Efficient Data Extraction From Multiple Datasources

Alberto Mendelzon Workshop 2007
Punta del Este - Uruguay

Ing. Daniel Ferreira Zoppis
Lic. Mario Casaretto
Agenda

• Study Case
• Business Intelligence Systems
• Extraction, Transformation and DW Loading process (ETL)
• Commercial ETL Tools
• Federated Database Systems (FDBS)
• Commercial ETL Tools & FDBS
• Distributed Remote Extraction Interface (DREI)
• Integration DREI - ETL Tools
• Conclusions
• References
• About TCS
• TCS BI Practice Profile
• TCS BI Methodology
• TCS BI Centre of Excellence
Study Case

- Telecommunications Company with more than 2.5M customers
- 60% of national telecom market
- Billing system: more than 90 Databases distributed across the all territory
- Unix Environment
- Periodically loading of invoicing information
  - 12M invoices
  - 10GB
- Restriction: 1 hour for entire ETL process!!
Business Intelligence System (a brief review)

BI back-end
- Staging
- Datawarehouse
- ETL Tools
  - DB 1
  - DB 2
  - ... (DB n)
  - App. 1
  - App. 2
  - ... (App. n)

BI front-end
- OLAP Tools
- Relational Tools
- Reporting Tools
  - ...
Extraction, Transformation and DW Loading

Source Systems
- Oracle
- SQL Server
- .xls
...

Staging Area

DW
- Corporative DW
- Departamental Data Marts

Extraction
Data Profiling
Data Cleansing
...
Data Transformation
Data Integration
...
Commercial ETL Tools

- Microsoft Integration Services
- Oracle Warehouse Builder
- COGNOS DecisionStream
- Informatica PowerCenter
- Business Objects Data Integrator
- Pentaho Data Integration (Kettle)

✓ Very pretty interfaces
✓ A lot of Marketing around
✓ Workflow process, exception handling
✓ Metadata
✓ Data Profiling & Cleansing
✓ Heterogeneous connectivity via ODBC, JDBC, OLEDB, etc.
→ Design and implementation time reduction
→ Maintenance support via Metadata
Federated Database Systems

FDBMS

Service Layer

Transaction Coordinator

Master Data

Integration Layer

FDB 1

App. 1

Cons. 1

... Cons. k

FDB 2

App. 2

FDB n

App. n

Optimal for transactional usage, where we query a lot of data and retrieve a few records
Commercial ETL Tools & FDBS

... and bandwidth usage??

ODBC
OLEDB
JDBC
...

FDBMS
(Source System)

Service Layer

Transaction Coordinator

Integration Layer

FDB 1
App. 1

FDB 2
App. 2

...

FDB n
App. n

Datawarehouse

DW (Target)

Master Data

(Source System)

ODBC
OLEDB
JDBC
...

... and bandwidth usage??

1 November 2007
Distributed Remote Extraction Interface (DREI)

Source FDBS

FDB 1

... ...

FDB n

Network

DREI

ETL

Staging

Datawarehouse

DW

Dataflow:

Execution process flow

Source FDBS

Network

DREI

Staging

ETL

Datawarehouse

FDB 1

FDB n
Integration commercial ETL - DREI

shell: proc(db_group, sql, ...)

DREI Metadata

IP Address
O/S Connection parameters
DB Connection parameters
Remote command-shell Script Template
Conclusions

• ETL
  • Very useful to implement ETL process
  • Focus on heterogeneous data source integration
  • Not focus on performance issues of ETL process
  • Additional effort required to implement a complete performing solution

• Distributed Remote Extraction Interface
  • At the end we always implement something like that
  • Successful integration at O/S level
  • Non dependencies with commercial ETL Tool
  • Good feature to add by commercial ETL Tools
References

*Bases de datos federadas*
http://www.tonahtiu.com/notas/BD/BDF.htm

*Bases de Datos Heterogéneas*
Sandra Navarro, Carlos Castellano.

*Overview of optimizing remote data sources with DB2 Data Warehouse Edition*

*Enterprise Information Integration is a Strategy, not a Technique.*
S. Radhakrishnan.

*How to Obtain Flexible, Cost-Effective Scalability and Performance through Pushdown Processing*
An Informatica white paper, Apr. 2006.

