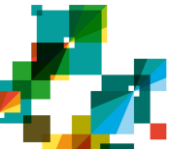
Netezza S's

Robert Hartevelt

31 October 2012

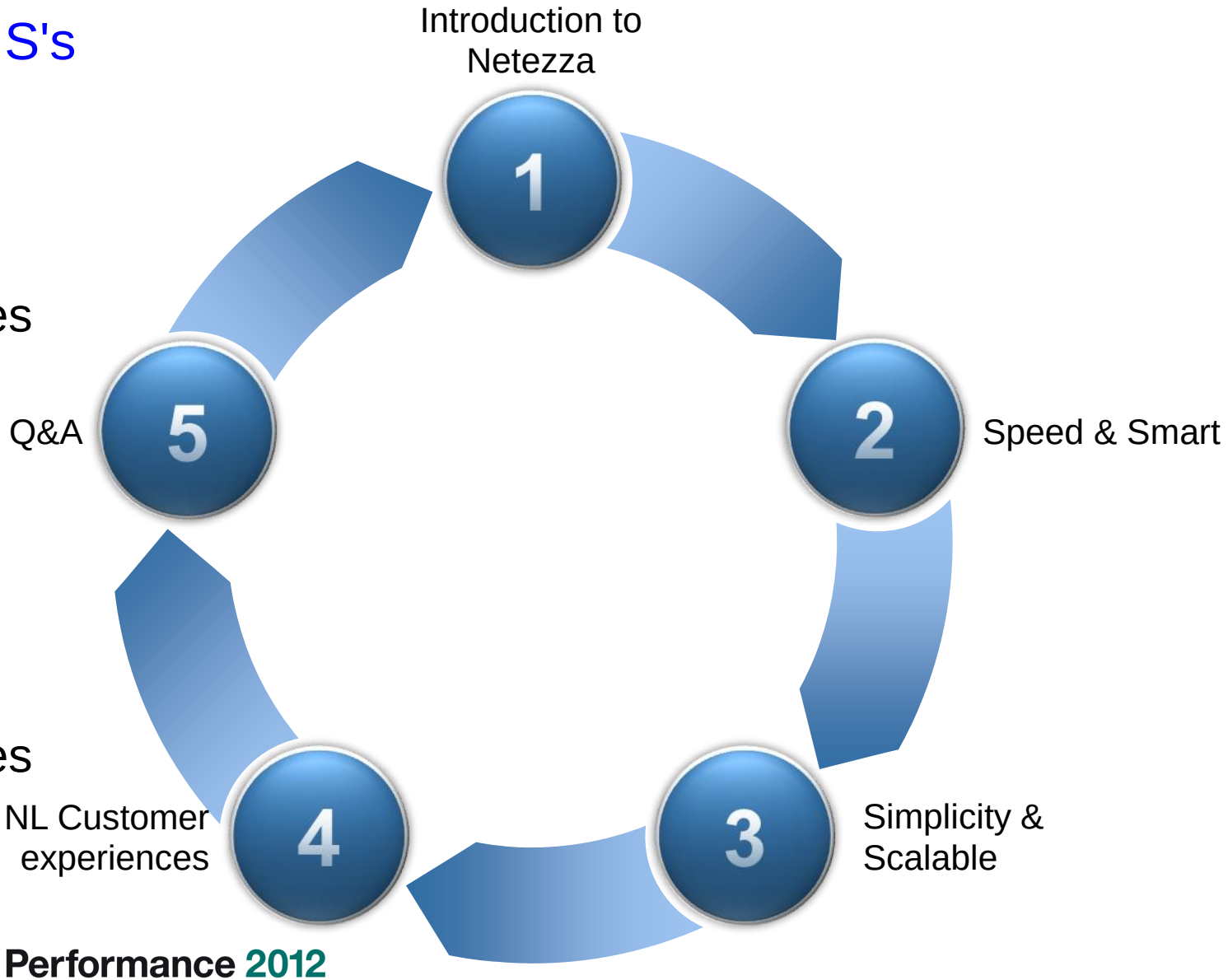


© 2012 IBM Corporation

Netezza S's

10 minutes

30 minutes



Netezza S's

Introduction to Netezza

Why workload optimized systems for Analytics?

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

Traditional data warehouses are just too complex

They are based on databases optimized for transaction processing—
NOT to meet the demands of advanced analytics on big data.

- Too complex an infrastructure
- Too complicated to deploy
- Too much tuning required
- Too inefficient at analytics
- Too many people needed to maintain
- Too costly to operate

Too long to get answers




Netezza's revolutionary approach

The Appliance

Simpler, faster, more accessible analytics

"This is what Netezza has done in the data warehousing market: It has **totally changed the way we think** about data warehousing."

- Philip Howard, Bloor Research

Appliance

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

The IBM Netezza Appliance – 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

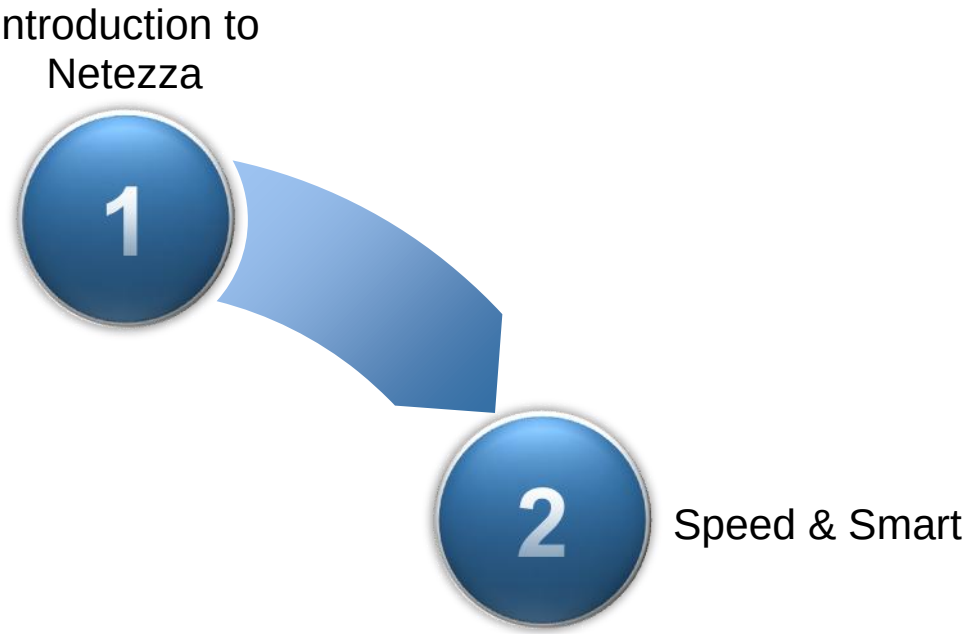
Netezza S's

Introduction to Netezza

So What?

