

ORACLE®



Building Of Data Integration Competency Center Master-Class

Guideline To

Health-Check Of Total Economic Impact

And Business Value Assessment

Milomir Vojvodic

Senior Business Development Manager



Agenda

- Data Integration Competency Center
And Total Economical Impact
- Data Integration Trends
- Oracle GoldenGate
- Oracle Data Integrator
- Enterprise Data Quality

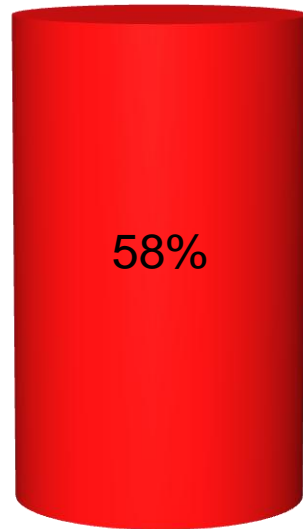
**Data Integration Competency
Center**

And

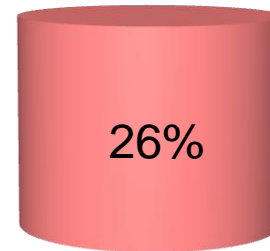
Total Economical Impact

Highest Friction Points To DI Adoption

■ Getting Budget Approval



■ Developer Resistance



Source : Info-Tech Group "Friction Points To DI Adoption"

TEI extends the concepts of ROI and TCO

ROI = 178%
Payback = 6.6 months

ROI = 167%
Payback = 6.9 months

Traditional TCO

TEI

IT impact

IT costs



IT cost savings



Business impact

User efficiency



Business effectiveness



Risk/uncertainty

Risk mitigation



Risk versus reward

Strategic impact

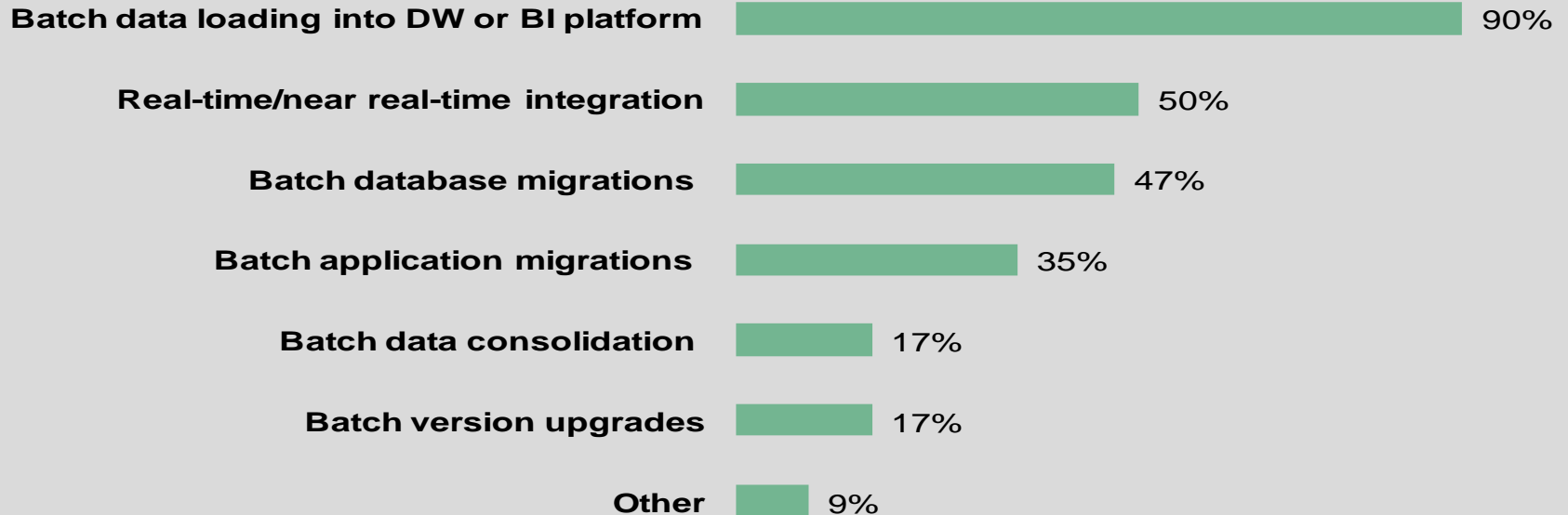
Scalability



Flexibility

Integration Scenarios

*“What are you currently using your ETL tool to support?”
(Select all that apply)*



Base: 173 professionals whose organizations have purchased ETL software
(multiple responses accepted)

Source: October 2011 Global ETL Forrester Online Survey

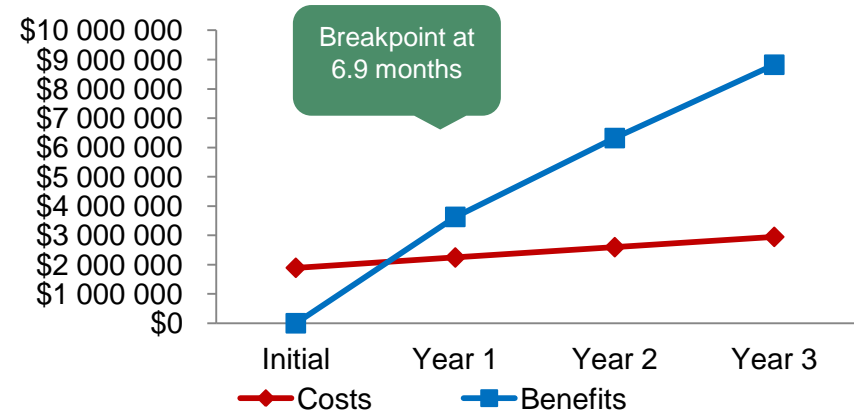
ORACLE

Executive summary

In April 2011, Oracle commissioned Forrester Consulting to examine the potential return on investment of Oracle Data Integration

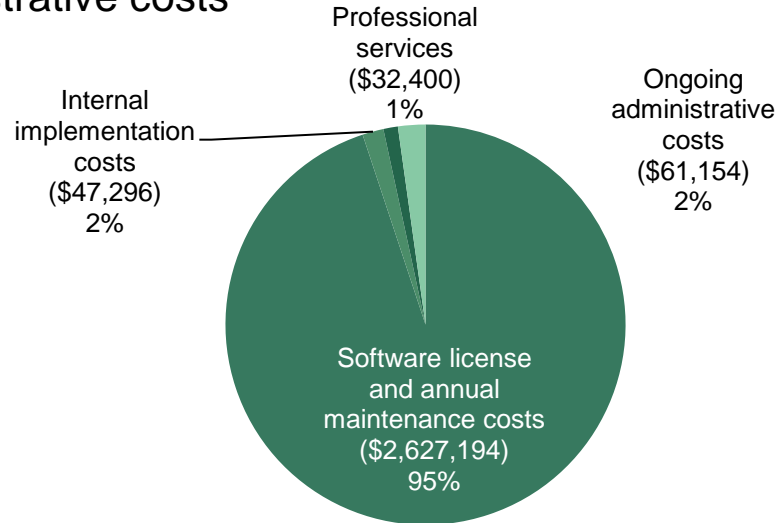
Financial metrics summary:

- ▶ Achieved risk-adjusted three year ROI of 167%
- ▶ Reached breakeven on the investment in 6.9 months
- ▶ Realized total (risk-adjusted) Net Present Value of \$4,636,329
- ▶ Used a triangular distribution low, medium, and high to risk-adjust costs and benefit estimates
- ▶ The study found that by spending **(\$2,768,044)** over three years the organization realized risk-adjusted benefits of \$7,404,373 resulting from:
 - Improvement in project completion
 - Shorter reporting cycle
 - Overall administrative and third party training cost savings
 - Deferring hardware upgrades



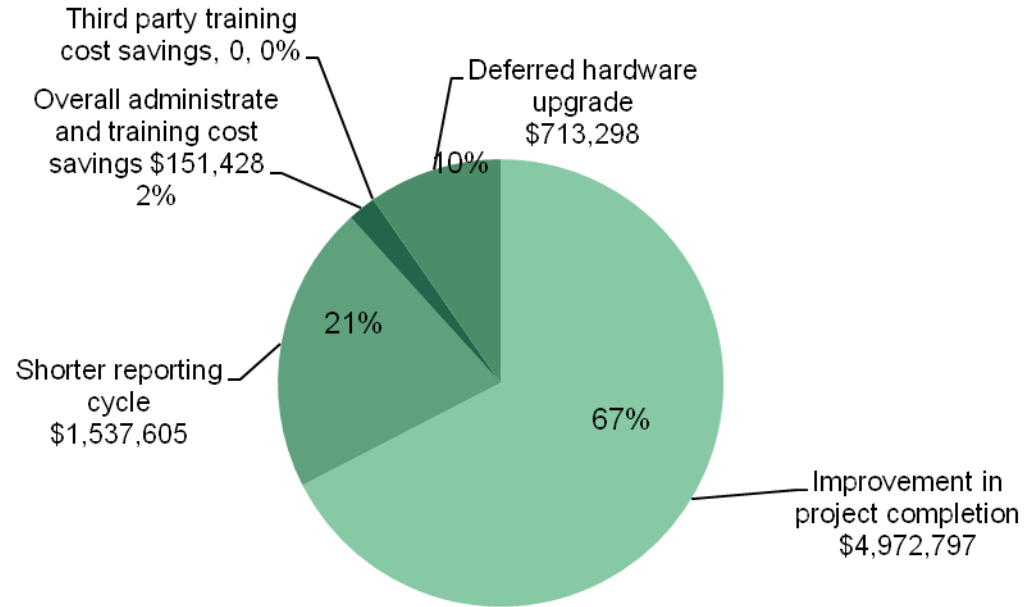
Overview of cost categories

- ▶ Software license and annual maintenance costs
- ▶ Internal implementation costs
- ▶ Professional services
- ▶ Ongoing administrative costs



Overview of benefit categories

- ▶ Improvement in project completion
- ▶ Shorter reporting cycle
- ▶ Overall administrative and savings
- ▶ Third party training cost
- ▶ Deferring hardware upgrades



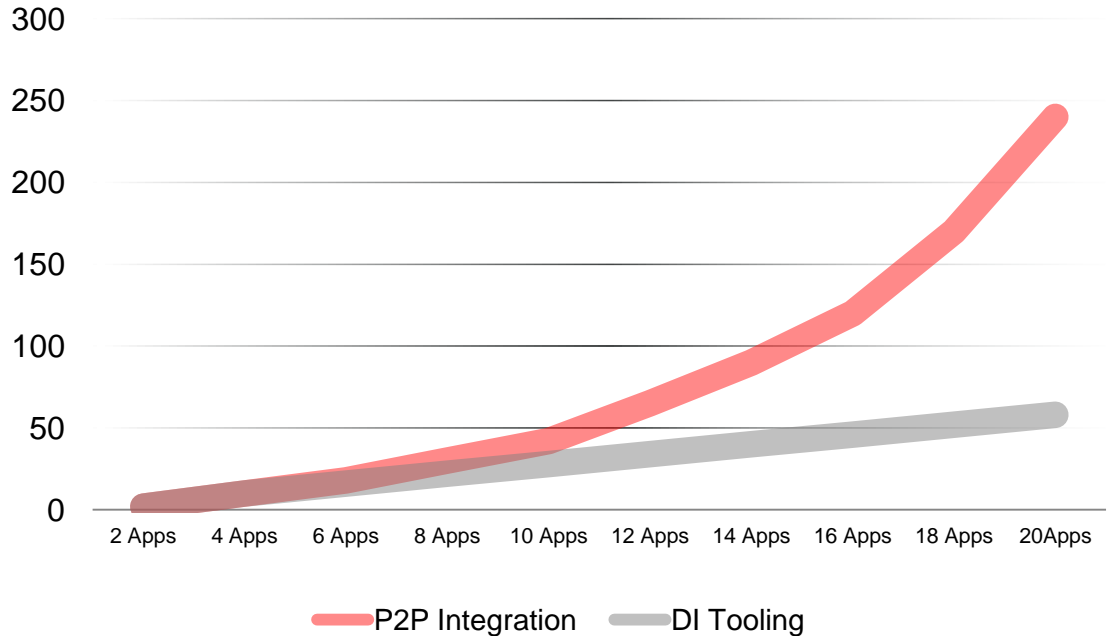
Data Integration Trends

Exponential Chaos

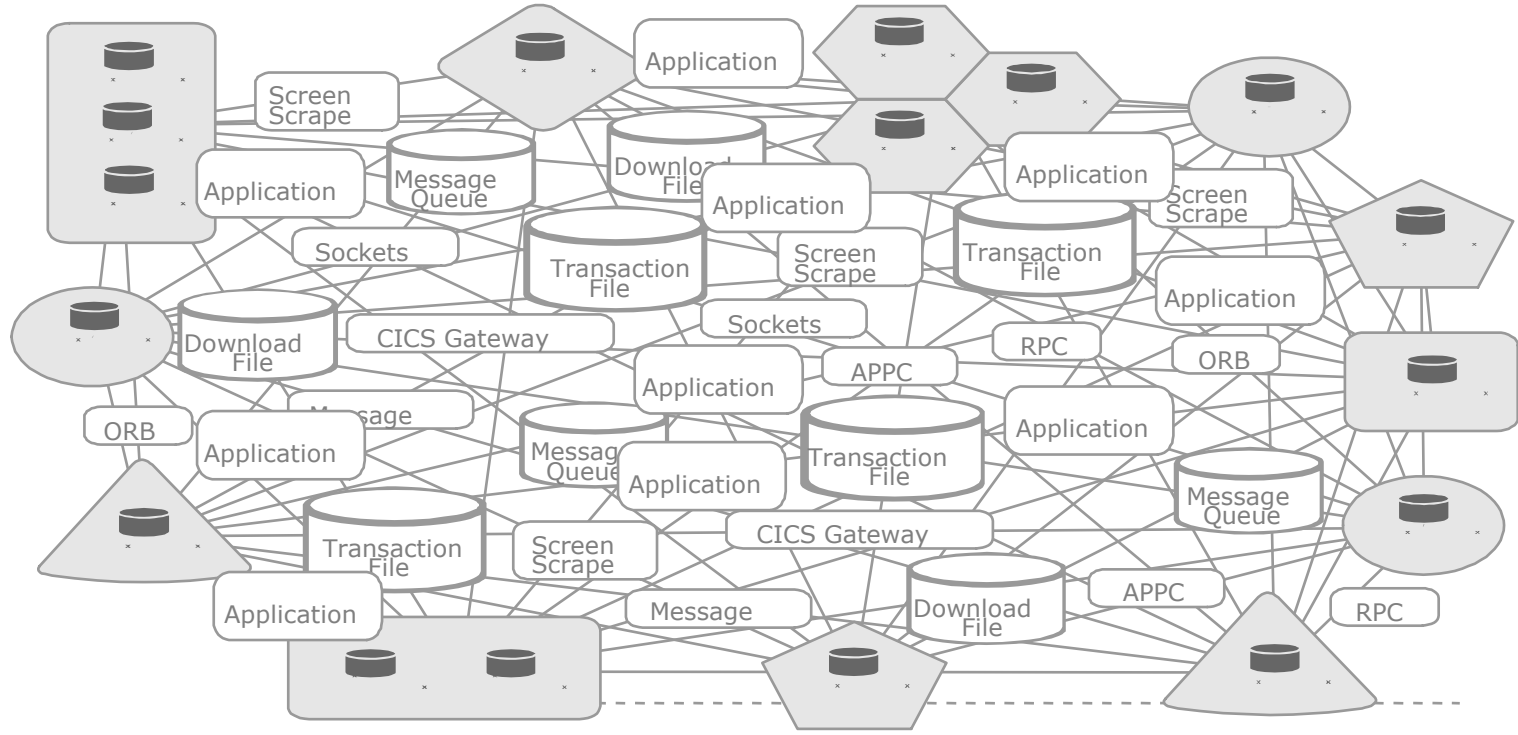
of integration scenarios

Point-to-point integration architectures cannot keep up with business growth

and represent weak links in mission critical integration scenarios.

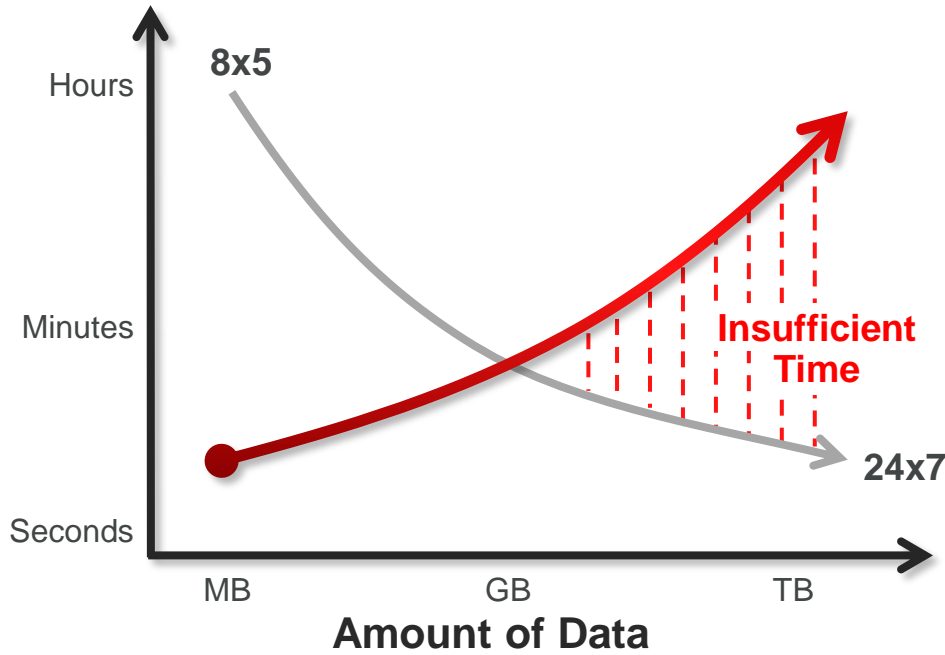


All Businesses Share Similar Challenges



Change is Inevitable and Unstoppable

Batch Processing Windows Are Shrinking



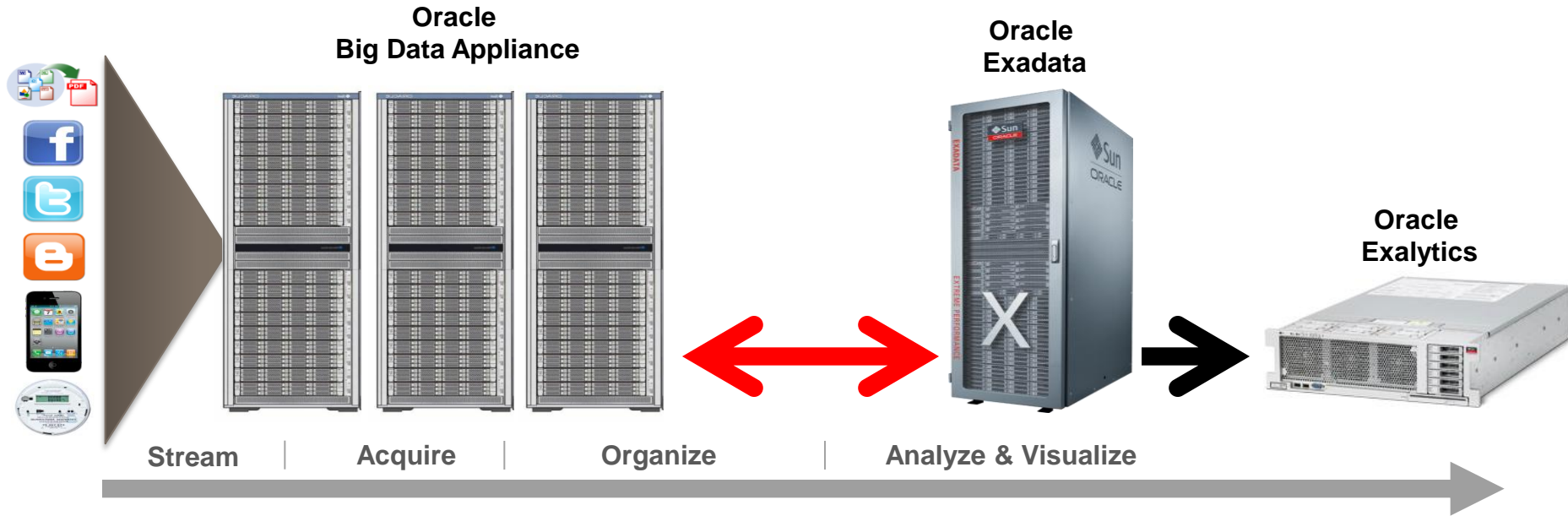
Batch processing takes too long

Global 24 x7 operations limit downtime

Processing volumes exceed batch windows

Even the best hardware will struggle eventually

Integrate Big Data with DW and Transactional Data Stores



- Load from big data processing into your data warehouse for further analysis
- Access your customer information while you process through your big data in order to look for patterns

More Of DI Trending

- Mergers And Acquisitions
- Trading Partner Network
- Business Process Automation
- Data Sharing
- Enterprise Risk And Compliance
- Application Data Migrations
- Enterprise Data Integration For Applications
- Application Data Migration

Oracle Data Integration

Oracle Data Integration Solutions

Sample list of Sources and Targets

- IBM DB2 UDB
- IBM DB2 z Series
- IBM DB2 i Series
- Enscribe
- SQL/MP
- SQL/MX
- MySQL
- Sybase ASE
- Informix
- JMS
- Teradata
- Netezza
- Hadoop
- Sybase IQ
- TimesTen,
- PostgreSQL
- Greenplum,
- HP Neoview
- SAS
- Salesforce
- SAP ERP & BW
- Generic SQL
- Hypersonic SQL
- Microsoft SQL
- Oracle Database
- Oracle Exadata
- Oracle Big Data Appliance
- Enterprise Data Quality
- Oracle E-Business Suite
- JD Edwards Enterprise One
- Oracle Enterprise Service Bus
- Oracle Hyperion
- Oracle OLAP
- Oracle PeopleSoft
- Oracle Siebel CRM
- Oracle Communications BRM
- Microsoft Access
- Oracle BI EE
- Oracle BI Apps
- Linux
- Sun Solaris
- Windows 2000, 2003, XP
- HP NonStop
- HP-UX
- IBM AIX
- zLinux
- ...

Roadmap For Streams & OWB

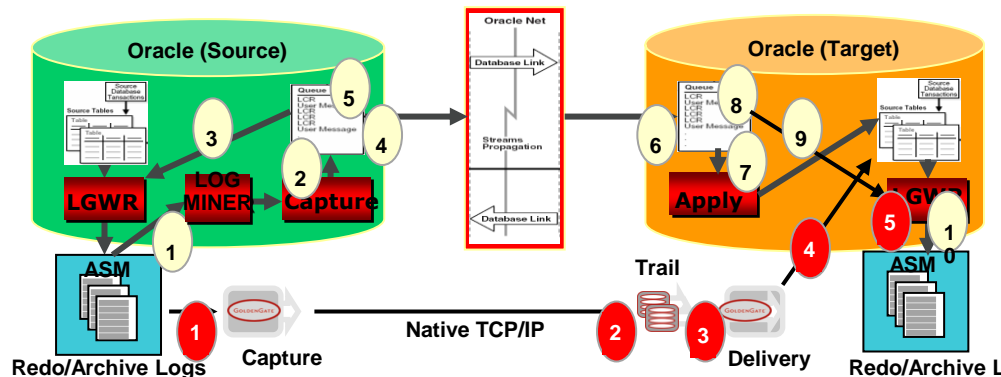
Streams 8000 throughput in 6 seconds. OGG 2.7 seconds!

Features	OWB basic	ODI-EE
Oracle DB Targets	Y	Y
Non-Oracle Sources	Oracle Gateways	Y
Slowly Changing Dimensions (type 1)	Y	Y
OLAP Targets (Oracle + Essbase)	Oracle OLAP only	Y
Built-in Scheduler	Oracle DBMS only	Y
Non-Oracle DB Targets	-	Y
Changed Data Capture (including OGG integration)	-	Y
Web Services (API & Src/Trgt) & XML Targets	-	Y
Data Lineage / Impact Analysis	-	Y
Declarative Design Environment	-	Y
Slowly Changing Dimensions (2 & 3)	-	Y
Load Plans, Parallel Execution and Restartability	-	Y
JEE Deployment , Clustering and High Availability	-	Y
Java based SDK for full Programmatic Control	-	Y
Unified Management and Administration with OEM	-	Y
Pre-built Integration with OBI-EE, Hyperion, SOA/AIA, etc.	-	Y

Source Server (CPU %)	Through-put (ops/sec)	Streams Latency (secs)	GG5 Latency (secs)
~80%	8,000	6	2.7
~45%	5,000	2.7	1.5
~20%	2,500	2	1.3

Streams 9 steps configuration. OGG 5 steps!

OWB missing these functionalities. ODI has them!

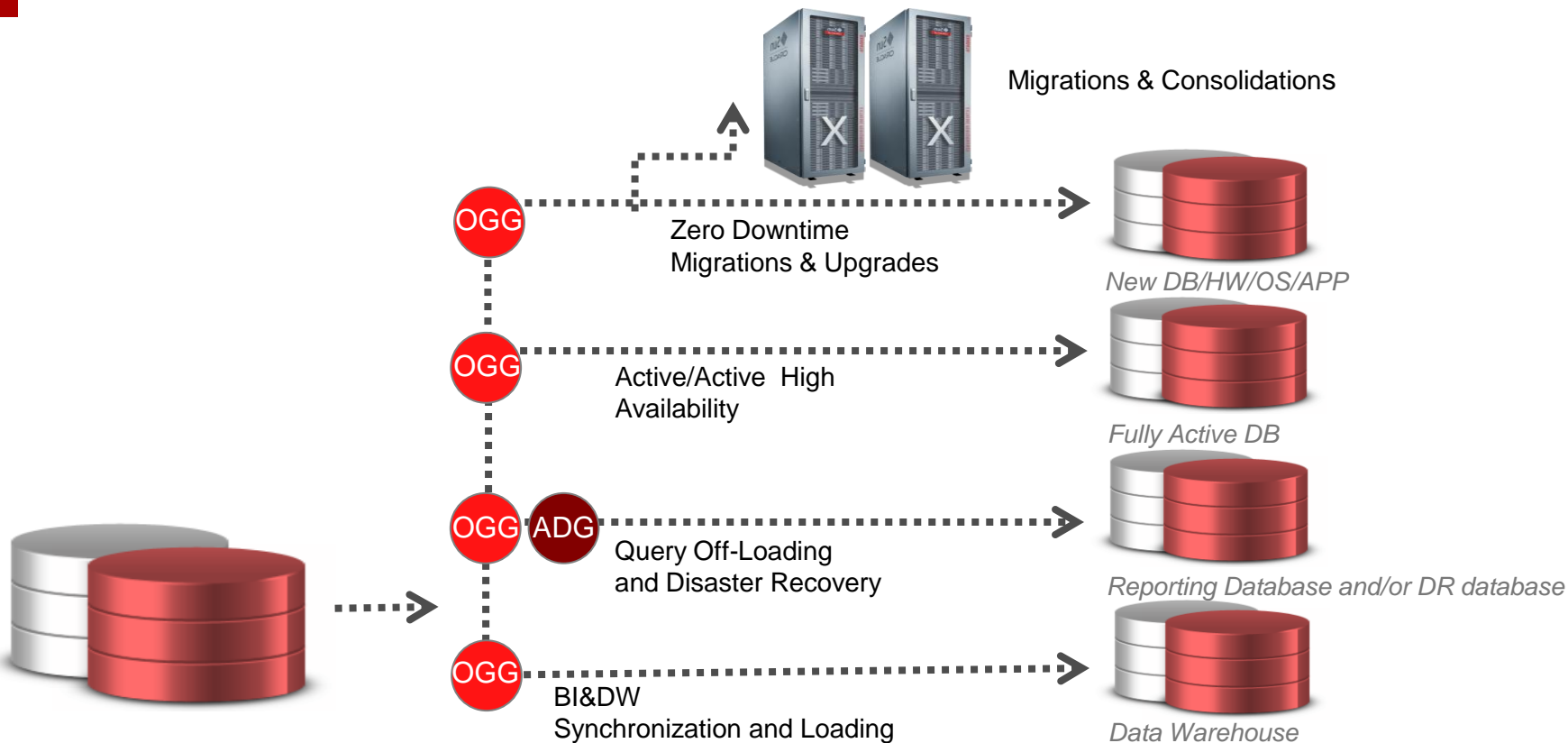


Oracle GoldenGate

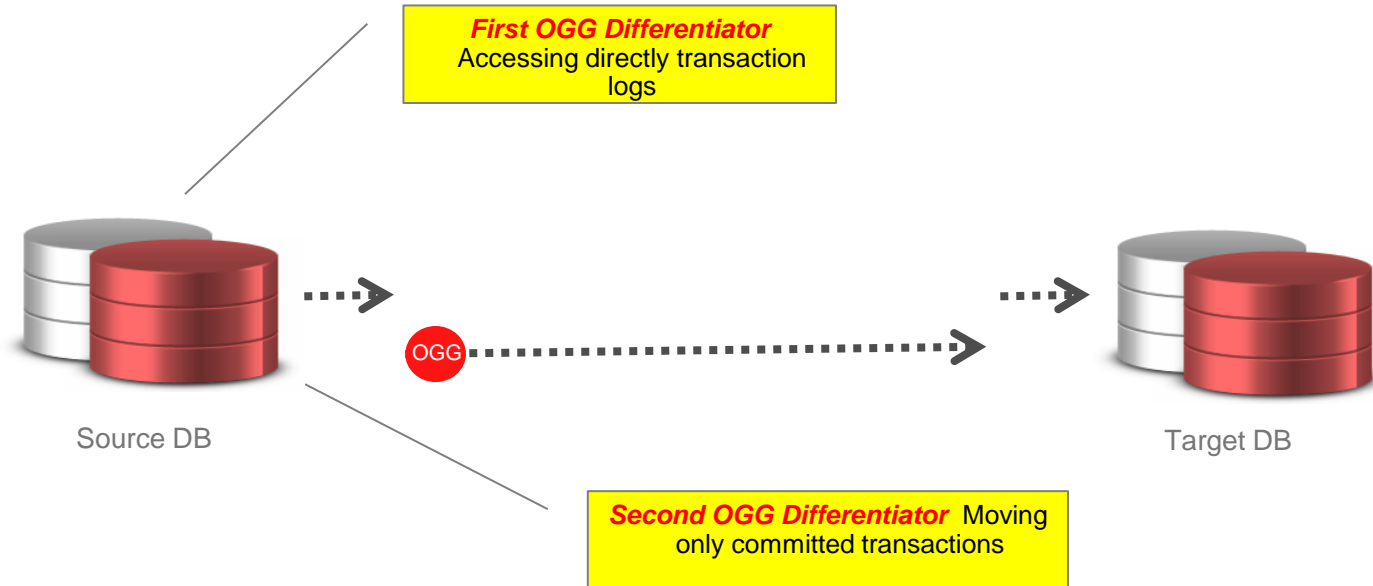
What is OGG?



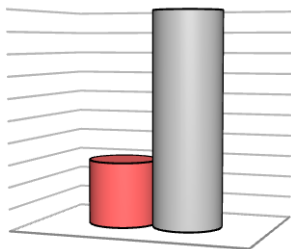
Use Cases Of OGG



Why Is OGG Different?



Alternative To Batch Window



Reduce source system overhead (and costs for stronger HW) by 70%

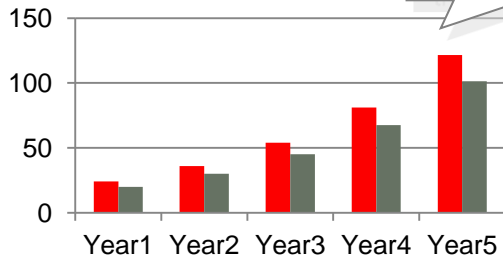
■ After OGG
■ Before OGG

OR

First OGG Differentiator
Accessing directly transaction logs

TIME REQUIRED FOR THE END OF DAY PROCEDURE

Hours

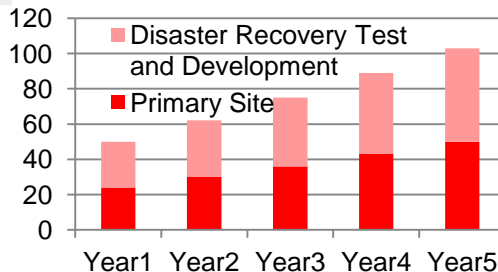


Daily load time can reach 5 days with the current HW

Currently during the End Of Day utilizes the Server CPU by 40-50% and the IO by 90%. Probably the IO is the bottleneck.

