



ORACLE®

Beyond ETL: Next Generation Data Integration Architectures with Oracle Data Integrator Enterprise Edition

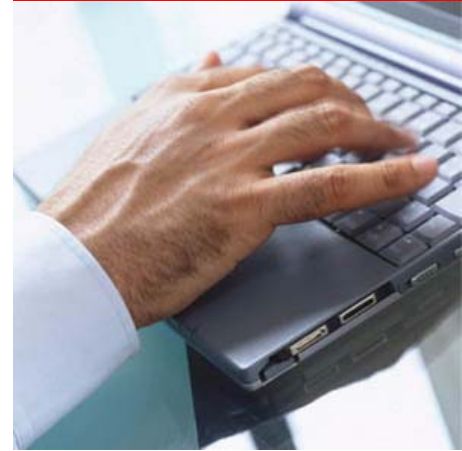
Ugo Pollio

Principal Sales Consultant

Oracle EMEA Data Integration Solutions Team

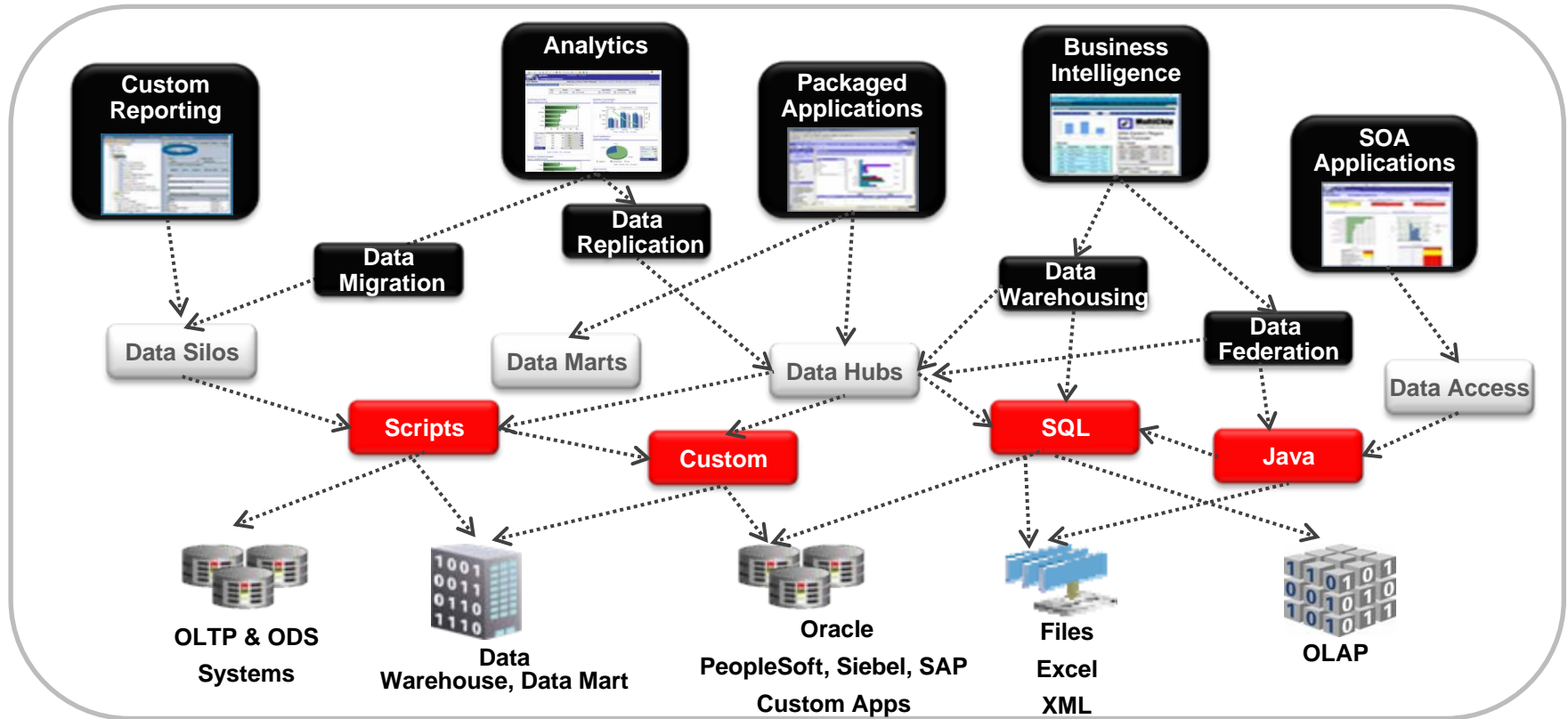
Data Integration: **Agenda**

- Obstacles to Unifying Enterprise Information
- Oracle's Data Integration Strategy
- What's new in ODI-EE 11g?
- ODI-EE Architectural Advantages
- Why Data Quality
- Oracle DQ offerings in DI
- DQ Lifecycle management
 - Investigate
 - Improve
 - Govern
- Oracle Data Integrator in Action
- How to get started