Purpose build Appliance

Netezza S's



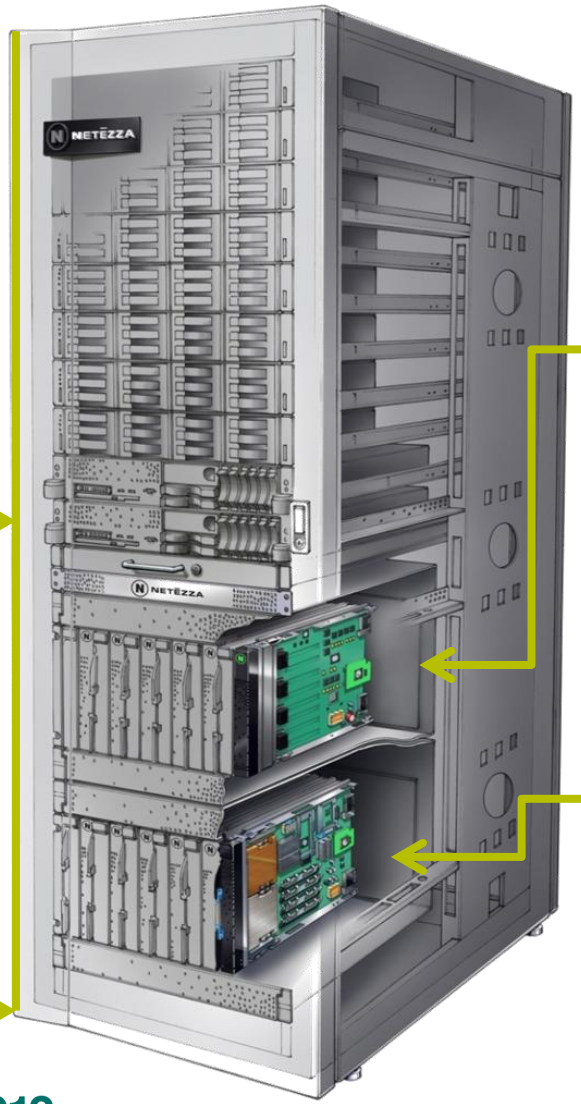
The IBM Netezza Architecture

Optimized Hardware + Software

- Purpose-built for high performance analytics
- Requires no tuning

True MPP

- All processors fully utilized
- Maximum speed and efficiency



Streaming Data

- Hardware-based query acceleration
- Blistering fast results

Deep Analytics

- Complex analytics executed in-database
- Deeper insights

Massively Parallel Processing Architecture

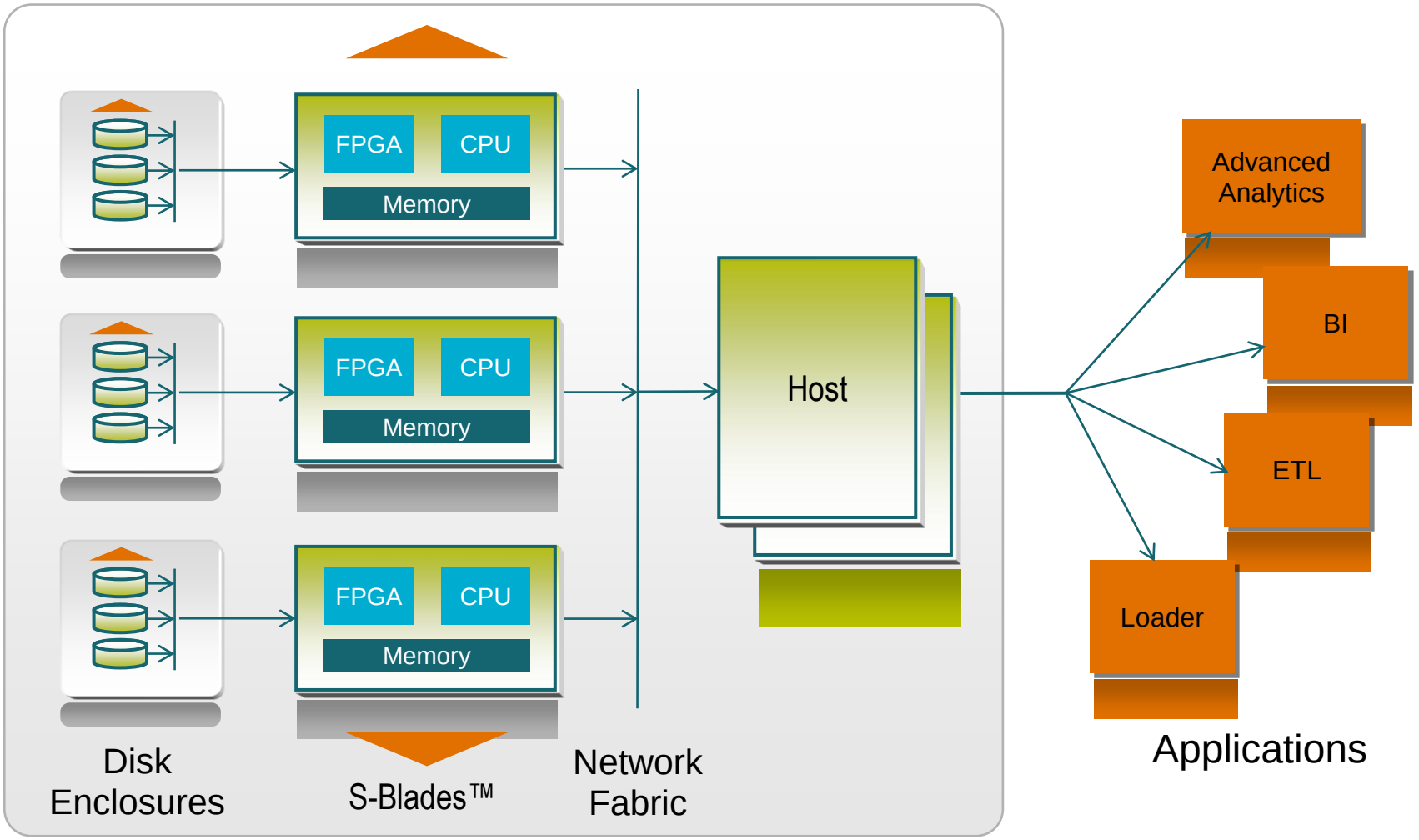
"Divide and conquer"