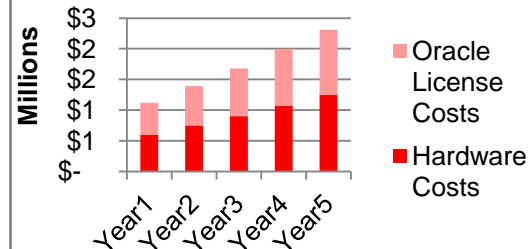
NO OF CPUs REQUIRED FOR SAME PERFORMANCE*

No Of Required CPUs



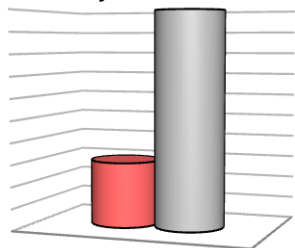
ESTIMATED COSTS FOR SERVER AND LICENSE**

Estimated Cost of Purchase in USD

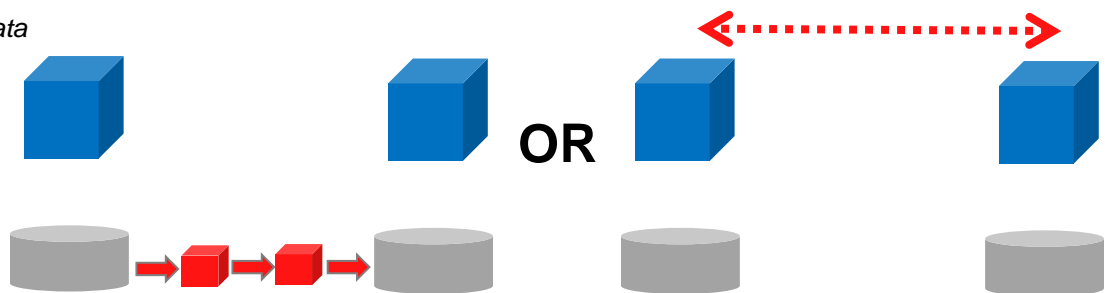


Alternative To Storage Replica

Reduce costs and efforts of data loss by 70%



■ After OGG
■ Before OGG



Second OGG Differentiator
Moving only committed transactions

	Traditional	OGG
Platform	Like-to-like	Heterogeneous databases and platforms
Topology	One-to-one topology	Many-to-many topology
Integrity	No integrity	Guaranteed data integrity and optimized data movement
Data corruption	Data corruption related to target	Data corruption is isolated at source
Rollback	No rollback	Selective and dynamic rollback with advanced point-in-time recovery
Target system	Target system unavailable	Both source and target systems are available
Data transformation	No data transformation capability	Data transformation and enrichment
Geographic distance	Geographic distance between data source and target	No distance limitations

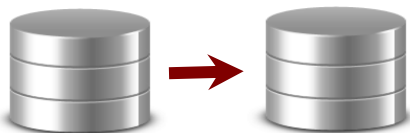
Oracle GoldenGate Supported Platforms

Databases	O/S and Platforms
<p>Oracle GoldenGate Capture:</p> <ul style="list-style-type: none"> ▪ Oracle ▪ DB2 for v 9.7 ▪ DB2 for v 10 on z/OS NEW ▪ Microsoft SQL Server 2008 ▪ Sybase ASE, 15 ▪ Teradata ▪ Enscribe ▪ SQL/MP ▪ SQL/MX ▪ MySQL ▪ JMS message c <p>Oracle GoldenGate Replicat:</p> <ul style="list-style-type: none"> ▪ All listed above ▪ TimesTen, IBM System I, Netezza & Greenplum ▪ ETL product 	<ul style="list-style-type: none"> Linux Sun Solaris <div data-bbox="459 380 1537 882" style="background-color: red; color: white; padding: 10px; text-align: center;"> <p>New in Oracle GoldenGate 11g Release 2</p> <ul style="list-style-type: none"> ▪ IBM i Series ▪ Postgres ▪ MySQL v 5.5 ▪ IBM AIX v7.1 ▪ Sun Solaris v11 </div>

Directions

Unidirectional

Query Offloading
Zero-Downtime Migration



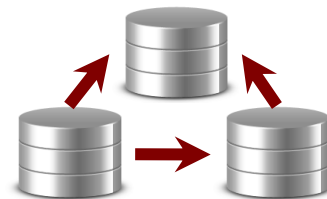
Bi-Directional

Hot Standby or
Active-Active for HA



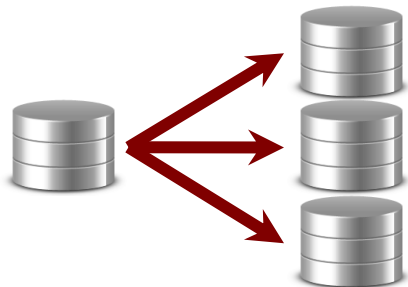
Peer-to-Peer

Load Balancing
Multi-Master



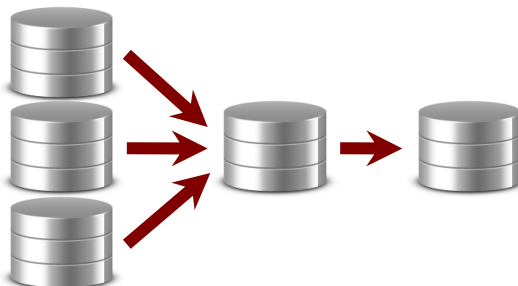
Broadcast

Data Distribution



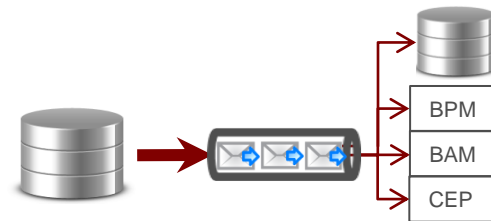
Integration/Consolidation

Data Warehouse

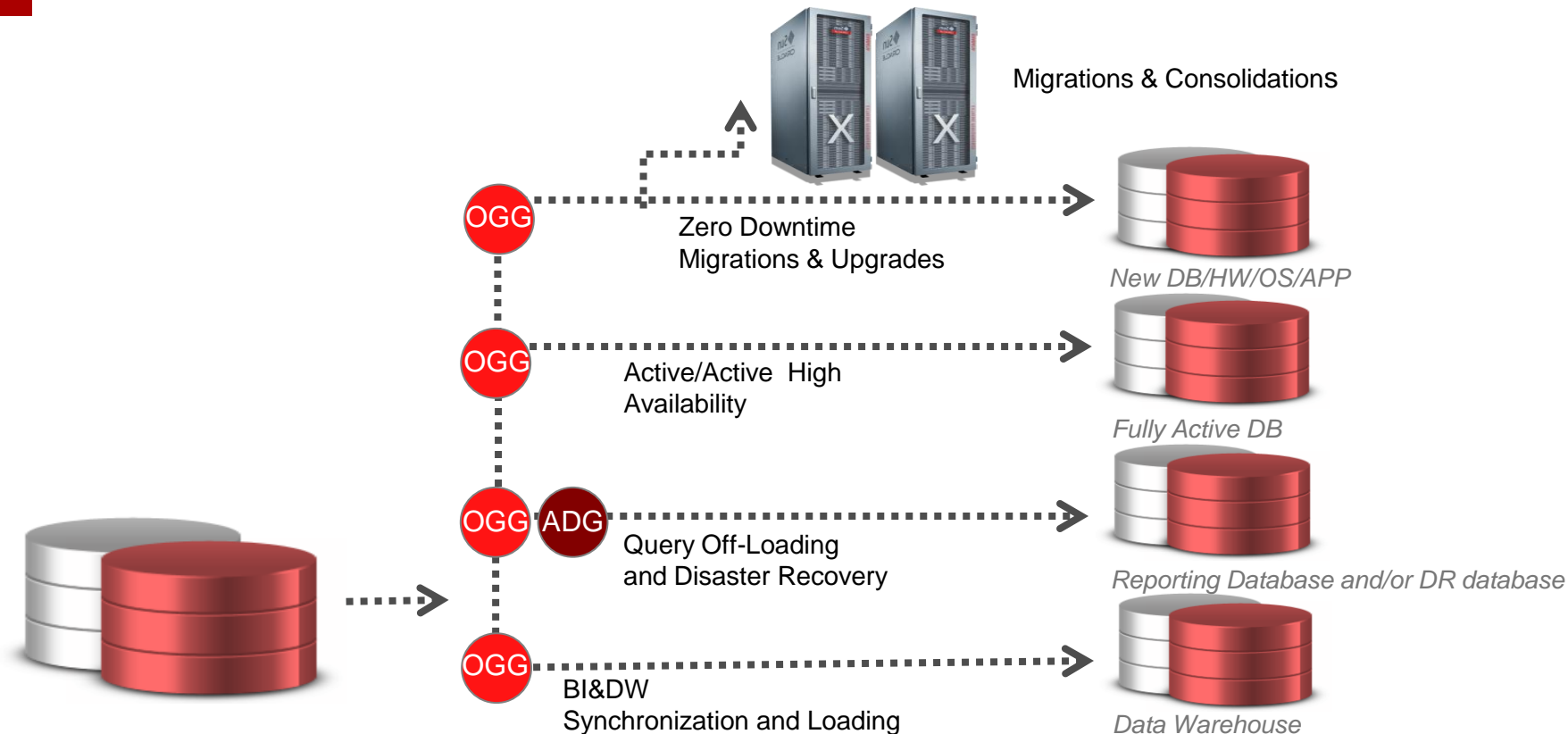


Data Distribution

via Messaging

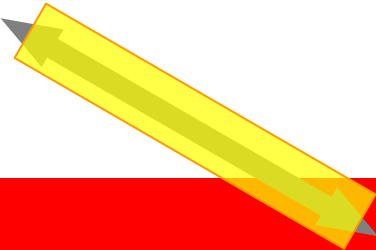
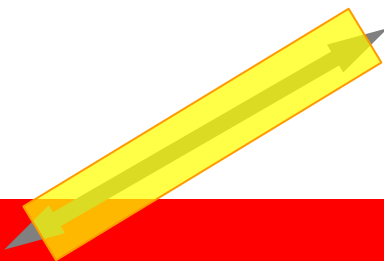


Use Cases Of OGG



Reusability Of DIS Licenses

Product Needs To Be Available – No Issues



- ***Need Planned Outage***

- Upgrade
- Migration
- Maintenance

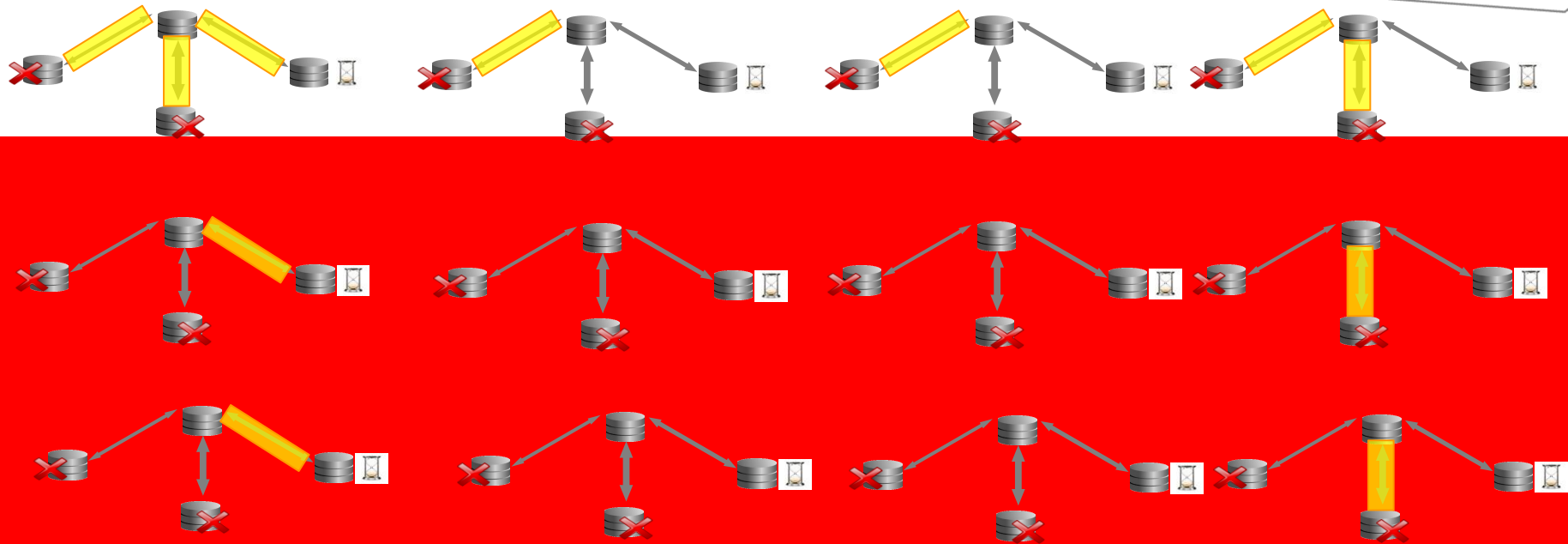
- ***Need Additional Computing Resources***

- Performance issues
- Growth

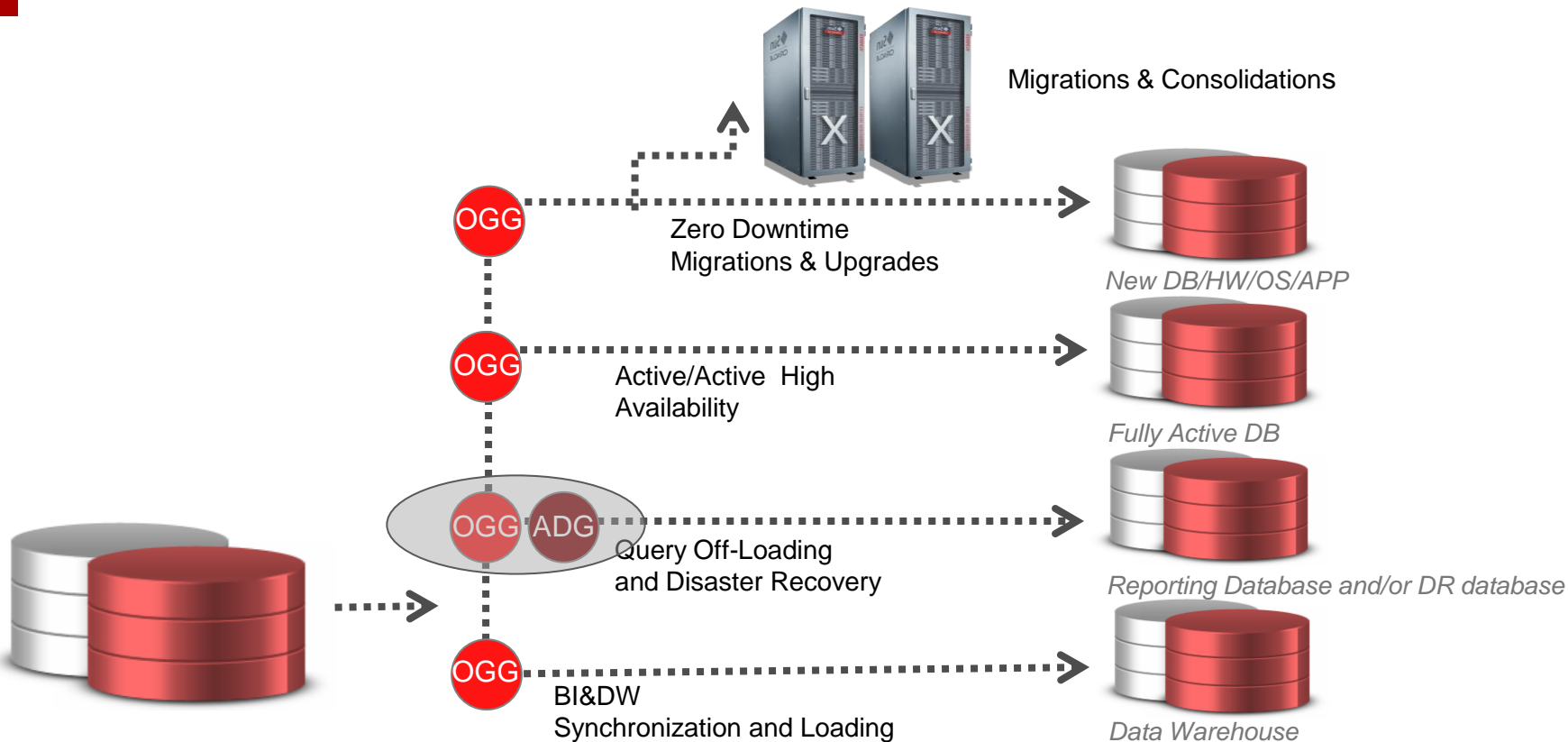
- ***Protect From Unplanned Outage***

- System Failure
- Data Failure

Reusability Of DIS Licenses



Use Cases Of OGG



OGG Offload Reporting Benefits

Reduce the workload on production

- Annual amortization of HW purchase 250 000 USD (1 000 000 USD per 4 years)
- Annual HW maintenance 100 000 USD
- Annual SW maintenance 200 000 USD
- Energy, floor space 200 000 USD
- Overall 800 000 USD, as there are 36 CPUs, Annual cost 22 000 USD per CPU

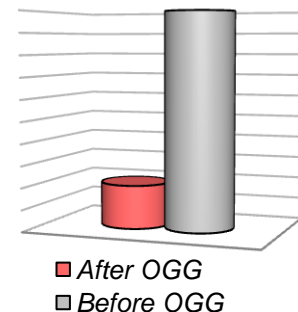
- Overhead 8% with other solution, overhead 3% with GoldenGate
- Cost of overhead 64 000 USD vs 24 000 USD
- Savings 40 000 USD per year

Alternative option for target server

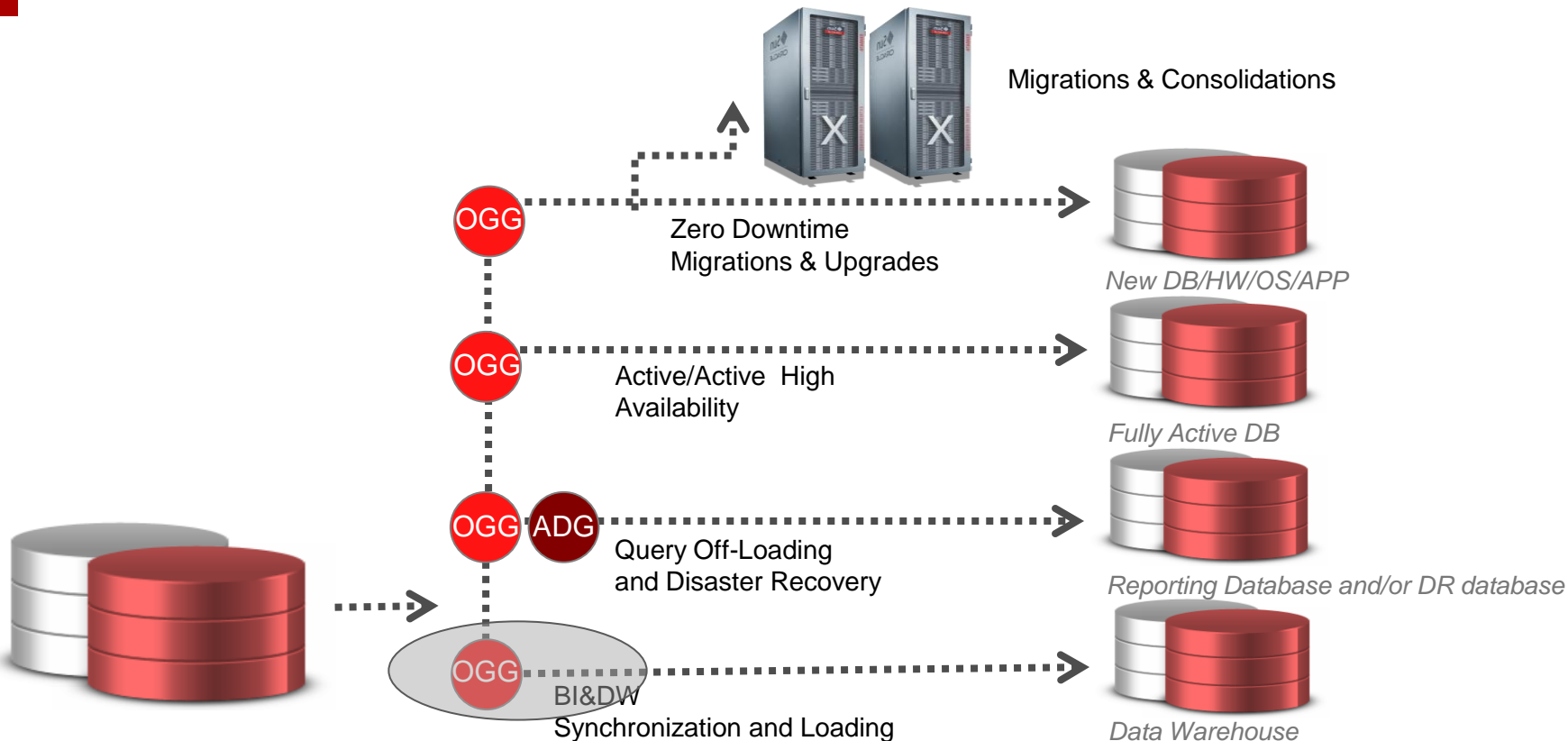
- DB software first option 100 000 USD, DB software second option 50 000 USD,
- Annual DB software support 22 000 USD vs 11 000 USD



Decrease TCO of report server
by 80%



Use Cases Of OGG



Example Of DI Health Check In Telco

- 1-lowest impact on production (transactional logs, no batch windows)
- 2-performance and low bandwidth (filtering-even on delivery, compression, playing with modularity-grouping-splitting)
- 3-flexibility in secondary server configuration
- 4-logical replica

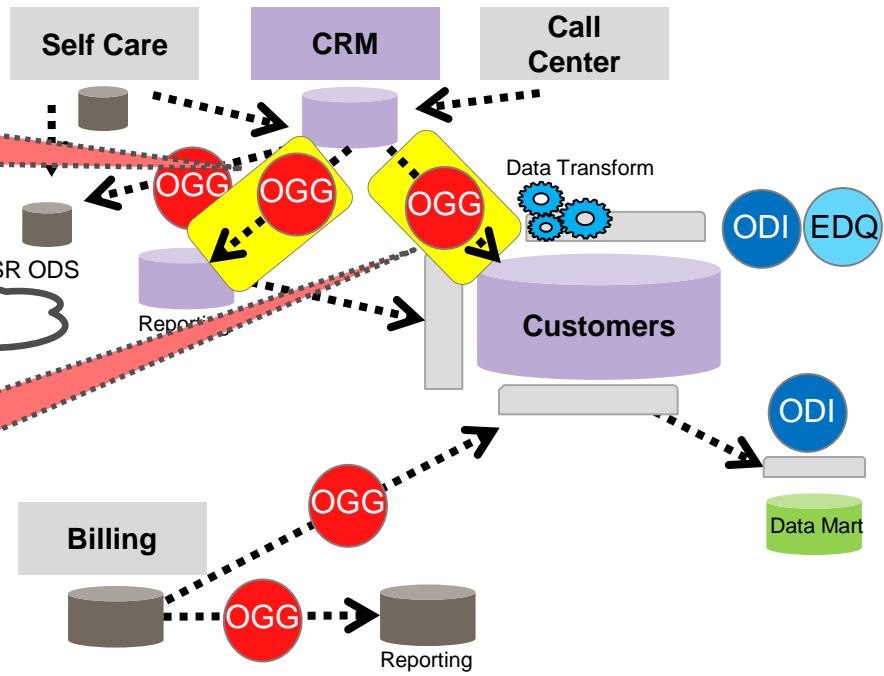


- 5-lowest impact on production
- 6-moving data without errors (no data failure causes replicate-f.e.block level failures or corrupted data)
- 7-can stand such highly transactional systems (trail files-no cashing, playing with modularity)

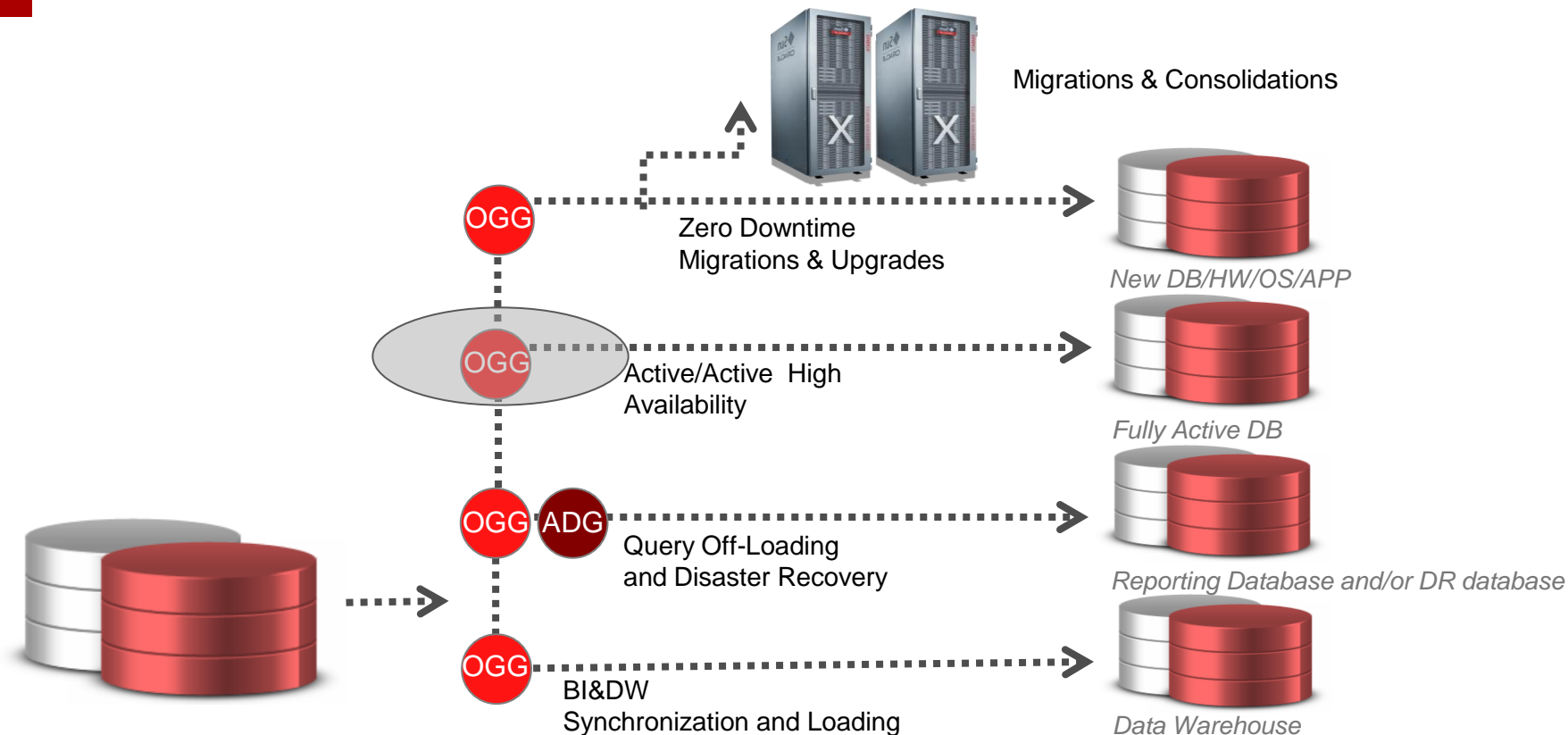
Customer Care

Collect Usage

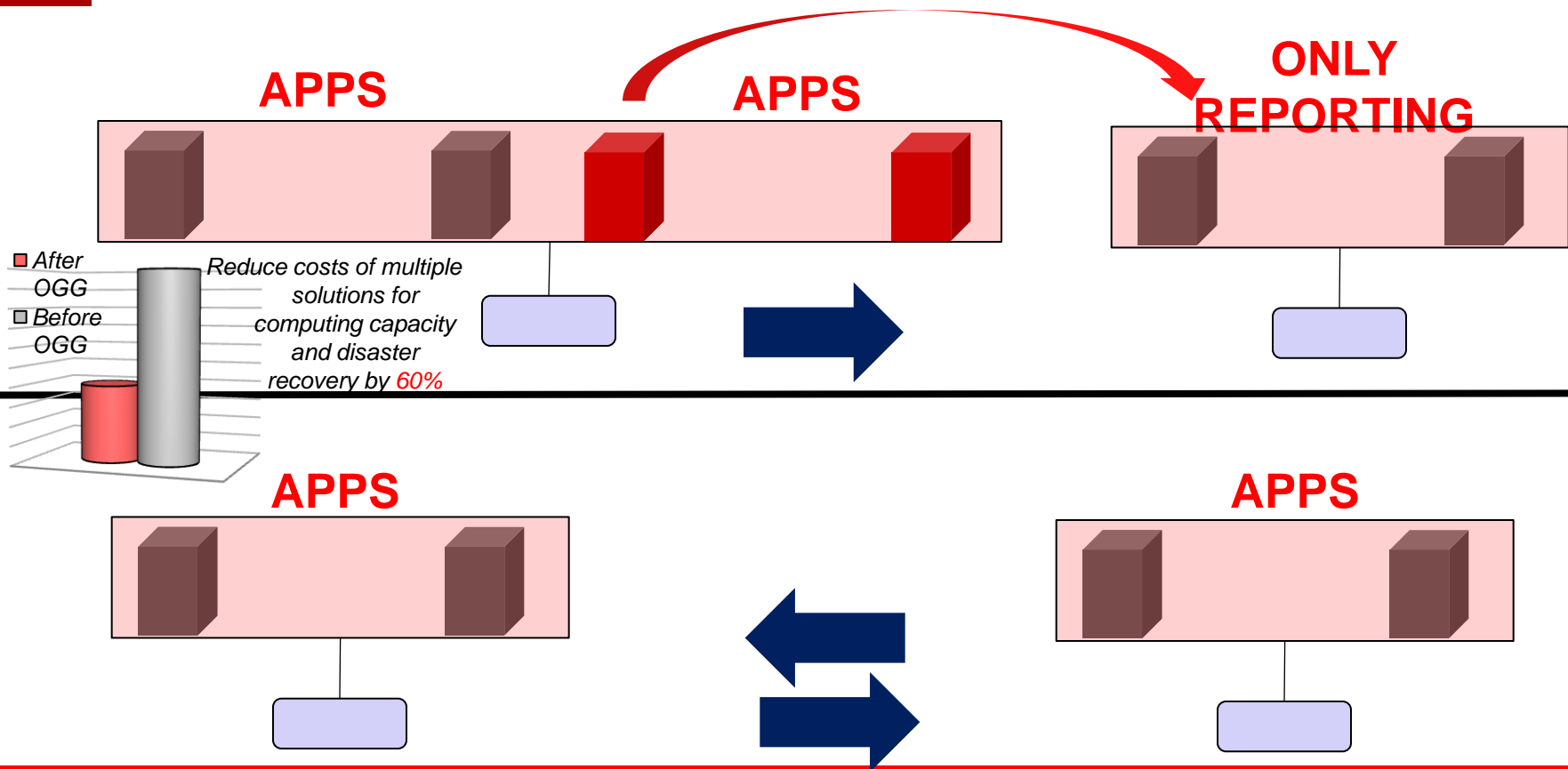
Billing



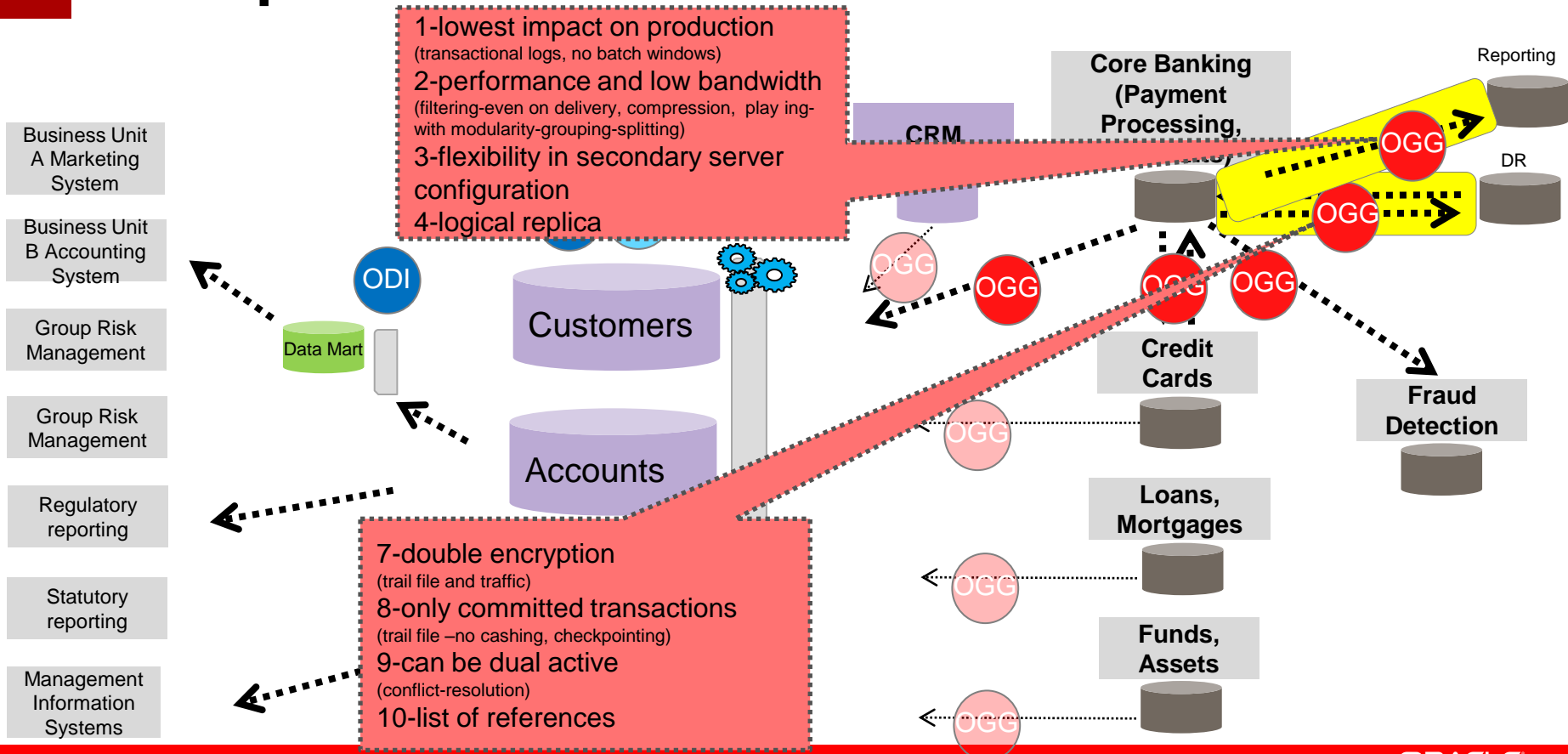
Use Cases Of OGG



OGG Benefits For Dual Active DBs

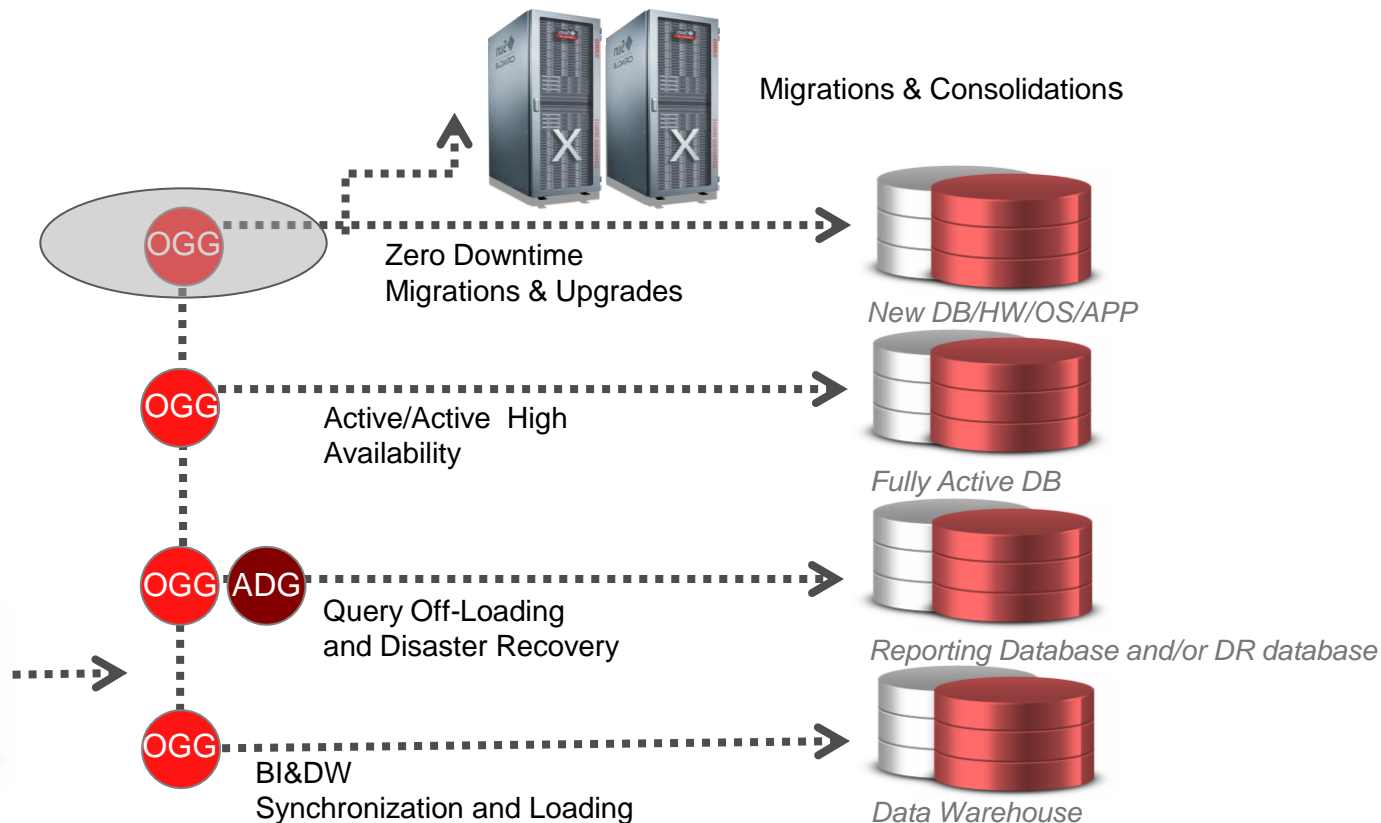
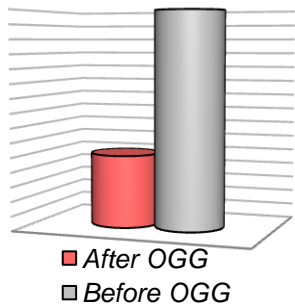