Unifying Information in a Mixed Landscape

Maximize the Value from BI/DW, SOA, and Business Apps



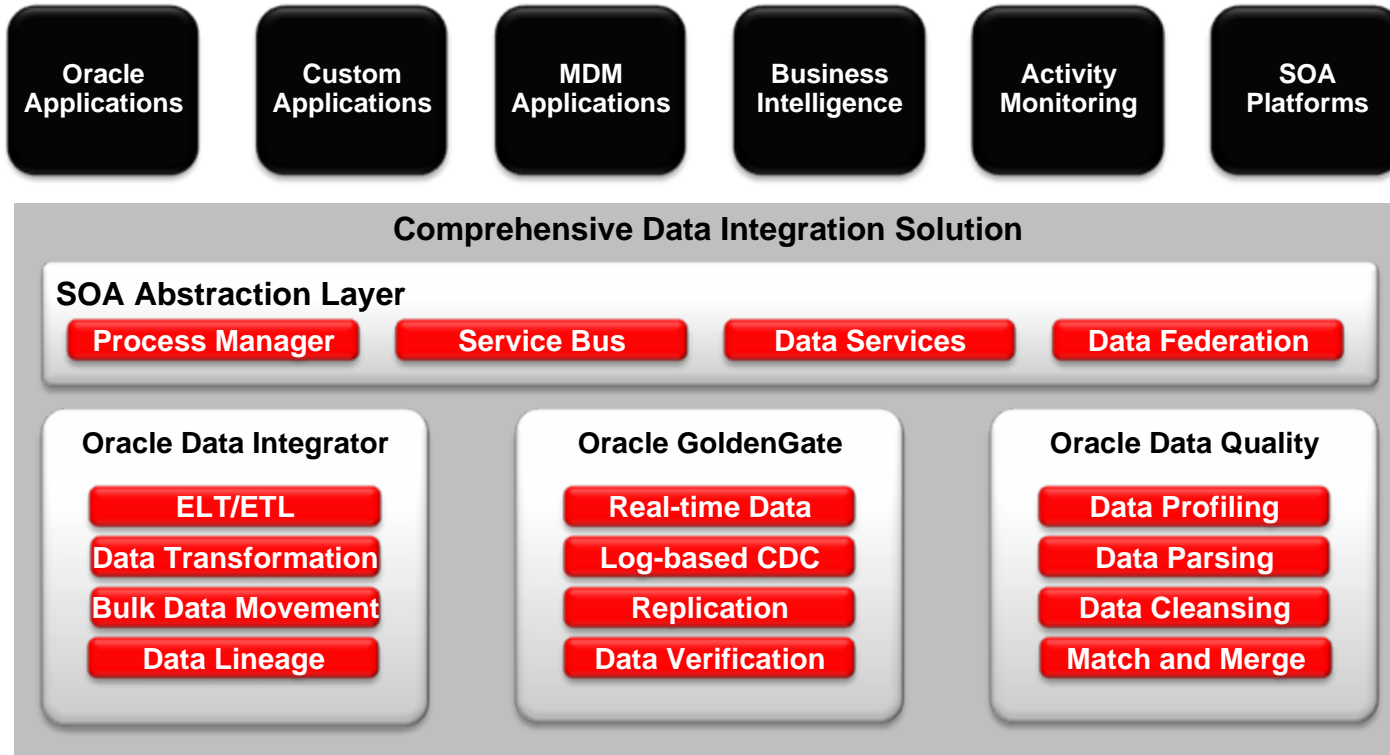
High Cost of Custom Coding

Lack of clean, consistent data

Multiple standards, disciplines

Comprehensive Approach to Data Integration

Combining Operational and Analytical Data Integration



Storage



Data Warehouse/
Data Mart



OLTP
System



OLAP Cube



Flat Files



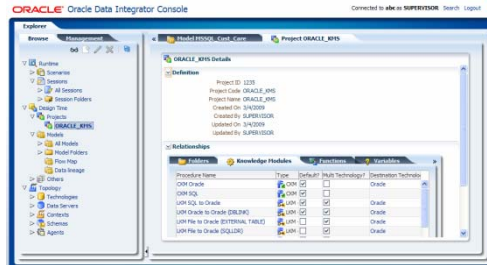
Web 2.0



Web and Event
Services, SOA

Oracle Data Integrator 11g (ODI-EE)

Fastest E-LT, Lowers TCO, Speeds Time-to-Value



Legacy Sources



Application Sources



OLTP DB Sources



Oracle Data Integrator Enterprise Edition

Best in class performance, scalability

Lowered TCO, lightweight deployment

Improved usability and manageability

Heterogeneous, Open, Standards-based

Log-based CDC with Oracle GoldenGate

Any Data Warehouse

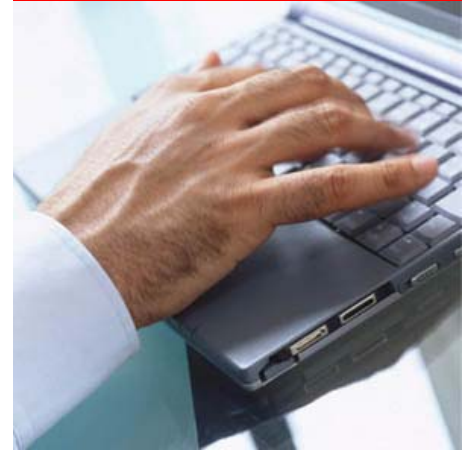


Any Planning System



Data Integration: **Agenda**

- Obstacles to Unifying Enterprise Information
- Oracle's Data Integration Strategy
- **What's new in ODI-EE 11g?**
- ODI-EE Architectural Advantages
- Why Data Quality
- Oracle DQ offerings in DI
- DQ Lifecycle management
 - Investigate
 - Improve
 - Govern
- Oracle Data Integrator in Action
- How to get started



What's New in ODI-EE 11g

Comprehensive Solution Goes Beyond Traditional ETL

Enterprise Deployment

- Integrated to WebLogic, Application Grid deployments, and Coherence, OPSS LDAP
- Provides Improved HA, Scalability, Security

Enhanced Productivity

- IDE based on JDeveloper, new interface editor and improved graphical design environment

Flexible Connectivity

- Supports SAP integrations, OLAP, Oracle Apps
- Extended API for custom integrations

Improved Manageability

- Integrated to Oracle Enterprise Manager
- Enhanced session control and error management