*Teradata Load and Unload Utilities*

*Building Secure Data Warehouse Schemas From Federated Information Systems*
Félix Saltor, Marta Oliva, Alberto Abelló, José Samos.
http://www-lsi.ugr.es/~bdf/Trabajos/codata'00abs.pdf
Question & Answers

Daniel.Ferreira@tcs.com
Mario.Casaretto@tcs.com
About TCS Iberoamerica

Operation across 15 countries in America, Spain, Portugal:

- Founded in 2002, 5 years ago
- 2400 professionals, 5000+ by FY2007
- Strong local country staffing (over 98%)
- Experienced Latin region management team
- 2 CMM Level 5 centres in Uruguay & Brazil
- BPO centres in Chile and Colombia
- Serving over 150 clients from Latin America, Europe and the US
**Uruguay GDC: Their Largest IT Center**

<table>
<thead>
<tr>
<th>History</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Founded in May 2002</td>
</tr>
<tr>
<td>• Inaugurated by Dr. Jorge Batlle, President of Uruguay</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Infrastructure</th>
</tr>
</thead>
<tbody>
<tr>
<td>• + 700 seat office with modern infrastructure</td>
</tr>
<tr>
<td>• Located in Zonamerica, a free trade zone technology park with World class network, internet and telecom facilities</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Personnel</th>
</tr>
</thead>
<tbody>
<tr>
<td>• 800 people</td>
</tr>
<tr>
<td>• 95% local Spanish speaking personnel, 5% experts from India</td>
</tr>
<tr>
<td>• Initial &amp; continuous training for all personnel</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quality Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Certified at SEI CMM Level 5 in Oct 2003</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clients</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Served over 30 clients in Latin America, Europe and the US</td>
</tr>
</tbody>
</table>
Latin America as an offshore / nearshore location

1. Spanish/Portuguese /English languages related services
2. Time Zone Advantage
3. Lower risk – offers low cost offshore alternative to India
4. Physical, strategic & cultural proximity to the US / Europe
5. Many of their global clients have operations in the region, needing nearshore services

TCS sees Latin America not just as a market for IT Services, but also a strategic base for serving customers in the US and Europe
BI Practice Profile

Track Record
- 215+ active customer relationships
- 90% customer satisfaction index
- Revenue Growth (Y-o-Y) of 65%
- 345+ active BI Projects

Global Presence
- 33 Delivery Centers in 10 countries
- Regional BI Practice Directors in the Americas, UK and Europe
- Regional CoEs of BI Technology in Latin America, Australia, China and Hungary

Vision
Global among Global Top 5 by 2010

Mature Delivery Process
- Leverage TCS Quality Models - CMMI, PCMM, Six Sigma, ISO, BS7799 etc.
- Leverage Networked Delivery
- Delivery Assurance Groups
- End to end service maturity

Alliances & Partnerships
- Global Partner Relationship with IBM, Oracle, Microsoft, NCR Teradata, Informatica, Ab Initio, Business Objects, Cognos, Hyperion, SAS, SAP, Siebel
- Joint go-to-market strategy
- Participation in beta Programs
- Bundled offerings of Product Sale & Service

Depth of Experience
- 3,850+ experienced consultants
- Proven BI Program Definition and Implementation Solutions- BIDS™
- Rigorous internal competency development and evaluation program

Technology Excellence
- Mature CoE for focus technologies
- Achieve Excellence with TCS Technology Maturity Model
- Industry Specific Offerings for BFSI, Manufacturing, Telecom, Retail
TCS BI Services, among the leaders in North America

- Effective use of global delivery (price and staffing)
- Technical resource consistency
- Uruguay Delivery Centre provide BI services to North America market as Near Shore

Figure 1. Magic Quadrant for Business Intelligence Services, North America, 2007

Source: Gartner (March 2007)
BIDS™ Methodology

A Global Conceptualization and Implementation Methodology for Enterprise Business Intelligence Solutions
Related Solutions Offerings

- BIBP
- BIHC
- BICC
- KDD
- BCM

Metadata Management
Performance Management
Data Quality Management
Dashboard Management
Maintenance
Migration
- Application
- Data
- M2X

Multisite Rollout
Due-diligence
Data Marts Consolidation
Financial Consolidation
Upgrade
Master data Management

Tata Consultancy Services
TCS Uruguay BI CoE

• Global Delivery Model for USA, Europe and Latin America
• Global Knowledge Database
• TCS BI Forums
• Expertise in Oracle, Cognos, Microstrategy, Business Object and Microsoft BI tools
• Knowledge Development Centre (Iberoamerica educational centre)
• Internal Technologies Days (Tech Day and Tactics)