Example Of DI Health Check In Bank

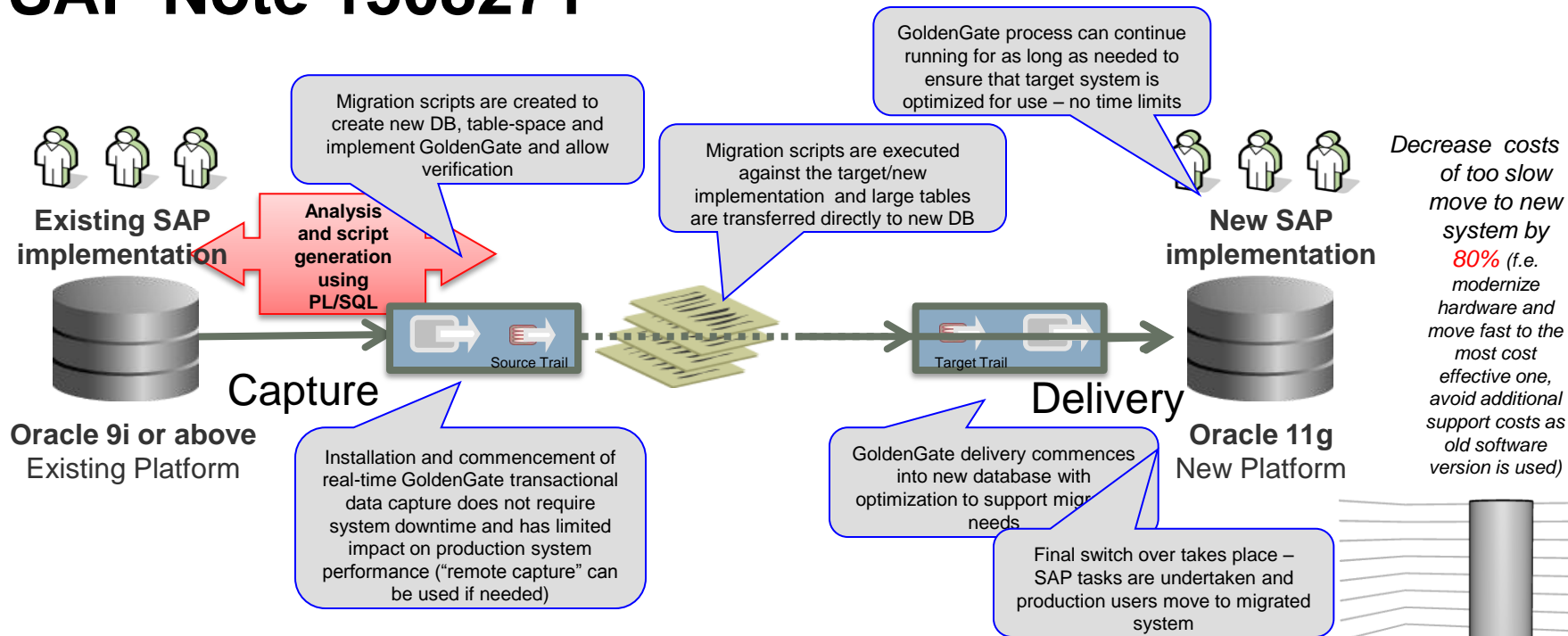


Use Cases Of OGG

Reduce migration efforts
and costs by **80%**

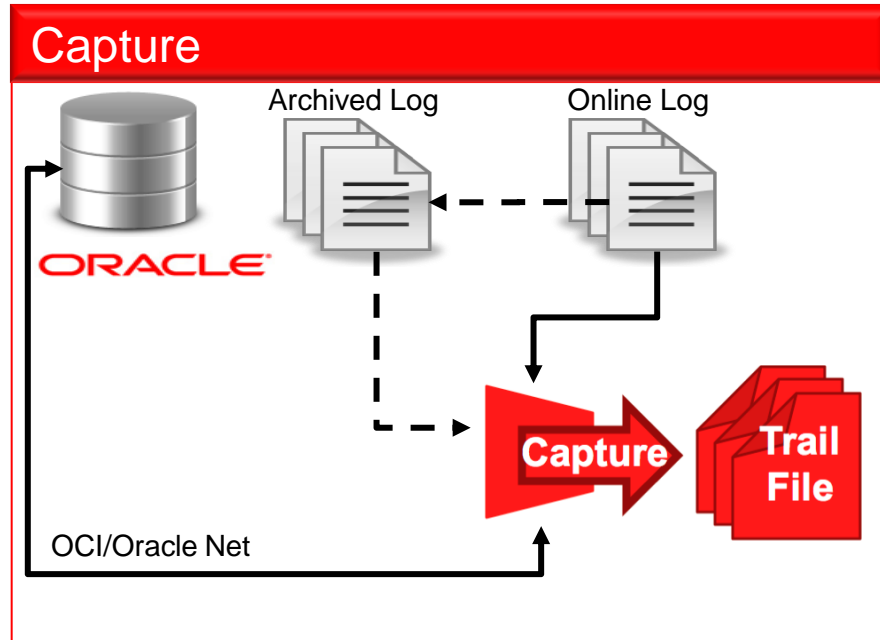


SAP Note 1508271



- Minimum Oracle database version for the production system is 9.2.0.8
- Oracle version on the target can be 10.2 or 11.2 and any UNIX, Linux or Windows platform is supported
- All SAP products and solutions where Oracle database is certified are supported (such as R/3, BW, CRM or XI)
- Supplemental Logging must be enabled on the source database prior to Oracle GoldenGate implementation
- The source Oracle database cannot contain any OLTP compressed tables

OGG In Oracle Strategy



- Support for Oracle RAC
- Support for ASM
- Support for index-organized tables (IOTs) with overflow
- Support for clustered tables
- Support for object tables
- Support for object types (UDTs)
- Support for DDL operations
- Support for direct loads (append hint)
- Archived log only (ALO) mode
- Off-platform capture (LOGSOURCE)
- Multi-threaded capture

OGG Director

Oracle GoldenGate Director:
Has basic configuration,
management, monitoring,
and alerting.

Based on legacy pull infrastructure.
Works on legacy
GoldenGate instances.

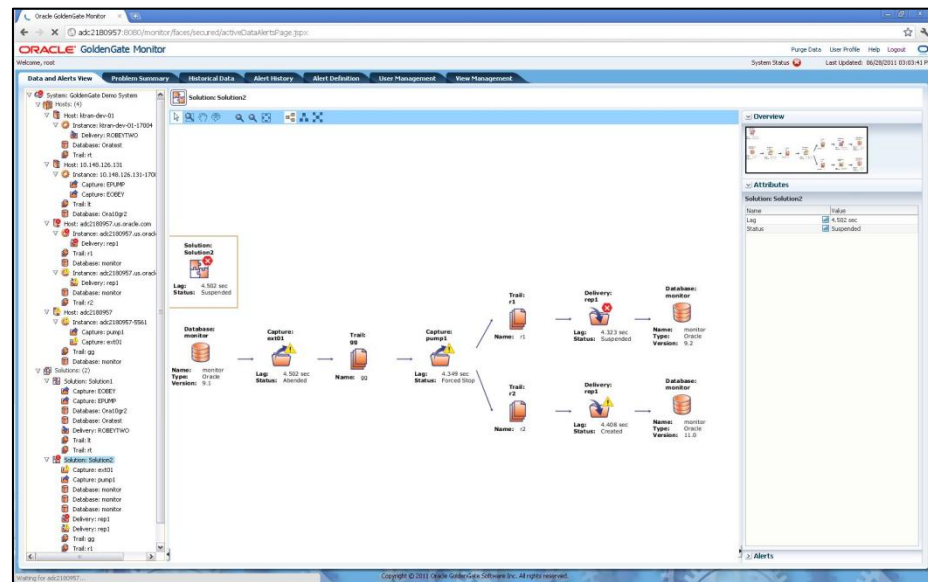
The screenshot displays the GoldenGate Director 1.0 interface. The top window shows a data flow diagram for a 'DirectorBackup' test. The flow starts with a 'Data Source LocalOracle', goes through a 'Capture: DIRBACKUP LocalOracle' (Lag: 00:00:00), then through a 'Map for DIRBACKUP Trail: ...\ogg-odbc\Vendor\O' and 'LocalOracle' to a 'Delivery: DIRREP LocalOracle' (Lag: 00:00:00). From there, it goes through another 'Map for DIRREP LocalOracle' to a 'Data Source LocalOracle'. A second path goes from the 'Delivery: DIRREP' to a 'Delivery: SCNDEL LocalOracle' (Lag: 00:00:00), which then goes through a 'Map for SCNDEL LocalOracle' to another 'Data Source LocalOracle'. The bottom window shows a log table with the following data:

Level	Source	Process	Type	Date/Time	Message
INFO	LocalOdbc	port 8324	SERVER	2005-05-17 11:50:21	GoldenGate Collector, port 832
INFO	LocalOdbc	port 8324	SERVER	2005-05-17 11:50:16	GoldenGate Collector, port 832
INFO	LocalOdbc	MANAGER	MANAGER	2005-05-17 11:50:16	GoldenGate Manager for ODBC
INFO	LocalOdbc	MANAGER	MANAGER	2005-05-17 11:50:16	GoldenGate Manager for ODBC
INFO	LocalOracle	MANAGER	MANAGER	2005-05-17 11:50:16	GoldenGate Manager for ODBC
INFO	LocalOracle	DIRBACKUP	EXTRACT	2005-05-17 11:50:16	GoldenGate Capture for Oracle
ERROR	LocalOdbc	DIRREP	REPLICAT	2005-05-17 11:50:12	GoldenGate Delivery for ODBC
ERROR	LocalOdbc	DIRREP	REPLICAT	2005-05-17 11:50:12	GoldenGate Delivery for ODBC
WARN	LocalOdbc	DIRREP	REPLICAT	2005-05-17 11:50:12	GoldenGate Delivery for ODBC
WARN	LocalOdbc	DIRREP	REPLICAT	2005-05-17 11:50:12	GoldenGate Delivery for ODBC
INFO	LocalOracle	DIRBACKUP	EXTRACT	2005-05-17 11:50:12	GoldenGate Capture for Oracle
INFO	LocalOdbc	DIRREP	REPLICAT	2005-05-17 11:50:11	GoldenGate Delivery for ODBC
INFO	LocalOdbc	MANAGER	MANAGER	2005-05-17 11:50:11	GoldenGate Manager for ODBC
INFO	LocalOdbc	GGSCI	GGSCI	2005-05-17 11:50:11	GoldenGate Command Interpreter

OGG Monitor

Oracle GoldenGate Monitor:
Has advanced monitoring,
alerting (SNMP support),
lag graphs,
and historical repository.

Based on new agent
infrastructure,
which will be the foundation for all future integration projects.



OGG Economic Impact - Reduce Negatives

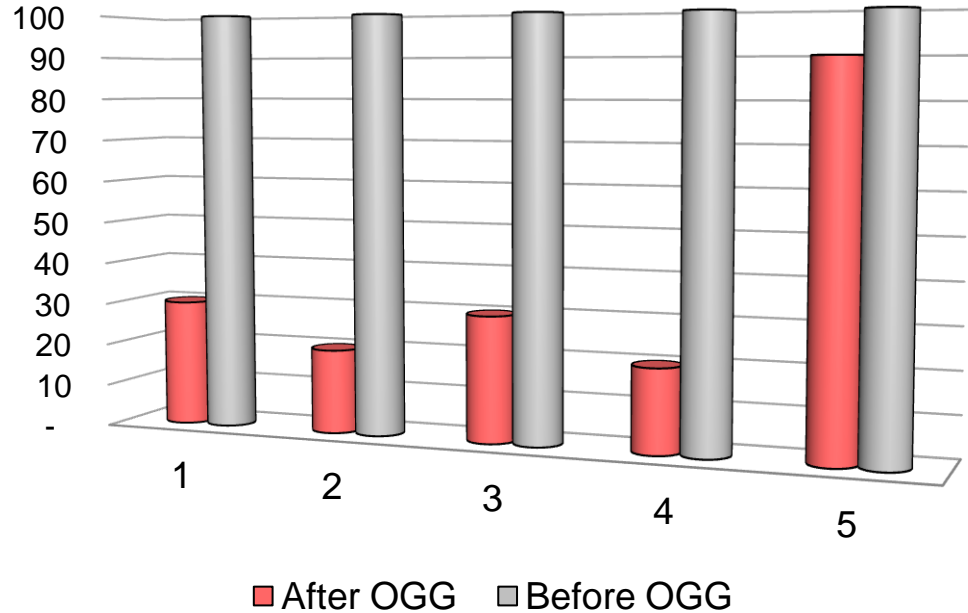
1-Reduce downtime (end user productivity and customer satisfaction) by **70%**

2-Reduce migration efforts by **80%**

3-Reduce source system overhead (and costs for stronger HW) by **70%**

4-Decrease TCO of report server by **80%**

5-Communication cost savings and no need for expensive network infrastructure upgrades by **10%**

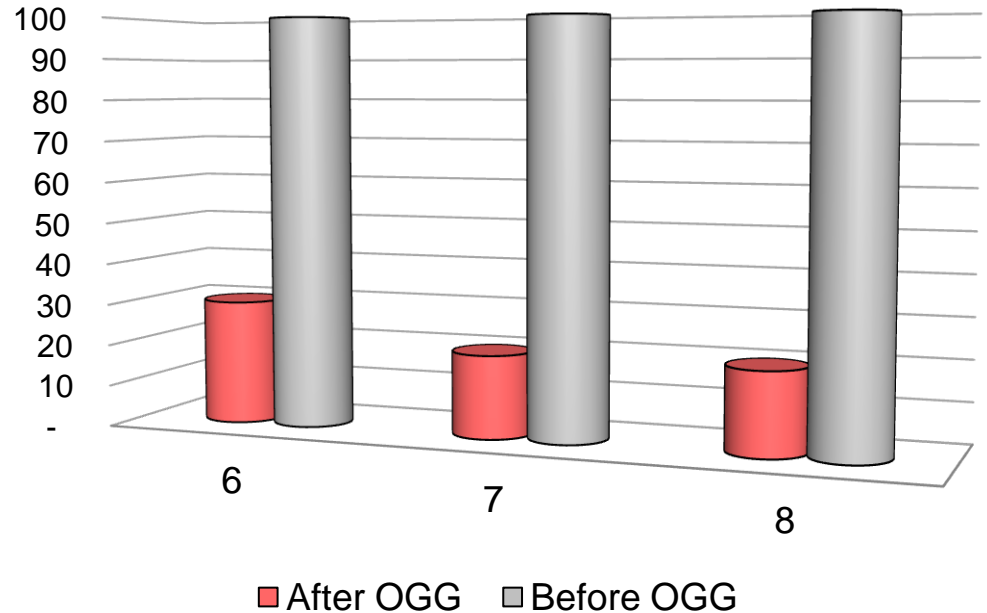


OGG Economic Impact - Reduce Negatives

6-Reduce costs and efforts of data loss by **70%**

7-Decrease the manpower needed for system maintenance and recovery by **80%**

8-Decrease costs of too slow move to new system by **80%** (f.e. modernize hardware and move fast to the most cost effective one, avoid additional support costs as old software version is used)

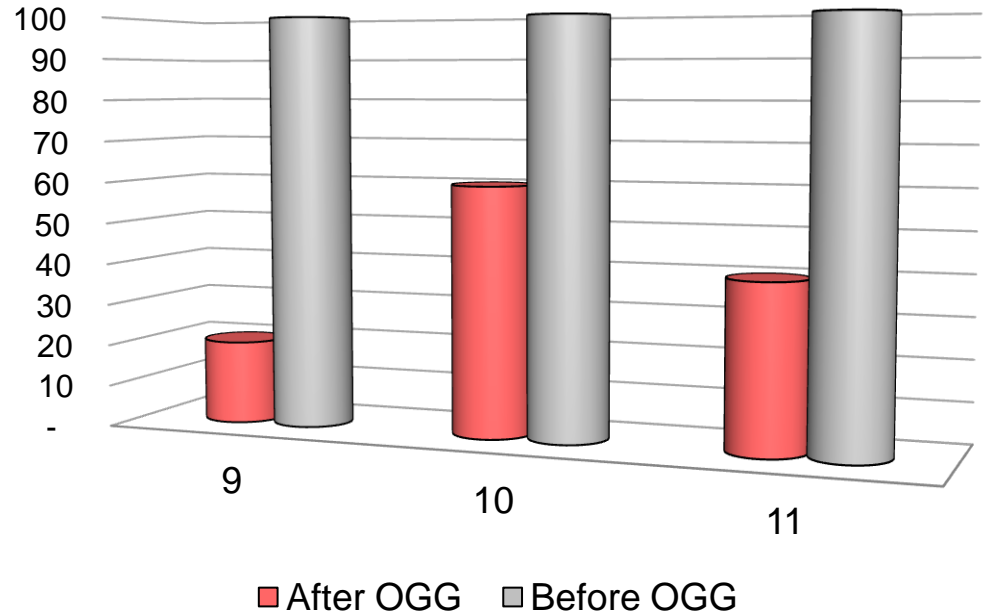


OGG Economic Impact – Reduce Negatives

9-Reduce the manpower needed for batch loading by **80%**

10-Reduce costs of expensive and heavy read only queries that are taking place on OLTP system by **40%**

11-Reduce costs of multiple solutions for computing capacity and disaster recovery by **60%**

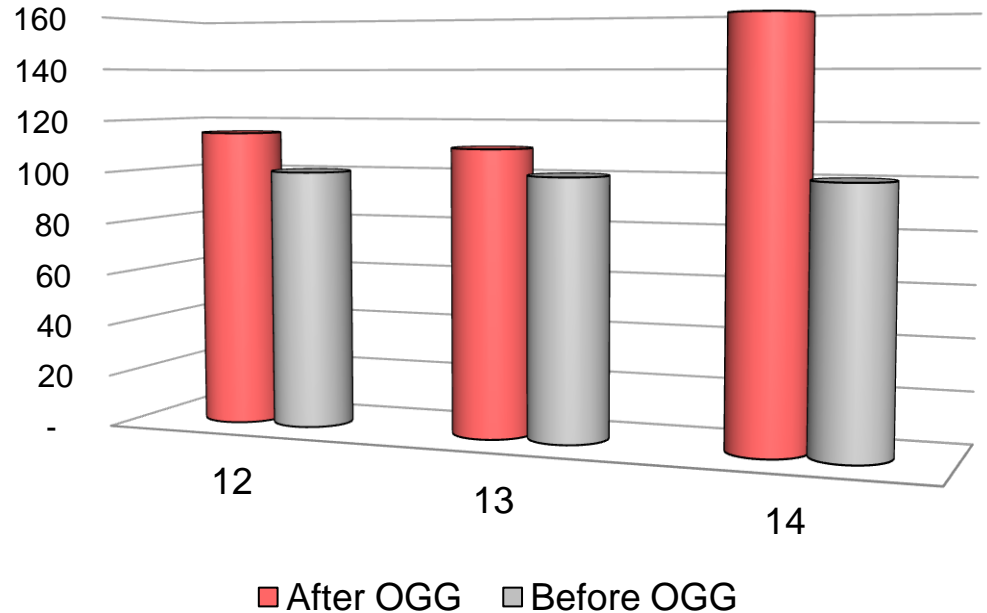


OGG Economic Impact – Improve Positives

12- Increase number of new customers as improve customer loyalty and brand equity by **15%**

13-Gain by **10%** more revenue due to agility to quickly react on ongoing market opportunities

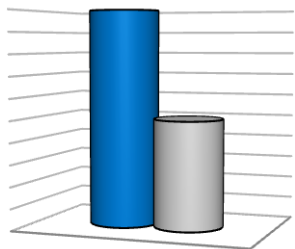
14-Improve speed of implementation and go-live with TDM solution by **60%** (as so fast and easy deployment)



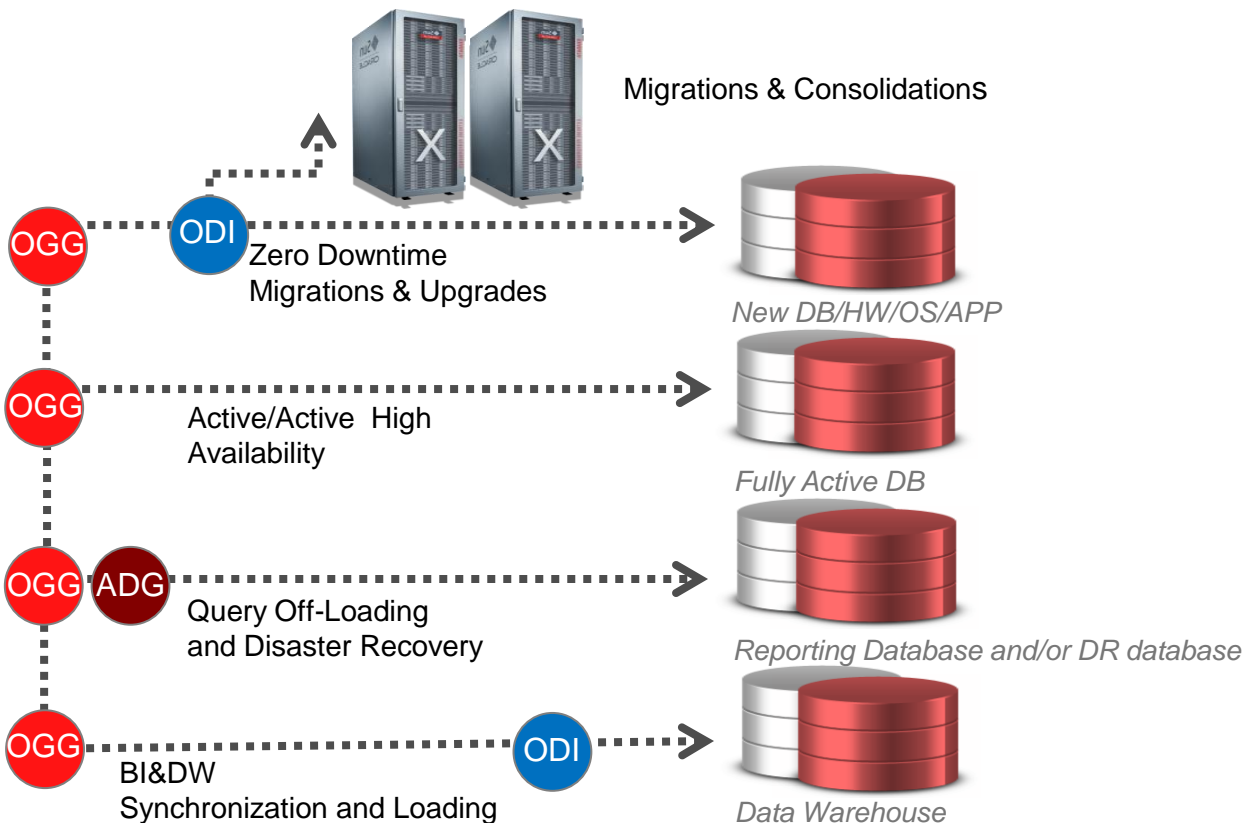
Oracle Data Integrator

Use Cases Of ODI

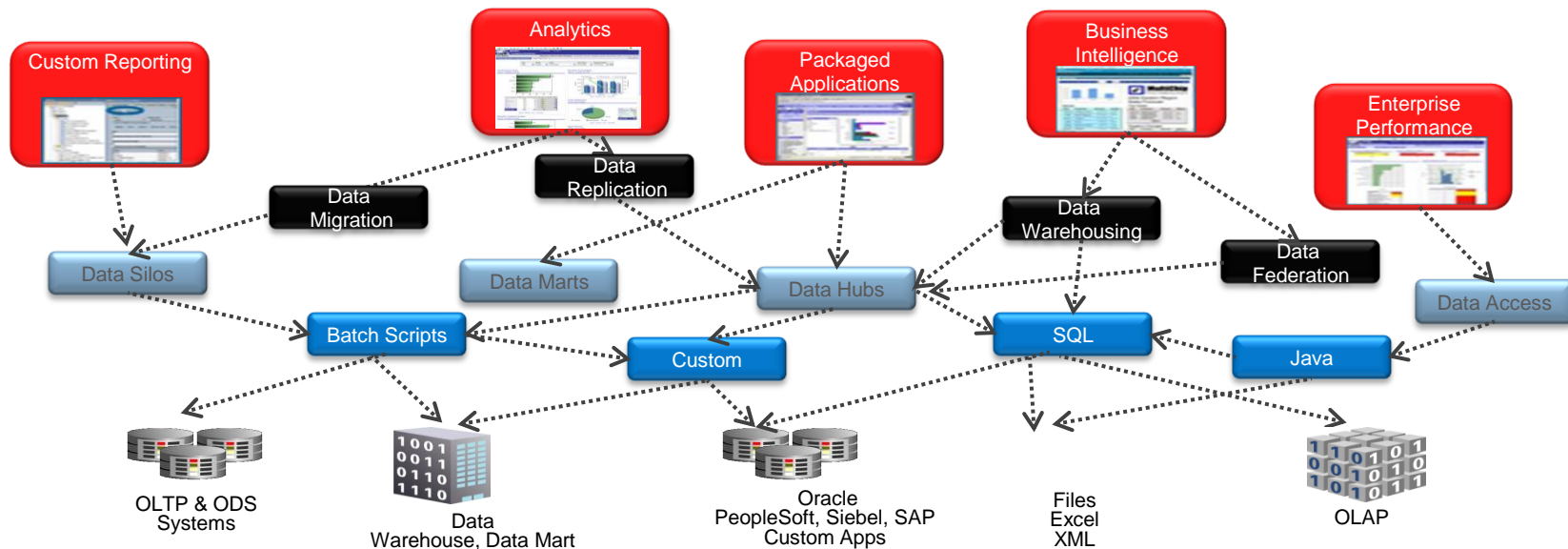
Improve the speed of handling data by 50%



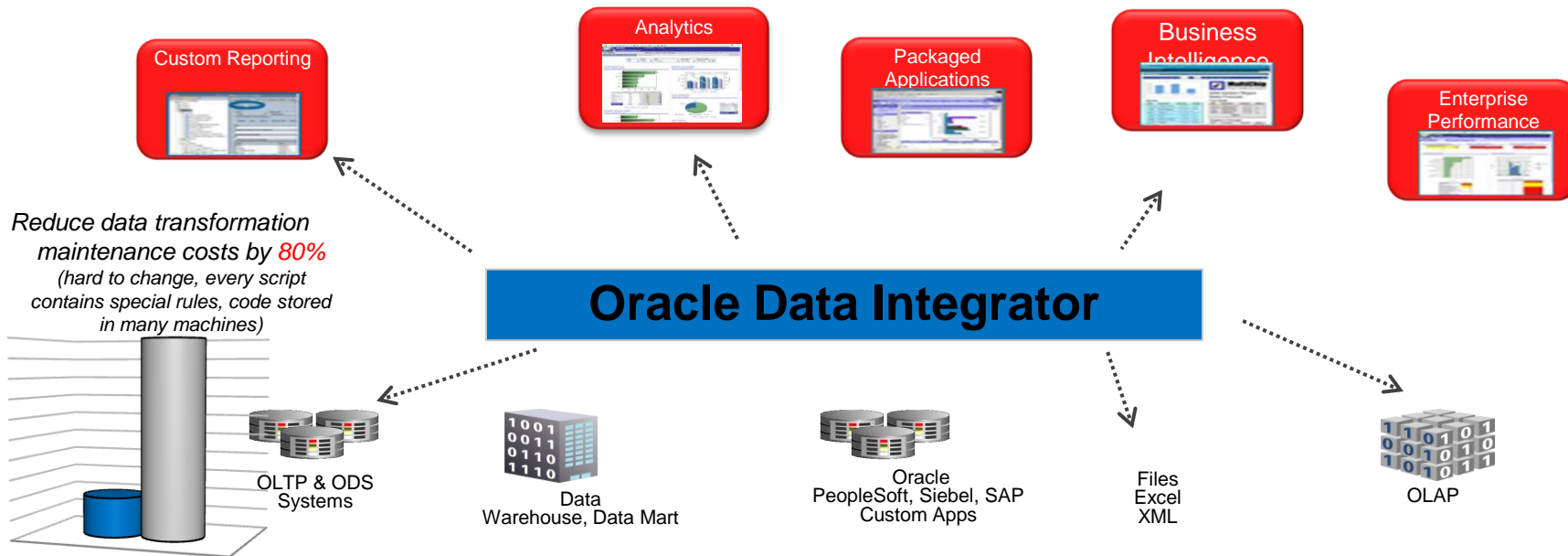
■ After ODI □ Before ODI



Replacing Manual Coding

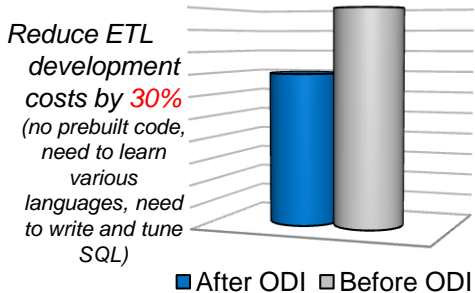


Replacing Manual Coding

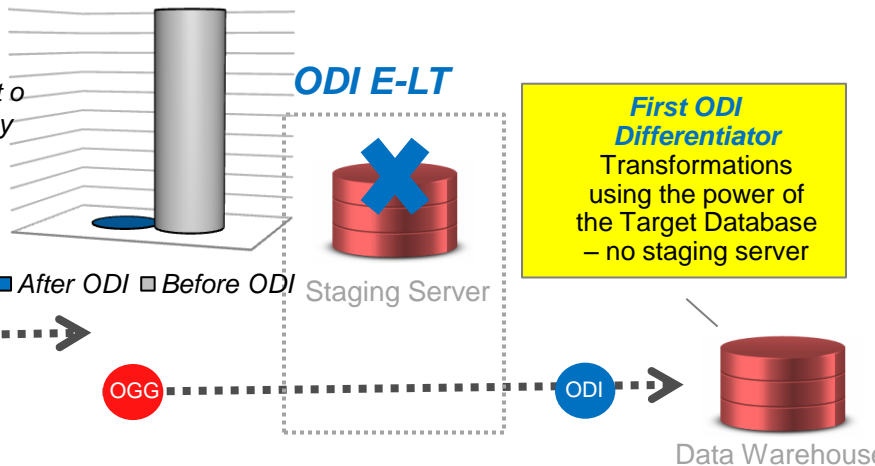


■ After ODI □ Before ODI

Why Is ODI Different?