- MPP
  - "Shared Nothing" concept
  - Divides the work in smaller tasks
    - A big task is sliced vertically into a series of smaller tasks
    - The smaller tasks run independently
    - The work is automatically balanced among the tasks to minimize the time to complete
    - Each task is assigner the same amount of physical resources
    - Communication between is made only at the beginning and end of the task

- Benefits
  - A large task completes in a short elapsed time
  - Maximizes use of resources

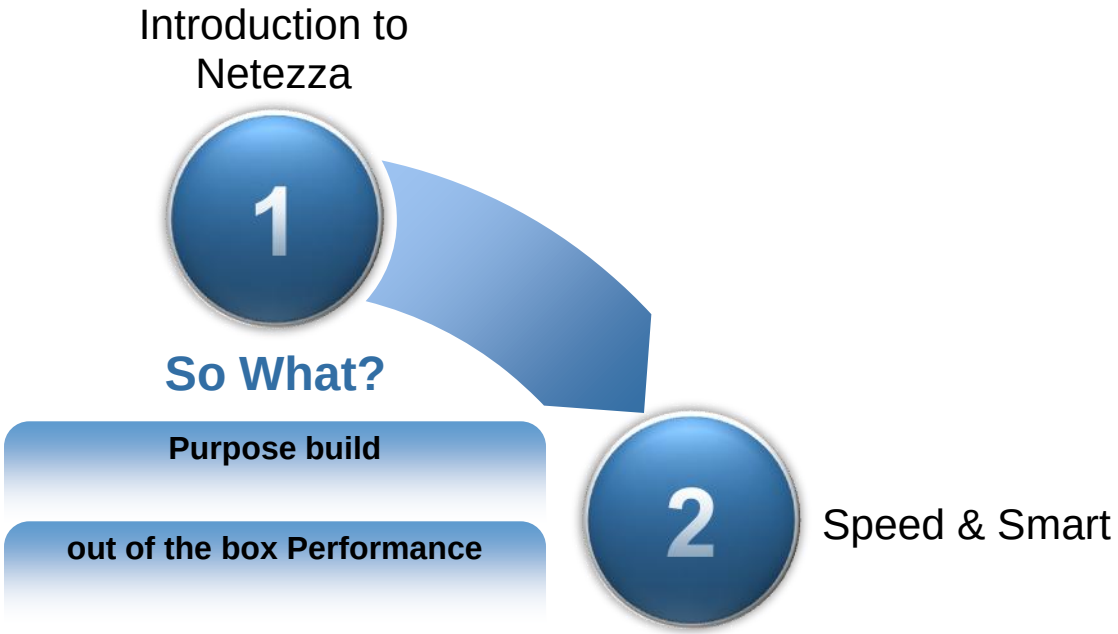
- Points of Attention
  - Complexity on administration and management
  - Communication bottlenecks

