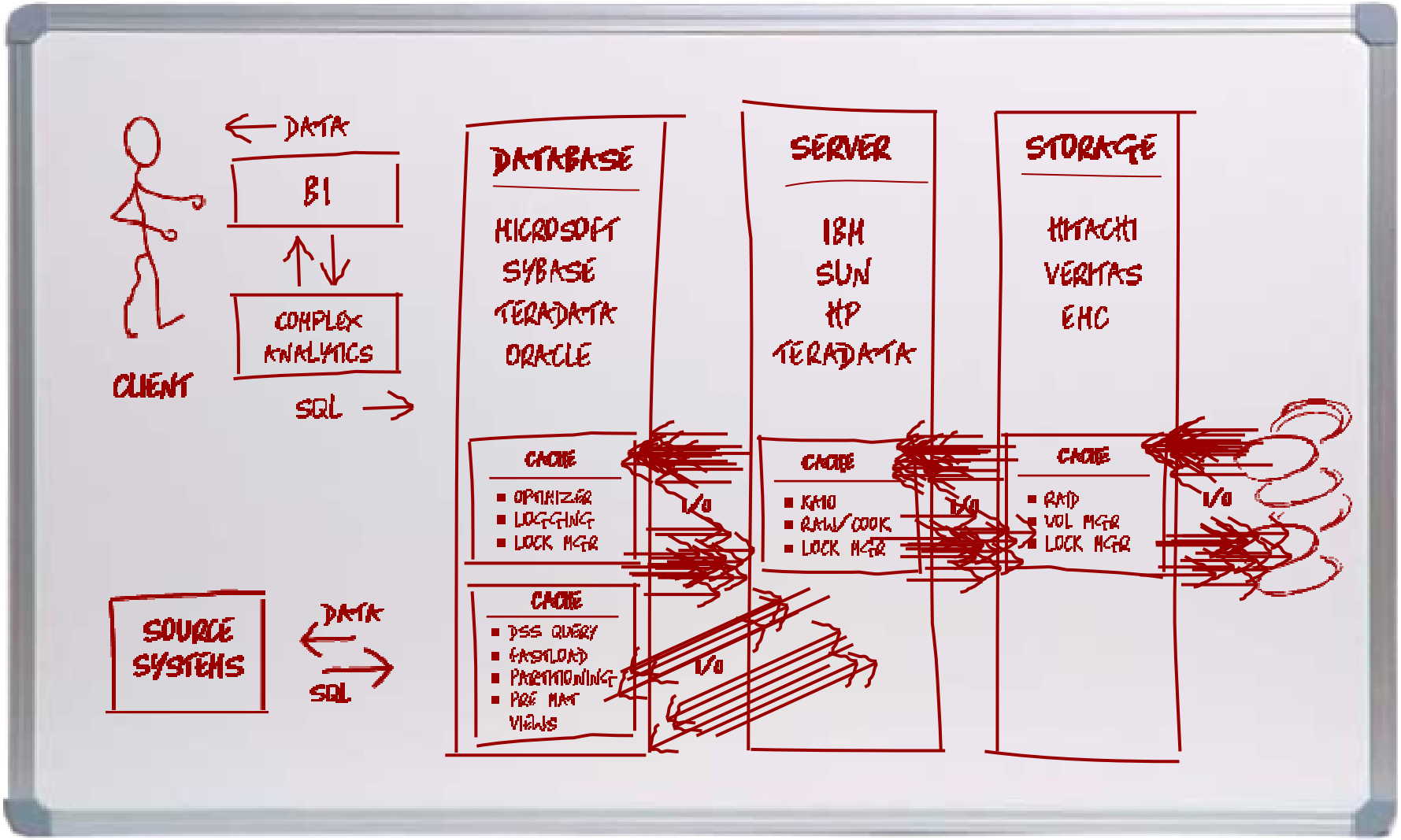
Revolutionizing Analytics

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

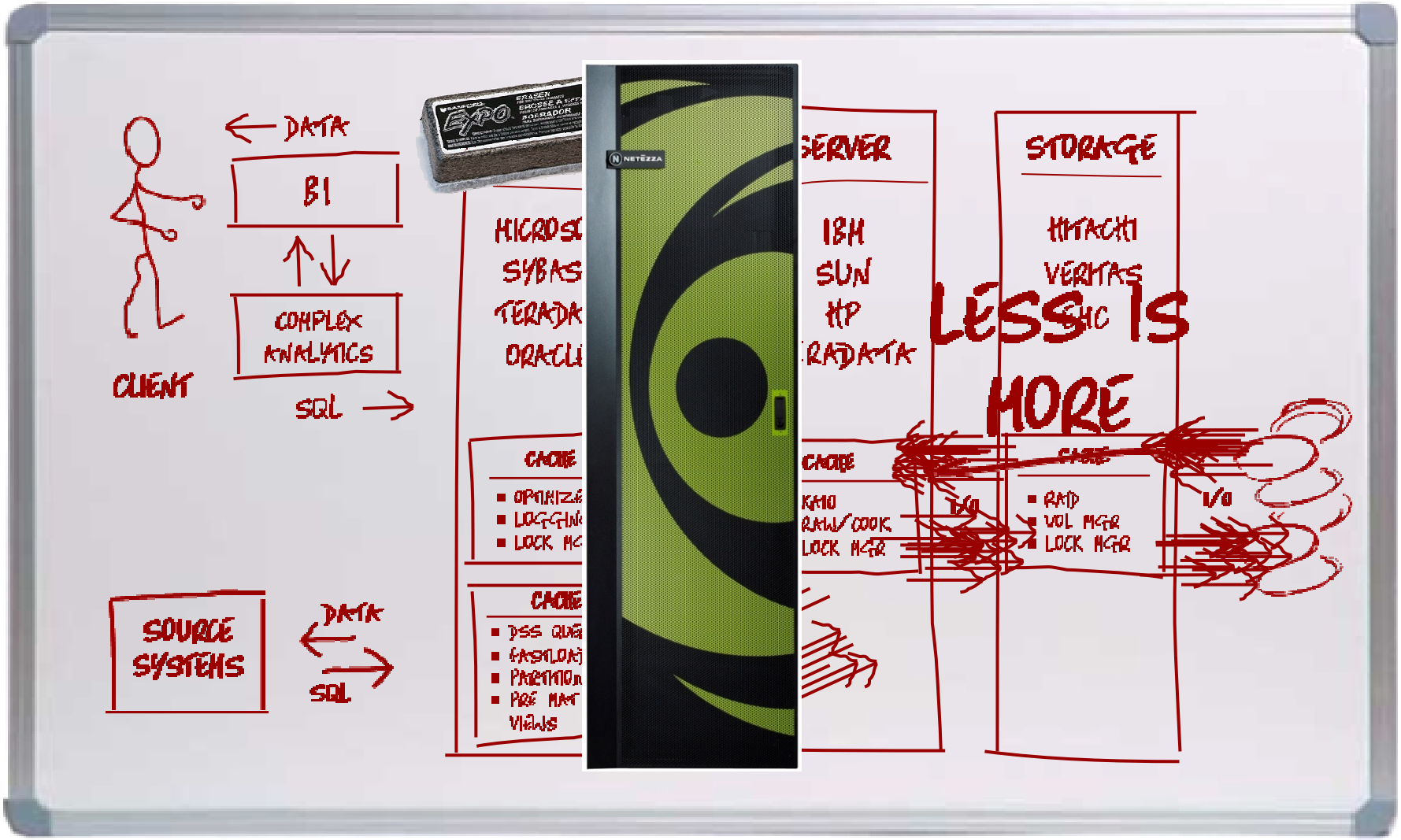
- **Speed**: 10-100x faster than traditional systems
- **Simplicity**: Minimal administration and tuning
- **Scalability**: Peta-scale user data capacity
- **Smart**: High-performance advanced analytics

How does it work to achieve Appliance Simplicity

Traditional Data Warehouse Complexity



Data Warehousing – Simplified



Managing The Netezza Appliance

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

Resources become Data Managers instead of Database Administrators

Traditional Complexity … Netezza Simplicity

0. CREATE DATABASE TEST LOGFILE 'E:\OraData\TEST\LOG1TEST.ORA' SIZE 2M, 'E:\OraData\TEST\LOG2TEST.ORA' SIZE 2M, 'E:\OraData\TEST\LOG3TEST.ORA' SIZE 2M, 'E:\OraData\TEST\LOG4TEST.ORA' SIZE 2M, 'E:\OraData\TEST\LOG5TEST.ORA' SIZE 2M EXTENT MANAGEMENT LOCAL MAXDATAFILES 100 DATAFILE 'E:\OraData\TEST\SYS1TEST.ORA' SIZE 50 M DEFAULT TEMPORARY TABLESPACE temp TEMPFILE 'E:\OraData\TEST\TEMP.ORA' SIZE 50 M