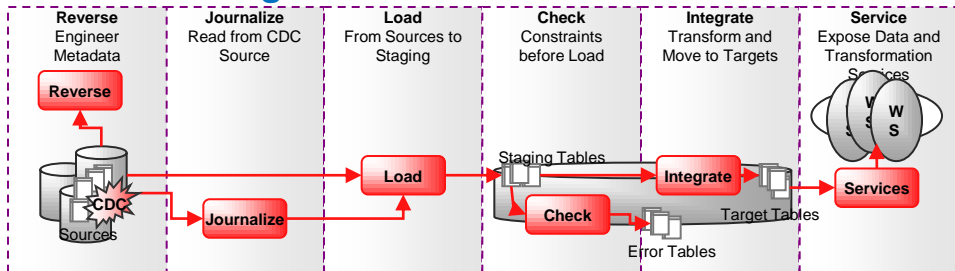


Decrease the cost of ETL HW by **100%**

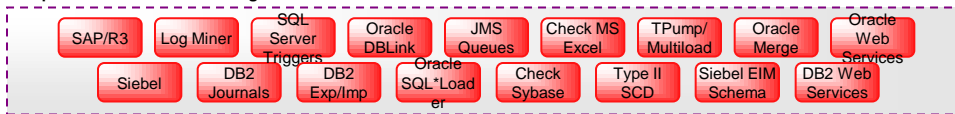


Second ODI Differentiator
ODI Declarative Design and ODI Knowledge Modules for reusing already written down level SQL code

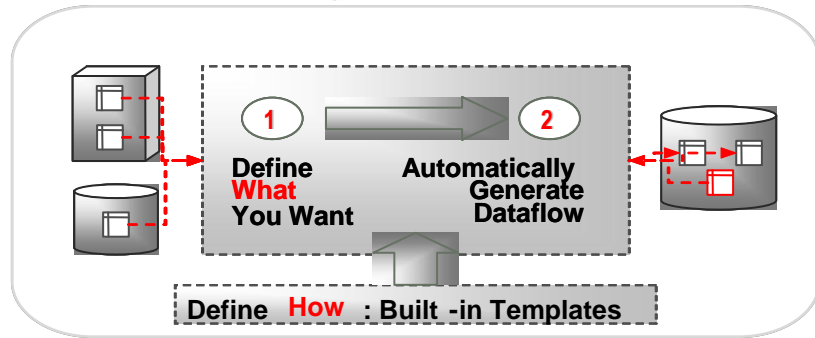
ODI Knowledge Modules



Sample out-of-the-box Knowledge Modules



ODI Declarative Design

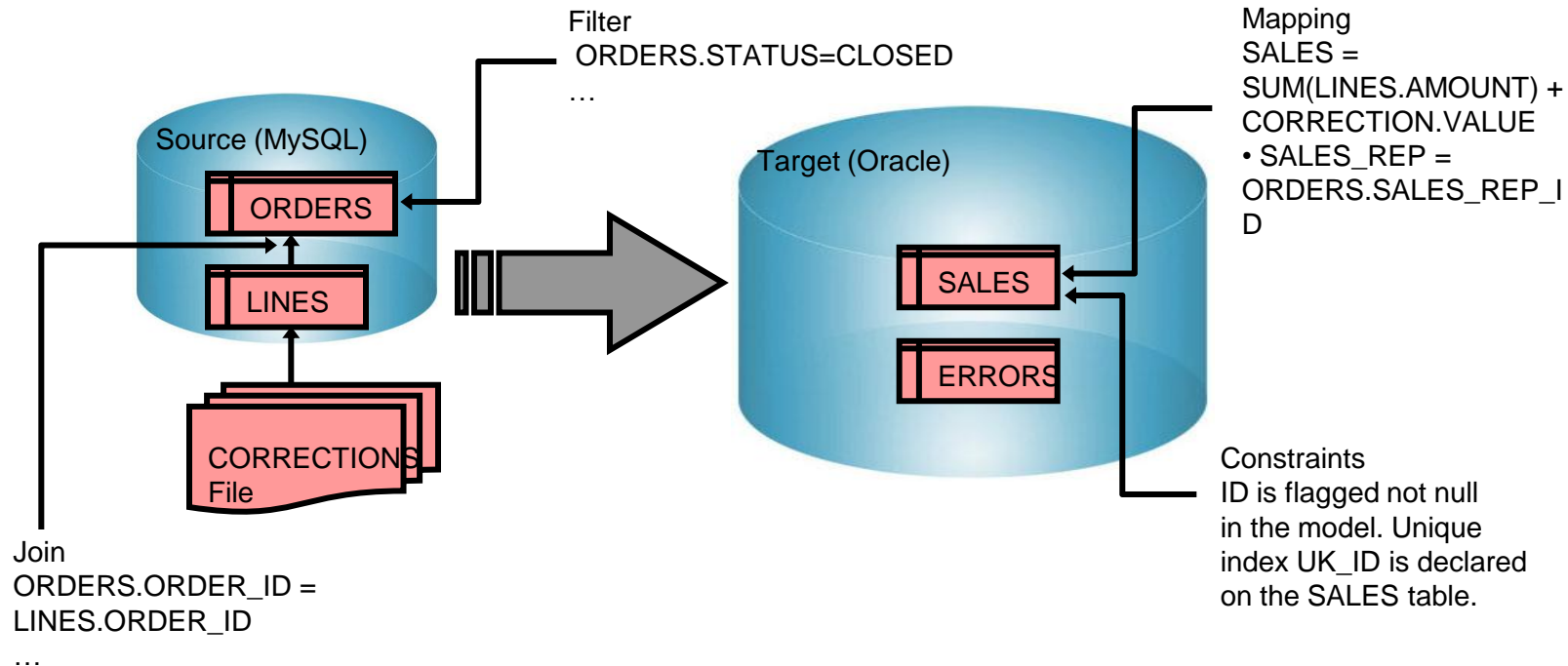


Benefits

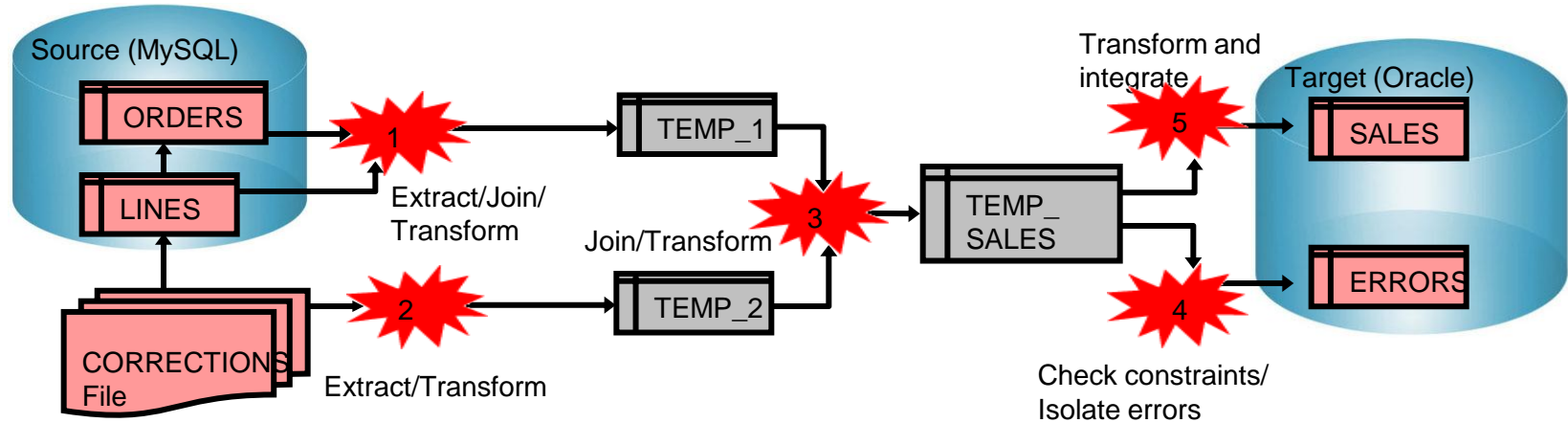
ORACLE

Implementing The Rules

Business rules implemented in SQL



Process Implementation Without ODI



STEP	DESCRIPTION	EXAMPLE OF CODE
1	Execute the join between ORDERS and LINES as well as the filters on the source Microsoft SQL Server database using a database view.	Create view C\$_SALES As select ... from ORDERS, LINES where ORDERS.STATUS = 'CLOSED' and ORDERS.ORDER_ID = LINES.ORDER_ID
	Extract the content of the view into a flat file using the BCP utility.	bcp C\$_SALES out c\$_sales_extract.bcp -c -S... -U... -P... -t\b
	Use the SQL*Loader utility to load the temporary BCP file into the TEMP_1 Oracle table.	sqlldr control=TEMP_1.ctl log=logfile.log userid=.../...
2	Use the SQL*Loader utility to load the CORRECTIONS ASCII file into the TEMP_2 Oracle table.	sqlldr control=TEMP_2.ctl log=logfile.log userid=.../...
3	Join, transform and aggregate the 2 temporary tables TEMP_1 and TEMP_2 and load the results into a 3 rd table (TEMP_SALES) using SQL	insert into TEMP_SALES (...) select SUM(TEMP_1.AMOUNT+TEMP_2.VALUE), TEMP1.SALES_REP_ID,

4

Check Unique constraints using SQL and insert the errors into the Errors table

```
insert into Errors(...)
select ... from TEMP_SALES
where ID in (select ID
            from TEMP_SALES
            group by ID
            having count(*) > 1)
```

Check Reference constraints using SQL and insert the errors into the Error table

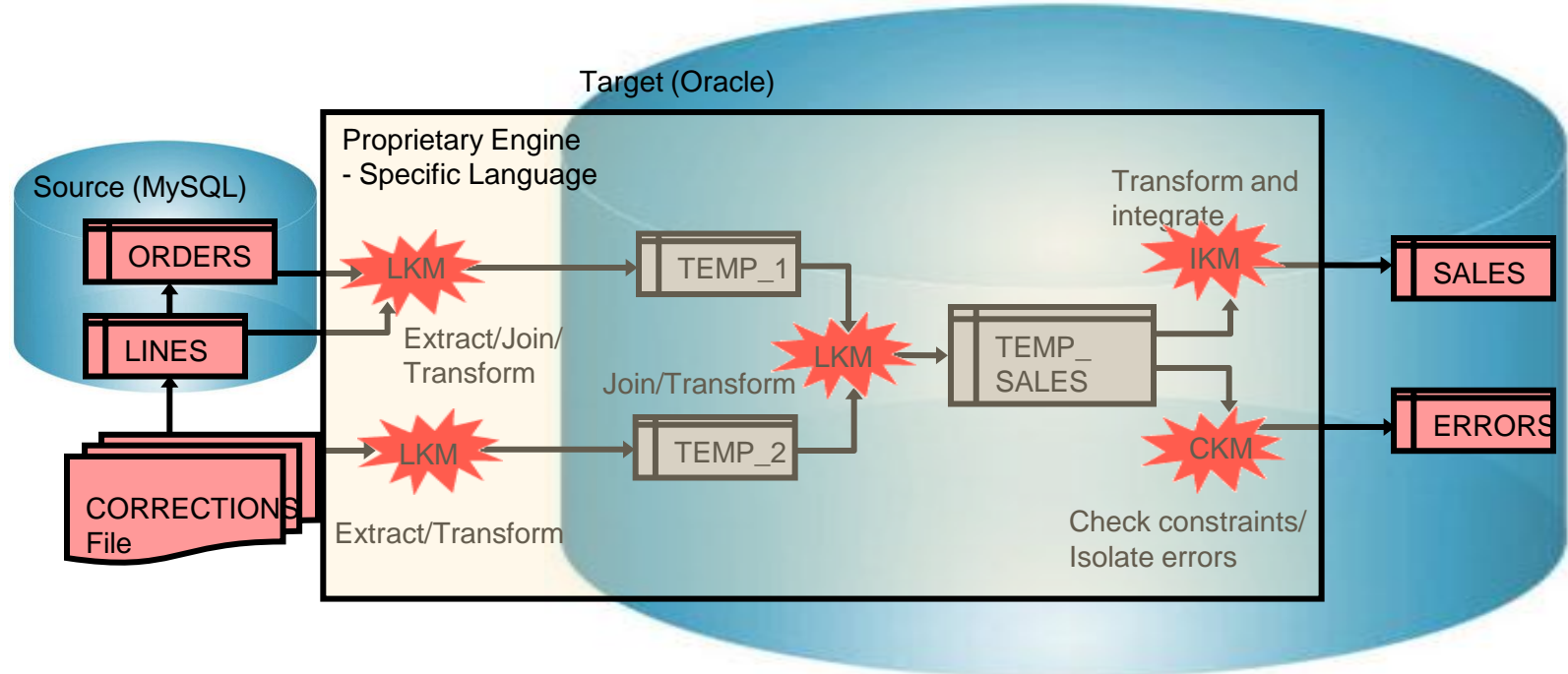
```
insert into Errors(...)
select ... from TEMP_SALES
where SALES_REP not in
(select SALES_REP_ID from SALES_REP)
```


5

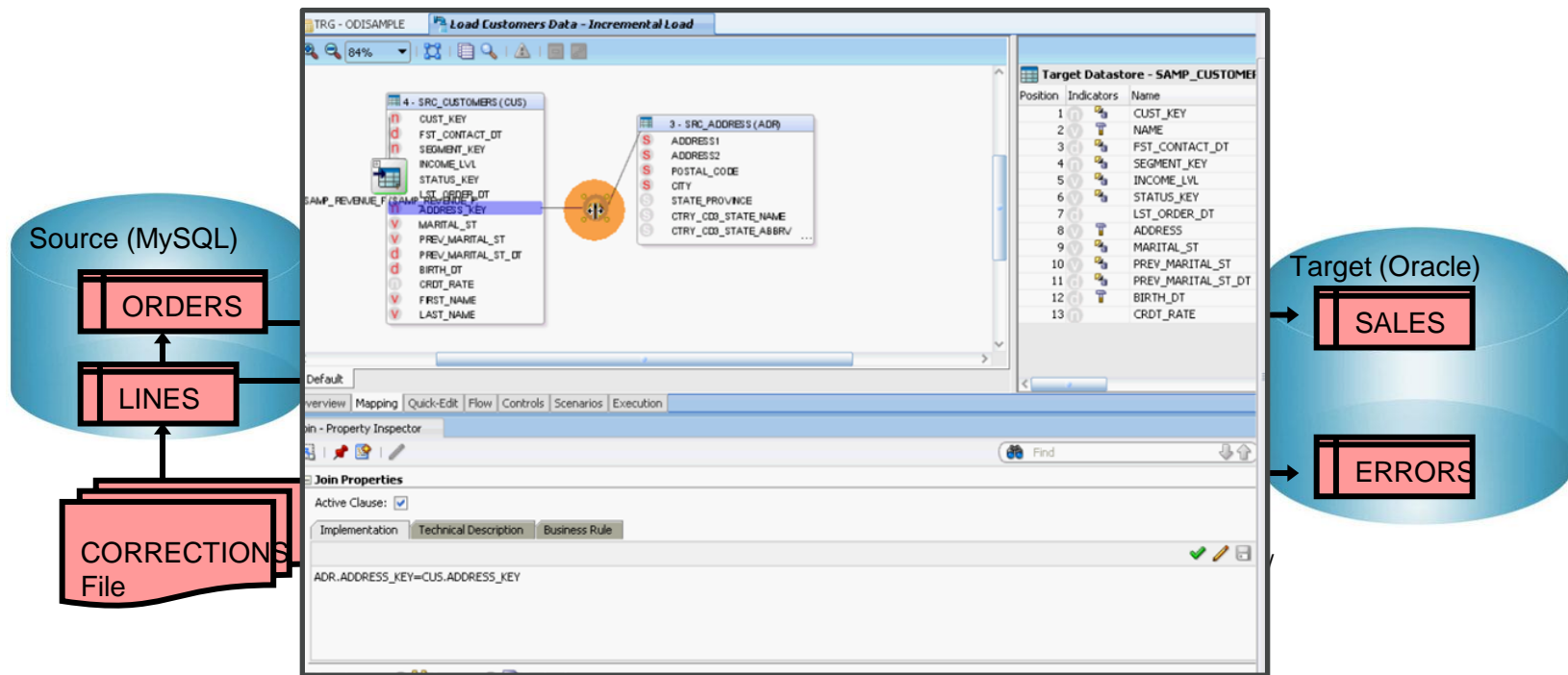
Finally, use SQL logic to insert / update into the target SALES table using a query on TEMP_SALES

```
update SALES set ...  
from ...  
where ID in  
(select ID  
from TEMP_SALES  
where IND_UPDATE='U')  
  
insert into SALES (...)  
select ...  
from TEMP_SALES  
where IND_UPDATE='I'  
...
```