The IBM Netezza AMPP™ Architecture

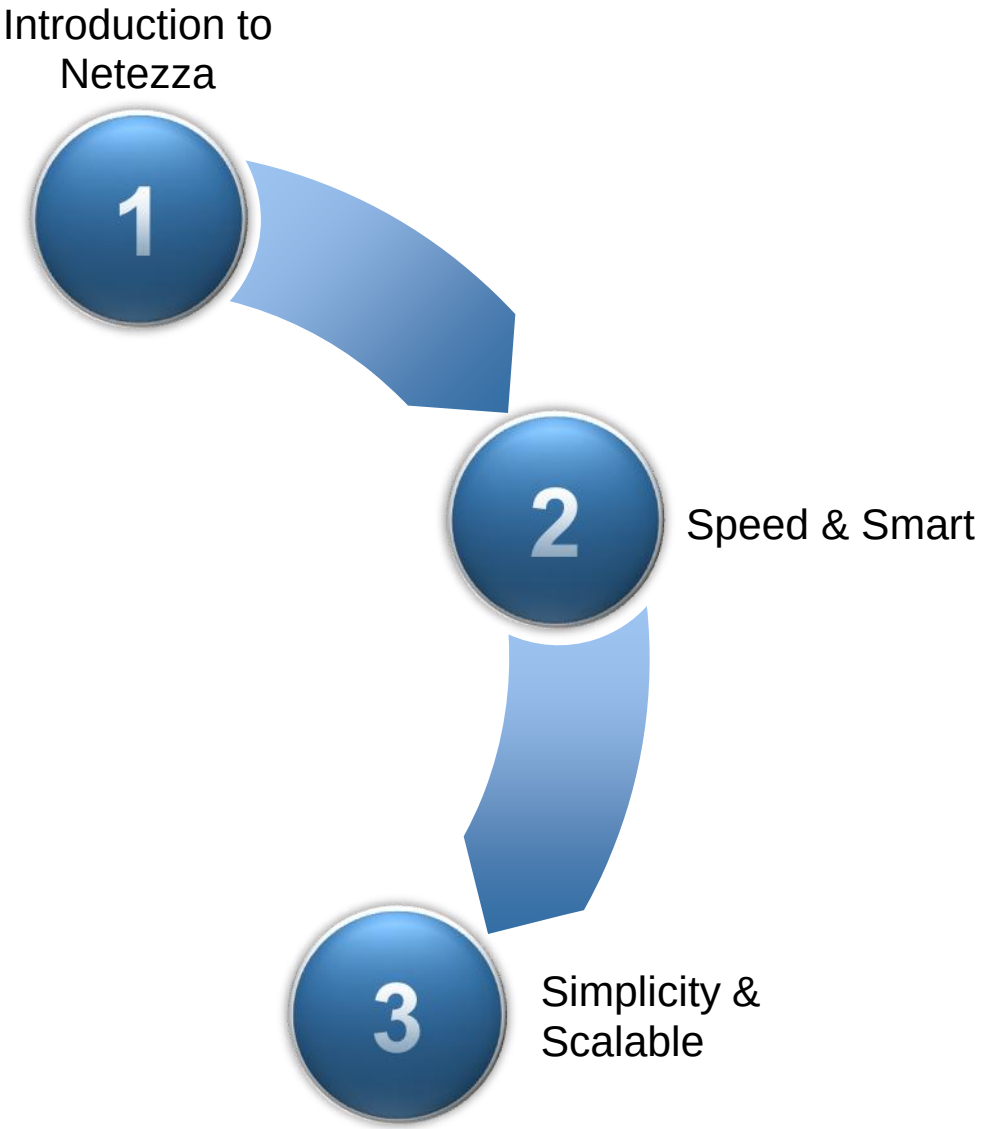


IBM Netezza Appliance

Netezza S's



Netezza S's

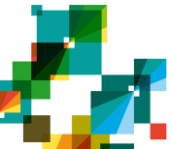


Appliance Simplicity

Resources become Data Managers instead of Database Administrators

- 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

Expert level within 2 trainings, takes you 1 week.



Traditional Complexity ... Netezza Simplicity

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
PARALLEL ( DEGREE 4 INSTANCES 1) LOCAL(PARTITION "PART1" 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, PARTITION "PART2"
PCTFREE 10 INITRANS 6 MAXTRANS 255 STORAGE(INITIAL 4194304 NEXT 4259840
MINEXTENTS 1 MAXEXTENTS 100000 PCTINCREASE 0 FREELISTS 10 FREELIST GROUPS
1 BUFFER_POOL DE
PARTITION "PART3
4194304 NEXT 425
FREELISTS 10 FRE
"DDM_DATAMART_IN
MAXTRANS 255 STO
100000 PCTINCREA
TABLESPACE "DDM_
INITRANS 6 MAXTR
MAXEXTENTS 10000
DEFAULT) TABLESP
PCTFREE 10 INITR
MINEXTENTS 1 MAX
1 BUFFER_POOL DE
```

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  smalli[illegible] NOT NULL ,
  CD_BOOK_SOURCE CHAR(1) NO[illegible]ULL)
distribute on random;
```

