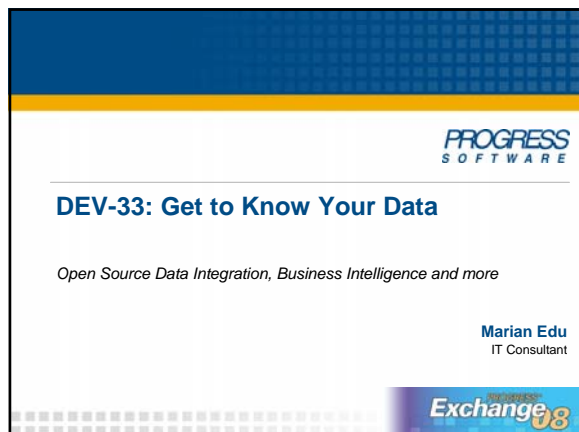
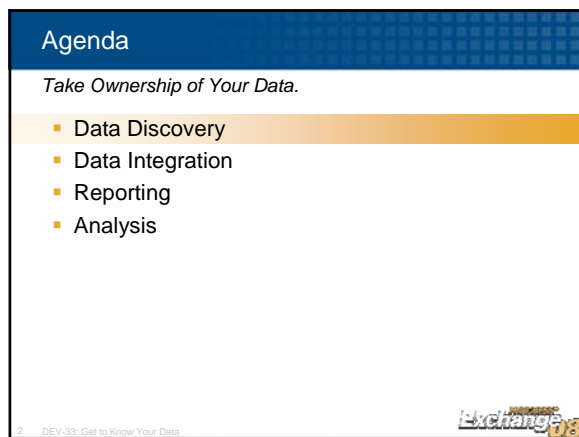
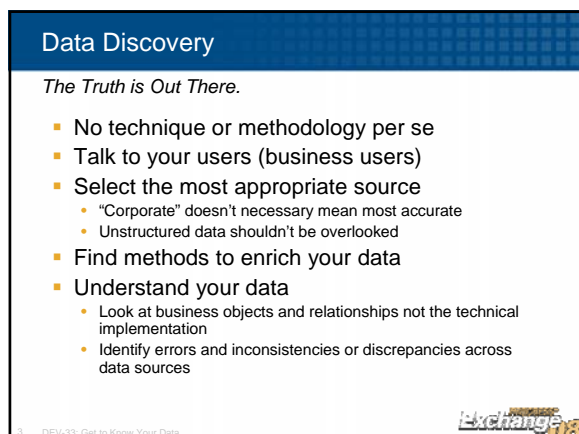


DEV-33: Get to Know Your Data – Open Source Data Integration, Business Intelligence and more

Marian Edu







Data Integration

- Extract
- Transform
 - Cleansing
 - Consolidation and Enrichment
 - Transform to adapt to required/optimal structure (from transaction optimized to reporting and analyze)
- Load

DEV-33: Get to Know Your Data

Exchange 08

Data Integration

Data Cleansing

"No enterprise wants, needs or will pay for perfect quality data."

John Morris,
iergo ltd

DEV-33: Get to Know Your Data

Exchange 08

Data Integration

ETL Tools

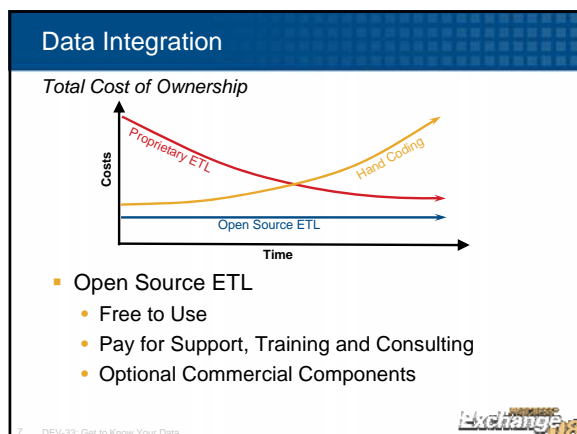
- Remove the need of "Hand Coding"
- Various Data Sources and Targets
 - Relational database
 - Flat files
 - Internet
- Data Quality
- Logging and Audit
- Metadata and Documentation

DEV-33: Get to Know Your Data

Exchange 08

DEV-33: Get to Know Your Data – Open Source Data Integration, Business Intelligence and more

Marian Edu



Data Integration

Open Source ETL

- Pentaho Kettle
- Talend
- Apartar

Exchange 2008

Data Integration

Pentaho Kettle

- Easy to use User Interface (Eclipse plug-in)
- Connectivity to a very wide variety of databases (JDBC), flat files and web services
- Comprehensive collection of Jobs and Transformations
- Metadata and Business Model Repository
- Data Quality
- Practical Monitoring and Logging
- Stand-alone Java™ Engine