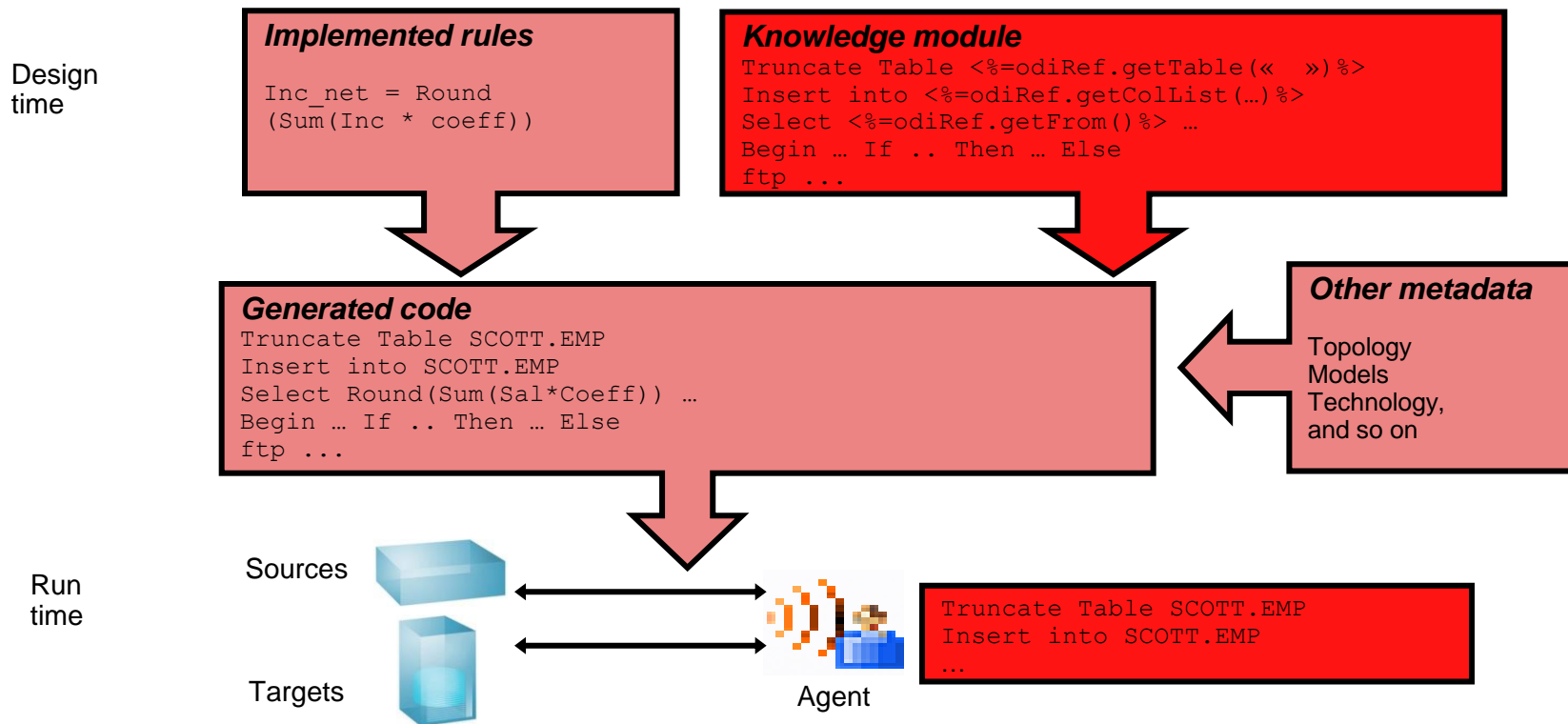
Process Implementation With ODI

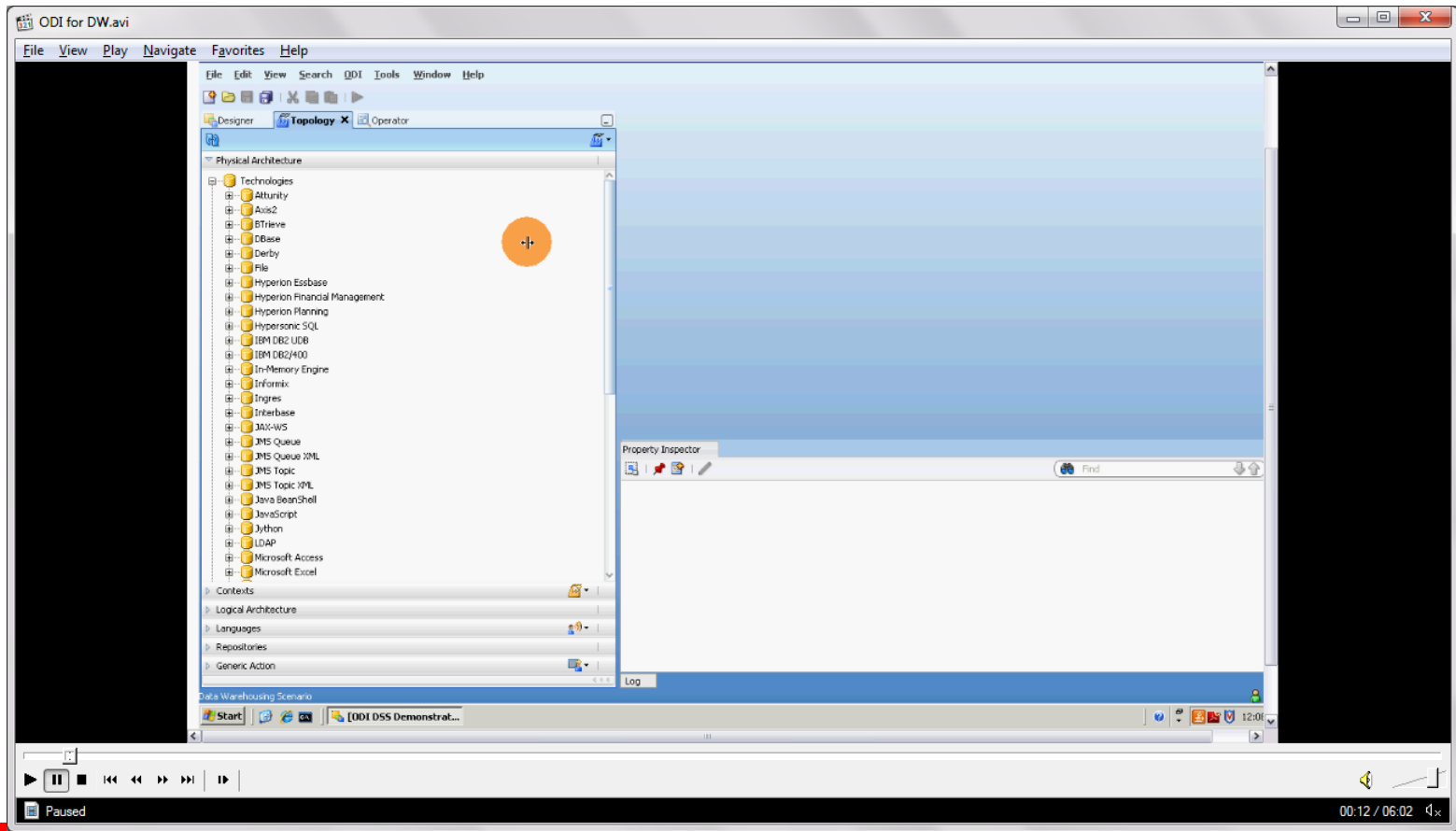


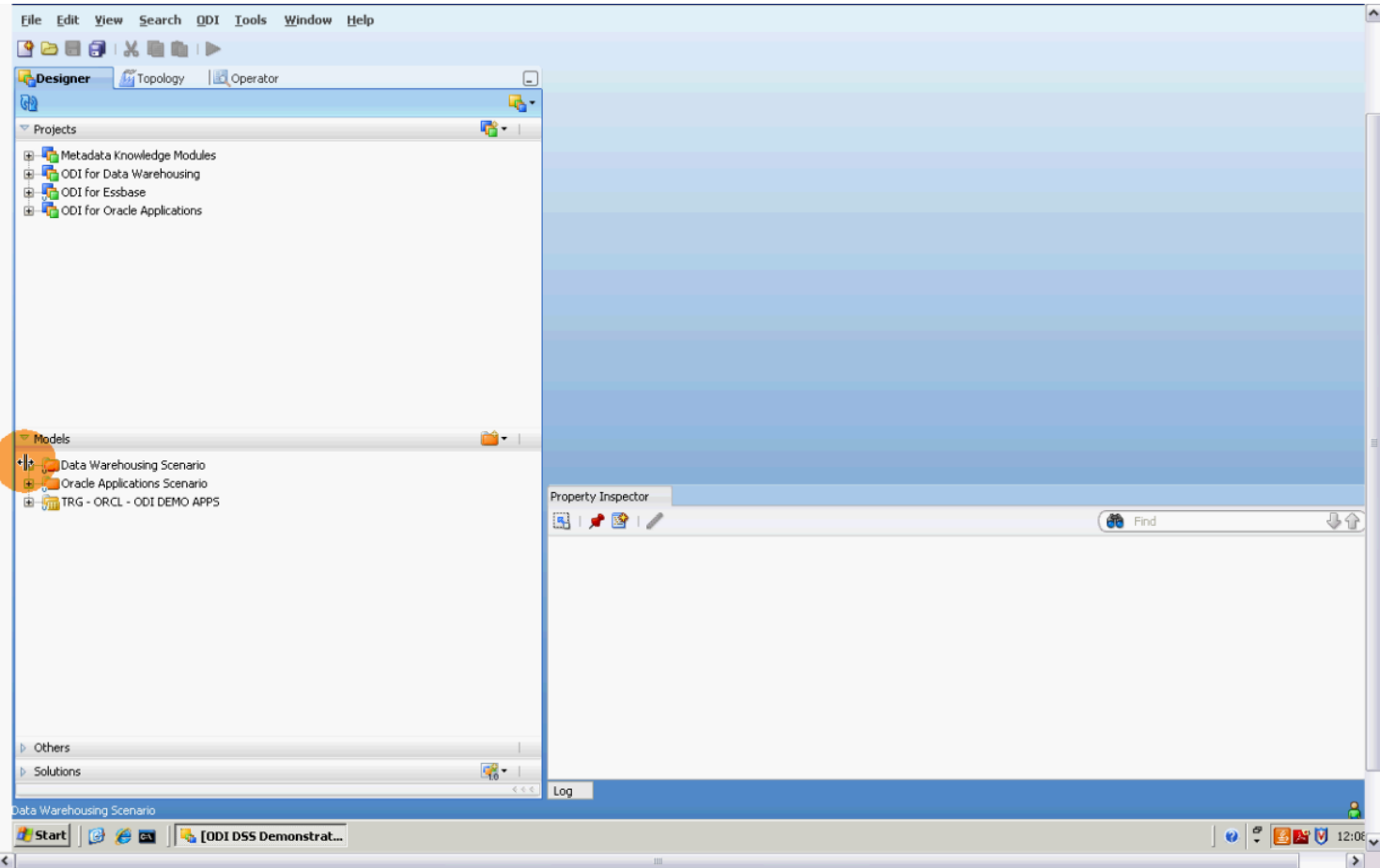
Process Implementation With ODI

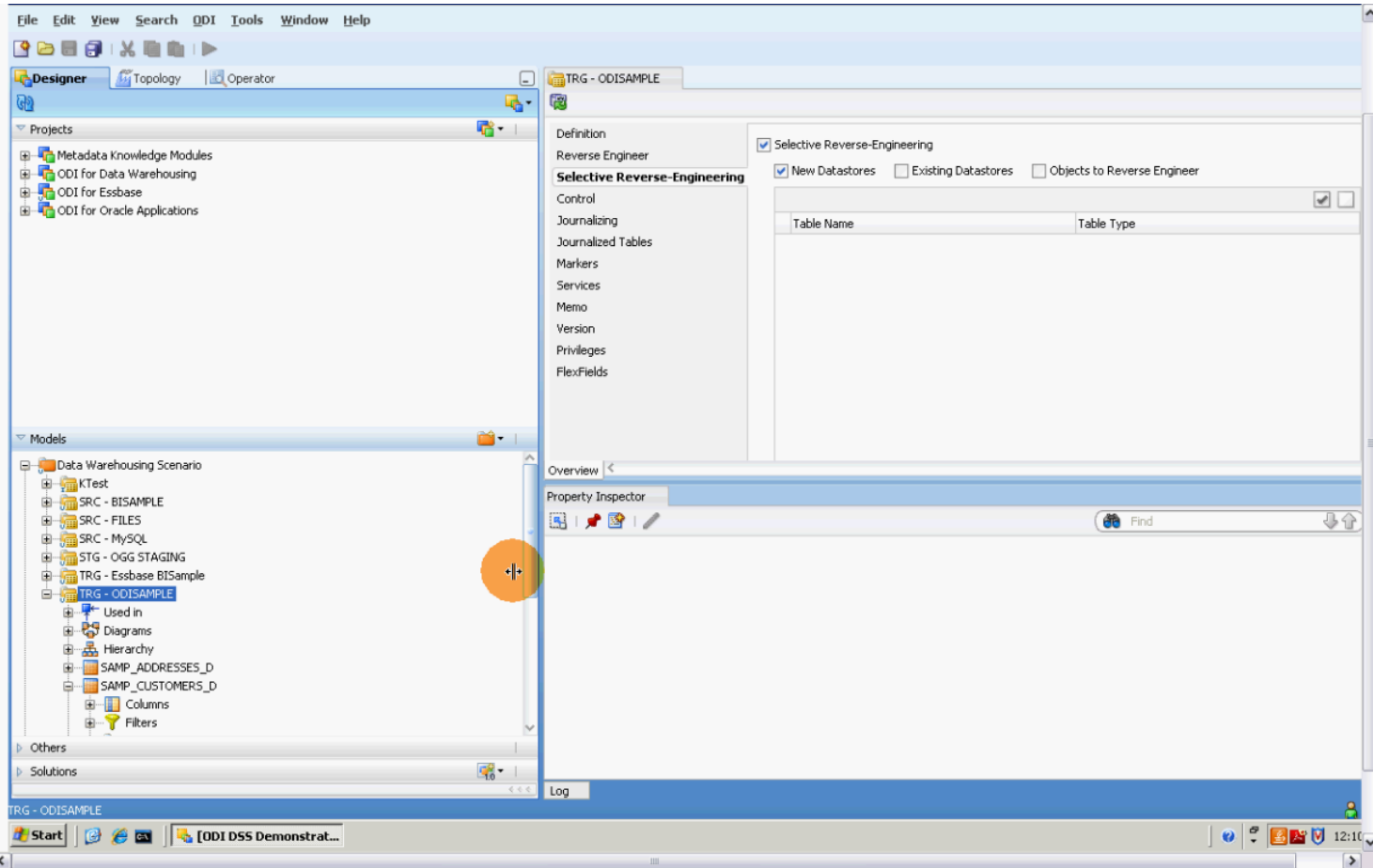


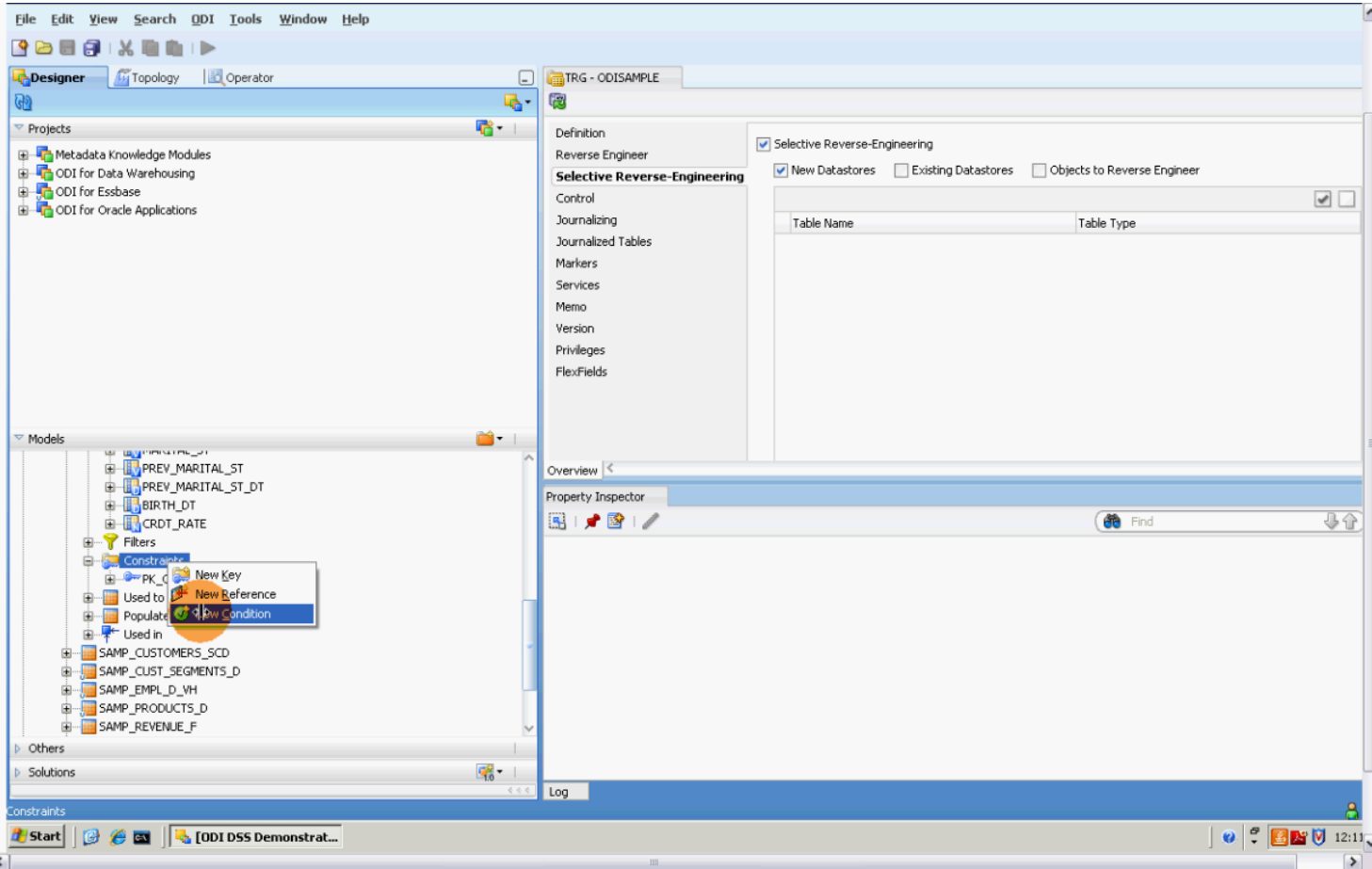
Code Generation

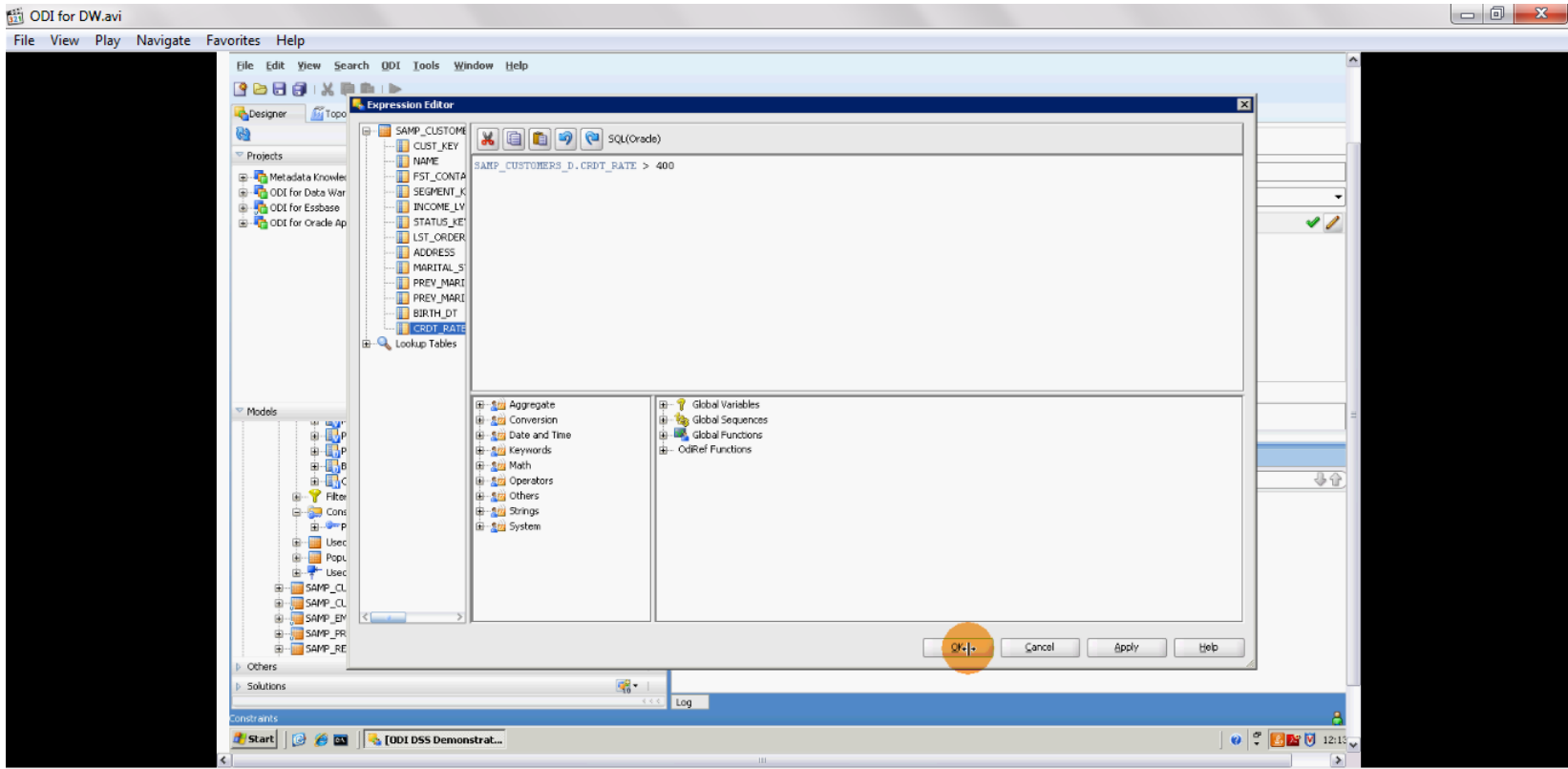




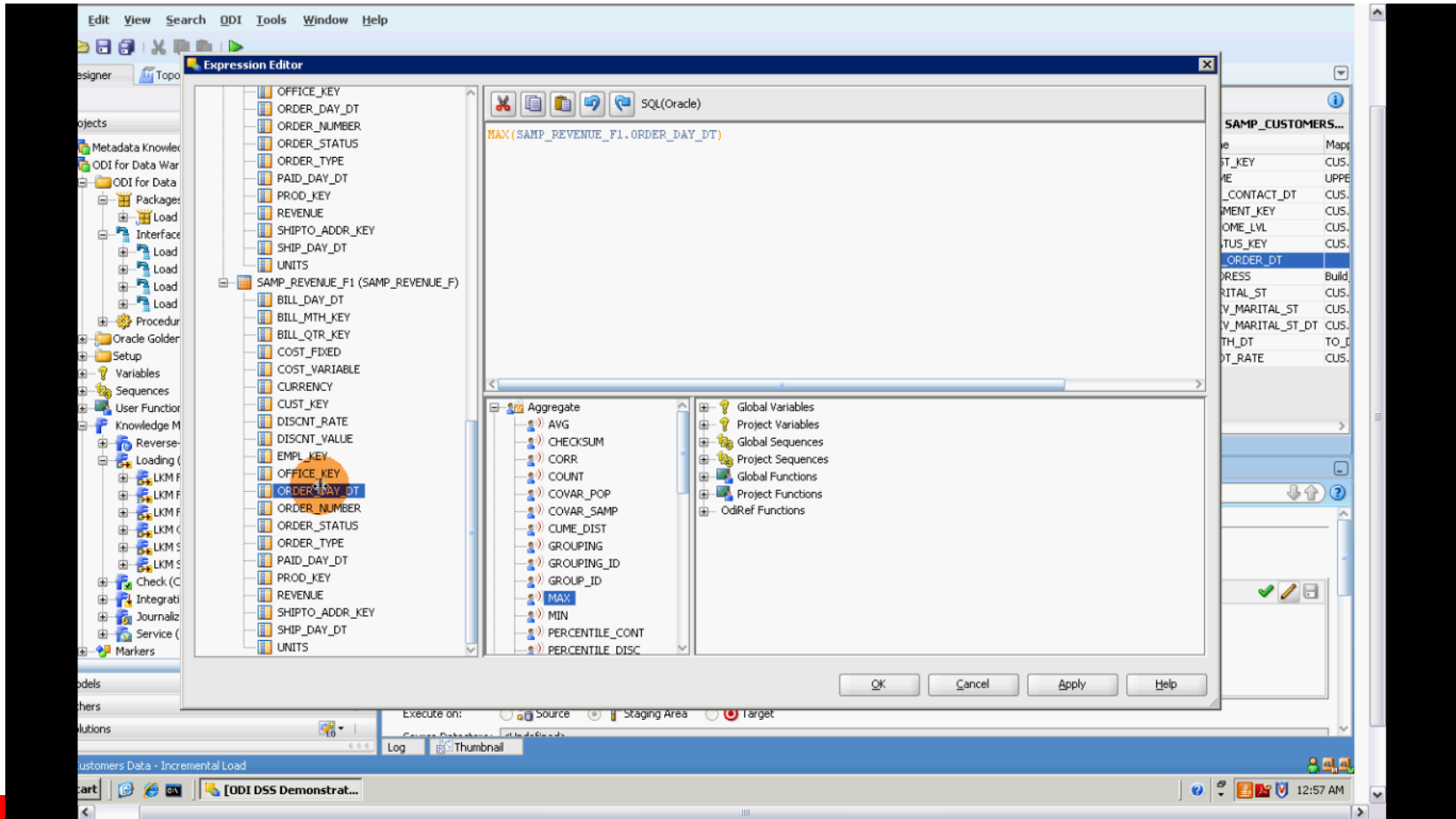








The screenshot displays the Oracle ODI 11g Designer interface. The left-hand pane shows a project tree for 'TRG - ODISAMPLE', with the 'Interfaces' folder expanded to show 'Load Customers Data - Incremental Load'. The main workspace is titled 'Load Customers Data - Incremental Load' and shows the 'Definition' tab for the selected interface. The 'Name' field is set to 'Load Customers Data - Incremental Load', and the 'Optimization Context' is set to 'Global'. The 'Oracle' field is set to 'Oracle: ODISAMPLE_LogicalSchema'. The 'Description' field is empty. Below the workspace, the 'Property Inspector' is visible, and the 'Log' window is at the bottom. The Windows taskbar at the bottom shows the Start button and several open applications, including 'ODI DSS Demonstrat...'. The system clock in the bottom right corner indicates the time is 12:15.



The screenshot displays the Oracle Data Integrator (ODI) Designer interface. The main window shows a process diagram for a package named "Load Customers and Products Data". The diagram consists of the following steps:

- Wait for Input Files**: The starting point of the process.
- Prepare Environment**: A step that follows the initial wait.
- Load Addresses data...**: A data load step.
- Load Products Data**: A data load step.
- Load Customers Da...**: A data load step.
- Retrieve Number of...**: A step that likely involves a query or table operation.
- Evaluate Number...**: A decision step that branches based on the result of the previous step.
- Alert on Error**: A terminal step reached if any of the load steps fail (indicated by red "ko" arrows).
- Alert on Success**: A terminal step reached if the process completes successfully (indicated by green "ok" arrows).
- Alert on Error**: A terminal step reached if the evaluation step fails.

The interface includes a menu bar (Edit, View, Search, ODI, Tools, Window, Help), a toolbar, a left-hand "Objects" tree, a central "Toolbox" with various ODI components, and a "Property Inspector" at the bottom. The status bar at the very bottom shows the time as 1:04 AM.

ORACLE
D: 8238966 Playing
Mute
Time 00:55:34/01:03:01

Rewind the recording

orcl_source.pdf - Adobe Reader
ODI+ODQ 11g - VMwar... File VM Help

File Edit View Document Tools Window Help

1 / 9 81% Find

Model Oracle_Source

Model Id: 1000
Model name: Oracle_Source
Technology: Oracle
Logical Schema: ORCL_SOURCE

Description:
Creation date: Dec 18, 2009 11:53:37 AM
Created by: SUPERVISOR
Last modification: Apr 7, 2011 9:48:35 AM
Modified by: SUPERVISOR

Number of Sub-Models: 1

Sub-Models list:

Sub-Model Global

Sub-Model Id: 1000
Sub-Model name: Global

Creation date: Dec 18, 2009 11:53
Created by: SUPERVISOR

Sub-Model Id: 1000
Sub-Model name: Global
Technology: Oracle
Logical Schema: ORCL_SOURCE
Number of Tables: 4

Creation date: Dec 18, 2009 11:53:37 AM
Created by: SUPERVISOR
Last modification: Dec 18, 2009 11:53:37 AM
Modified by: SUPERVISOR

Table list:

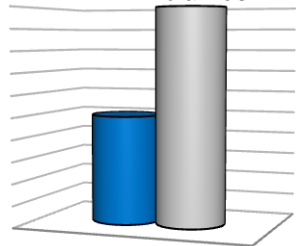
Name	Resource name	Type	No. of Rows	Modification date
SRC_CITY	SRC_CITY	Table		Dec 18, 2009 11:53:48 AM
SRC_CUSTOMER	SRC_CUSTOMER	Table		Dec 18, 2009 11:53:48 AM
SRC_REGION	SRC_REGION	Table		Dec 18, 2009 11:53:48 AM
SALES	SALES	Table		Dec 18, 2009 12:08:43 PM

Bookmarks

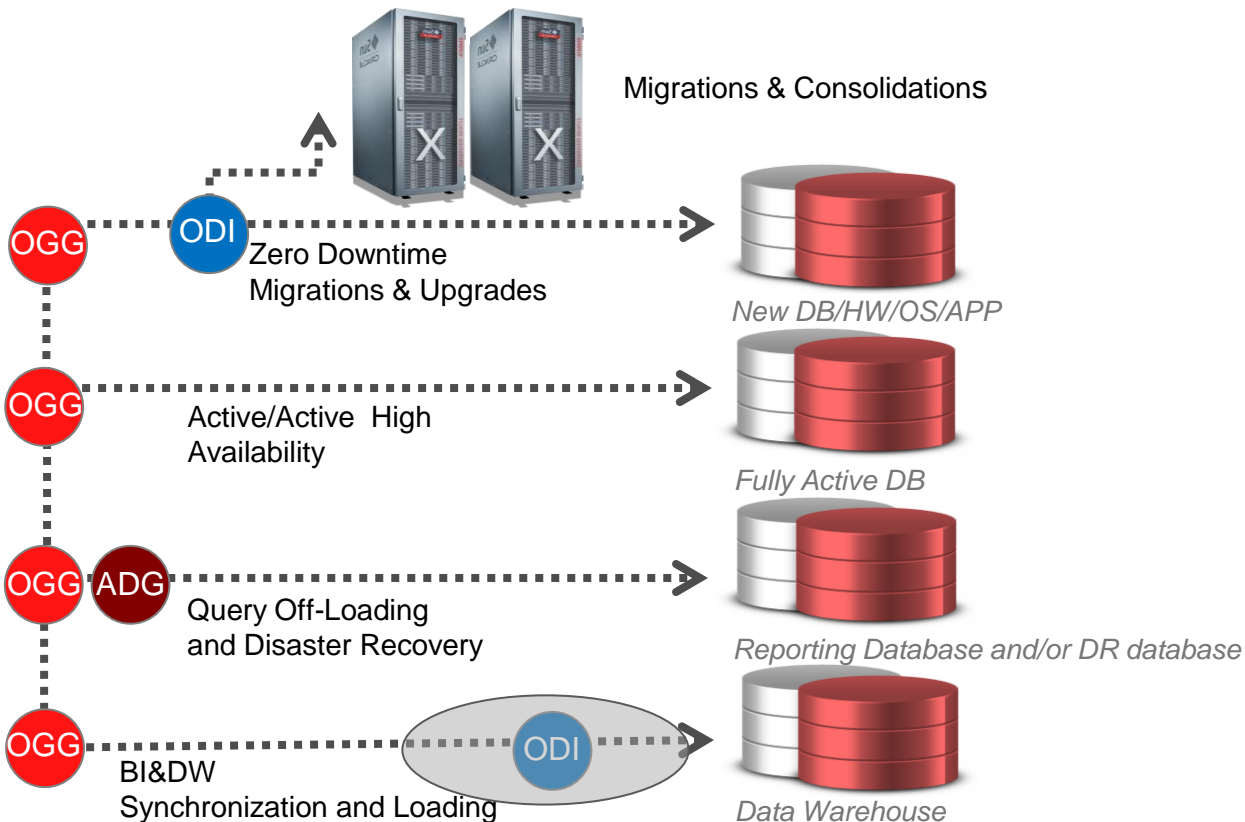
- Model Oracle_Source
 - Sub-Model Global
 - Table SRC_CITY
 - Table SRC_CUSTOMER
 - Table SRC_REGION
 - Table SALES

Use Cases Of ODI

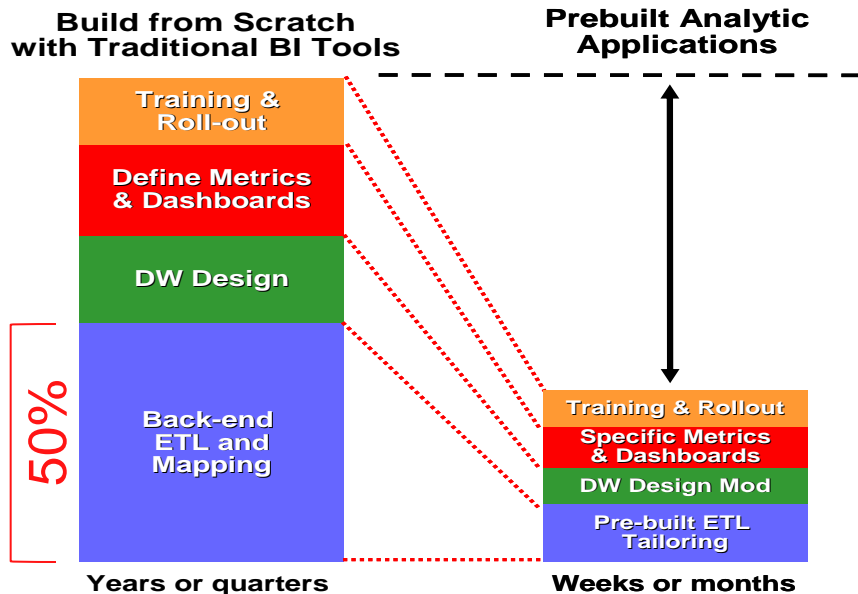
Deferring hardware upgrades by **50%** and still run ETL jobs with increasing data volumes



■ After ODI □ Before ODI



ODI For Data Warehouse Projects



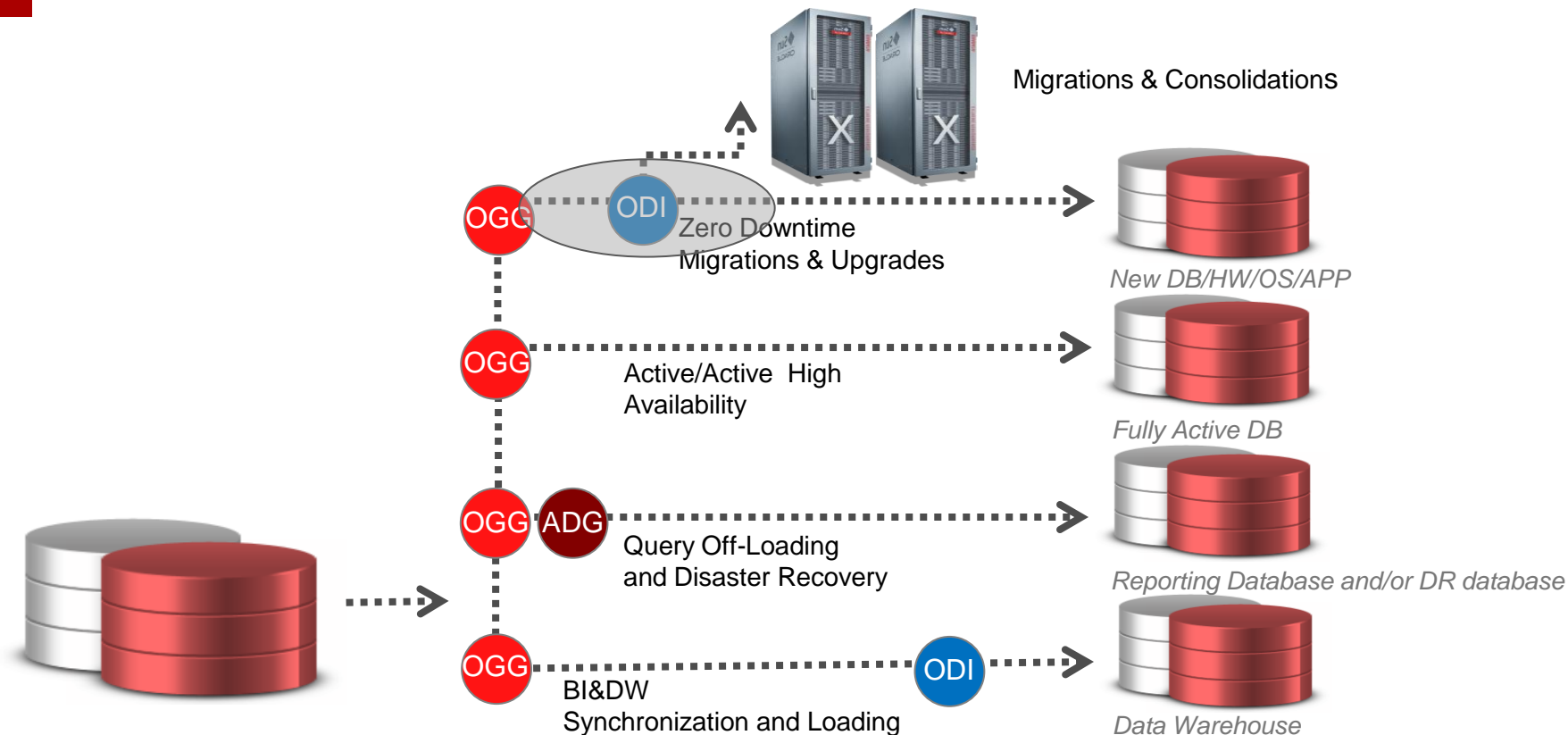
INFA

Number of Setup Steps	10
Number of Servers	3
Number of connections	7

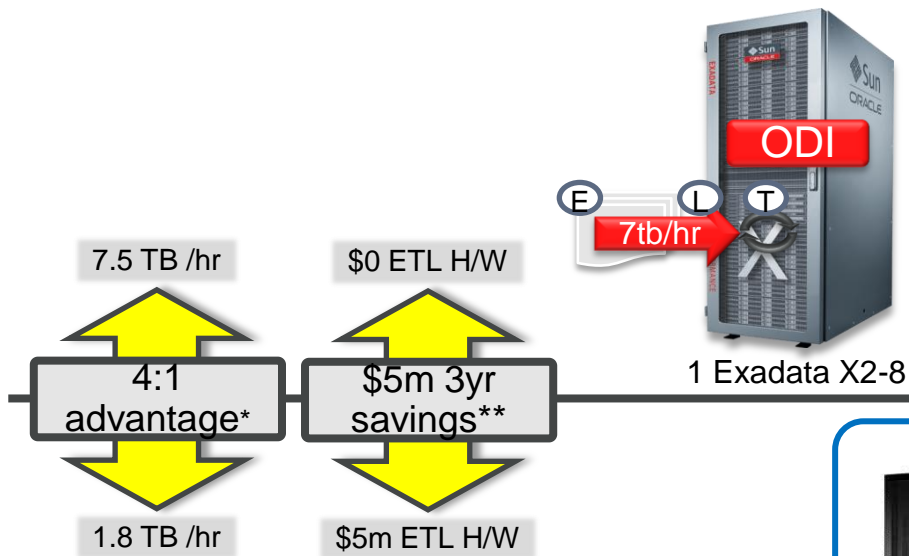
ODI

Number of Setup Steps	7
Number of Servers	1
Number of connections	3

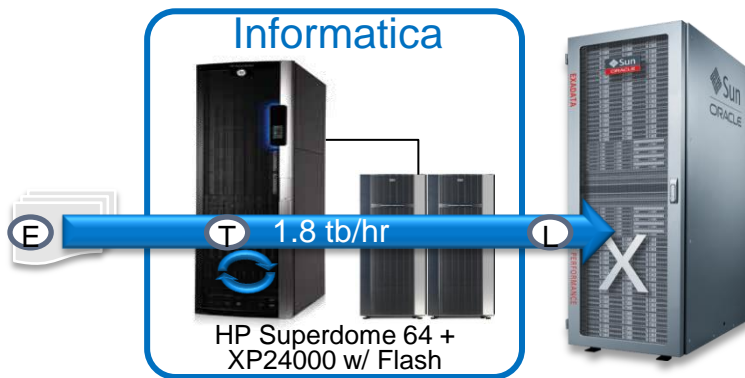
Use Cases Of ODI



ODI And Exadata

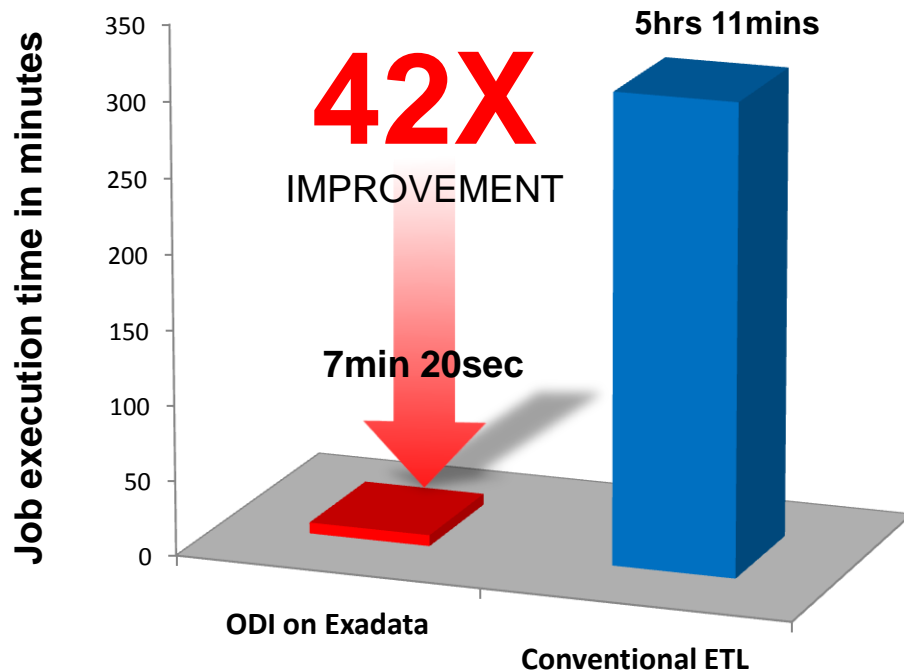


- Run ODI Directly on Exadata
- Complex Data Transformations
- Linear ETL Scalability
- Fully Leverages DBFS/Infiniband, Smart Storage, and Advanced Compression






* TPC-H data sets with transformations
** Production hardware savings
(not including Dev + Test environments,
management costs or software savings)

ODI And Exadata



- A complex branch of the customer's tax allocation process runs 5 hrs 11 mins during quarter close
- Exadata and ODI (E-LT) combined is able to execute the process **42X faster** (7mins 20 secs)

Oracle DIS And Exadata

-  Oracle GoldenGate
-  Oracle Data Int.
-  Active Data Guard



Any Source Systems (to be constantly synchronized with Exadata)



ODI is the fastest tool for loading Exadata!



For Active Stand By



For Read Only Stand By



OGG for fully active stand by or OGG/ADG for only read-capable standby system that can be used for reporting and for pre-production testing!



Old Systems (to be migrated on Exadata)

Only GG can do easy Zero Downtime Exadata migration with the lowest possible overhead to production!

Sharing Common Features

Application and Data Integration

Oracle SOA Suite

- Message-based transformations
- Mediation
- Service composition
- Orchestration
- Event processing

Common Features

- Standardized development environment
- Security
- Management & Administration
- Communication & Messaging
- Transformation
- Adapters

Oracle Data Integrator

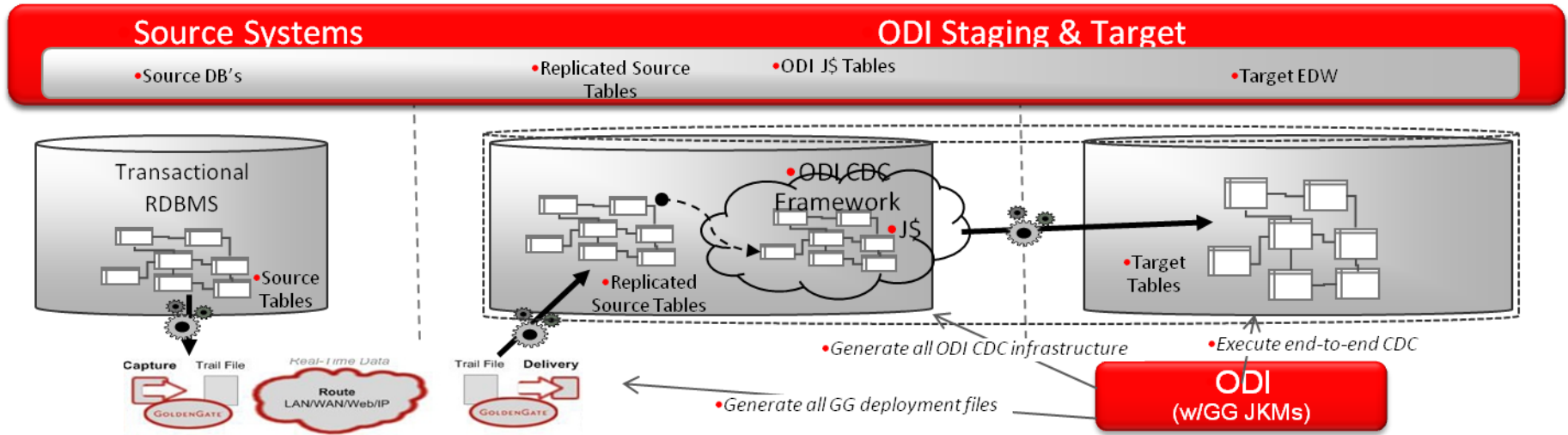
- Set-based transformations
- Data Semantics and formats
- Data Quality
- Bulk data movement and transfer

- Sophisticated data transformations to merge, split or standardize data
- Expose data integration tasks as reusable services
- Reduce cost of building and maintaining separate integration services
- DI layer provides a unified, logical view of data- ready for data warehouse loads, etc.

OGG&ODI

- *Key Benefits:*

1. Eliminate Overhead → no need for DB API overhead on the Source, or the invasiveness of the ODI J\$ objects on the Source system,
2. Automate GoldenGate → automation of GG deployment *directly from* ODI GUI
3. Provide Common DW Pattern → supplies a common pattern for mini-batch style (non-real-time) DW aggregate loads



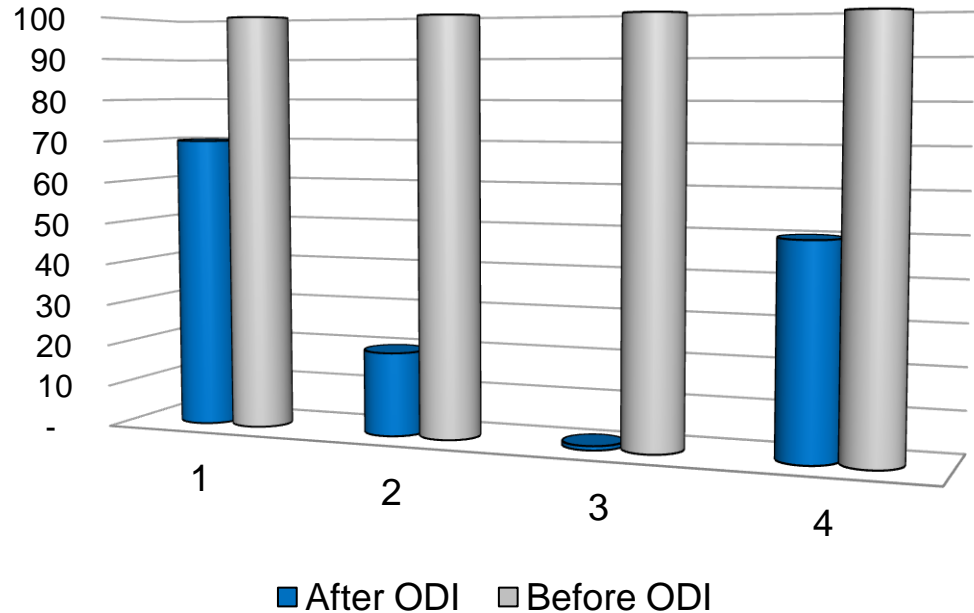
ODI Economic Impact – Reduce Negatives

1-Reduce ETL development costs by **30%** (no prebuilt code, need to learn various languages, need to write and tune SQL)

2-Reduce data transformation maintenance costs by **80%** (hard to change, every script contains special rules, code stored in many machines)

3-Decrease the cost o of ETL HW by **100%**

4-Decrease TCO (because of unified platform and single vendor) **by 50%**



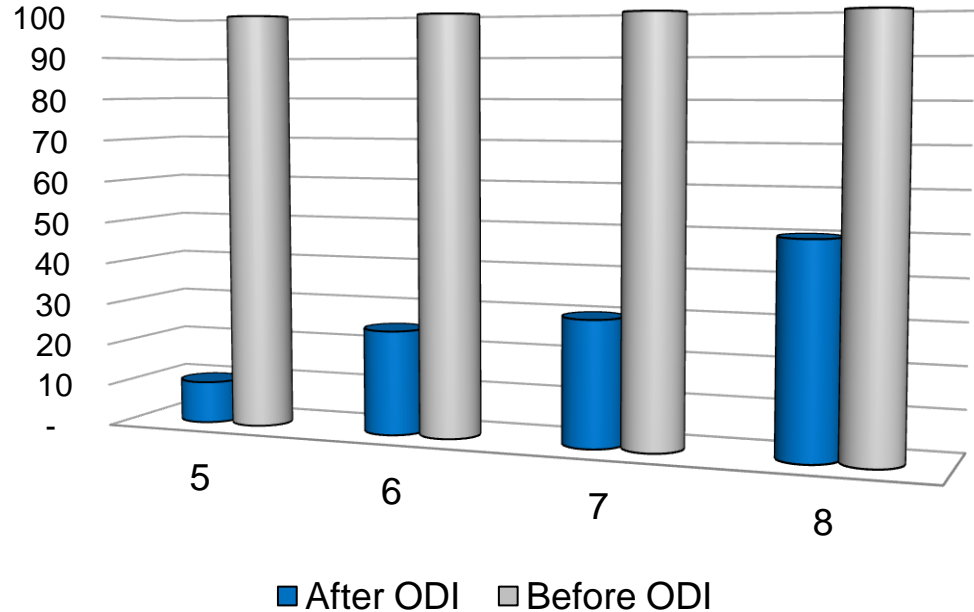
ODI Economic Impact – Reduce Negatives

5-Eliminating ongoing third-party training costs **by 90%**

6-Shorten reporting cycle that helps business users make decisions faster **by 75%**

7- Cut management and administrative efforts of loading and unloading data **by 70%**

8-Deferring hardware upgrades **by 50%** and still run ETL jobs with increasing data volumes



ODI Economic Impact – Improve Positives

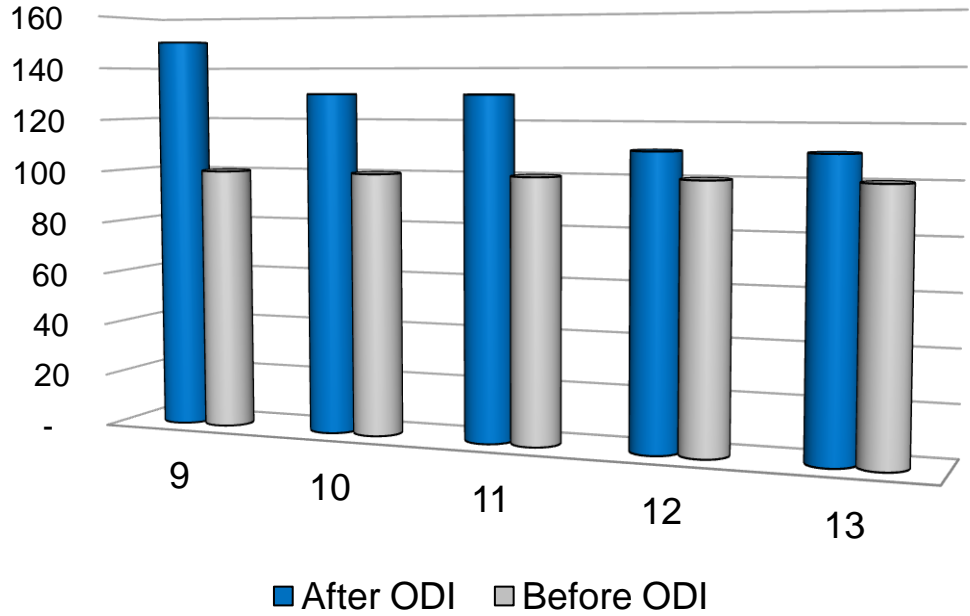
9-Improve the speed of handling data by **50%**

10-Improve business process execution times by **30%**

11-Improve speed of delivering projects by **30%**

12- Increase revenue by **10%** from smart decisions around centralized customer view

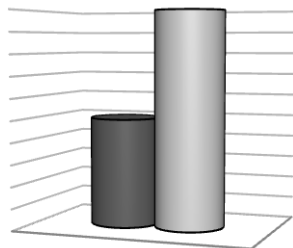
13- Increase revenue by **10%** from gaining new customers and customer satisfaction.