No indexes

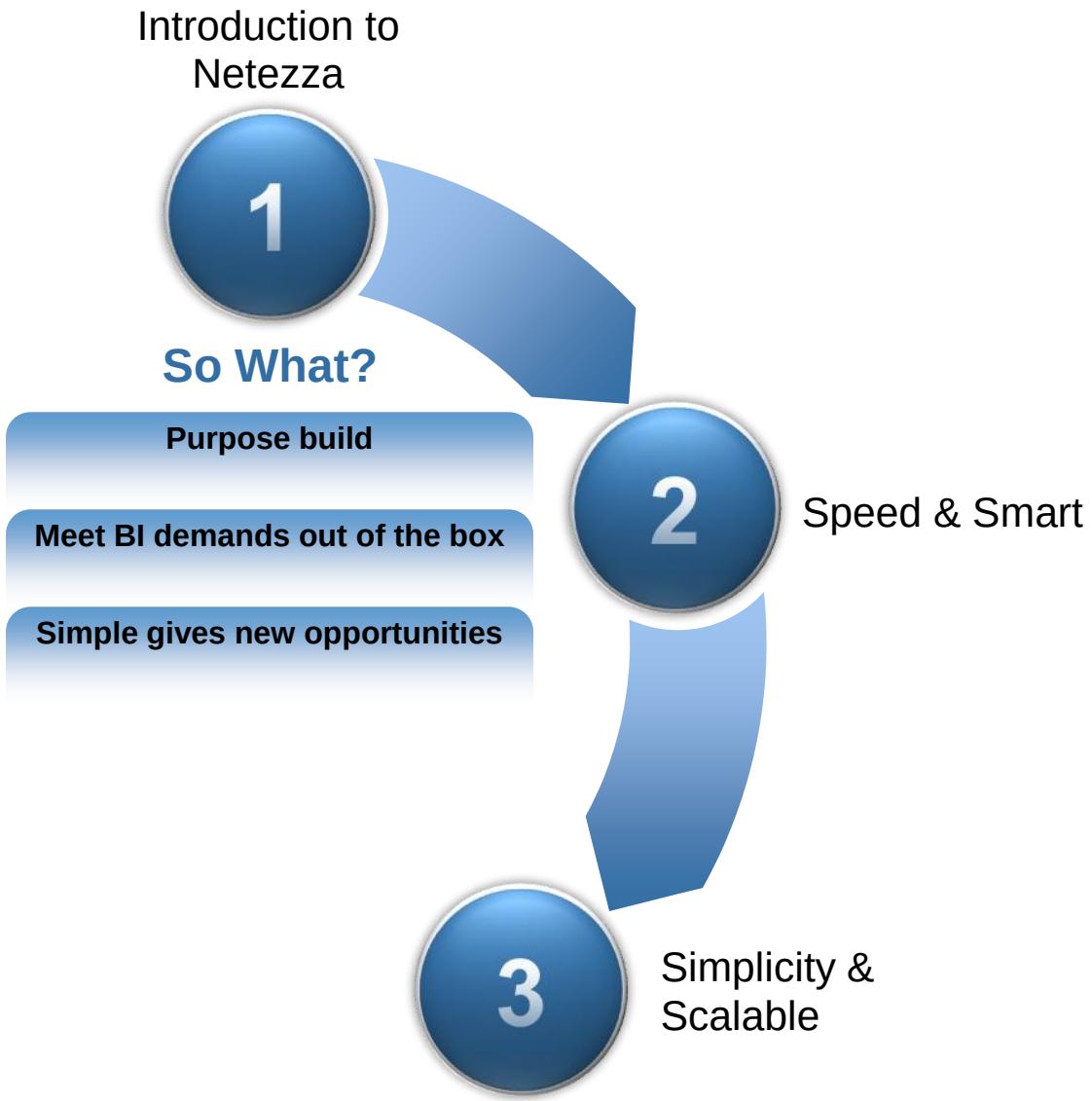
No Physical Tuning/Admin

Stripe data randomly, or by Columns

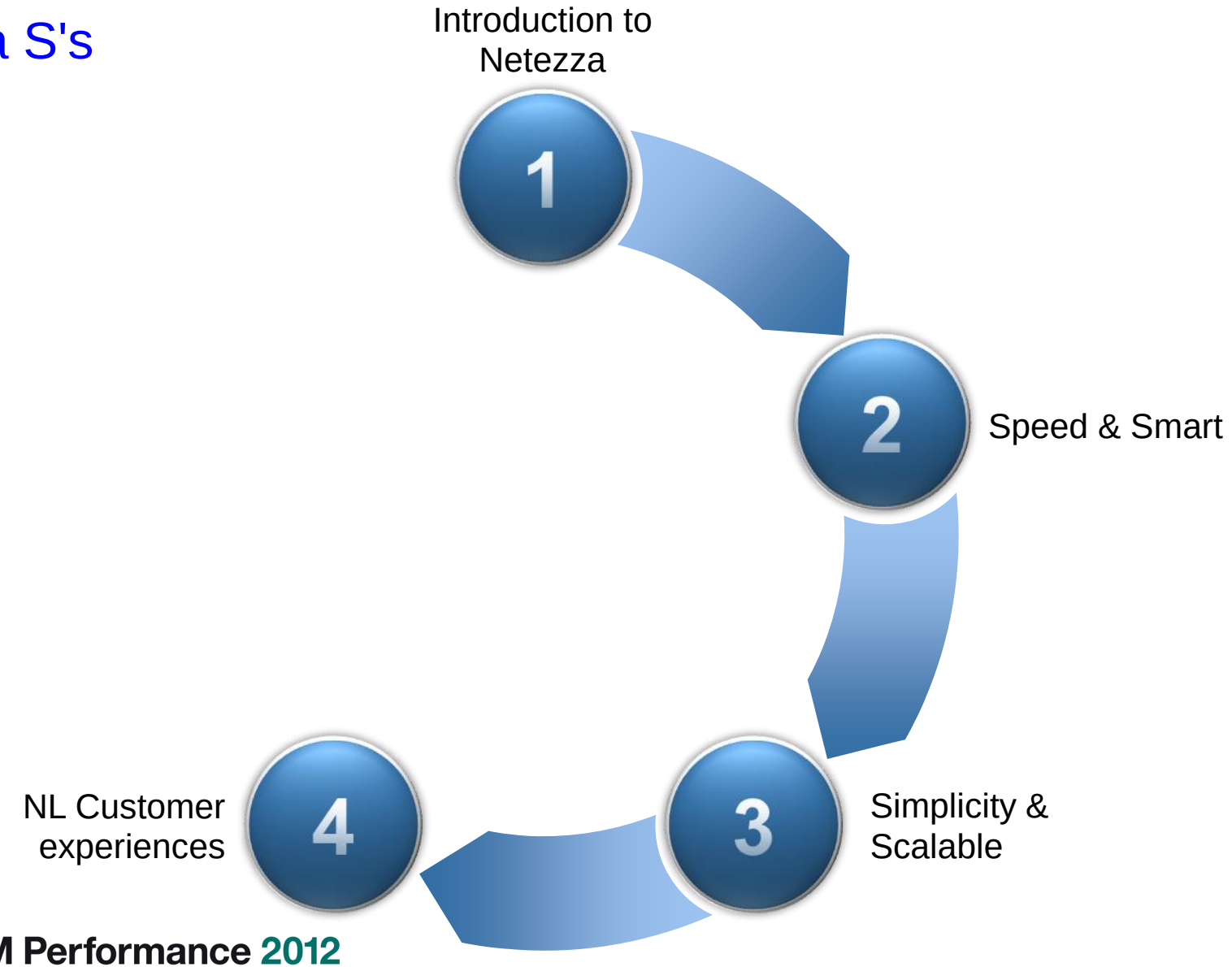
IBM Netezza Models

Capacity	= User Data space
Compression	= Effective User Data Space

Netezza S's



Netezza S's



Netezza4ING

NetInsight/Netezza Implementation 2012

F Deden

31-10-2012

NetInsight Overview

- NetInsight is ING's main Web Analytics tool from IBM
- NetInsight is used to measure daily clicks on ING portal
- Increasing importance of Web Analytics for:
- Customized Campaigns
- Fraud detection
- Analyzing purposes

NetInsight Oracle – History

First NetInsight – Oracle release December 2010

- No State of the Art hardware
- Oracle 10
- Increasing number of clicks

Issues

- Slow performance
- Long running night batch
- No pre-generated reports
- One active profile

Second NetInsight – Oracle release December 2011

- State of the Art hardware
- Oracle 11g
- Increasing number of clicks

Issues

- Recurring Unique constraint error
- Unacceptable slow ad-hoc reporting times (up to 12 hours)
- Backup sizing and time to restore too large
- Deleting of historical data with NetInsight functionality not possible
- One active profile

ING

NetInsight Netezza POC

- **Proof of Concept with Netezza from February 2012 to March 2012**
- **Followed by Project, ending 1 August 2012**
- **Acceptance criteria**