UNDO TABLESPACE undo DATAFILE 'E:\OraData\TEST\UNDO.ORA' SIZE 50 M NOARCHIVELOG CHARACTER SET WE8ISO8859P1;

1. Oracle* table and indexes
2. Oracle tablespace
3. Oracle datafile
4. Veritas file
5. Veritas file system
6. Veritas striped logical volume
7. Veritas mirror/plex
8. Veritas sub-disk
9. SunOS raw device
10. Brocade SAN switch
11. EMC Symmetrix volume
12. EMC Symmetrix striped meta-volume
13. EMC Symmetrix hyper-volume
14. EMC Symmetrix remote volume (replication)
15. Days/weeks of planning meetings

Netezza: Low (ZERO) Touch:

CREATE DATABASE my_db;

Traditional Complexity ... Netezza Simplicity

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
NULL ENABLE, "CD_RATE_TYPE" CHAR(1) NOT NULL ENABLE,
"CD_MARKET_SEGMENT" CHAR(2) NOT NULL ENABLE, "ID_CONFO_NUM_ORIG"
NUMBE
NULL
"ID_D
NUMBE
ENABL
"CU_R
"QY_D
CHAR(
STORA
PARTI
THAN
STORA
"DDM_
LESS
STORA
"DDM_
LESS
STORA
"DDM_
LESS
STORA
"DDM_
LESS
STORA
"DDM_
LESS
STORAGE(INITIAL 16777216 FREELISTS 6 FREELIST GROUPS 1) TABLESPACE
"DDM_ROOMS_SOLD_DATA" NOLOGGING NOCOMPRESS ) ;
```

ORACLE Indexes

```
CREATE INDEX "MRDWDDM"."RDWF_DDM_ROOMS_SOLD_IDX1" ON "RDWF_DDM_ROOMS_SOLD"
("ID_PROPERTY" , "ID_DATE_STAY" , "CD_ROOM_POOL" , "CD_RATE_PGM" ,
"CD_RATE_TYPE" , "CD_MARKET_SEGMENT" ) PCTFREE 10 INITRANS 6 MAXTRANS 255
STORAGE( FREELISTS 10) TABLESPACE "DDM_DATAMART_INDEX_L" NOLOGGING
PARALLE
INITRA
MAXEXT
DEFAULT
PCTFREE
MINEXTE
1 BUFFE
PARTITI
4194304
FREELIS
"DDM_DA
MAXTRA
100000
TABLESP
INITRA
MAXEXT
DEFAULT
PCTFREE 10 INITRANS 6 MAXTRANS 255 STORAGE(INITIAL 4194304 NEXT 4259840
MINEXTENTS 1 MAXEXTENTS 100000 PCTINCREASE 0 FREELISTS 10 FREELIST GROUPS
1 BUFFER_POOL DEFAULT) TABLESPACE "DDM_DATAMART_INDEX_L" NOLOGGING ) ;
```

ORACLE Bitmap index

```
CREATE BITMAP INDEX "CRDBO"."SNAPSHOT_MONTH_IDX13" ON
"SNAPSHOT_OPPTY_MONTH_HIST" ("SNAPSHOT_YEAR" ) PCTFREE 10 INITRANS 2
MAXTRANS 255 STORAGE(INITIAL 4194304 NEXT 4194304 MINEXTENTS 2 MAXEXTENTS
2147483645 PCTINCREASE 0 FREELISTS 1 FREELIST GROUPS 1 BUFFER_POOL
DEFAULT) TABLESPACE "SFA_DATAMART_INDEX" NOLOGGING ;
```

ORACLE Table Clusters