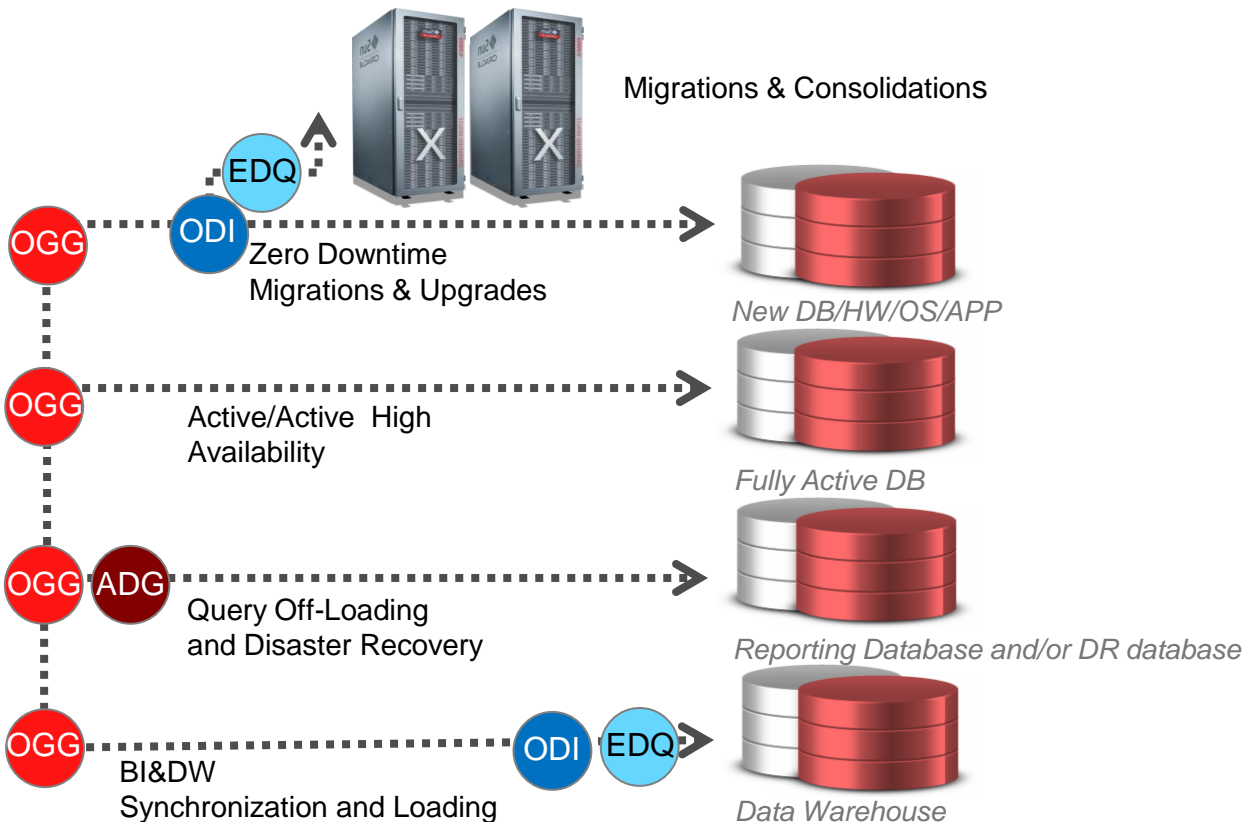
Oracle Enterprise Data Quality

Use Cases Of EDQ

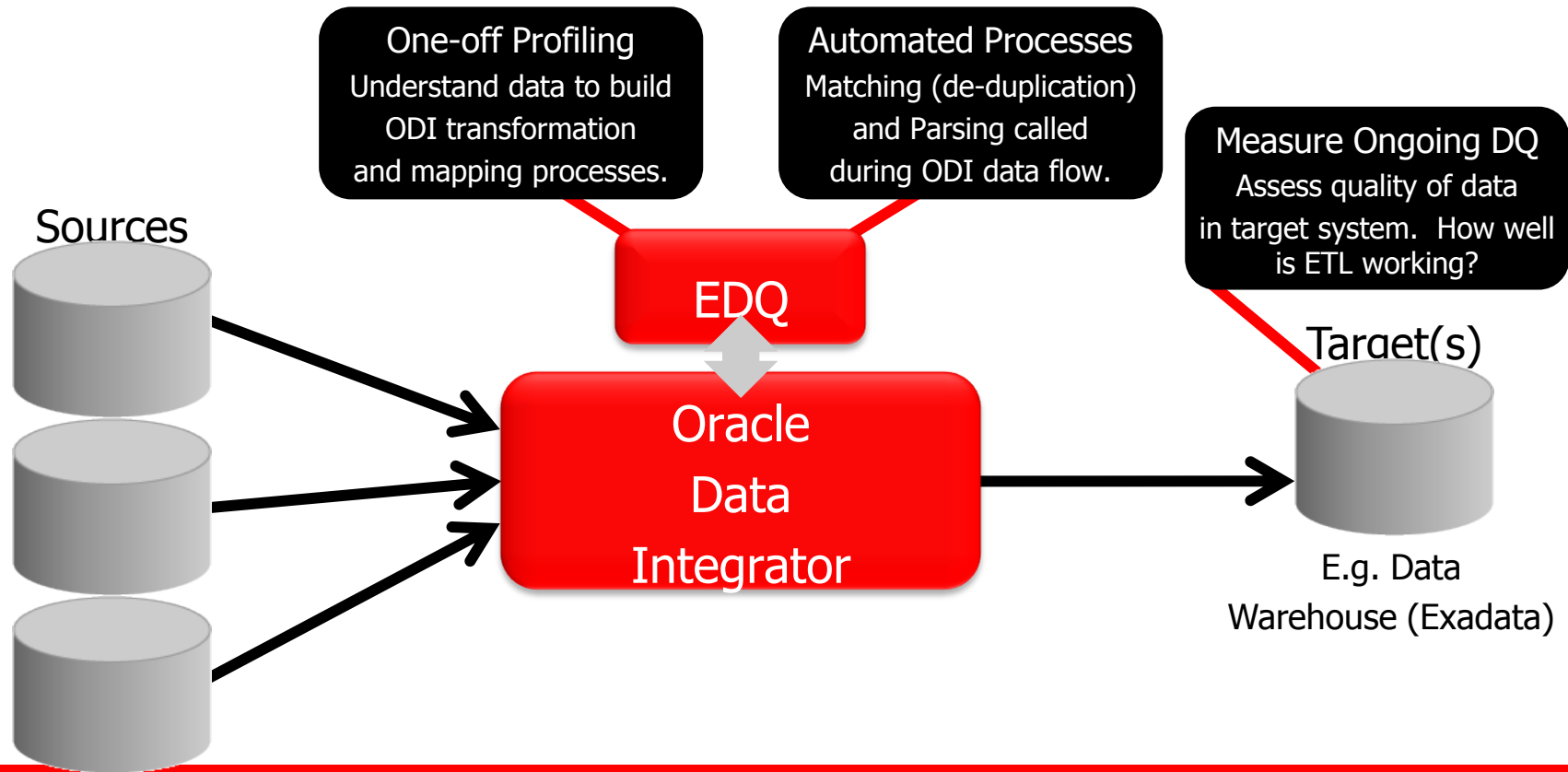
Decrease TCO (because of unified platform and single vendor) **by 50%**

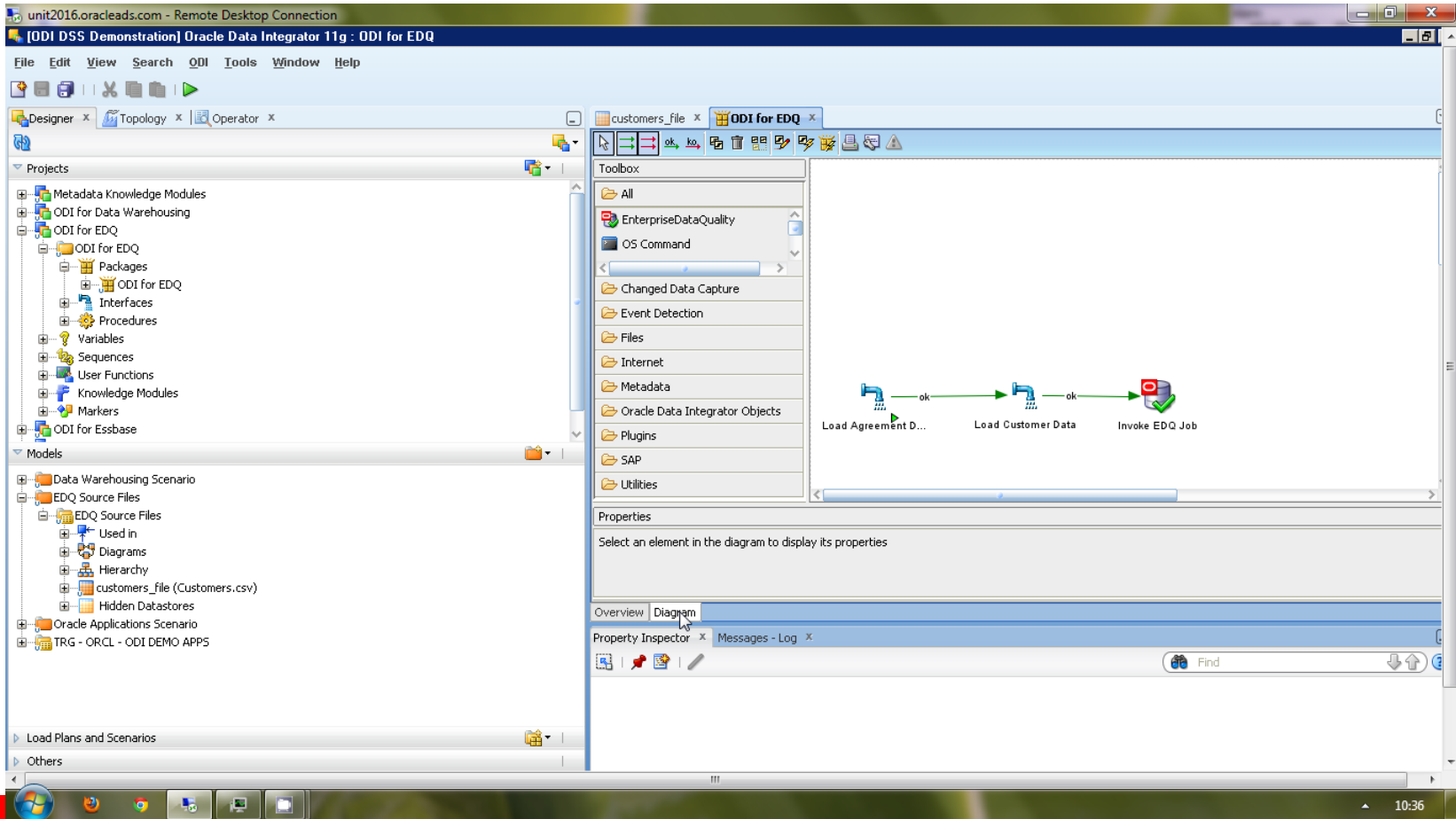


■ DIS Platform
□ No DIS Platform



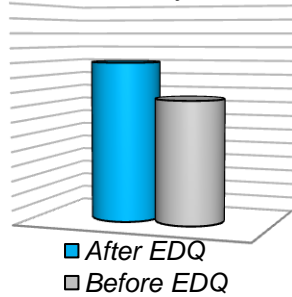
Data Flows And EDQ



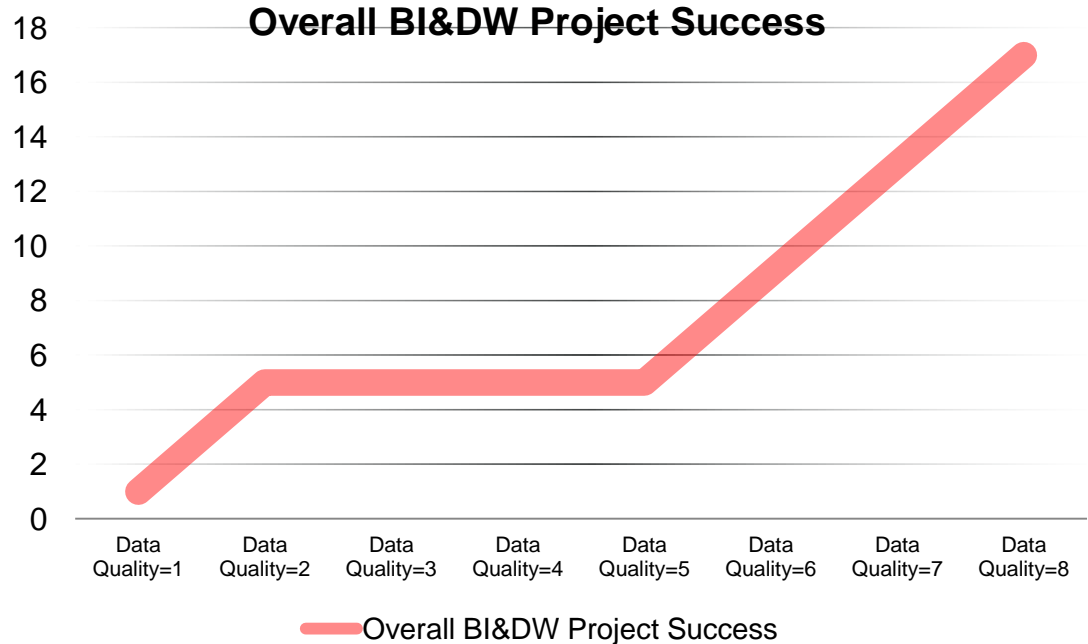


Failed BI&DW Projects

Increase BI, DW, CRM, Apps
ROI by 15%



BI success is directly related to the quality of the underlying data from HC investment.



Why EDQ?

Inconsistent formats

Abbreviations
(often ambiguous)

Attributes non-standard,
missing or invalid

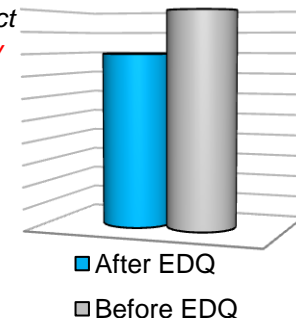
Customer ID	Customer Name	Address 1	Address 2	City	State	Zip	Country	Birth Date	Gender
AD23298	Mr Peter Mayhew	9407 Main St		Fairfax	VA	22031-4001	USA	02/23/61	M
VS38611	Dr Ellen Van Der Heijde 1	144 E Grove St		Kingston	PA	18704	US	07/12/57	
DC18223	Jalila Abdul-Alim (Do Not Call) 2	4548 Pennsylvania Ave	Apt 205	Kansas City	MO	64111-3349	USA	02/23/63	F
CO9387A	Tayside Computers Inc. 3	4912 E 41st N		Idaho Falls	ID	83401	USA	31/03/2007 7	
TZ35019	Mr Zachary P Jahn	98-1731 Ipuala Loop	Aiea	Hawaii		96701 5	United States	06/12/86	Male
CB27843	Mrs Edith Y Baba Junior	Baba Real Est. Corp. 3	209 Stony Point Trl	Webster	NY		USA	11/17/1971	M 6
OX80306	Andrew & Mary Baxter 4	14 Oxbridge Way		Milfrod 6	NH	03055-4614	US	05/28/67	F
JP70210	Mr RJ & Mrs FB MacDonald 4	57 Hadleigh Close	Westlea	Swindon SN5 9BZ	MA				Y
RD48107	Mr Andy Baxter	14 Oxbridge Wy		Milford	NH	3056	USA	01/01/01 8	M

Widespread duplication
(often hard to spot)

- 1** Compound Names
- 2** Embedded Additional Information
- 3** Mixed Business & Personal Names
- 4** Multiple Names

- 5** Mis-Filed Data
- 6** Erroneous Data
- 7** International Date Formats
- 8** Default or Dummy Data

Avoid error costs (incorrect orders, inventory etc.) by **20%**



Companies

In one hour...

- 240 businesses will change addresses
- 150 business telephone numbers will change or be disconnected
- 112 directorship (CEO, CFO, etc.) changes will occur
- 20 corporations will fail
- 12 new businesses will open their doors
- 4 companies will change their name

Individuals

In one hour...

- 5,769 individuals in the US will change jobs
- 2,748 individuals will change address
- 515 individuals will get married
- 263 individuals will get divorced
- 186 individuals will declare a personal bankruptcy

Products

In one year...

- On average 20% duplicates in product data
- 90% product introductions fail
- Retailers lost 40 billion or 3.5% of total sales lost each year due to item info inefficiencies
- 60% error rate for all invoices generated
- Global Data Sync will realize 30% lower IT costs

Master data changes at rate of 2% per month

Compounded, 2% monthly change is 27% per year, 61% in two years, 104% in three years!!!

Source: D&B, US Census Bureau, US Department of Health and Human Services, Administrative Office of the US Courts, Bureau of Labor Statistics, Gartner, A.T Kearney, GMA Invoice Accuracy Study

Why EDQ?

Variation or Error	Example
Sequence errors	<ul style="list-style-type: none">• Mark Douglas or Douglas Mark
Involuntary corrections	<ul style="list-style-type: none">• Browne – Brown
Concatenated names	<ul style="list-style-type: none">• Mary Anne, Maryanne
Nicknames and aliases	<ul style="list-style-type: none">• Chris – Christine, Christopher, Tina
Noise	<ul style="list-style-type: none">• Full stops, dashes, slashes, titles, apostrophes
Abbreviations	<ul style="list-style-type: none">• Wlm/William, Mfg/Manufacturing
Truncations	<ul style="list-style-type: none">• Credit Suisse First Bost
Prefix/suffix errors	<ul style="list-style-type: none">• MacDonald/McDonald/Donald
Spelling & typing errors	<ul style="list-style-type: none">• P0rter, Beht

Variation or Error	Example
Transcription mistakes	<ul style="list-style-type: none">• Hannah, Hamah
Missing or extra tokens	<ul style="list-style-type: none">• George W Smith, George Smith, Smith
Foreign sourced data	<ul style="list-style-type: none">• Khader AL Ghamdi, Khadir A. AlGamdey
Unpredictable use of initials	<ul style="list-style-type: none">• John Alan Smith, J A Smith
Transposed characters	<ul style="list-style-type: none">• Johnson, Jhonson
Localization	<ul style="list-style-type: none">• Stanislav Milosovich – Stan Milo
Inaccurate dates	<ul style="list-style-type: none">• 12/10/1915, 21/10/1951, 10121951, 00001951
Transliteration differences	<ul style="list-style-type: none">• Gang, Kang, Kwang
Phonetic errors	<ul style="list-style-type: none">• Graeme – Graham

Why EDQ?



10hp motor 115V Yoke mount

MOT-10,115V, 48YZ,YOKE

mtr, ac(115) 10 horsepower 115volts

This 10hp yoke mounted motor is rated for 115V with a 5 year warranty

10 Caballos, Motor, 115 Voltios

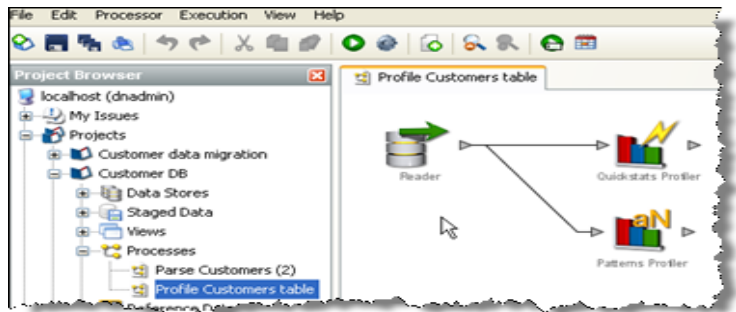
TEAO HP = 10.0 1725RPM 115V 48YZ YOKE MTR

Motor, TEAO, 1725 RPM, 48YZ, 15 Voltios, Montaje de Yugo, hp = 10

Item	Motor
Classification	26101600
Power	10 horsepower
Voltage	115
Mounting	Yoke

Product data are much more variable and unpredictable than other data types

EDQ Profiling



TITLE	Count	%
Mr	816	40.8
Ms	468	23.4
Mrs	309	15.4
Miss	251	12.5
Dr	15	0.7
Rev	1	<0.1
Prof.	1	<0.1
Col.	1	<0.1

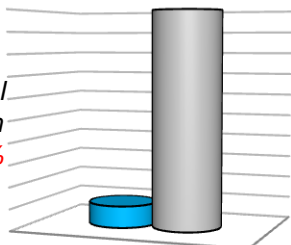


Identify and quantify issues
in data

Input field	Without data	Singleton	Duplications	Distinct values	Comment
CU_NO	1	1997	3	1998	Potentially damaged key; Investigate nulls; Investigate duplicates
CU_ACCOUNT	1	2000	0	2000	Potentially damaged key; Investigate nulls
TITLE	139	3	1859	8	
NAME	1	1980	20	1990	Potentially damaged key; Investigate nulls; Investigate duplicates
GENDER	148	0	1853	2	
BUSINESS	331	1629	41	1649	Investigate duplicates
ADDRESS1	2	1926	73	1954	Potentially damaged key; Investigate nulls; Investigate duplicates
ADDRESS2	80	554	1367	839	Investigate nulls
ADDRESS3	969	278	754	379	
POSTCODE	239	1604	158	1672	
AREA_CODE	117	64	1820	270	
TEL_NO	7	1875	119	1934	Potentially damaged key; Investigate nulls
EMAIL	65	1904	32	1920	Potentially damaged key; Investigate nulls; Investigate duplicates
ACC_MGR	5	0	1996	30	Investigate nulls
DT_PURCHASED	3	1090	908	1499	Investigate nulls
DT_ACC_OPEN	3	1093	905	1500	Investigate nulls
DT_LAST_PAYMENT	4	1026	971	1425	Investigate nulls
DT_LAST_PO_RAISED	3	1003	995	1433	Investigate nulls
BALANCE	2	7	1992	10	Investigate nulls

EDQ Parsing (Standardization), Matching

Avoid data remediation costs (manual effort, custom code) **by 80%**



■ After EDQ
□ Before EDQ

Title: Mr
First: Robert
Last: Fulmar
Gender: Male
DoB: 12/05/1978
Phone: 555-120-1329
Address:
9405 Main St
Fairfax
Virginia
22030

First: Bob
Last: Fulmar
Gender: Male
Email: chem291_rjf@barker.edu

Title: Dr
First: Robert
Last: Fulmar
Gender: Male
DoB: 12/05/1978
Email: chem291_rjf@barker.edu
Phone: 555-120-1329
Address:
9407 Main St
Fairfax
VA
22031-4001

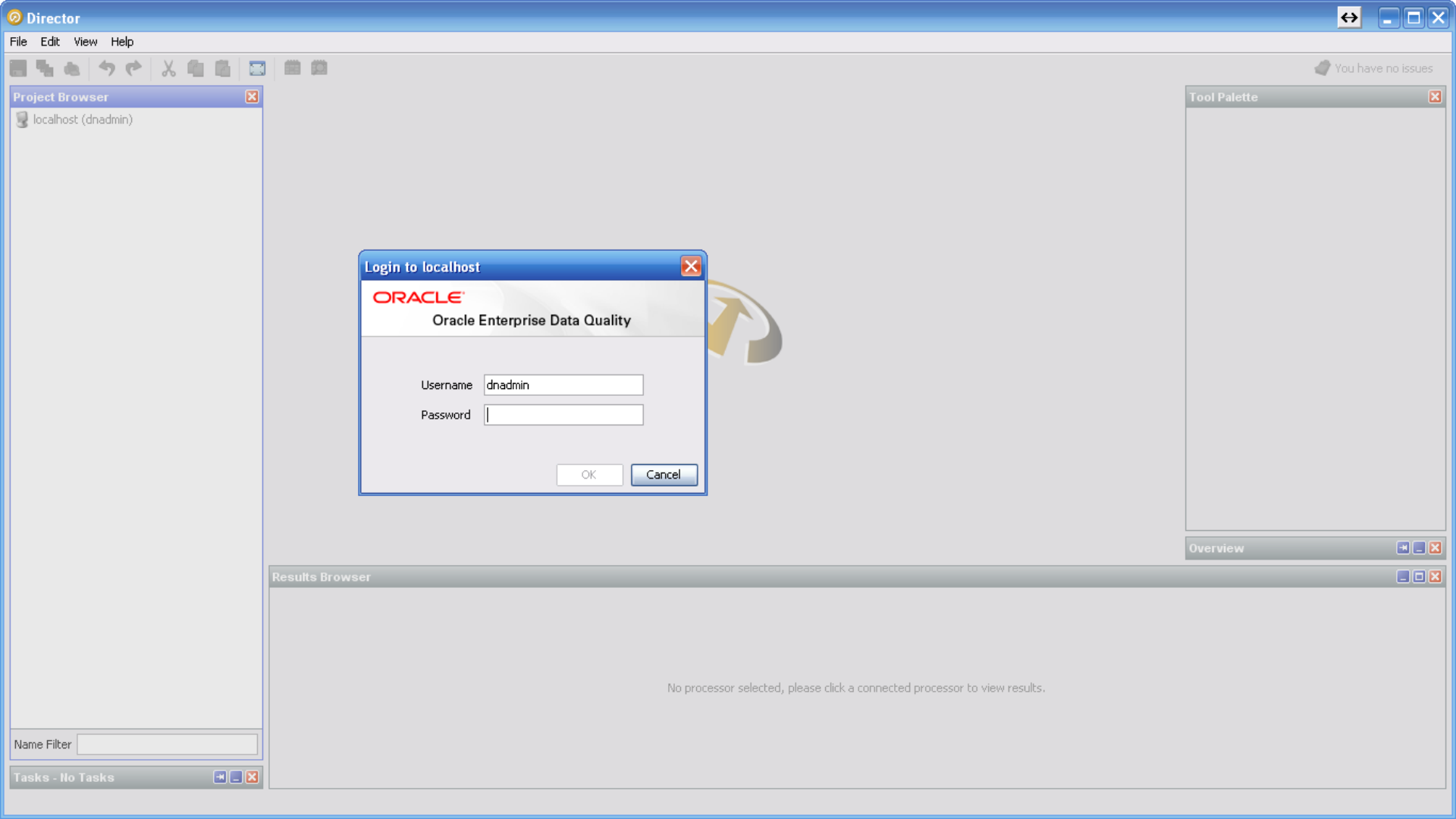
Entry Address :
"300 Berry #1210 SF California"

Parse

Validate

PremiseNumber	300	300
ThoroughfareName	Berry	Berry St
SubPremise	#1210	Unit 1210
Locality	SF	San Francisco
AdministrativeArea	California	CA
PostCode		94158-1670

Title: Dr
First: R
Last: Fulmer
DoB: 01/01/1978
Email: chem291_rjf@barker.edu
Address:
9407 Main Street
Fairfax
VA
22031-4001



Project Browser

localhost (dnadmin)

Tool Palette

You have no issues

Login to localhost



Username dnadmin

Password

OK

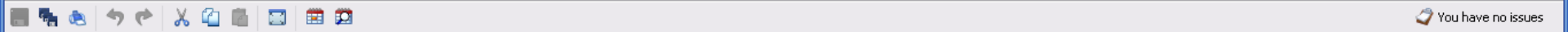
Cancel

Results Browser

No processor selected, please click a connected processor to view results.

Name Filter

Tasks - No Tasks



Project Browser

- localhost (dnadmin)
 - Projects
 - CDEP Processors on canvas
 - Compass Data Quality
 - Customer data demo - US
 - Demo
 - Data Stores
 - DemoData
 - Staged Data
 - US Customers
 - Views
 - Processes
 - Exploring US Customers
 - Reference Data
 - Results Books
 - External Tasks
 - Jobs
 - Exports
 - Web Services
 - Notes
 - Web Service demo
 - Reference Data
 - Data Stores
 - Published Processors

Exploring US Customers

```

    graph LR
      A[Read US Customers] --> B[Quickstats Profiler]
  
```

Process

Tool Palette - Profiling

- Date Profiler
- Equal Attributes Profiler
- Frequency Profiler
- Length Profiler
- Max/Min Profiler
- Number Profiler
- Patterns Profiler
- Quickstats Profiler**
- Record Completeness Profiler
- Record Duplication Profiler
- RegEx Patterns Profiler

Search

Overview

Name Filter

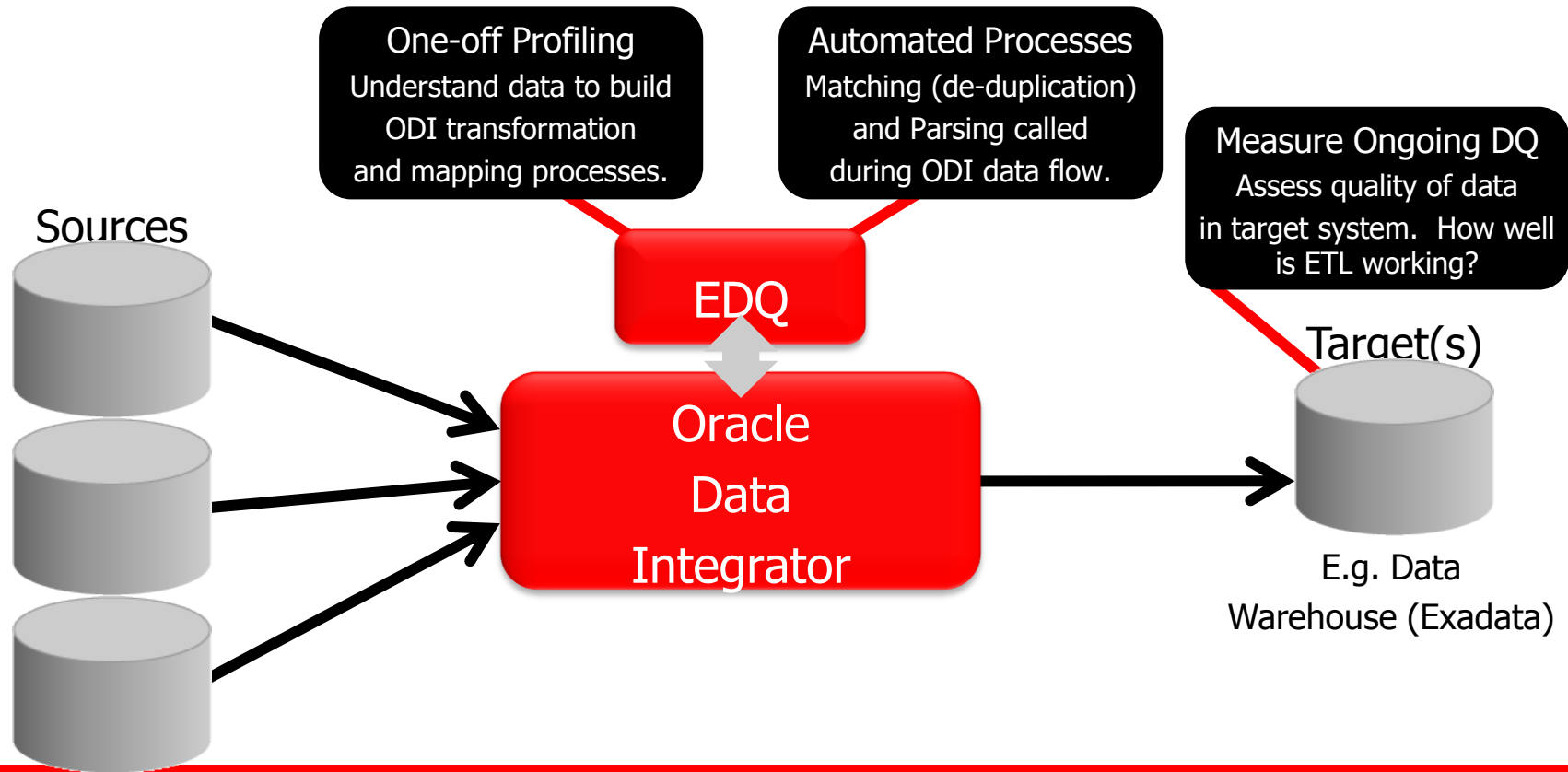
Results Browser

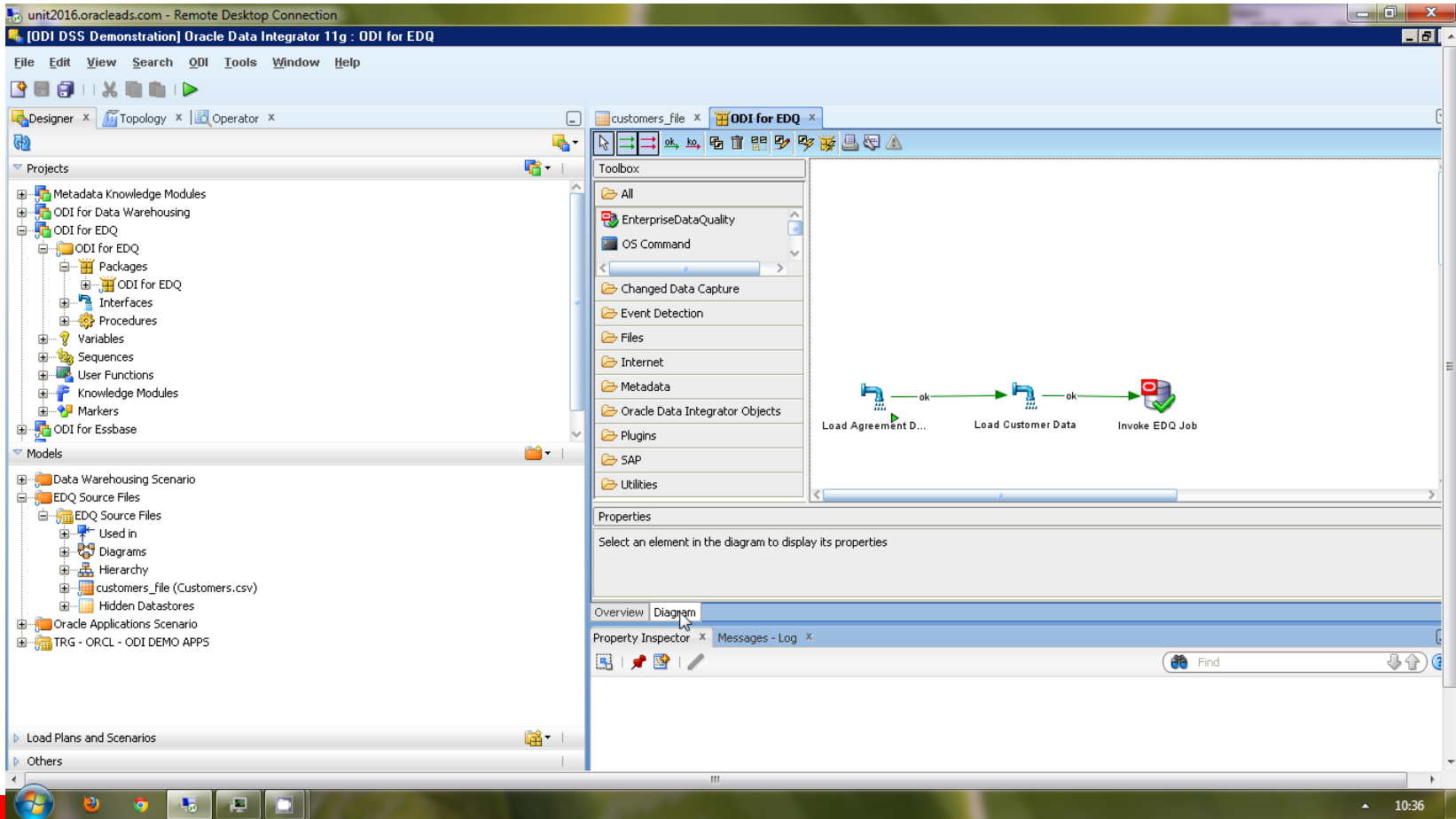
Job: Exploring US Customers Latest Run: 20-Aug-2011 15:12:02 - 15:12:17