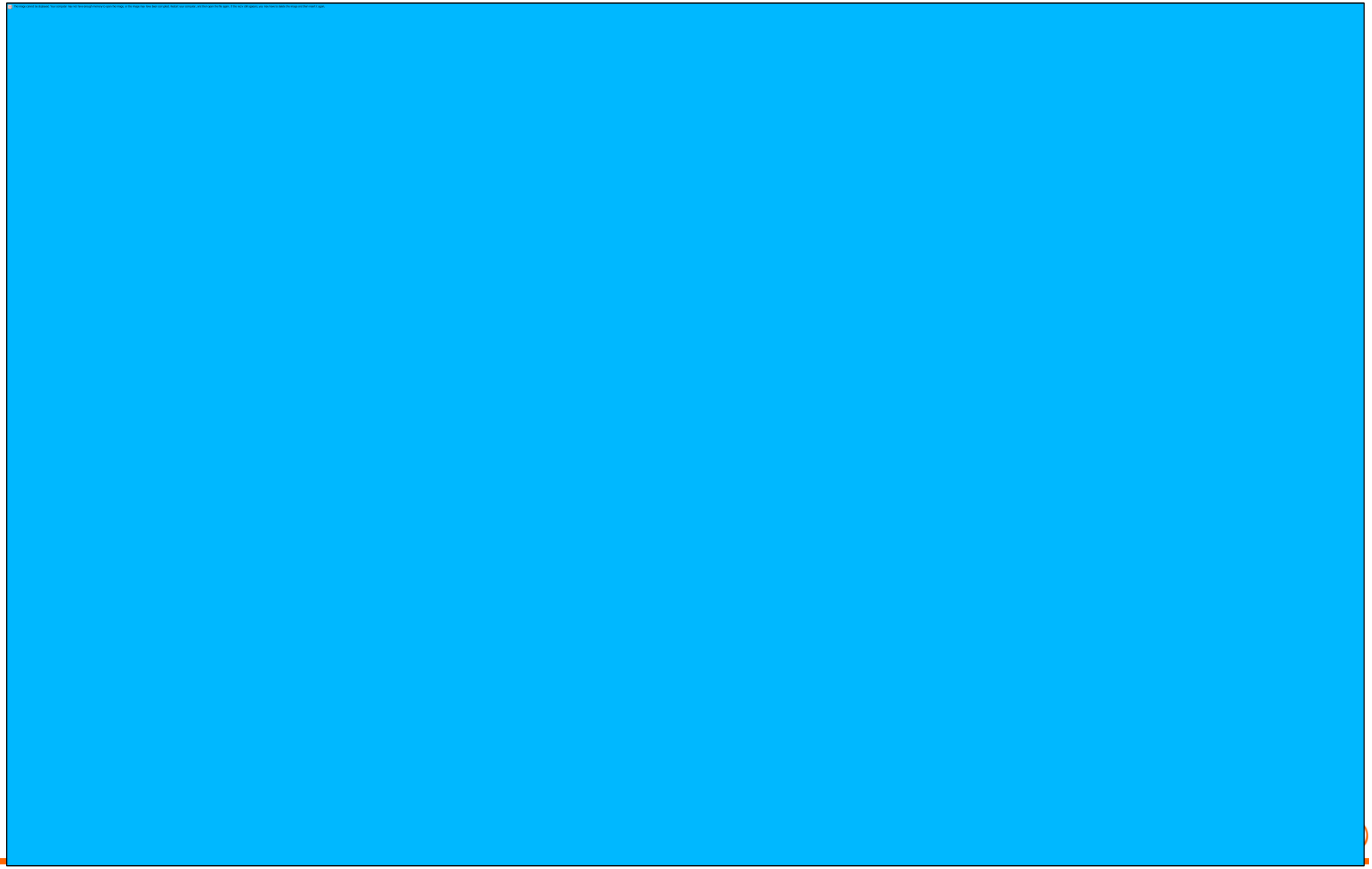
- Processing the internet clicks with sufficient throughput to have a manageable batch window.
- Webanalytics standard and ad-hoc reporting with sufficient performance to make it usable.
- Demonstrate that “Unique constraint” doesn’t appear in this configuration
- Deleting a day of data can be done within 2 hours
- Demonstrate that data can be deleted within 4 hours
- Demonstrate that data can be backed up and restored
- The Unique constraint error doesn’t appear in the POC setup.
- It can perform a (daily) batch load (against requirements of 2013 (50 million clicks) with a throughput of 10 million clicks per hour.
- It can perform the “Reporting Requirements”.

- **Hardware**

- Blade P70, 8 CPU’s, 64 Gbyte memory, no extra storage

ING

POC/Project Results



Plus Retail Challenge

Ruud Stroet, Architect Plus Retail

De vraag vanuit de Analisten:
laat mij queries uitvoeren op de beschikbare datamarts

De vraag vanuit de Ondernemers:
standaard rapportage kunnen opvragen

De uitdaging:
Analisten zullen zware queries uitvoeren waarbij het risico op een instabiele performance voor de Ondernemersrapportage het gevolg kan zijn.

Hierop moeten we als IT organisatie voorbereid zijn.

Plus Retail Action

Ruud Stroet, Architect Plus Retail

We hebben samen met onze partner AXI gekeken welke stappen genomen moeten worden om de uitdaging te kunnen pareren.

De vraagstelling op een Datamart betreft een andere invalshoek dan we gewend zijn.

Na een zoektocht vond ik een interessante technologie waarbij data opgedeeld wordt in behapbare blokken.

Deze opdeling van data wordt toegekend aan een Netezza TwinFin, welke in staat is om middels ANSI standaard instructies snel uit te kunnen laten voeren.