```
CREATE CLUSTER "MRDW"."CT_INTRMDRY_CAL" ("ID_YEAR_CAL" NUMBER(4, 0),
"ID_MONTH_CAL" NUMBER(2, 0), "ID_PROPERTY" NUMBER(5, 0)) SIZE 16384
PCTFREE 10 PCTUSED 90 INITRANS 3 MAXTRANS 255 STORAGE(INITIAL
83886080 NEXT 41943040 MINEXTENTS 1 MAXEXTENTS 1017 PCTINCREASE 0
FREELISTS 4 FREELIST GROUPS 1 BUFFER_POOL RECYCLE) TABLESPACE
"TSS_FACT" ;
```

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  smallin  NOT NULL ,
  CD_BOOK_SOURCE CHAR(1) NO  LL)
distribute on random;
```

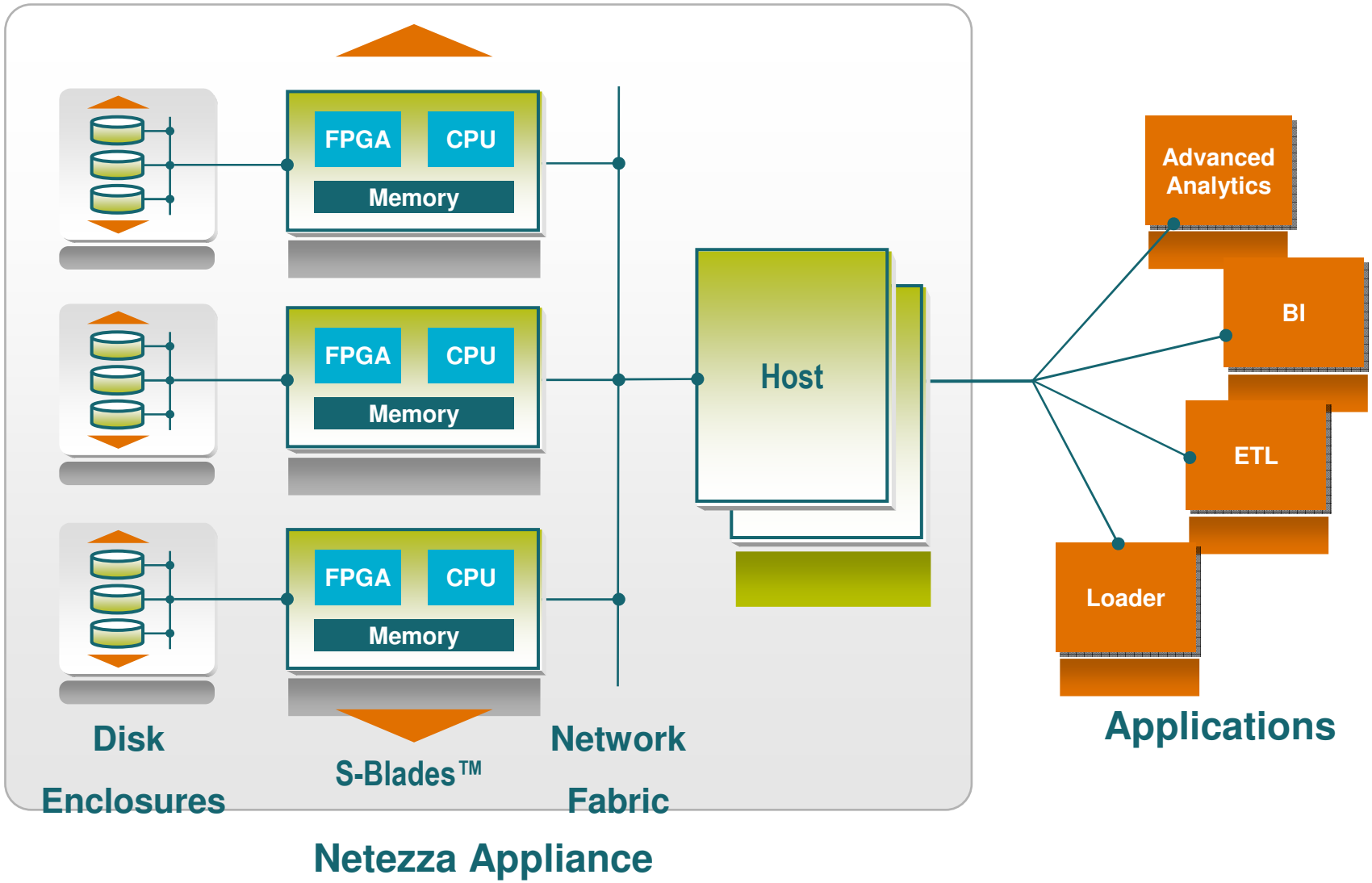
- No indexes
- No Physical Tuning/Admin
- Stripe data randomly, or by Columns