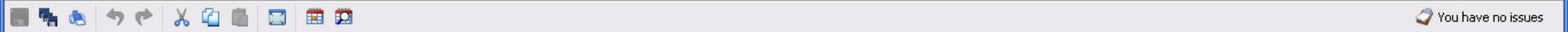
Viewing all 17 records

Input Field	Record Total	With Data	Without Data	Singleton	Duplicates	Distinct Values	Comment
ID	5438	5438	0	5438	0	5438	Complete; Possible key
Name	5438	5438	0	5327	111	5380	Complete; Potentially damaged key; Investigate duplicates
Street	5438	5438	0	5319	119	5376	Complete; Potentially damaged key; Investigate duplicates
City	5438	5438	0	396	5042	1232	Complete
State	5438	5438	0	12	5426	65	Complete
ZIP	5438	5438	?	400	4048	1823	Investigate blanks

Data Flows And EDQ







Project Browser

- localhost (dnadmin)
 - Projects
 - CDEP Processors on canvas
 - Compass Data Quality
 - Customer data demo - US
 - Demo
 - Data Stores
 - DemoData
 - Staged Data
 - US Customers
 - Views
 - Processes
 - Exploring US Customers
 - Reference Data
 - Results Books
 - External Tasks
 - Jobs
 - Exports
 - Web Services
 - Notes
 - Web Service demo
 - Reference Data
 - Data Stores
 - Published Processors

Exploring US Customers

```

    graph LR
      A[Read US Customers] --> B[Quickstats Profiler]
  
```

Process

Tool Palette - Profiling

- Date Profiler
- Equal Attributes Profiler
- Frequency Profiler
- Length Profiler
- Max/Min Profiler
- Number Profiler
- Patterns Profiler
- Quickstats Profiler**
- Record Completeness Profiler
- Record Duplication Profiler
- RegEx Patterns Profiler

Search

Overview

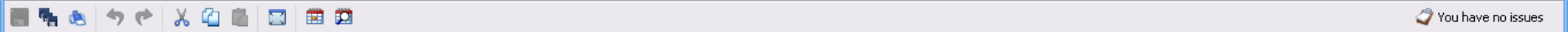
Name Filter

Results Browser

Job: Exploring US Customers Latest Run: 20-Aug-2011 15:12:02 - 15:12:17

Viewing all 17 records

Input Field	Record Total	With Data	Without Data	Singleton	Duplicates	Distinct Values	Comment
ID	5438	5438	0	5438	0	5438	Complete; Possible key
Name	5438	5438	0	5327	111	5380	Complete; Potentially damaged key; Investigate duplicates
Street	5438	5438	0	5319	119	5376	Complete; Potentially damaged key; Investigate duplicates
City	5438	5438	0	396	5042	1232	Complete
State	5438	5438	0	12	5426	65	Complete
ZIP	5438	5438	?	400	4048	1823	Investigate blanks



Project Browser

- localhost (dnadmin)
 - Projects
 - CDFP.Processors.on.canvas
 - Exploring US Customers**

Results Browser - Exploring US Customers - Quickstats Profiler

Job: Exploring US Customers Latest Run: 20-Aug-2011 15:12:02 - 15:12:17

Viewing all 17 records

Input Field	Record Total	With Data	Without Data	Singleton	Duplicates	Distinct Values	Comment
ID	5438	5438	0	5438	0	5438	Complete; Possible key
Name	5438	5438	0	5327	111	5380	Complete; Potentially damaged key; Investigate duplicates
Street	5438	5438	0	5319	119	5376	Complete; Potentially damaged key; Investigate duplicates
City	5438	5438	0	396	5042	1232	Complete
State	5438	5438	0	12	5426	65	Complete
ZIP	5438	5436	2	490	4948	1823	Investigate blanks
Country	5438	3641	1797	1	5437	10	
Phone	5438	5422	16	5214	224	5247	Potentially damaged key; Investigate blanks ; Investigate duplic
Cell	5438	2350	3088	2346	3092	2349	
Work	5438	1156	4282	1154	4284	1156	
eMail	5438	2531	2907	2325	3113	2430	
DoB	5438	5326	112	3336	2102	4220	Investigate blanks
Gender	5438	4380	1058	0	5438	4	
Active	5438	5124	314	0	5438	5	
CreditLimit	5438	5438	0	0	5438	329	Complete
StartDate	5438	3865	1573	0	5438	38	

Summary statistics view **Data**

Tool Palette - Profiling

- Date Profiler
- Equal Attributes Profiler
- Frequency Profiler
- Length Profiler
- Max/Min Profiler
- Number Profiler
- Patterns Profiler
- Quickstats Profiler**
- Record Completeness Profiler
- Record Duplication Profiler
- RegEx Patterns Profiler

Search:

Overview

🔍

Name Filter

Tasks - No Tasks

Job: Exploring US Customers Latest Run: 20-Aug-2011 15:12:02 - 15:12:17

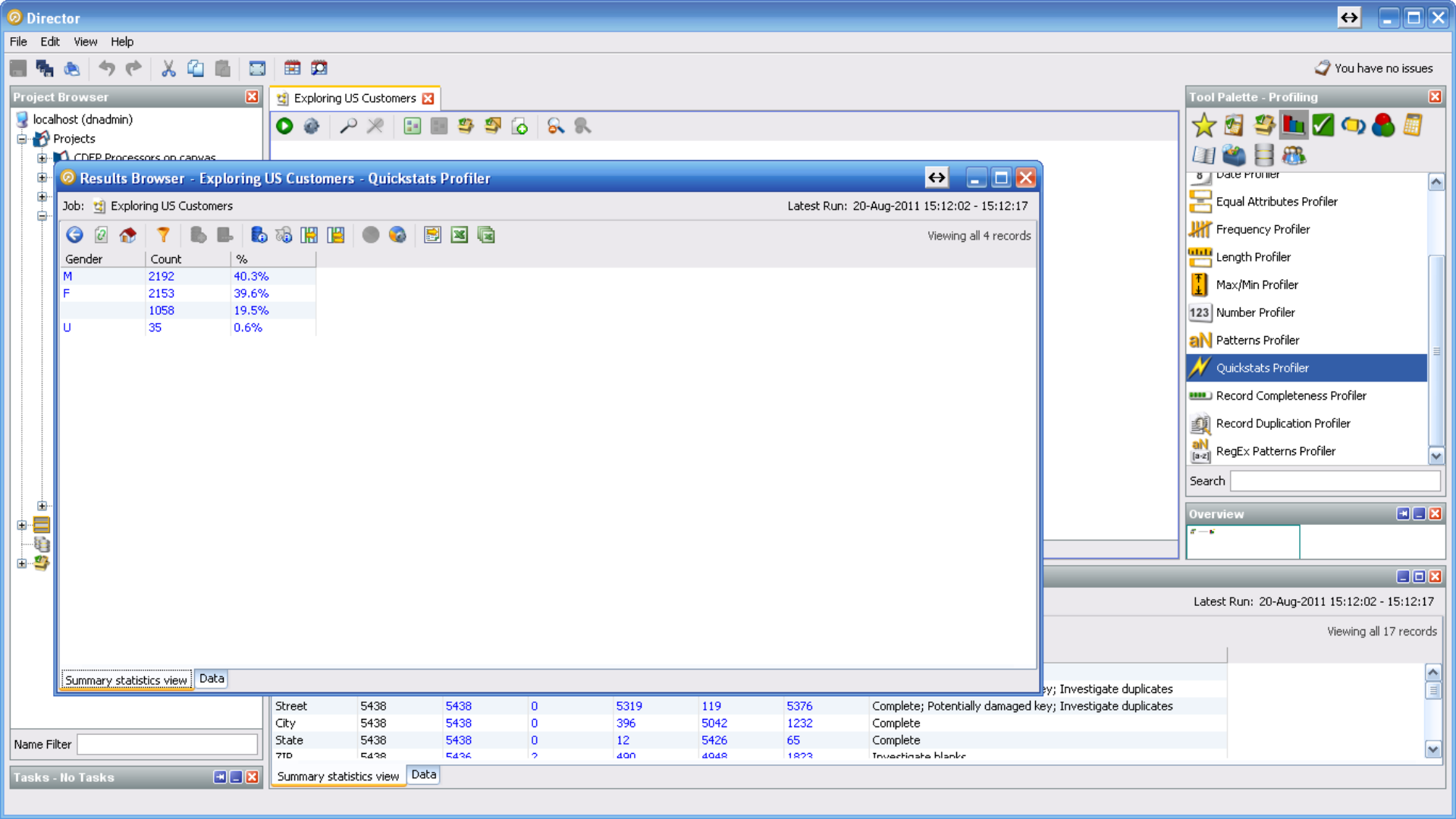
Viewing all 17 records

Input Field	Record Total	With Data	Without Data	Singleton	Duplicates	Distinct Values	Comment
ID	5438	5438	0	5438	0	5438	Complete; Possible key
Name	5438	5438	0	5327	111	5380	Complete; Potentially damaged key; Investigate duplicates
Street	5438	5438	0	5319	119	5376	Complete; Potentially damaged key; Investigate duplicates
City	5438	5438	0	396	5042	1232	Complete
State	5438	5438	0	12	5426	65	Complete
ZIP	5438	5436	2	490	4948	1823	Investigate blanks

Summary statistics view **Data**

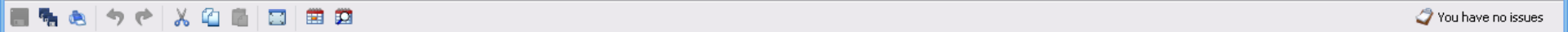
Overview

🔍



Gender	Count	%
M	2192	40.3%
F	2153	39.6%
	1058	19.5%
U	35	0.6%

Street	5438	5438	0	5319	119	5376	Complete; Potentially damaged key; Investigate duplicates
City	5438	5438	0	396	5042	1232	Complete
State	5438	5438	0	12	5426	65	Complete
ZIP	5438	5438	?	400	4048	1822	Investigate blanks



Project Browser

- localhost (dnadmin)
 - Projects
 - CDPE Processors on canvas
 - Exploring US Customers**

Exploring US Customers

🔍 ✂️ 📄 📁 🔍 🗑️ 📄 📄 🔍 🗑️

Results Browser - Exploring US Customers - Quickstats Profiler

Job: Exploring US Customers Latest Run: 20-Aug-2011 15:12:02 - 15:12:17

Viewing all 35 records

Gender	ID	Name	Street	City	State	ZIP	Country	Phone
U	MBH421308	John Mitchell	380 Beinoris Drive	WOOD DALE	IL	60191	USA	(140) 2
U	PUW442949	Irma Bailey	2101 New Beaver Avenue	PITTSBURGH	PA	15233	USA	(121) 8
U	VRL714090	Gregory Smith	3932 Ridgeoak Way	FARMERS BRANCH	TX	75234		(144) 2
U	JEH471585	Marion Chan	Old River Road	MARCY	NY	13403		(901) 3
U	YJU651180	Amanda Brown	4205 Jasper Court	ROWLETT	TX	75088	USA	(561) 8
U	BX5475882	Cecilia Ong & Charles Archer	1854 East Broadway	ALTON	IL	62002	USA	(421) 6
U	DRC711036	Victor McCoy	9300 East Smith Road	DENVER	CO	80207-1757		(753) 1
U	KB5609457	Eileen Person	3015 S Valley Avenue	MARION	IN	46953		(773) 5
U	B5C590670	Rosalind Decker	179 Darnell Lane	DRESDEN	TN	38225		(165) 3
U	PBQ406254	Dr. Jessica Frailey	Route 120 and Wilson Road	ROUND LAKE	IL	60073	U.S.A	(629) 5
U	FLA554976	Alan Hammond Associates	889 West Johnson Avenue	TERRE HAUTE	IN	47802	USA	(447) 3
U	PHF441856	Dr Jeannette Sather	1935 Motor Street	DALLAS	TX	75235		(294) 5
U	HJG532356	Ayesha Holmes	729 Navco Drive	LAFAYETTE	IN	47905	USA	(367) 6
U	MZK411179	Mary Collins	1901 N. Fountain Green Road	BEL AIR	MD	21015-1411		(561) 2
U	RSM597707	Tamara Pearson	1 Mirror Lake Drive	LAKE PLACID	NY	12946	USA	770-29
U	NYN525227	Joyce Perry	710 South 9th Street	GUNNISON	CO	81230	USA	(878) 3
U	WQG650768	Kenneth Thomas	9009 W Shawnee Mission Parkway	MERRIAM	KS	66202	USA	(153) 3
U	QSN402369	Lessie Sanchez	26673 Lawrence	CENTER LINE	MI	48015	USA	(909) 3
U	LAZ660348	Kevin Britt	12814 W Denton Avenue	LITCHFIELD PARK	AZ	85340	USA	(651) 3
U	HEJ720083	Mr Amado Faison & Ms Pat Butcher	845 Larch Avenue	ELMHURST	IL	60126	USA	(638) 8
U	ANJ609945	Ryan Arter & Emily May	4128 Rockford Road	Reno	NV	89501	USA	901-79
U	AQM457231	Ira Dudley	3348 Honeysuckle Lane	Vancouver	WA	98686		443-29

Summary statistics view | Data

Street	5438	5438	0	5319	119	5376	Complete; Potentially damaged key; Investigate duplicates
City	5438	5438	0	396	5042	1232	Complete
State	5438	5438	0	12	5426	65	Complete
710	5438	5438	?	400	4048	1823	Investigate blank

Name Filter

Tasks - No Tasks

Summary statistics view | Data

Tool Palette - Profiling

- Date Profiler
- Equal Attributes Profiler
- Frequency Profiler
- Length Profiler
- Max/Min Profiler
- Number Profiler
- Patterns Profiler
- Quickstats Profiler**
- Record Completeness Profiler
- Record Duplication Profiler
- Regex Patterns Profiler

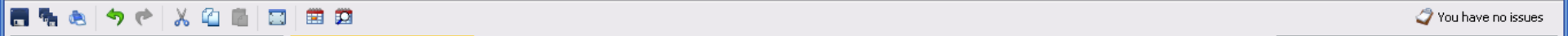
Search

Overview

Latest Run: 20-Aug-2011 15:12:02 - 15:12:17

Viewing all 17 records

Investigate duplicates

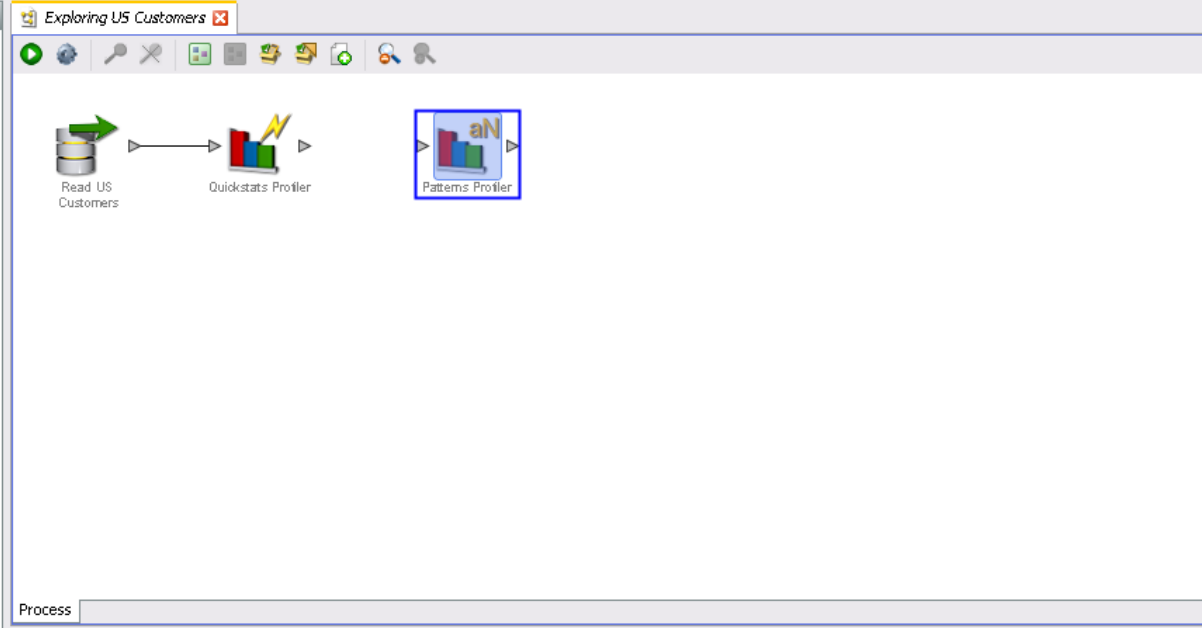


Project Browser

- localhost (dnadmin)
 - Projects
 - CDEP Processors on canvas
 - Compass Data Quality
 - Customer data demo - US
 - Demo
 - Data Stores
 - DemoData
 - Staged Data
 - US Customers
 - Views
 - Processes
 - Exploring US Customers
 - Reference Data
 - Valid Genders
 - Results Books
 - External Tasks
 - Jobs
 - Exports
 - Web Services
 - Notes
 - Web Service demo
 - Reference Data
 - Data Stores
 - Published Processors

Name Filter

Tasks - No Tasks



Tool Palette - Profiling

You have no issues

- Date Profiler
- Equal Attributes Profiler
- Frequency Profiler
- Length Profiler
- Max/Min Profiler
- Number Profiler
- aN Patterns Profiler**
- Quickstats Profiler
- Record Completeness Profiler
- Record Duplication Profiler
- RegEx Patterns Profiler

Search

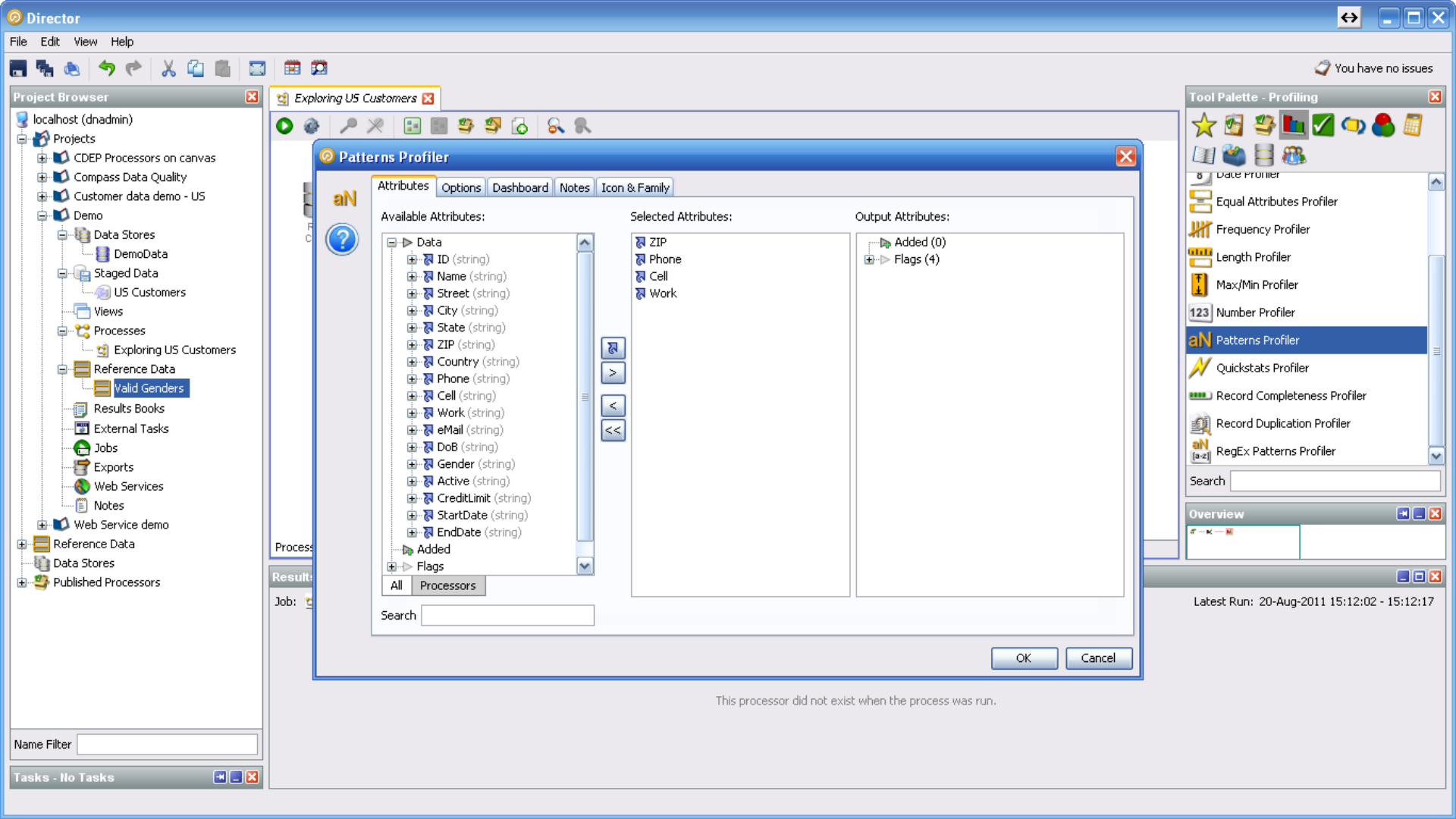
Overview

⏪ ⏩ ⏴ ⏵

Results Browser

Job: Exploring US Customers Latest Run: 20-Aug-2011 15:12:02 - 15:12:17

This processor did not exist when the process was run.



Project Browser

- localhost (dnadmin)
- Projects
 - CDEP Processors on canvas
 - Compass Data Quality
 - Customer data demo - US
 - Demo
 - Data Stores
 - DemoData
 - Staged Data
 - US Customers
 - Views
 - Processes
 - Exploring US Customers
 - Reference Data
 - Valid Genders
 - Results Books
 - External Tasks
 - Jobs
 - Exports
 - Web Services
 - Notes
 - Web Service demo
- Reference Data
- Data Stores
- Published Processors

Exploring US Customers



Patterns Profiler

- Attributes
- Options
- Dashboard
- Notes
- Icon & Family

Available Attributes:

- Data
 - ID (string)
 - Name (string)
 - Street (string)
 - City (string)
 - State (string)
 - ZIP (string)
 - Country (string)
 - Phone (string)
 - Cell (string)
 - Work (string)
 - eMail (string)
 - DoB (string)
 - Gender (string)
 - Active (string)
 - CreditLimit (string)
 - StartDate (string)
 - EndDate (string)
- Added
- Flags

Selected Attributes:

- ZIP
- Phone
- Cell
- Work

Output Attributes:

- Added (0)
- Flags (4)

Buttons: [Add], [Remove], [Left], [Right], [Double Left], [Double Right]

Search:

All Processors

OK Cancel

This processor did not exist when the process was run.

Tool Palette - Profiling



- Date Profiler
 - Equal Attributes Profiler
 - Frequency Profiler
 - Length Profiler
 - Max/Min Profiler
 - Number Profiler
 - Patterns Profiler**
 - Quickstats Profiler
 - Record Completeness Profiler
 - Record Duplication Profiler
 - Regex Patterns Profiler
- Search:

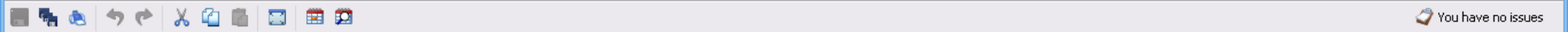
Overview



Latest Run: 20-Aug-2011 15:12:02 - 15:12:17

Name Filter:

Tasks - No Tasks [Navigation icons]



Project Browser

- localhost (dnadmin)
 - Projects
 - CDEP Processors on canvas
 - Exploring US Customers

ZIP Phone Cell Work Data

Results Browser - Exploring US Customers - Patterns Profiler

Job: Exploring US Customers Latest Run: 20-Aug-2011 15:19:55 - 15:19:57

Viewing all 8 records

Pattern	Length	Count	%
NNNNN	5	5172	95.1%
NNNNN-NNNN	10	190	3.5%
aNa_NaN	7	50	0.9%
NNNN	4	16	0.3%
aNaNNaN	6	4	<0.1%
aaN_Naa	7	3	<0.1%
	0	2	<0.1%
aNa-NaN	7	1	<0.1%

ZIP Phone Cell Work Data

Tool Palette - Profiling

You have no issues

- Date Profiler
- Equal Attributes Profiler
- Frequency Profiler
- Length Profiler
- Max/Min Profiler
- Number Profiler
- aN Patterns Profiler**
- Quickstats Profiler
- Record Completeness Profiler
- Record Duplication Profiler
- RegEx Patterns Profiler

Search

Overview

Name Filter

Tasks - No Tasks

Results Browser - Exploring US Customers - Patterns Profiler

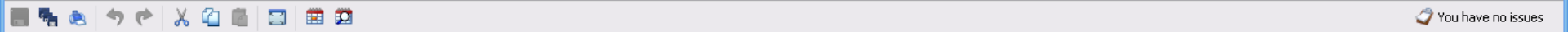
Job: Exploring US Customers Latest Run: 20-Aug-2011 15:19:55 - 15:19:57

Viewing all 8 records

Pattern	Length	Count	%
NNNNN	5	5172	95.1%
NNNNN-NNNN	10	190	3.5%
aNa_NaN	7	50	0.9%
NNNN	4	16	0.3%
aNaNNaN	6	4	<0.1%
aaN_Naa	7	3	<0.1%

ZIP Phone Cell Work Data

Overview



Project Browser

- localhost (dnadmin)
 - Projects
 - CDPEProcessors.on.ca.was
 - Exploring US Customers**

Results Browser - Exploring US Customers - Patterns Profiler

Job: Exploring US Customers Latest Run: 20-Aug-2011 15:19:55 - 15:19:57

Viewing all 50 records

ZIP	Phone	Cell	Work	ID	Name	Street	City	Sta
VOR 2W0	(138) 347 3124	(313)	(735)	PSN568898	Mr Dorian Davenport	2783 GREGORY RD RR1	SHAWNIGAN LAKE	BC
A1C 5X3	(525) 813 7072	(740)	(282)	PYC446898	Mr Reilly Shaw	34 Glencoe Drive	St. John's	NL
R2G 4E9	(518) 861 6975	(590)	(390)	PXA422898	Mr Braxton Nuyeng	1795 Henderson Hwy	Winnipeg	MB
N2H 6M3	(697) 636 6472	535 4714	358 7537	PSU592898	Ms Makaila Sevigne	640 TRILLIUM DRIVE	KITCHENER	ON
P1B 8K1	(122) 125 2385	286 3351		PTZ458898	Mrs Juliette Garrett	710 MCKEOWN AVE	NORTH BAY	ON
V8W 9V1	(852) 288 7982	139 8768	803 2263	PTL676898	Mr Jordan Sturgess	4000 SEYMOUR PLACE	VICTORIA	BC
X0A 1H0	(213) 907 8837	(361) 546 5849	(603)	PVC625898	Mr Cortez Herchy	APT # 302 PAUNA PLACE	IQUALUIT	NT
K8V 5R5	(634) 511 8366	(222) 314 5047	(810)	PTV412898	Mrs Cassidy Queener	9 RIVERSIDE DRIVE	TRENTON	ON
M4K 3Z3	(831) 707 7808			PVD717898	Mrs Janessa Van Deventer	1032 Pape Ave	Toronto	ON
N8S 4W1	(750) 207 5990	(209)		PVQ413898	Mr Colby Roberto	5745 WYANDOTTE ST E	WINDSOR	ON
V9R 5N3	(411) 780 3073			PYB574898	Mr Alec Kiddley	6250 HAMMOND BAY ROAD	NANAIMO	BC
MSK 1J5	(444) 189 2936	(129)		PUE649898	Mrs Angelina Merlo	222 Bay St. TD Centre	Toronto	ON
V8W 9W6	(446) 834 2038	679 7717		PSW682898	Mr Maximus Balmin	2975 Jutland Road, 2nd Floor	Victoria	BC
L4R 4K6	(314) 903 3727	(693)	(510)	PXM602898	Ms Sarah Shieh	250 Second Street	Midland	ON
H4N 3J1	(609) 594 6485	478 8445		PYH478898	Mrs Aleah Sankaranarayanan	237 COTE-VERTU	SAINT-LAURENT	QC
M4P 1P9	(182) 572 6590	(725) 214 7761		PWG452898	Mr Jakob Brandie	7053 Longwoods Road	London	ON

ZIP Phone Cell Work Data

Tool Palette - Profiling

- Date Profiler
- Equal Attributes Profiler
- Frequency Profiler
- Length Profiler
- Max/Min Profiler
- 123 Number Profiler
- aN Patterns Profiler**
- Quickstats Profiler
- Record Completeness Profiler
- Record Duplication Profiler
- RegEx Patterns Profiler

Search

Overview

Name Filter

Tasks - No Tasks

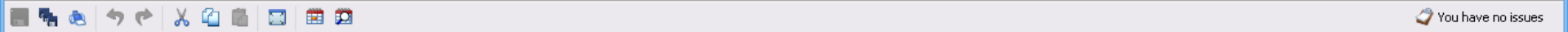
Job: Exploring US Customers Latest Run: 20-Aug-2011 15:19:55 - 15:19:57

Viewing all 8 records

Pattern	Length	Count	%
NNNNN	5	5172	95.1%
NNNNN-NNNN	10	190	3.5%
aNa_NaN	7	50	0.9%
NNNN	4	16	0.3%
aNaNaN	6	4	<0.1%
NaN NaN	7	3	<0.1%

ZIP Phone Cell Work Data

Overview



Project Browser

- localhost (dnadmin)
 - Projects
 - CDFP Processors on canvas
 - Results Browser - Exploring US Customers - Patterns Profiler

Exploring US Customers

Job: Exploring US Customers Latest Run: 20-Aug-2011 15:19:55 - 15:19:57

Viewing all 29 records

Pattern	Length	Count	%
(NNN)_NNN_NNNN	14	5006	92.1%
NNN-NNN-NNNN	12	226	4.2%
aa_aaa_aaaa	11	68	1.3%
NNN	3	43	0.8%
aa_aaaaa	8	25	0.5%
	0	16	0.3%
.	1	12	0.2%
a	1	8	0.1%
NNN_NNN_NNNN	12	6	0.1%
+N_NNN_NNN_NNNN	15	4	<0.1%
(NNN)_NNN_NNN	13	3	<0.1%
NNNN-NNN-NNNN	13	2	<0.1%
(NNN)_NNN-NNNN	14	2	<0.1%
NNNNNN_NNNN	11	2	<0.1%
+N_NNN)_NNN_NNNN	16	1	<0.1%
(NNN)_NNNNNNN	13	1	<0.1%
NN-NNNN-NNN-NNN	15	1	<0.1%

ZIP Phone Cell Work Data

Tool Palette - Profiling

You have no issues

- Date Profiler
- Equal Attributes Profiler
- Frequency Profiler
- Length Profiler
- Max/Min Profiler
- Number Profiler
- aN Patterns Profiler**
- Quickstats Profiler
- Record Completeness Profiler
- Record Duplication Profiler
- RegEx Patterns Profiler

Search

Overview

Name Filter

Tasks - No Tasks

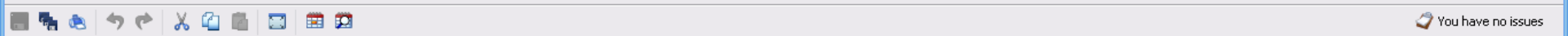
Job: Exploring US Customers Latest Run: 20-Aug-2011 15:19:55 - 15:19:57

Viewing all 8 records

Pattern	Length	Count	%
NNNNN	5	5172	95.1%
NNNNN-NNNN	10	190	3.5%
aNa_NaN	7	50	0.9%
NNNN	4	16	0.3%
aNaNaN	6	4	<0.1%
aaAl_Naa	7	3	<0.1%

ZIP Phone Cell Work Data

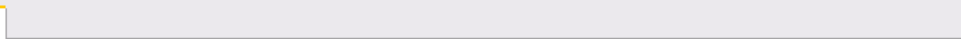
Overview



Project Browser

- localhost (dnadmin)
 - Projects
 - CDFP.Processors.on.canvas

Exploring US Customers



Results Browser - Exploring US Customers - Patterns Profiler

Job: Exploring US Customers Latest Run: 20-Aug-2011 15:19:55 - 15:19:57

Viewing all 68 records

Phone	ZIP	Cell	Work	ID	Name	Street	City	State
do not call	46208			TD2630563	Mrs Stacey Ocampo	1000 W 42nd Street	INDIANAPOLIS	IN
do not call	15108			HMW7272725	Mrs Rebecca Koon	100 Airside Drive	MOON TOWNSHIP	PA
do not call	79720			YUQ599956	Mrs Lisa Baker	1001 N Birdwell Lane	BIG SPRING	TX
do not call	65708			QMX547701	Mrs Karen Salas	10 Dairy Street	MONETT	MO
do not call	38305			PSE620551	Mr Blake Owens	1008 Old Hickory Boulevard	JACKSON	TN
do not call	35811			MHF413435	Mr Paul Robinson	1008 Oakwood Avenue	HUNTSVILLE	AL
do not call	21030			XPV448872	Mrs Ruth Reese	10 North Park Drive	HUNT VALLEY	MD
do not call	15239			LFY468588	Mrs Helen Mills	1001 Millers Lane	PITTSBURGH	PA
do not call	64060			YKQ546540	Mr John Prater	1000 West 92 Highway	KEARNEY	MO
do not call	33025			UMM550083	Mr David Harrison	1000 SW 84 Avenue	PEMBROKE PINES	FL
do not call	12701			WEH463268	Mr Marvin Pacheco	100 North Street	MONTECELLO	NY
do not call	76177			HPA565680	Mrs Brenda Anderson	1005 Railroad Drive	FORT WORTH	TX
do not call	60030			QDB690101	Mr Stuart Foster	100 Library Lane	GRAYSLAKE	VA
do not call	22812			TJZ731131	Mrs Temeka Halcomb	100 Quality Street	BRIDGEWATER	IL
do not call	48043			LFR175958	Mr Tristan Chambers	10 North Main	MOUNT CLEMENS	MI
do not call	66048			GXM679114	Ms. Sarah Reinke	100 N 5th Street	FAVEMWORTH	KS

ZIP Phone Cell Work Data

You have no issues

Tool Palette - Profiling

- Date Profiler
- Equal Attributes Profiler
- Frequency Profiler
- Length Profiler
- Max/Min Profiler
- Number Profiler
- Patterns Profiler**
- Quickstats Profiler
- Record Completeness Profiler
- Record Duplication Profiler
- Regex Patterns Profiler

Search

Overview

Job: Exploring US Customers Latest Run: 20-Aug-2011 15:19:55 - 15:19:57

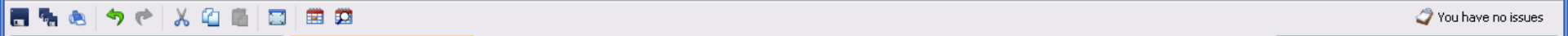
Viewing all 8 records

Pattern	Length	Count	%
NNNNN	5	5172	95.1%
NNNNN-NNNN	10	190	3.5%
aNa_NaN	7	50	0.9%
NNNN	4	16	0.3%
aNaNNaN	6	4	<0.1%
NaN NaN	7	3	<0.1%

Name Filter

Tasks - No Tasks

ZIP Phone Cell Work Data

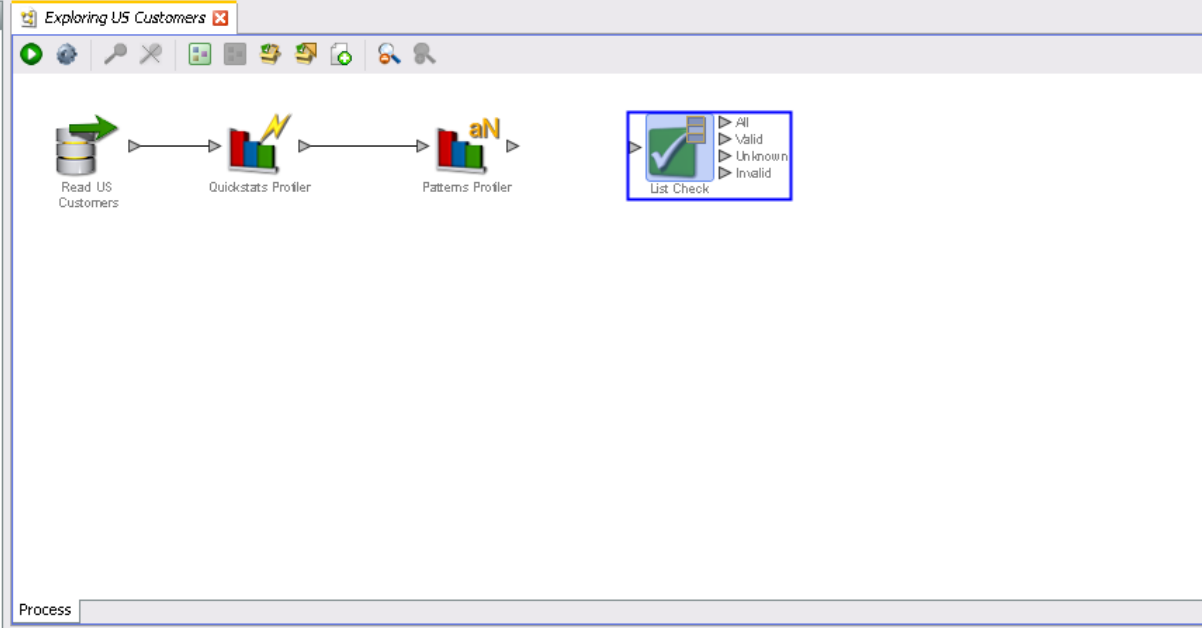


Project Browser

- localhost (dnadmin)
 - Projects
 - CDEP Processors on canvas
 - Compass Data Quality
 - Customer data demo - US
 - Demo
 - Data Stores
 - DemoData
 - Staged Data
 - US Customers
 - Views
 - Processes
 - Exploring US Customers
 - Reference Data
 - Valid Genders
 - Results Books
 - External Tasks
 - Jobs
 - Exports
 - Web Services
 - Notes
 - Web Service demo
 - Reference Data
 - Data Stores
 - Published Processors

Name Filter

Tasks - No Tasks



Tool Palette - Audit

You have no issues

- Business Rules Check
- Cross Attribute Check
- Data Type Check
- Duplicate Check
- Email Check
- GBR Postcode Format Check
- Invalid Character Check
- Length Check
- List Check
- Logic Check
- Lookup Check

Search

Overview

Navigation icons: back, forward, search, etc.