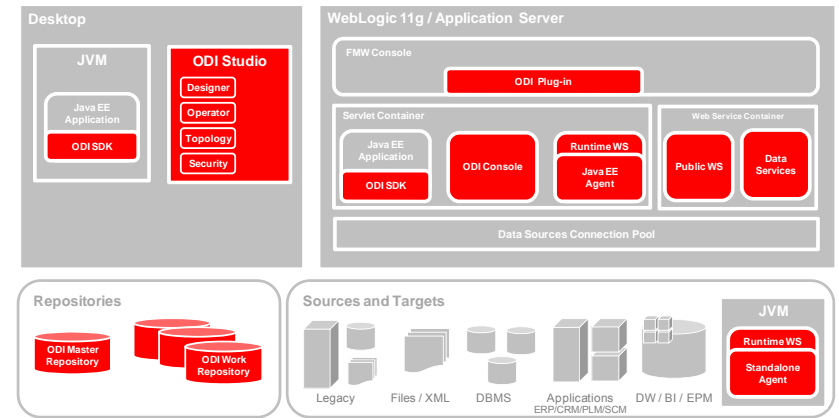
Real-Time CDC Integration

- Integrated to Oracle GoldenGate's CDC for real-time capture and delivery

Enterprise Deployment

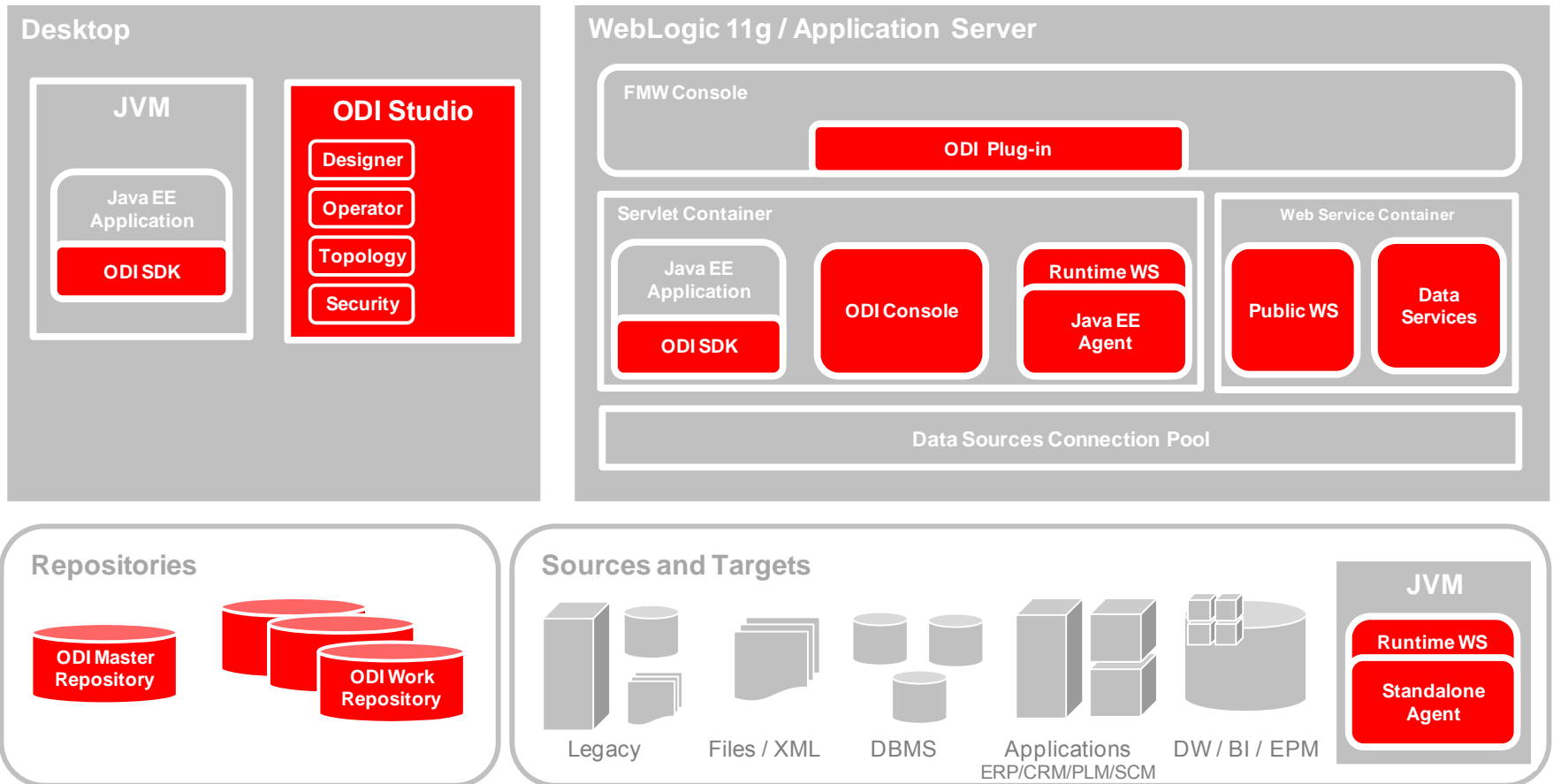
Provides Lowered TCO and Improved Efficiencies

- Run-time components deploy on Oracle WebLogic Suite, Coherence
- Options for High Availability deployments as standalone or JEE-based
- Connection pooling, load balancing
- Connection retry (Oracle RAC)
- Corporate Security Integration
- Lightweight E-LT agents for lowered TCO



Enterprise Deployment

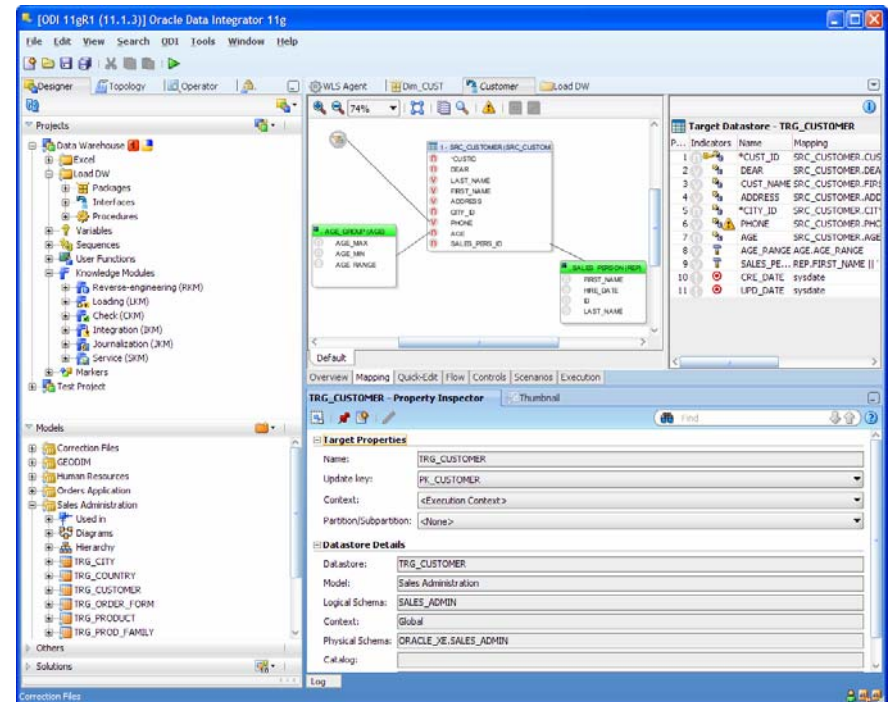
Provides Lowered TCO and Improved Efficiencies