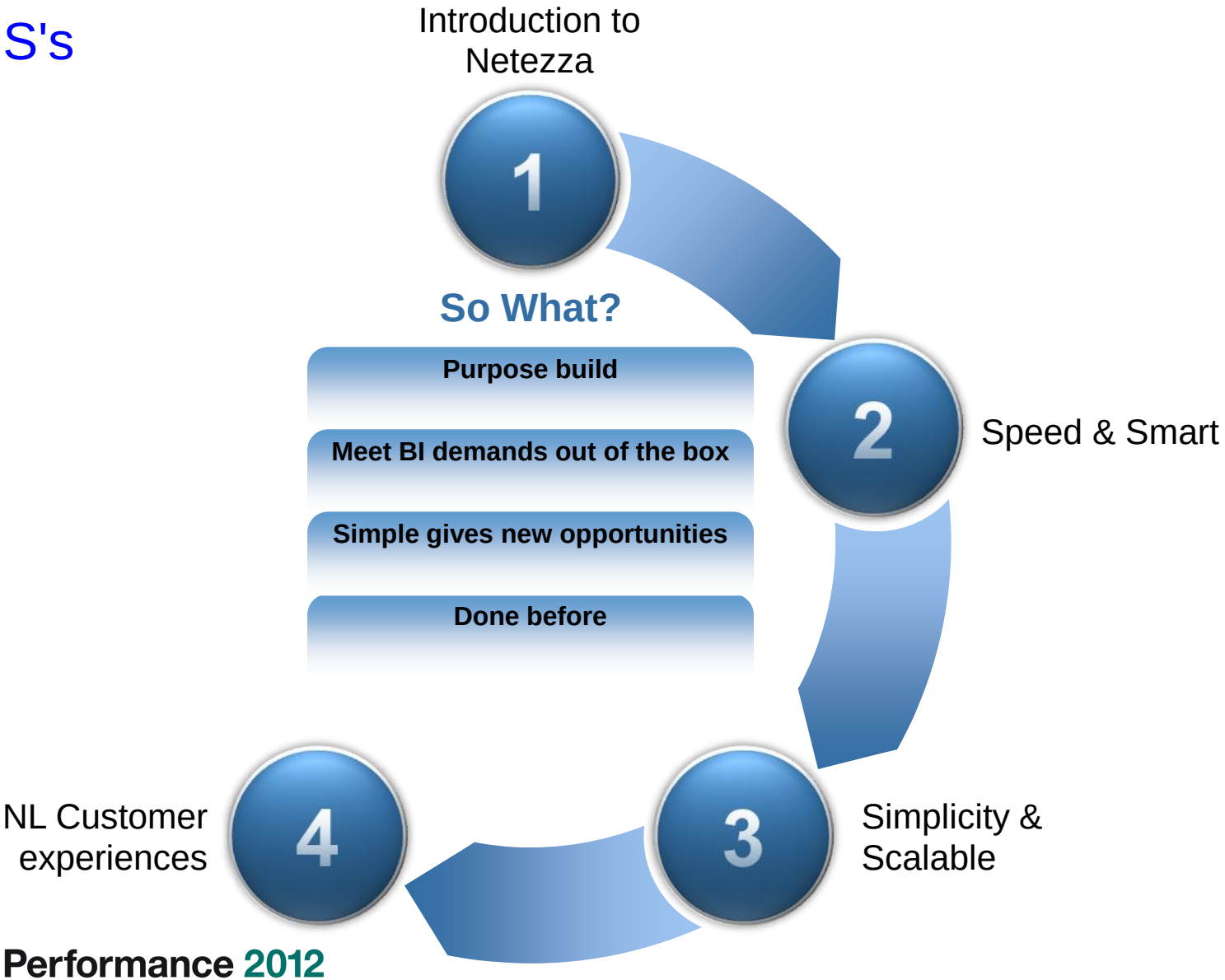
Plus Retail Result

Testen zijn nog in volle gang op dit moment.

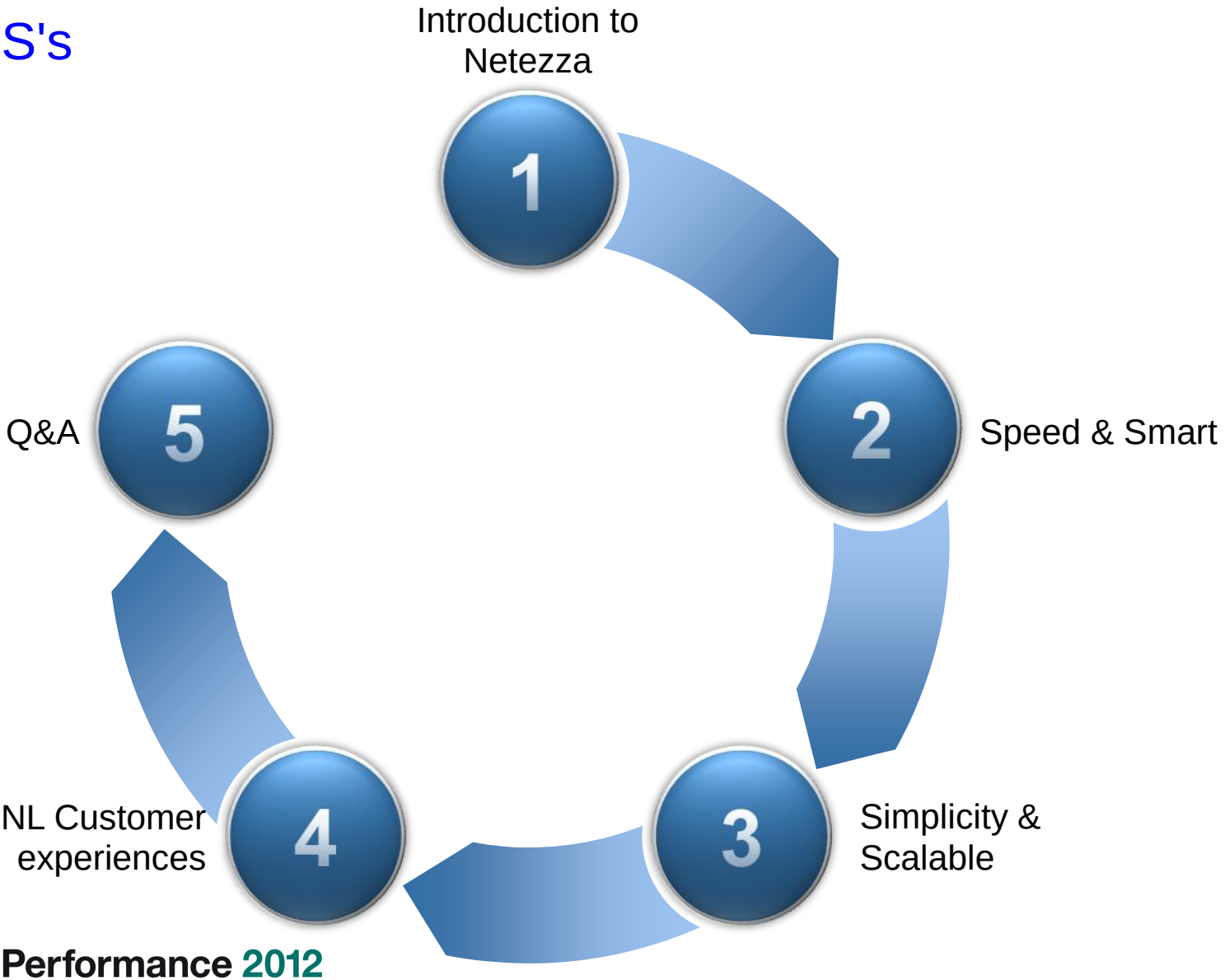
Maar zien er nu al veel belovend uit.