Results Browser

Job: Exploring US Customers

Latest Run: 20-Aug-2011 15:19:55 - 15:19:57

This processor did not exist when the process was run.

Director

File Edit View Help

You have no issues

Project Browser

- localhost (dnadmin)
- Projects
 - CDEP Processors on canvas
 - Compass Data Quality
 - Customer data demo - US
 - Demo
 - Data Stores
 - DemoData
 - Staged Data
 - US Customers
 - Views
 - Processes
 - Exploring US Customers
 - Reference Data
 - Valid Genders
 - Results Books
 - External Tasks
 - Jobs
 - Exports
 - Web Services
 - Notes
 - Web Service demo
 - Reference Data
 - Data Stores
 - Published Processors

Name Filter

Tasks - No Tasks

Exploring US Customers

Exploring US Customers - Gender Check - Summary

Exploring US Customers - Gender Check - Summary

Record Type	Count
Valid Records	4345
Unknown Records	~1000
Invalid Records	~350

Process

Results B

Job: E

Valid Recc
4345

Summary Data

Tool Palette - Audit

- Business Rules Check
- Cross Attribute Check
- Data Type Check
- Duplicate Check
- Email Check
- GBR Postcode Format Check
- Invalid Character Check
- Length Check
- List Check
- Logic Check
- Lookup Check

Search

Overview

Latest Run: 20-Aug-2011 15:25:20 - 15:25:21

Viewing all 1 records

Director

File Edit View Help

You have no issues

Project Browser

- localhost (dnadmin)
- Projects
 - CDEP Processors on canvas
 - Compass Data Quality
 - Customer data demo - US
 - Demo
 - Data Stores
 - DemoData
 - Staged Data
 - US Customers
 - Views
 - Processes
 - Exploring US Customers
 - Reference Data
 - Valid Genders
 - Results Books
 - External Tasks
 - Jobs
 - Exports
 - Web Services
 - Notes
 - Web Service demo
- Reference Data
- Data Stores
- Published Processors

Name Filter

Tasks - No Tasks

Exploring US Customers

Exploring US Customers - Gender Check

Exploring US Customers - Gender Check

Gender	Count	Percentage
F	2153	39.6%
M	3087	60.4%

Process

Results B

Job: E

Value

M

F

Summary Data

Tool Palette - Audit

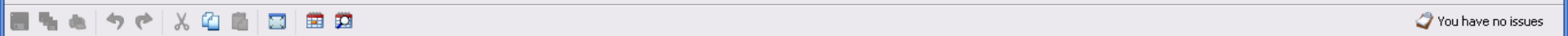
- Business Rules Check
- Cross Attribute Check
- Data Type Check
- Duplicate Check
- Email Check
- GBR Postcode Format Check
- Invalid Character Check
- Length Check
- List Check
- Logic Check
- Lookup Check

Search

Overview

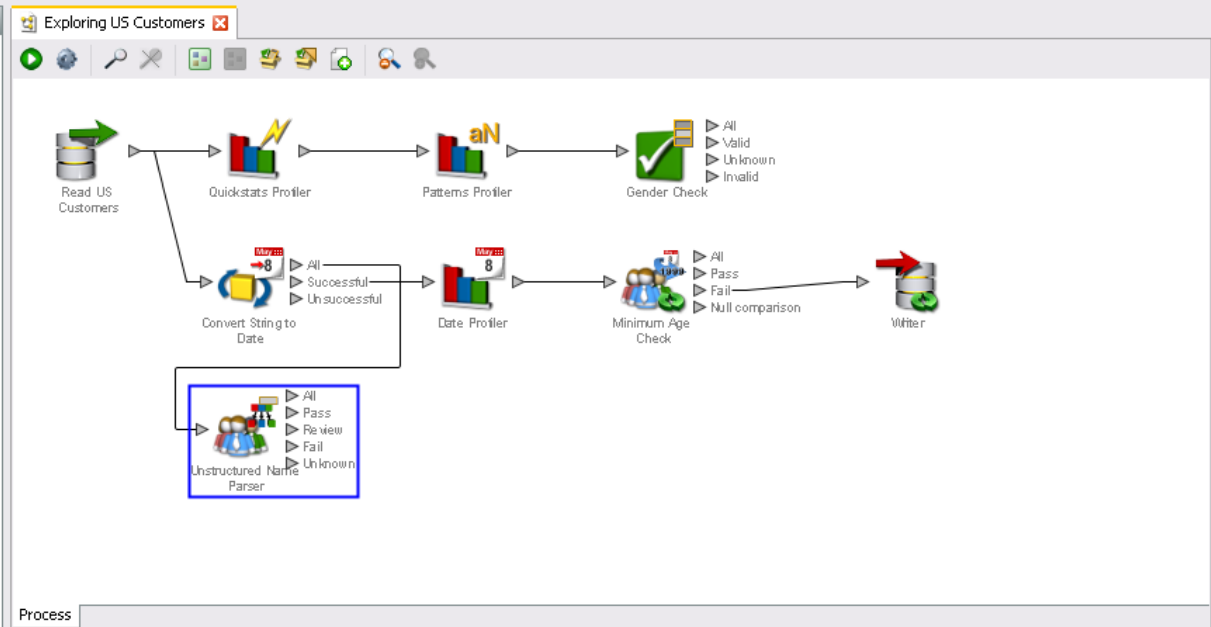
Latest Run: 20-Aug-2011 15:25:20 - 15:25:21

Viewing all 2 records



Project Browser

- localhost (dnadmin)
 - Projects
 - CDEP Processors on canvas
 - Compass Data Quality
 - Customer data demo - US
 - Demo
 - Data Stores
 - DemoData
 - Staged Data
 - Underage Customers
 - US Customers
 - Views
 - Processes
 - Exploring US Customers
 - Reference Data
 - Valid Genders
 - Results Books
 - External Tasks
 - Jobs
 - Exports
 - Web Services
 - Notes
 - Web Service demo
 - Reference Data
 - Data Stores
 - Published Processors



Tool Palette - Customer Data

- GeoNames Country Codes from City
- Get Year from Date
- Match Entities
- Match Households
- Match Individuals (Name, Address, DoB)
- Match Individuals (Name, Address)
- Profile Entity Names
- Standardize Country Names
- Standardize Entity Names
- Structured Name Parser

Search

Overview

Name Filter

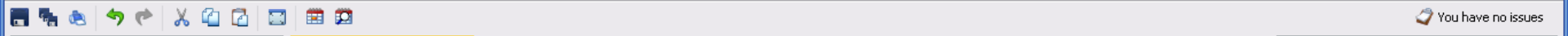
Results Browser

Job: Exploring US Customers Latest Run: 20-Aug-2011 15:52:11 - 15:52:45

Viewing 100 records of 5,438

Name	UnclassifiedData	P1_Prefix	P1_First	P1_Middle	P1_Last	P1_Suffix	A1_Prefix	A1_First	A1_Middle	A1_Last
Mr Everett Hughes & Miss Jane Young	&	Mr	Everett	Hughes	Young					
Mr Edward Yates & Mrs Evangelina Yates	&	Mr	Edward	Yates	Yates					
Mrs Maria Verdin and Dr John Roberts	and	Mrs	Maria	Verdin	Roberts					
Ryan Arter & Emily May	&		Ryan		Arter					
Mr Amada Espino & Mr Bob Butcher	&	Mr	Amada	Espino	Butcher					

Base Tokenization | Token Checks | Classification | Unclassified Tokens | Reclassification Rules | Reclassification | Selection | Resolution Rules | Results | **Data**

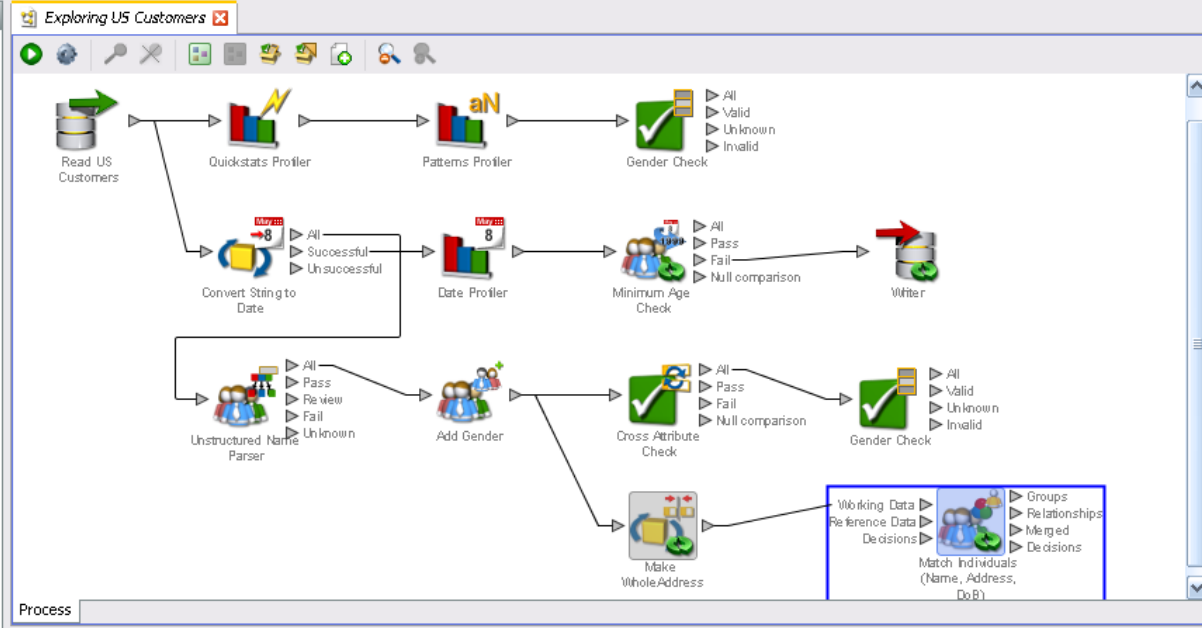


Project Browser

- localhost (dnadmin)
 - Projects
 - CDEP Processors on canvas
 - Compass Data Quality
 - Customer data demo - US
 - Demo
 - Data Stores
 - DemoData
 - Staged Data
 - Underage Customers
 - US Customers
 - Views
 - Processes
 - Exploring US Customers
 - Reference Data
 - Valid Genders
 - Results Books
 - Gender improvement
 - External Tasks
 - Jobs
 - Exports
 - Web Services
 - Notes
 - Web Service demo
 - Reference Data
 - Data Stores
 - Published Processors

Name Filter

Tasks - No Tasks



Tool Palette - Customer Data

- Match Entities
- Match Households
- Match Individuals (Name, Address, DoB)**
- Match Individuals (Name, Address)
- Profile Entity Names
- Standardize Country Names
- Standardize Entity Names
- Structured Name Parser
- Unstructured Name Parser
- URL Check

Search

Overview

Results Browser

Job: Exploring US Customers

Latest Run: 20-Aug-2011 16:02:34 - 16:02:50

This processor did not exist when the process was run.

Project Browser

- localhost (dnadmin)
 - Projects
 - CDEP Processors on canvas
 - Compass Data Quality
 - Customer data demo - US
 - Demo
 - Data Stores
 - DemoData
 - Staged Data
 - Underage Customers
 - US Customers
 - Views
 - Processes
 - Exploring US Customers
 - Reference Data
 - Valid Genders
 - Results Books
 - Gender improvement
 - External Tasks
 - Jobs
 - Exports
 - Web Services
 - Notes
 - Web Service demo
 - Reference Data
 - Data Stores
 - Published Processors

Name Filter:

Tasks - No Tasks

Exploring US Customers

Match Individuals (Name, Address, DoB)

Advanced Options | Assign Relationship Review
 Review Results | Assign Merged Review
 Configure Bulk Review Rules | View Match Statistics
 Delete Realtime Review Results | Delete Manual Decisions

Input Identify Cluster **Match** Merge

Rule	Priority	Decision
<input checked="" type="checkbox"/> DOB exact, Name, Address	0	MATCH
<input checked="" type="checkbox"/> DOB no data, Name, Address	0	MATCH
<input checked="" type="checkbox"/> DOB exact, Name, Postcode	0	MATCH
<input checked="" type="checkbox"/> DOB transposed, Name, Address	0	MATCH
<input checked="" type="checkbox"/> DOB close, Name, Address	0	MATCH
<input checked="" type="checkbox"/> YOBI, Name, Address	0	MATCH
<input checked="" type="checkbox"/> DOR exact, Name, Address	0	MATCH

Show Summary

Process: Match Individuals (Name, Address, DoB) X

Tool Palette - Customer Data

- GeoNames Country Codes from City
- Get Year from Date
- Match Entities
- Match Households
- Match Individuals (Name, Address, DoB)**
- Match Individuals (Name, Address)
- Profile Entity Names
- Standardize Country Names
- Standardize Entity Names
- Structured Name Parser
- Unstructured Name Parser
- URL Check

Search:

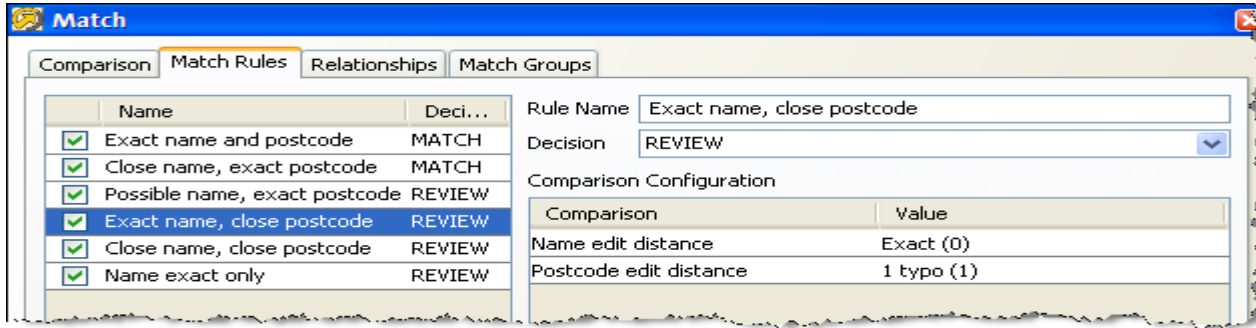
Overview

Results Browser

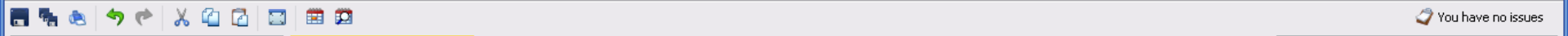
Job: Exploring US Customers

Latest Run: 20-Aug-2011 16:02:34 - 16:02:50

This processor did not exist when the process was run.



- A Match Rule is simply the combination of comparison results
- Rules are evaluated in order and if one hits, we stop
- Rules can be 'negative' to eliminate pairs that are too different with a 'No Match' rule
- Rules can easily be turned on & off during the tuning process

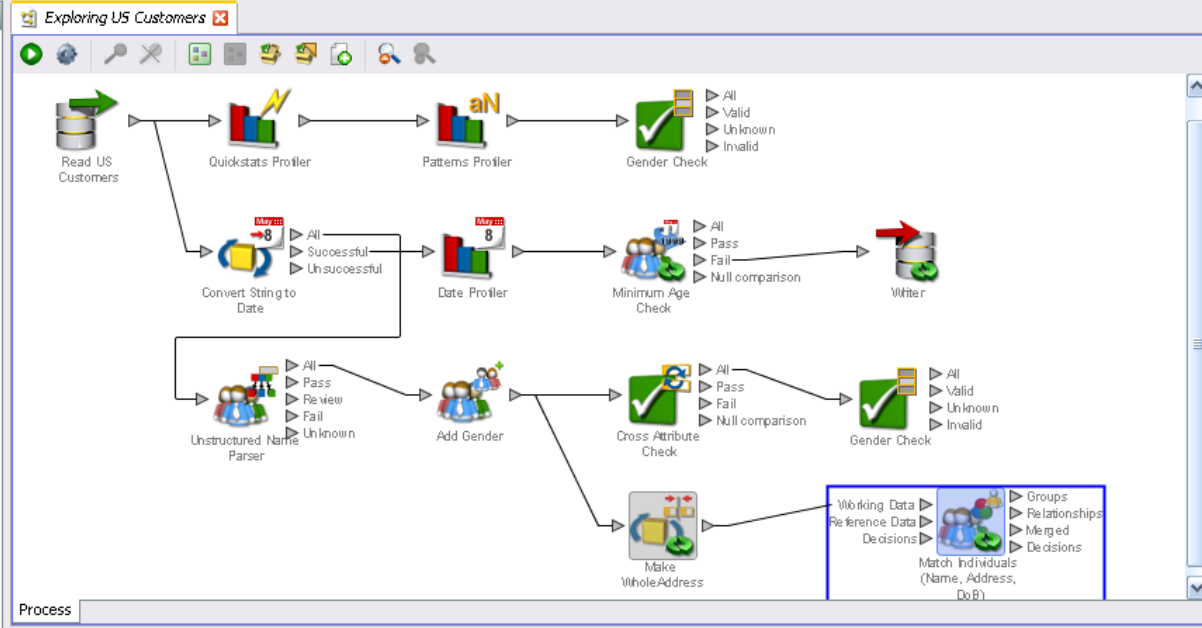


Project Browser

- localhost (dnadmin)
 - Projects
 - CDEP Processors on canvas
 - Compass Data Quality
 - Customer data demo - US
 - Demo
 - Data Stores
 - DemoData
 - Staged Data
 - Underage Customers
 - US Customers
 - Views
 - Processes
 - Exploring US Customers
 - Reference Data
 - Valid Genders
 - Results Books
 - Gender improvement
 - External Tasks
 - Jobs
 - Exports
 - Web Services
 - Notes
 - Web Service demo
 - Reference Data
 - Data Stores
 - Published Processors

Name Filter

Tasks - No Tasks



Tool Palette - Customer Data

- Match Entities
- Match Households
- Match Individuals (Name, Address, DoB)**
- Match Individuals (Name, Address)
- Profile Entity Names
- Standardize Country Names
- Standardize Entity Names
- Structured Name Parser
- Unstructured Name Parser
- URL Check

Search

Overview

Results Browser

Job: Exploring US Customers

Latest Run: 20-Aug-2011 16:02:34 - 16:02:50

This processor did not exist when the process was run.



Project Browser

- localhost (dnadmin)
 - Projects
 - CDEP Processors on canvas
 - Compass Data Quality
 - Customer data demo - US
 - Demo
 - Data Stores
 - DemoData
 - Staged Data
 - Underage Customers
 - US Customers
 - Views
 - Processes
 - Exploring US Customers
 - Reference Data
 - Valid Genders
 - Results Books
 - Gender improvement
 - External Tasks
 - Jobs
 - Exports
 - Web Services
 - Notes
 - Web Service demo
 - Reference Data
 - Data Stores
 - Published Processors

Name Filter

Matching Records Review - Demo\Exploring US Customers\Match Individuals (Name, Address, DoB) ↔ - □ ×

Viewing group 1 of 8 Additional groups are available for loading Review merged output

Review Review Merged Output

Filter Groups

Look For Search In Find Use OR logic

Review status = Awaiting Review Clear Case Sensitive

Exact Match

Highlight Differences

Records

	Internal ID	Match Group	Data Stream	Title	Given Names	Family Name	Full Name	DOB	YOB	Address1
R1	44	62	US Customers	Ms	Sheila	Gibson	Ms. Sheila Gibson	14-Nov-1959 00:00:00		11001 West 120th Avenue
R2	1242	1242	US Customers	Ms	Sheila	Gibson	Ms. Sheila Gibson	14-Nov-1959 00:00:00		111 Emerson Street

Relationships Auto select relationships based on search

Record 1	Group 1	Record 2	Group 2	Match Rule	Review Status	Decision	User	Date		
R1	62	R2	1242	*	DOB exact, Name	Awaiting Review	Possible Match			

Data Stream: US Customers First 4 surname meta + first 4 postcode First 2 forename stand + YOB + First 2 Postcode + First 3 surname Matching Rules

Review Status Match Groups Merge Summary Groups Output Relationships Output Merged Output

Tool Palette - Customer Data ↔ - □ ×

You have no issues

- GeoNames Country Codes from City
- Get Year from Date
- Match Entities
- Match Households
- Match Individuals (Name, Address, DoB)**
- Match Individuals (Name, Address)
- Profile Entity Names
- Standardize Country Names
- Standardize Entity Names
- Structured Name Parser
- Unstructured Name Parser
- URL Check

Search

Overview ↔ - □ ×

Latest Run: 20-Aug-2011 16:10:55 - 16:11:29

Viewing all 27 records

Matching Records Review - Demo\Exploring US Customers\Match Individuals (Name, Address, DoB)

Viewing group 1 of 8 Additional groups are available for loading

Review merged output

Review Merged Output

Filter Groups

Look For Search In Find Use OR logic

Match rule name = DOB exact, Name, Address (8) Clear Case Sensitive

Exact Match

Records Highlight Differences

	Internal ID	Match Group	Data Stream	Title	Given Names	Family Name	Full Name	DOB	YOB	Address1	Postcode	WholeAddress	Email
R1	82	1	US Customers	Mrs	Christine	Hunt	Mrs Christine Hunt	06-May-1985 00:00:00		11197 Leadbetter Road	23005	11197 Leadbetter Road ASHLAND VA 23005	Christine.G.Hunt@excite.co
R2	5417	1	US Customers	Mrs	Chris	Hunt	Mrs Chris Hunt	06-May-1985 00:00:00		11197 Leadbetter Rd	23005	11197 Leadbetter Rd ASHLAND VA 23005	Christine.G.Hunt@excite.co

Relationships Auto select relationships based on search

Record 1	Group 1	Record 2	Group 2	Match Rule	Review Status	Decision	User	Date			
R1	1	R2	1	* DOB exact, Name, Address	No Review Required	Match					

Project

local

Name File

Tasks

issues

DoB

records

Rules

11:29

What Is EDQ?

- Let's try the 'brute force' approach to **de-duplicating 10 million records**:
- Start at record 1 and compare it with:
 - Record 2, Record 3, ... , Record 10 million
- Now move onto record 2 and compare it with:
 - Record 3, Record 4, ... , Record 10 million

- This could take some time...
- The number of comparisons is about
 - Half of 10 million x 10 million, which is
 - 50,000,000,000,000!
- If a **server can do 100,000 per second** it will take
 - 500,000,000 seconds OR
 - 138889 hours OR
 - **15.85 years**
- Which is rather too long to wait!
- So we need to work a bit smarter...



Dashboard

Welcome, you are logged in as dnadmin.

[Administration](#) | [Customise](#) | [Logout](#) | [Help](#)

My Dashboard

Summaries

Status	Name	Rules	Red	Amber	Green
	Customer Name	4	0	1	3
	Marketability	4	1	0	3
	Mailing Address	6	5	1	0
	Demographic Data	5	0	0	5
	Compass Data Quality/Customer Name	6	3	2	1

8.1.6 (630)

[My Dashboard](#) > [Compass Data Quality/Customer Name](#)

Rules						
Status	Name	Checks	Passes	Issues	Pass Rate	
	Last Name Check	100	48	 48 (48%) 4 (4%)	48%	
	Title Check	100	81	 17 (17%) 2 (2%)	81%	
	Salutation Check	100	85	 10 (10%) 5 (5%)	85%	
	Multiple Names Check	100	91	 9 (9%) 0	91%	
	Given Names Check	100	93	 7 (7%) 0	93%	
	Business Name Check	100	94	 4 (4%) 2 (2%)	94%	

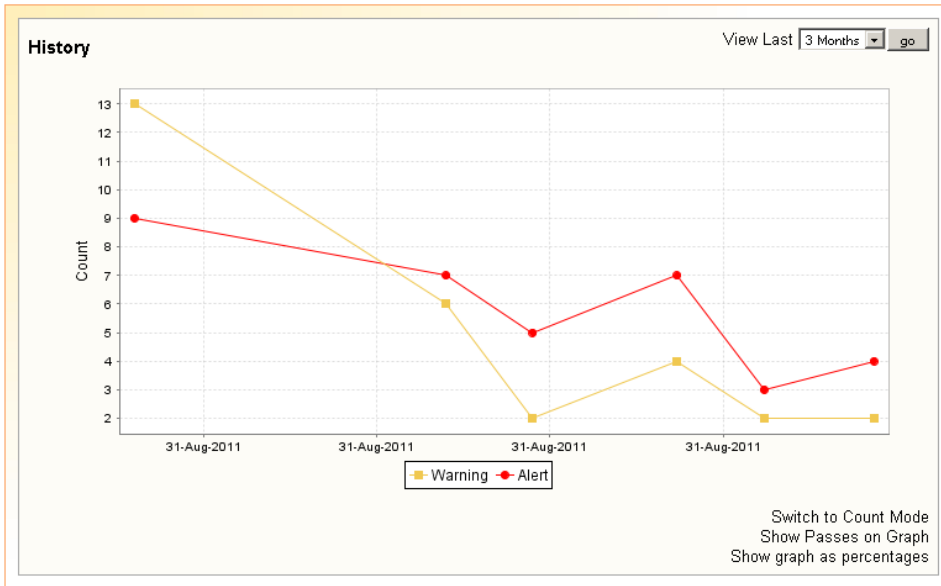
8.7.6(630)

Dashboard

Welcome, you are logged in as dnadmin.

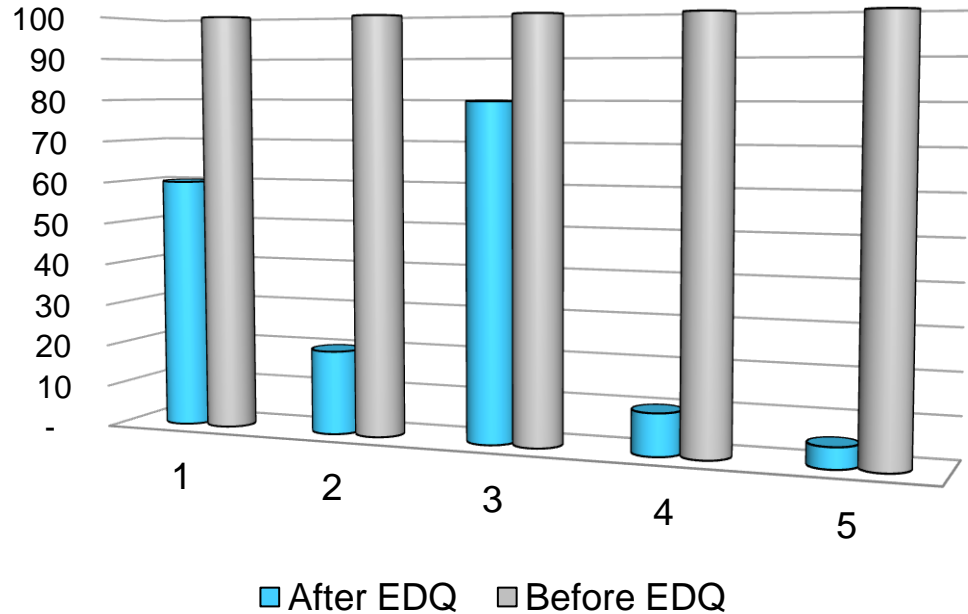
[Administration](#) | [Logout](#) | [Help](#)

[My Dashboard](#) > [Compass Data Quality/Customer Name](#) > [Business Name Check](#)



EDQ Economic Impact – Reduce Negatives

- 1-Reduce project risk **by 40%**
- 2-Avoid data remediation costs (manual effort, custom code) **by 80%**
- 3-Avoid error costs (incorrect orders, inventory etc.) **by 20%**
- 4-Decrease costs of unnecessary system changes due to data quality problem **by 90%**
- 5- Cut costs of handling duplicate data **by 95%**



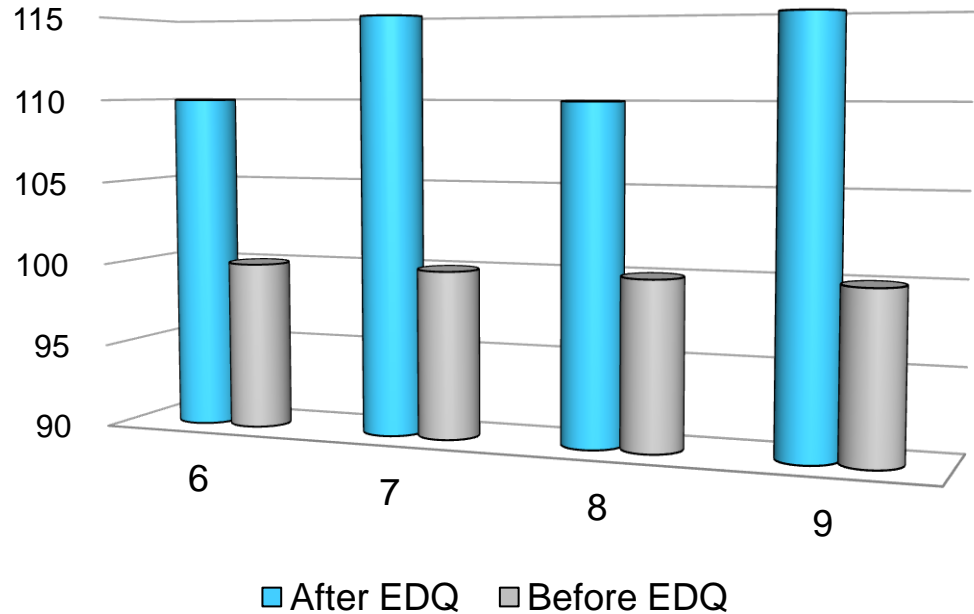
EDQ Economic Impact – Improve Positives

6-Speed up other systems, integration & processes **by 10%** (outputs and inputs in communication are correct and accurate data)

7-Increase BI, DW, CRM, Apps ROI **by 15%**

8-Gain **by 10%** more revenue due to agility to quickly reacting on ongoing market opportunities

9- Increase internal productivity and efficiency due to correct data **by 15%**



EDQ Economic Impact – Improve Positives

10- Increase revenue **by 10%** from gaining new customers and customer satisfaction Increase scalability

11- Improve speed of implementation and go-live with DQ solution by **50%** (as built-in specialized domain knowledge & expertise for rapid deployment)