Enhanced Productivity

Improves ease-of-use, speeds deployment

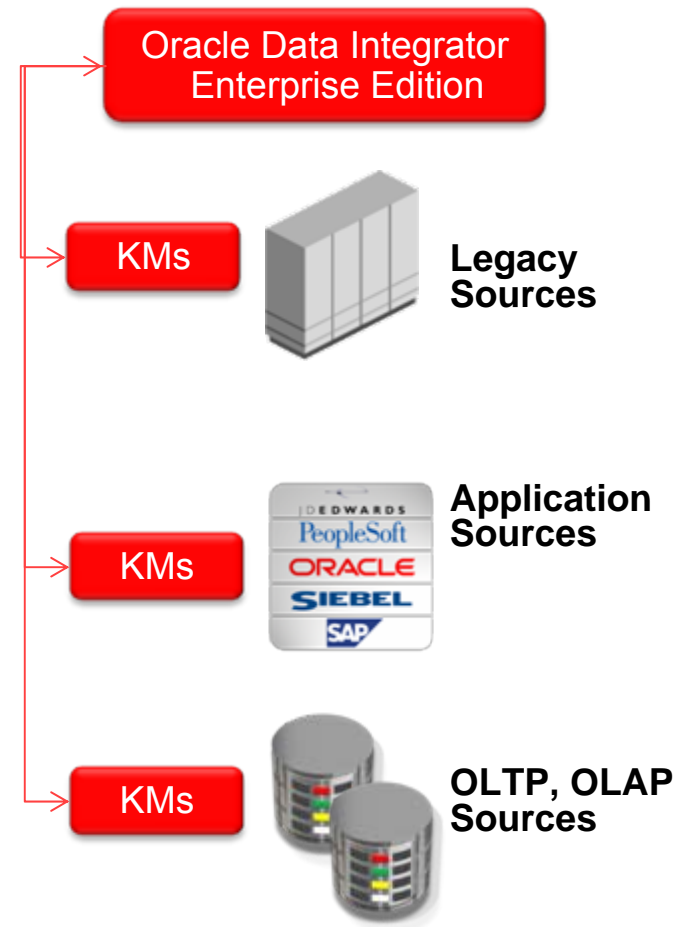
- New IDE based on JDeveloper
- Interface Editor leverages diagramming framework and includes Auto fixing
- Quick-Edit allows for faster interface design and mass updates
- Code Simulation



Flexible Connectivity

Improves ease-of-use, speeds deployment

- Application Adapters for Oracle E-Business Suite, PeopleSoft, Siebel, JD Edwards, SAP ERP and SAP BW.
- Oracle OLAP, Oracle CDC
- Hyperion Planning, Financial Management and Essbase
- Real-Time CDC with GoldenGate
- Optimizations for Teradata and multi-statements
- Oracle DB multi-table insert
- ADF-View Objects via OBI-EE



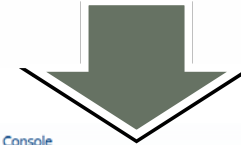
Improved Manageability

Improves ease-of-use, speeds deployment, lowers TCO

- Oracle Enterprise Manager for unified administration and management
- Enhanced error management – precise error messages, extended notifications
- Oracle Data Integrator Console manages run-time operations and browse run-time ,design-time artifacts, errors
- Enhanced Session Control - stop immediate stale sessions

The screenshot displays the Oracle Enterprise Manager Fusion Middleware Control 11g interface. The left pane shows a tree view of the topology, including the 'OracleDI Agent' under the 'edisonbase (AdminServer)' component. The main pane shows the 'OracleDI Agent' details, including 'General Info' (Agent Version: 11.1.1.3.0, Host and Port: dadmn0458:8001, Max Sessions: 1000) and 'Child Agents' (No Agents Found). Below this is a 'Sessions Notifications' table with columns for Status, Name, Origin, Work, Context, Message, and Updated.

Status	Name	Origin	Work	Context	Message	Updated
Success	BEEP	BEEP-001	WORKREP	GLOBAL		May 4, 2010 2
Success	BEEP	BEEP-001	WORKREP	GLOBAL		May 4, 2010 1
Warning	AUTO_DEBUG	AUTO_DEBUG...	WORKREP	GLOBAL	ODI-1217: Session AUTO_DEBUG...	Apr 27, 2010 1
Warning	Copy (2) of Copy of Co...	Copy (2) of Co...	WORKREP	GLOBAL	ODI-1217: Session Copy (2) of Co...	Apr 20, 2010
Warning	AUTO_DEBUG_NEW_JM	AUTO_DEBUG...	WORKREP	GLOBAL		Apr 20, 2010
Warning	AUTO_DEBUG	AUTO_DEBUG...	WORKREP	GLOBAL		Apr 20, 2010
Warning	KEYWORDTESTING	KEYWORDTEST...	WORKREP	GLOBAL		Mar 31, 2010



The screenshot displays the Oracle Data Integrator Console interface. The left pane shows a tree view of the project structure, including 'ORACLE_KHS'. The main pane shows the 'ORACLE_KHS Details' for a project with ID 1235, code ORACLE_KHS, and name ORACLE_KHS. Below this is a 'Relationships' table with columns for Procedure Name, Type, Default?, Multis Technology?, and Destination Technology.

