XLDB Challenges for Structural Data
Focus: Analytics @ Nokia

Pekka Barck
Head of Data Warehousing
Nokia
**Scale & Size**

- **100 TB +** centralized in EDW, ERP DW, Research Center, Services specific on Teradata, Oracle, Hadoop...

- **100 TB +** in hundreds decentralized / federated dw:s/marts, mostly Oracle + 10 + other technologies.

- **100 TB +** in application specific logs where history is discarded for cost reasons.

- **1 PB +** in multiple new sensor data applications.
Structures and Processing

• Data Structures
  • Commercial normalized - denormalized databases
  • ROLAP, MOLAP, HOLAP
  • Hadoop, Map/Reduce
  • Virtualization through cross-database drill-through
  • In-memory (tokenized) databases

• Data Preparation
  • Transformation tools with too high costs for reconciliation and audit trail
  • Business rule engine type of approach
  • Messaging software
  • Nokia custom made loading tools for data access from
The **Tools**

Data exploration/mining: 4 tools including SPSS, Minitab + custom made
Text mining: 3 tools including Attensity
Webanalytics: Omniture Site Catalyst, Hitbox
OLAP: 6 tools including Cognos, BO, OBI (Hyperion)
Reporting: 12 tools including Cognos, BO, OBI (Siebel Analytics), SAP products
Analytics Distribution

- Normal queries are noise and as such are not affecting computing capacity needs
- Precooked 70%, 2% of capacity
- Ad-hoc 30%, 55% of capacity
- Exploration 2% of users, 43% of capacity
The Challenges

1. AGILITY
Project Specific Solutions June 2008

**STRENGTHS**
- Sharing of Important Resources
- Reduced Data Movement
- Faster Delivery
- Increased Security

**CHALLENGES**
- Data Not Fully Integrated
- Workload Management
- Maintaining Solution Consistency
- Additional Resource Demands on Production

PRODUCTION

Presentation Layer

ADW

Solution Specific ADW

STAGING
**STRENGTHS**
- Quick Determination of Data Value
- Easy Access for Data Research
- More Efficient Use of Business Input
- Less Demand on Production System Resources
- Automated Loading

**CHALLENGES**
- Data Integrated
- Additional Data Movement
- Freshness of IT Data
- Not a Production Environment
**STRENGTHS**
- Reduced Data Movement
- Increased Data Freshness
- More Powerful Platform Resources
- Faster Development of Analytic Metrics

**CHALLENGES**
- All Data Must Enter Through Staging
- Mixed Workload Management
- Additional Resource Demands on Production
STRENGTHS
Some Data Loads Can Bypass Staging
Quick Access to External Data
Quick Recognition of Data Value

CHALLENGES
Some Data Not Fully Integrated
Mixed Workload Management
One Architecture Slide Removed
Ensuring **Business Value**

- Business is moving towards advanced analytics in **all** areas
  - Now 100 sources. 1 new source added every week.
  - Everything has to link together.
  - The data model becomes very complex.
  - Different data has different business value.
Other Challenges

- Efficient (open) compressions in databases -> Green IT
- Efficient quality procedures in databases / Integration tools
- We need more hours per day: We materialize for performance but reach a deadend.
More **Challenges**

- **Growing need for high volume/**low usage **XXL-database.** Volume estimates between 3 and 40 PB.
- Cross-database **drill-through** capabilities: Oracle, Teradata, Hadoop...
- **Integration costs** too high, tools and integrators have not brought enough significant improvements during past 10 years.
- Linguistic based **free text analytics**: 10+ most important languages.
Thank you
pektka.barck@nokia.com