We kijken nu ook naar SPSS omdat we de mogelijkheid hebben gecreeerd om door onze hele dataset te lezen op bon-niveau. Hiermee hopen we nog meer inzichten te verkrijgen uit onze data.

Netezza S's

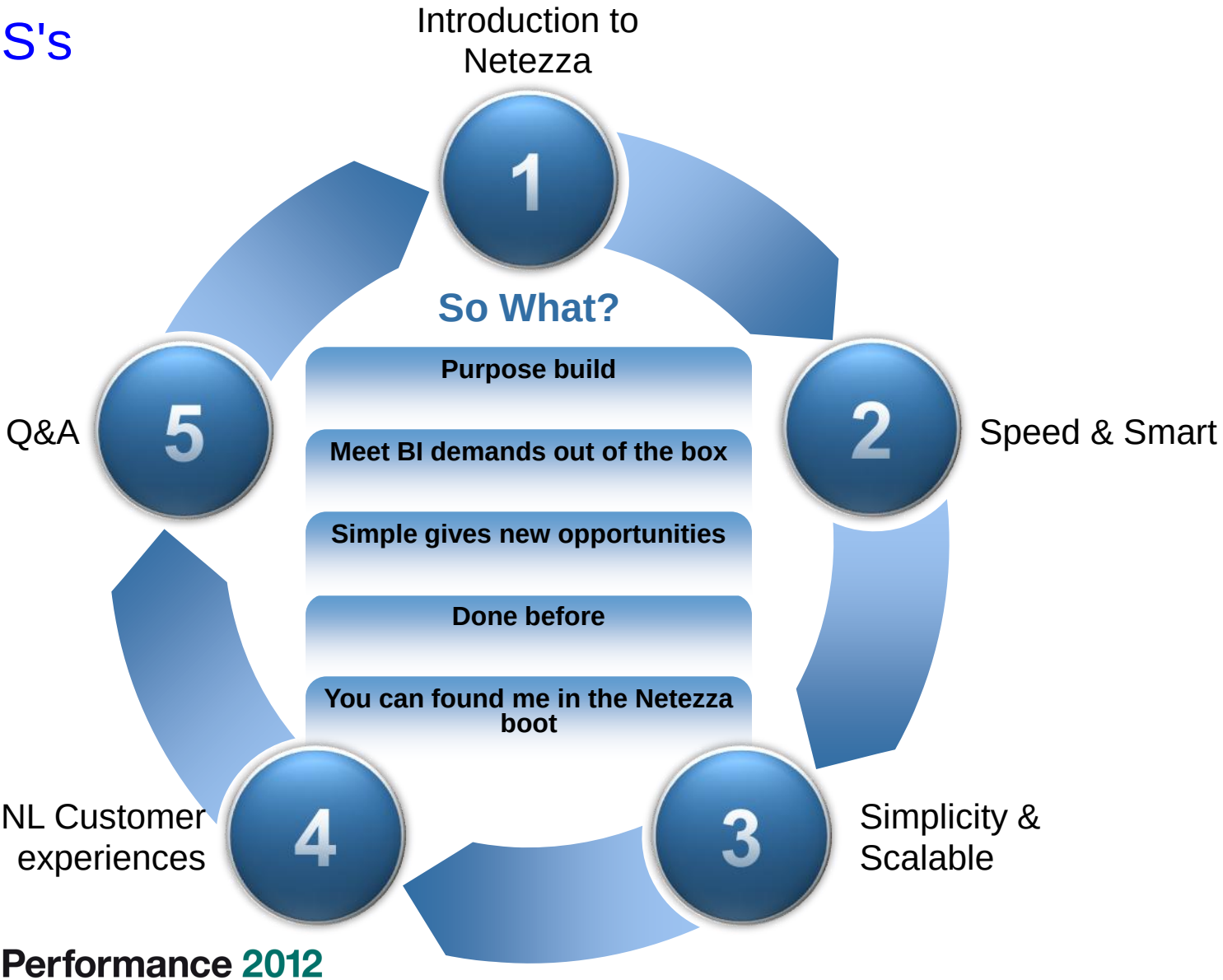


Netezza S's



Questions

Netezza S's



Get Proof. Not Promises.

Take the
Netezza TestDrive

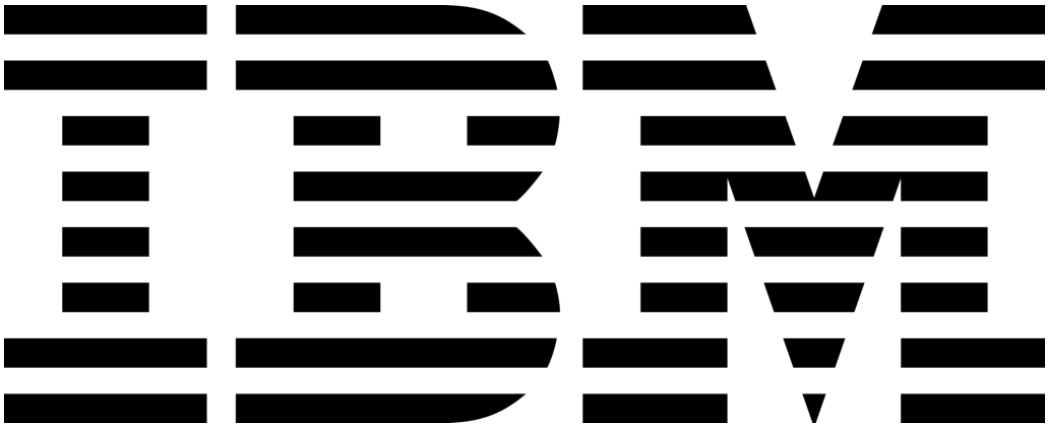
- ✔ Your Site
- ✔ Your Data
- ✔ Our Technology

A Proof of Concept in your own data center.

Netezza's data warehouse appliances are purpose-built to make advanced analytics simpler, faster and more accessible.

We'll bring it in.
You bring it on.

NETEZZA®
an IBM® Company


IBM