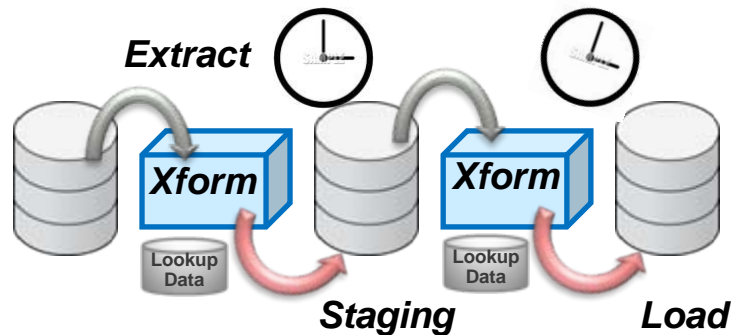
Procedure Name	Type	Default?	Multis Technology?	Destination Technology
ODM Oracle	ODM	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Oracle
ODM Sql	ODM	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Oracle
LOM SQL to Oracle (EBLDW)	LOM	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Oracle
LOM Oracle to Oracle (EXTERNAL TABLE)	LOM	<input type="checkbox"/>	<input checked="" type="checkbox"/>	Oracle
LOM File to Oracle (SQLDR)	LOM	<input type="checkbox"/>	<input checked="" type="checkbox"/>	Oracle

Real-time CDC Integration

Best-in-class solution for real-time with Oracle GoldenGate

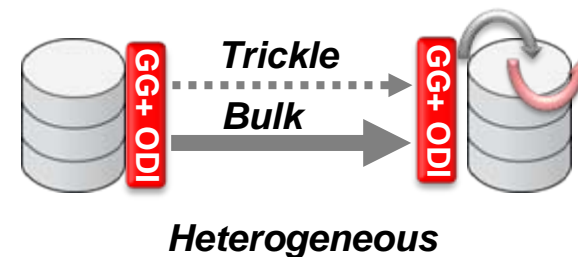
Traditional ETL + CDC

- Invasive Capture on OLTP systems using complex Adapters
- Transformations in ETL engine on expensive middle tier servers
- Bulk load to the data warehouse with large nightly/daily batch



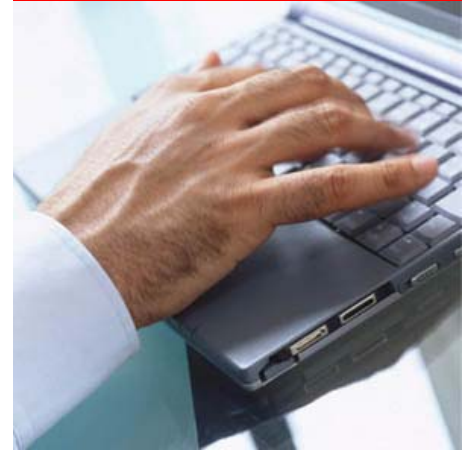
ODI + Oracle GoldenGate

- Continuous feeds from operational systems
- Non-invasive data capture
- Thin middle tier with transformations on the database platform (target)
- Mini-batches throughout the day or bulk processing nightly



Data Integration: **Agenda**

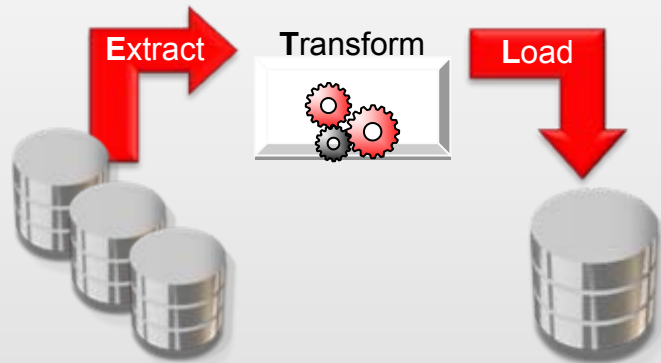
- Obstacles to Unifying Enterprise Information
- Oracle's Data Integration Strategy
- What's new in ODI-EE 11g?
- **ODI-EE Architectural Advantages**
- Why Data Quality
- Oracle DQ offerings in DI
- DQ Lifecycle management
 - Investigate
 - Improve
 - Govern
- Oracle Data Integrator in Action
- How to get started



Optimized Data Loading through E-LT

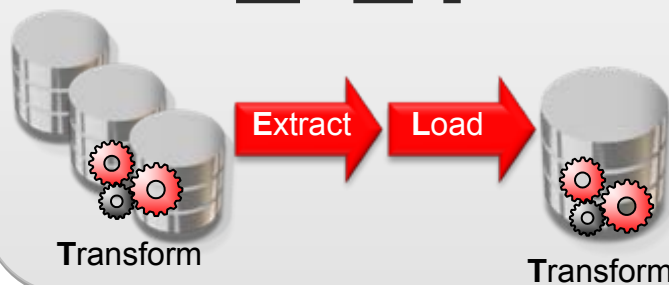
The key to improved performance and reduced costs

Conventional ETL Architecture



Next Generation Architecture

“E-LT”



- **E-LT** provides flexible architecture for optimized performance
- **Benefits:**
 - Leverage Set-based transformations
 - Improved performance for loading, no network hop
 - Takes advantage of existing hardware