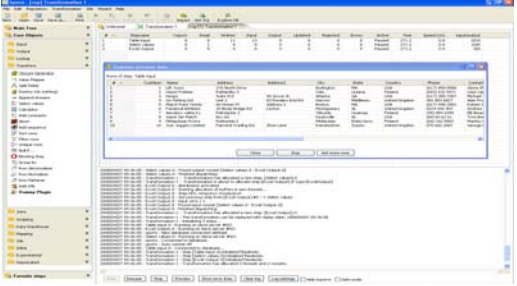
Exchange 2008

DEV-33: Get to Know Your Data – Open Source Data Integration, Business Intelligence and more

Marian Edu

Data Integration

Pentaho Kettle



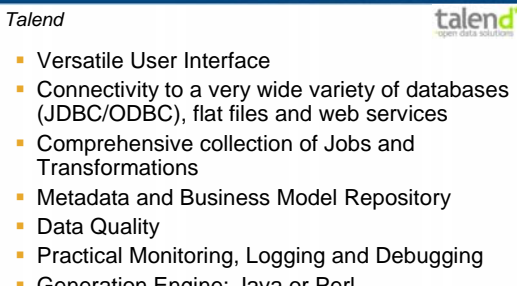
10 DEV-33: Get to Know Your Data

Exchange 2008

Data Integration

Talend

- Versatile User Interface
- Connectivity to a very wide variety of databases (JDBC/ODBC), flat files and web services
- Comprehensive collection of Jobs and Transformations
- Metadata and Business Model Repository
- Data Quality
- Practical Monitoring, Logging and Debugging
- Generation Engine: Java or Perl

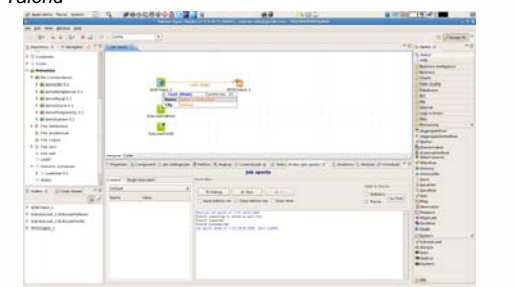


11 DEV-33: Get to Know Your Data

Exchange 2008

Data Integration

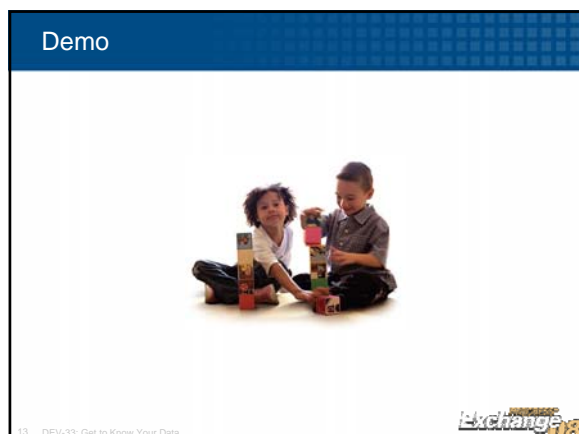
Talend



12 DEV-33: Get to Know Your Data

Exchange 2008

DEV-33: Get to Know Your Data – Open Source Data Integration, Business Intelligence and more Marian Edu



Data Integration

Apartar

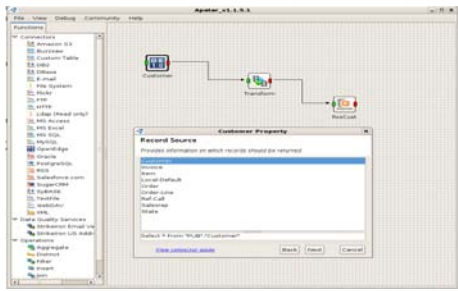
- Focus on “On-Demand” operational data integration
- Easy to use User Interface
- Target Audience is Business Users
- Decent connectivity options including databases (JDBC), flat files and Web 2.0
- Limited Transformations, user contributed DataMaps (reusable integrations jobs)
- Desktop application