What is the Secret
The Netezza Architecture

The Principle

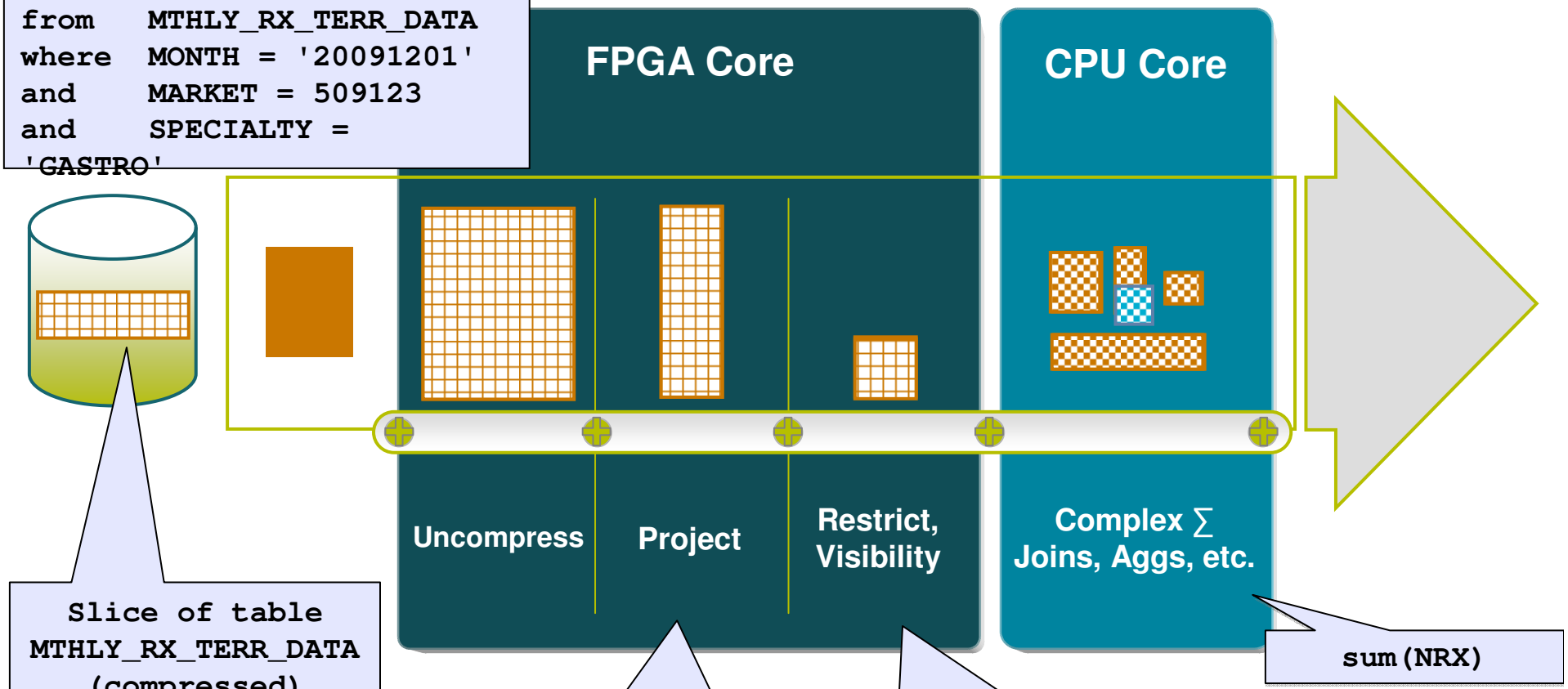
- **Processing close to the data source**
- **Balance massively parallel architecture**
- **Platform for advanced analytics**
- **Appliance simplicity**
- **Accelerated innovation and performance improvement**
- **Flexibile Configrations and extreme scalability**