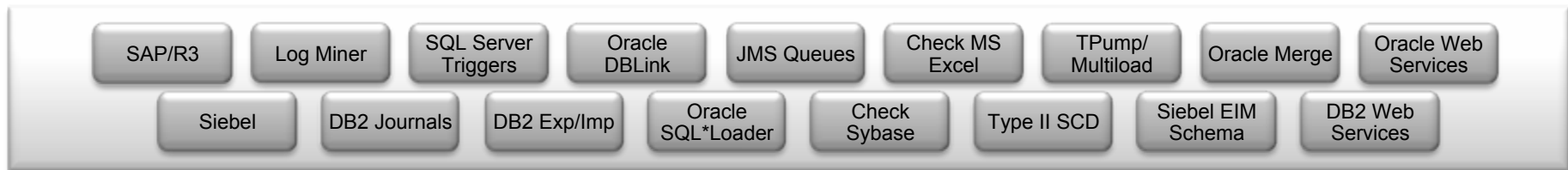
ODI-EE Knowledge Modules

Pluggable Architecture for Improved Flexibility

Pluggable Knowledge Modules Architecture



Sample out-of-the-box Knowledge Modules



- **Key Architecture Benefits:**
 - Ease of maintenance,
 - Flexibility, tailored to existing best practices,
 - Reduces cost of ownership

Leverage Pre-Packaged Integration Solutions

Direct Pre-Built Integrations"- Data Integration made simple

- ▶ Oracle CRM On Demand Integration to Siebel CRM
- ▶ Siebel Life Sciences Integration for Oracle Adverse Event Reporting System
- ▶ Demantra Integration Pack for Siebel CRM Consumer Goods
- ▶ Oracle Transportation Management (Glog) Integration to E-Business Suite

Cross Industry Process Integration Packs- Design, Sell, Plan and Execute

- ▶ Design to Release Agile Product Lifecycle Management integration Pack for Oracle E-Business Suite
- ▶ Opportunity to Quote Oracle CRM On Demand Integration Pack for Oracle E-Business Suite
- ▶ Order to Cash Siebel CRM Integration Pack for Oracle EBS Order Management
- ▶ Trade Promotion Mgmt Siebel CRM Integration Pack for Trade Promotion Management

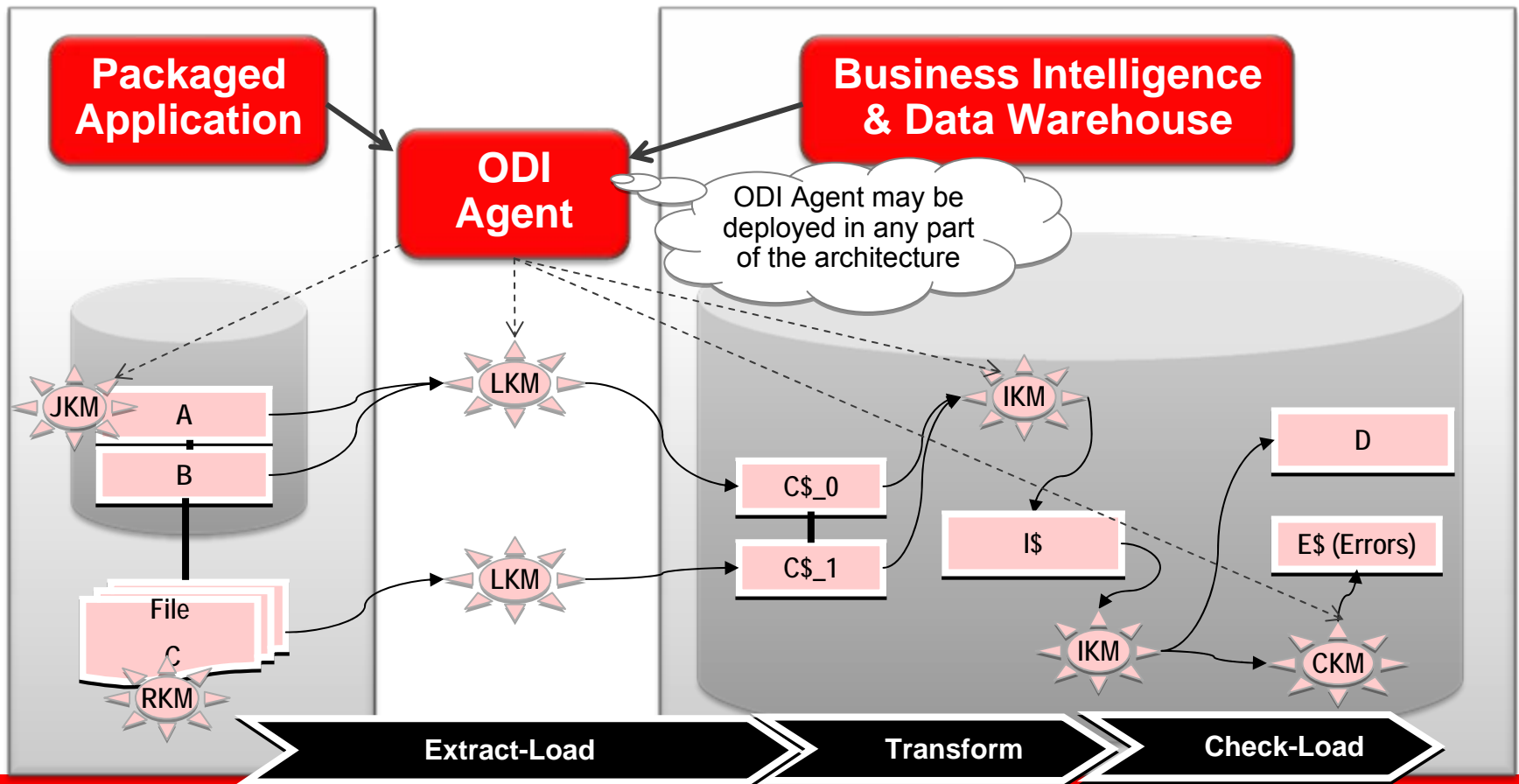
Industry Process Integration Packs- Transform key business processes

- ▶ **Comms** Order to Bill Siebel CRM Integration Pack for Oracle Comms Billing and Revenue Management: Order to Bill
- ▶ **Comms** Customer care Siebel CRM Integration Pack for Oracle Comms Billing & Revenue Management: Agent Assisted Billing Care
- ▶ **Comms** Revenue Acctng Oracle Comms Billing and Revenue Management Integration Pack for Oracle E-Business Suite: Revenue Accounting
- ▶ **Banking** Acct Originations Siebel CRM Integration Pack for i-flex FLEXCUBE Account Originations
- ▶ **Banking** Acct Originations Siebel CRM Integration Pack for Banking Account Originations