Exchange 2008

14 DEV-33: Get to Know Your Data

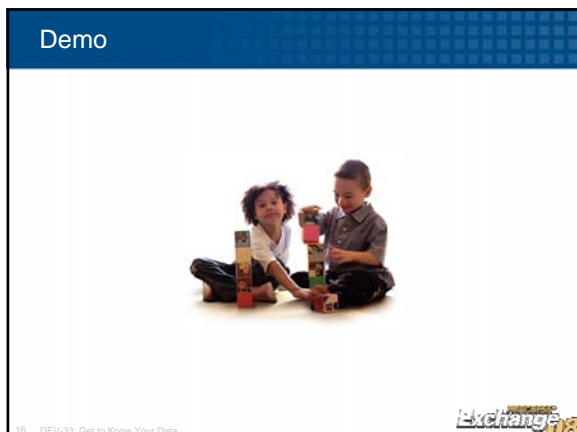
Data Integration

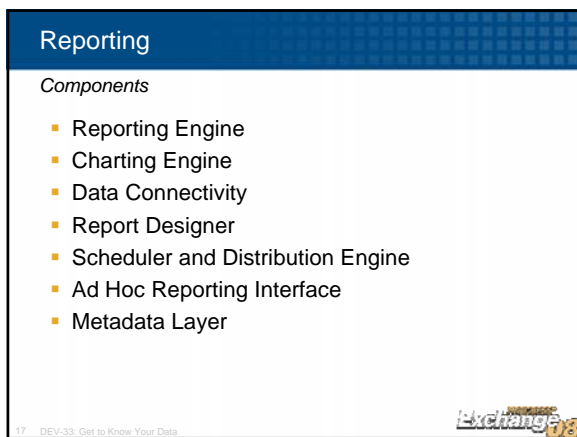
Apartar



Exchange 2008

15 DEV-33: Get to Know Your Data







DEV-33: Get to Know Your Data – Open Source Data Integration, Business Intelligence and more

Marian Edu

Reporting

Crystal Reports for Eclipse

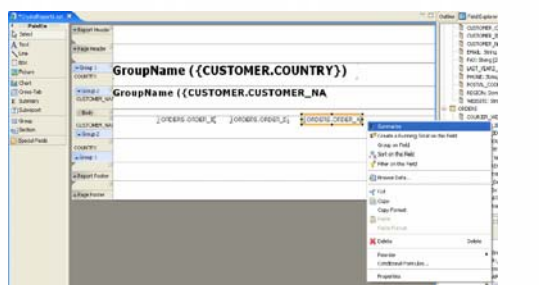
- Powerful Java Report Engine
- Report Designer build on Eclipse Framework
- Connectivity to a very wide variety of databases (JDBC), flat files and web services
- Web enablement made easy (JSP)
 - Web Project Wizard
 - Deployment Assistant
- OEM Partner
- Not integrated with Business Object Enterprise
 - Missing Central Repository, Scheduling, Bursting Distribution

19 DEV-33: Get to Know Your Data



Crystal Reports for Eclipse

Report Designer

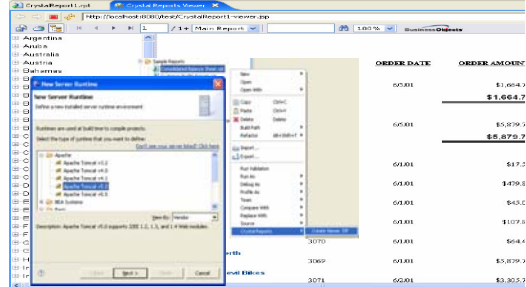


20 DEV-33: Get to Know Your Data



Crystal Reports for Eclipse

Report Designer



21 DEV-33: Get to Know Your Data



Reporting

Pentaho Reporting



- Powerful Java Report Engine
- Report Designer build on Eclipse Framework
- Broad Data Source support: JDBC, OLAP (XML/A, MDX), XML (XQuery)
- Integrated security, scheduling, portal integration, and metadata
- Web enablement made easy
- Add Hoc Reporting Interface (AJAX)
- Integrated Report Bursting



22 DEV-33: Get to Know Your Data

Reporting

Birt



- Powerful Java Report Engine
- Report Designer build on Eclipse Framework
- Connectivity to a very wide variety of databases (JDBC), flat files and web services
- Web Viewer
- Mozilla Rhino Javascript Engine
- Missing Central Repository, Scheduling, Bursting Distribution



23 DEV-33: Get to Know Your Data

Demo



24 DEV-33: Get to Know Your Data

Analysis

- Make it easy for knowledge workers to freely explore business information
 - dragging,
 - dropping,
 - drilling into, and
 - cross-tabulating data
- Provide speed-of-thought response times to complex analytical queries

Exchange 18

Analysis

Pentaho Mondrian

pentaho



- Sophisticated Relational OLAP (ROLAP) architecture with integrated, multi-level caching

Exchange 18

Analysis

Pentaho Mondrian

pentaho

- Supports leading OLAP standards (MDX, JOLAP)
- Large numbers of dimensions
- Calculations defined in the OLAP model, or requested at runtime
- Automatic conversion of OLAP requests into sophisticated SQL queries
- Aggregate-awareness
- Broad Data Source Support (JDBC 2.0)

Exchange 18

In Summary

- It's Your Data. Take Ownership.
- Not ready for the Data Warehouse? Build Data Marts.
- Centralized Reporting
- Analyze and OLTP not the best match
- Open Source and Commercial Open Source



Exchange 2008

28 DEV-33: Get to Know Your Data

For More Information, go to...

- Data Integration
 - kettle.pentaho.org
 - www.talend.com
 - www.apatar.com
- Reporting
 - diamond.businessobjects.com/creclipse
 - reporting.pentaho.org
 - www.eclipse.org/birt
- Analysis
 - mondrian.pentaho.org

marian.edu@ganimed.ro

Exchange 2008

29 DEV-33: Get to Know Your Data

Questions ?

Exchange 2008

30 DEV-33: Get to Know Your Data