The Netezza AMPP™ Architecture



Netezza Secret Sauce

```
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)

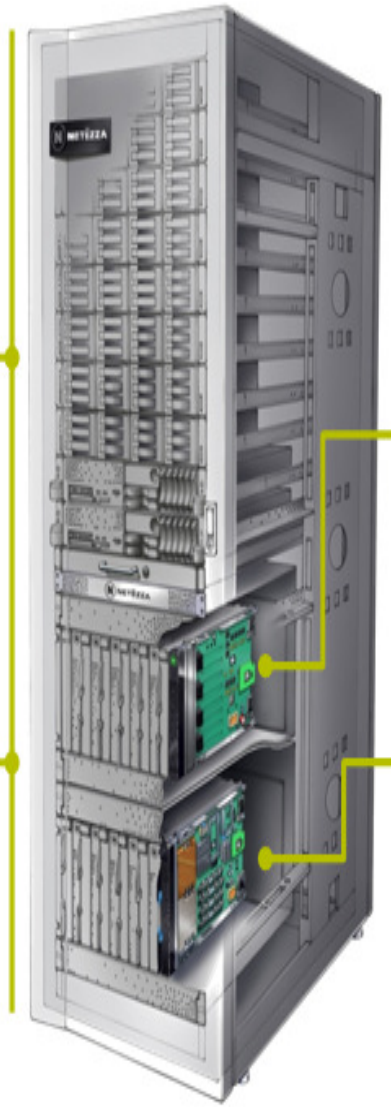
IBM Netezza TwinFin™

Optimized Hardware + Software

Purpose-built for high performance analytics; requires no tuning

True MPP

All processors fully utilized for maximum speed and efficiency



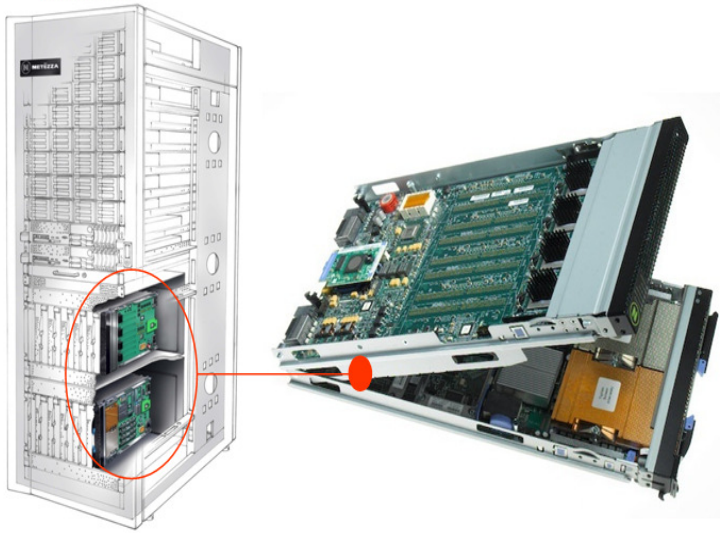
Streaming Data

Hardware-based query acceleration for blistering fast results

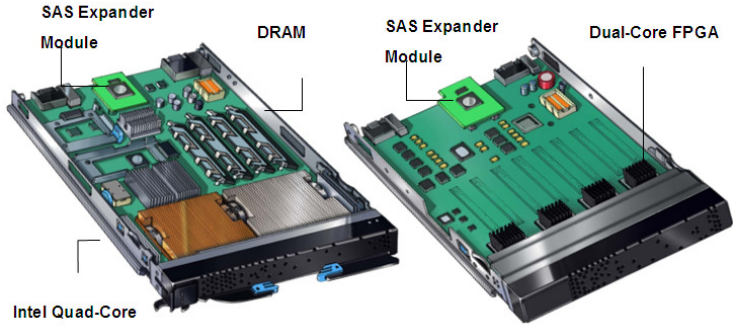
Deep Analytics

Complex analytics executed in-database for deeper insights

The Netezza S-Blade™



S-Blade™ Components



IBM BladeCenter Server

Netezza DB Accelerator

Appliance Family for Data Lifecycle Management

Skimmer	TwinFin	Cruiser
Development & Test System	Data Warehouse High Performance Analytics	Queryable Archiving Back-up/DR
1 TB to 10 TB	1 TB to 1.5 PB	100 TB to 10 PB

Into The Future

IBM

Thank You and Questions

November 11, 2011