BusinessConnect
A New Era of Smart
04/06/2014

Leveraging analytics @ De Lijn

Agenda

- VVM – De Lijn: The company
- VVM – De Lijn: BI in the organization / BI-track
- Why a DWH-Appliance @ De Lijn ?
- Why IBM PureDataSystem for Analytics (Netezza) @ De Lijn ?
- Netezza implementation
- Are the requirements met ?
- Next steps
- Q&A

VVM – De Lijn
The company

- Vlaamse Vervoermaatschappij
- Started in 1991
- Responsible for mobility policy and operations for trams and busses in Flanders
- Key-figures (2013)
  - Annual budget + € 1.088 million
  - 8.205 employees
  - 2.275 busses
  - 369 trams
  - 540 million passengers
  - 210 million KM's driven

  (50% sub-contracted)

VVM – De Lijn
BI in the organisation

- BICC (Business Intelligence Competence Center) since 2010
- ± 4 - 5 FTE
- Part of Finance, Group Controlling
- Service provider for all business units
- Management of BI-environment
  - Reporting & analysis
  - Planning & Budgeting
  - eDWH & data-integration

VVM – De Lijn
BI-track

- 2001 Basic standard reporting using Business Objects
- 2002 Start of Data Warehouse
- 2004 First release Balanced Scorecard

  Cost forecasting & simulation application
- 2007 First release of integrated Planning & Budgeting solution with IBM Cognos Planning
- 2008 Start of Data Warehouse 2.0 (DB2 on iSeries) using ETL (IBM Infosphere DataStage)
- 2008 Enterprise BI with IBM Cognos v 8
- 2010 First release of driver-based integrated Planning & Budgeting solution with IBM Cognos TM1
- 2012 Enterprise BI with IBM Cognos v 10

Why a DWH-Appliance @ De Lijn ?

- Focus on the value of existing data
- “New data” leading to new insights
- Projects which will exponentially increase the volume of data
  - Retibo
  - ERP
- Technological optimization of the DWH-platform
- Financial optimization of the DWH-platform
- Business satisfaction

Why Netezza @ De Lijn ?

- Public tender for “purchase & installation” of a DWH-Appliance
- Criteria:
  - Low TCO
  - “Ease of use”
  - High Performance on analytics

→ **Winner: Netezza N1001-002**

- Public tender for “implementation” of Netezza
- Criteria:
  - Certification & experience with Netezza

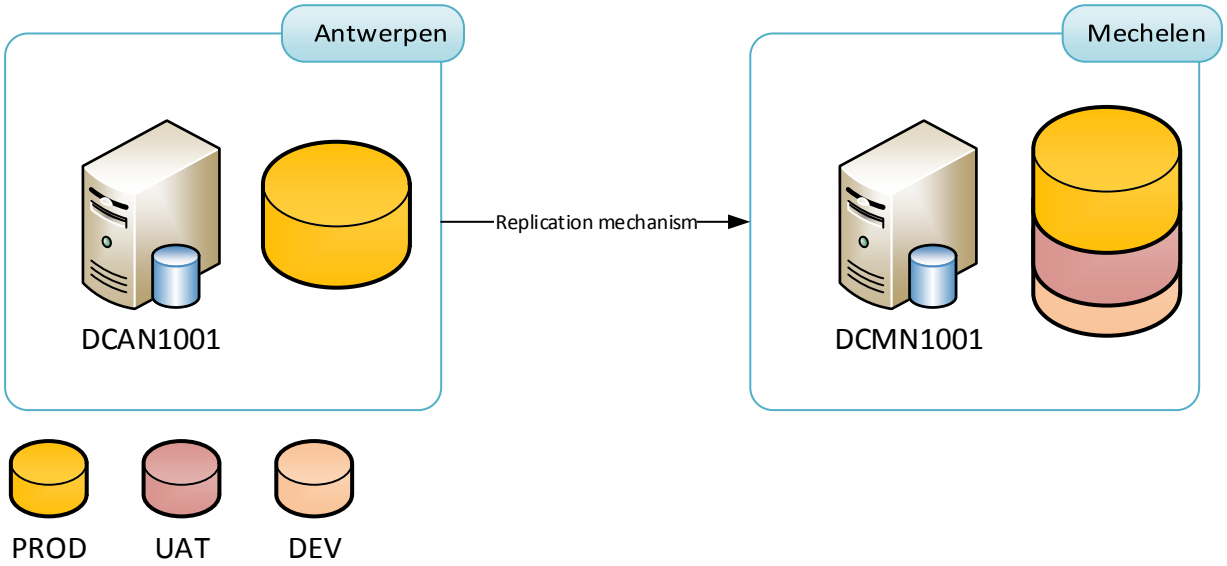
→ **Winner: LACO NV**

Netezza Implementation Objectives

- Get everything working on the Netezza box
- No disturbance of the production environment
- No interference with the existing processes
- Interface the box within the company framework
- Ensure performance and recovery

Netezza Implementation
Set up the environment

- Setting up databases, users and security levels
- Migrating Data Definition Language (DDL)
- Setting up ETL environment and connections



Netezza Implementation
Get everything in the box

- Data has been synchronized in two steps
  - "Hot" data
  - "Cold" data
- Hot data is reloaded every morning after the current datawarehouse loads
- Cold data was migrated slice by slice

Netezza Implementation
No disturbance of the production environment

- No resources available
- Lowest footprint possible on the existing environment
- Technology limitations
- Perform an intelligent database synchronization
  - Using Pentaho DI

Netezza Implementation
Interfacing the box

- Setting up a backup policy
- Defining scripts to do such backups
- Monitoring and scheduling scripts outside Netezza
- Connecting to consuming applications (Cognos)

Netezza Implementation Challenges

- Performance on loading
  - No particular bulk loading options to powercharge the box, will be improved when migrating to DataStage

- Think that's all ☺

Are the requirements met ?
Ease of use & low TCO

- System up-and-running in 12 hours

- DB Size-reduction
  - 10x space gain by full schemas (including index space) comparison
  - No need for indexes
  - No need for aggregate tables

- Availability
  - No maintenance performed since
  - System always online

Are the requirements met ?
Performance

- Out-of-the-box performance gains
  - Without the need for tuning mechanisms
    - No speeding up views
    - No aggregated tables
    - Simple distribution on 1^{sth} column (Id column)
  - Most on complex queries

- This means that performance can even be enhanced when needed !
  - Optimizing zonemaps
  - Doing ‘intelligent’ distribution

Are the requirements met? Some numbers

- **On average 92,01% faster compared to DB2** **(sample of 7 queries on "bus-stop-passing-times", +/- 870 mio records)**

E.g. Average delay on all bus-stops in 2013	Response Time
DB2/AS400	130 seconds
Netezza	11 seconds
E.g. Average delay on all bus-stops between 1/12/2013 & 31/12/2013	**Response Time**
DB2/AS400	27,5 seconds
Netezza	4 seconds
E.g. Complex join 1 dimension table (45k records) with 2 fact tables (800k & 870 mio records)	**Response Time**
DB2/AS400	No response
Netezza	580 seconds

- **Reports that were heavily filtered to enhance performance can now run on global data, complex queries run up to 40x faster.**

Next Steps

- Installing 2nd IPDA for Disaster Recovery
- Host Staging and Data Quality layers on IPDA
- Migrate & translate all (1500+) DataStage jobs
- Optimize IPDA structures for ultimate performance
- “Full” production mode estimated end of year

Q&A