Light E-LT Architecture with ODI-EE

High Performance, Flexible, Lightweight Architecture

- Key Architecture Benefits: 100% Java, Open APIs, fast E-LT



Best Data Warehouse Machine

Fastest E-LT Processing



OLAP



ELT



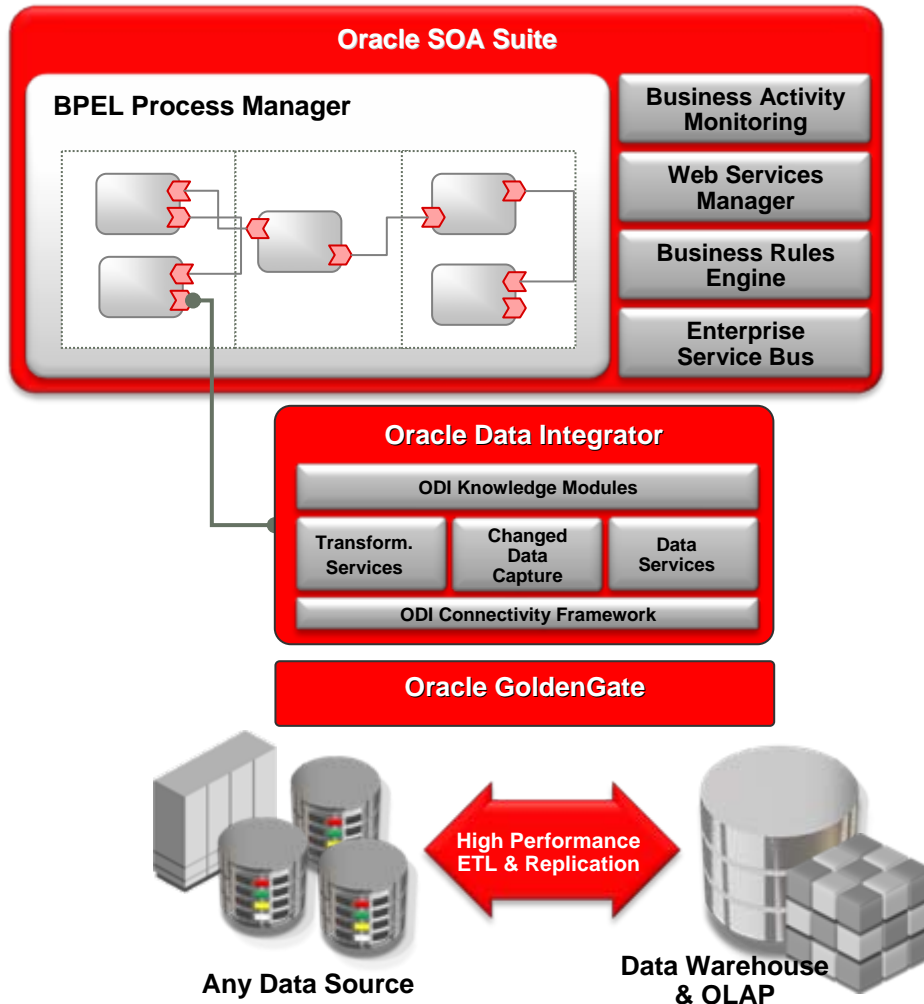
Data Mining



- Massively parallel high volume hardware to quickly process vast amounts of data
 - **Exadata runs data intensive processing directly in storage**
- Most complete analytic capabilities
 - *OLAP, Statistics, Spatial, Data Mining, Real-time transactional ETL, Efficient point queries*
- Powerful warehouse specific optimizations
 - *Flexible Partitioning, Bitmap Indexing, Join indexing, Materialized Views, Result Cache*
- **E-LT runs 20X faster only with Oracle**

Service-based Data Integration

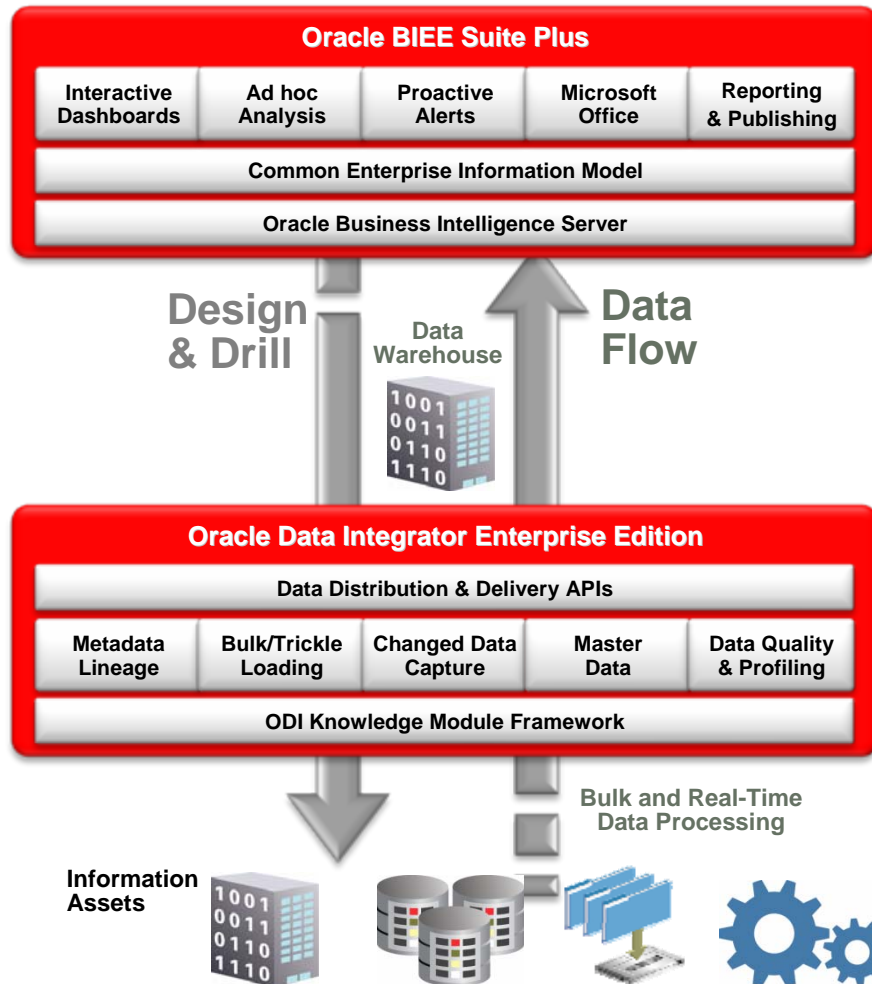
Best of Breed Data Integration as a SOA Service



- Decouple business apps from underlying data-centric applications
- Recombine existing services and app components instead of new development
- Orchestrate services to create custom view.
- **Benefit**
 - Enables efficiencies in transformation, data loading, data synchronization, all at improved performance

OBI EE Suite and ODI-EE

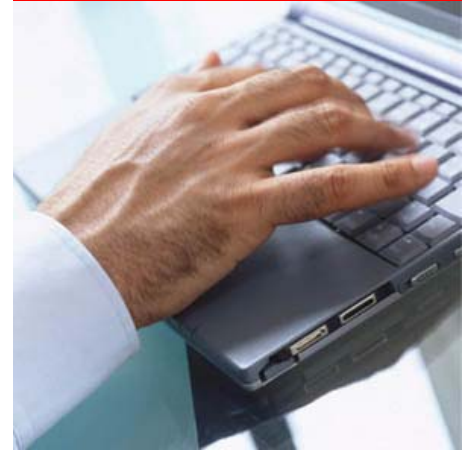
Data Lineage and Data Quality provides trusted data



- Support OLAP/R-OLAP sources and targets
- Report-to-source lineage
- Integrated Data Quality
- Common management: administration, monitoring, scheduling, auditing, exceptions
- Works with Oracle GoldenGate for improved real-time data access

Data Integration: **Agenda**

- Obstacles to Unifying Enterprise Information
- Oracle's Data Integration Strategy
- What's new in ODI-EE 11g?
- ODI-EE Architectural Advantages
- **Why Data Quality**
- Oracle DQ offerings in DI
- DQ Lifecycle management
 - Investigate
 - Improve
 - Govern
- Oracle Data Integrator in Action
- How to get started



Example of Data Quality Issues

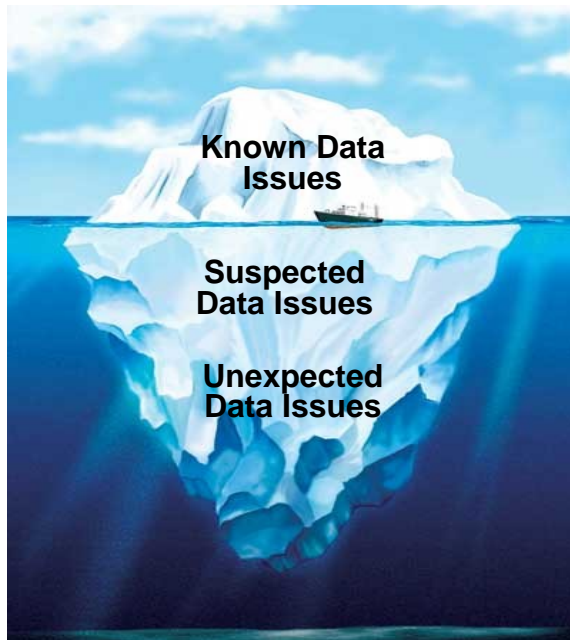
A Simple Customer Table Sample

Name	Address	City	State	Zip	Phone	Email
Bob Williams	36 Jones Avenue	Newton	MA	02106	617 555 000	bob.williams@yahoo.com
Robert Williams	36 Jones Av.	Newton	MA	02106	617555000	
Burkes, Mike and Ilda	38 Jones av.	Nweton	MA	02106	617-532-9550	mburkes@gmail.com
Jason Bourne, Bourne & Cie.	76 East 51 st	Newton	MA		617-536-5480	6175541329
...

Two Facts about Data Quality

Clear view on data

- The Data Quality Challenge is an iceberg
 - The biggest DQ threats are the ones we do not see.
 - ✓ **Data Profiling** lowers the water line and draws a clear view of the quality issues



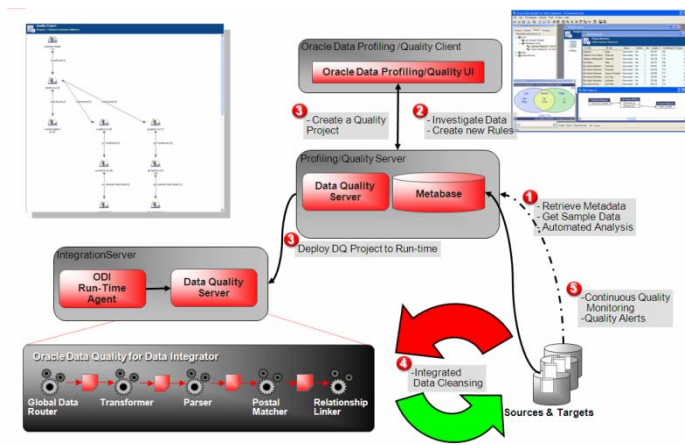
- Risk manageable
- Business rules tractable
- Expectations clear
- High business user involvement

- Risk unmanageable
- Business rules unknowable
- Missed expectations
- Little business user involvement

Two Facts about Data Quality

Data quality decays

- Data value decays
 - Data is an asset which value decays over time
 - Business events can make this worse
 - Quality is not a one shot process but a constant effort in the enterprise processes.
- ✓ **Data Quality** needs to be pervasive and **continuous**.



■ Pervasive

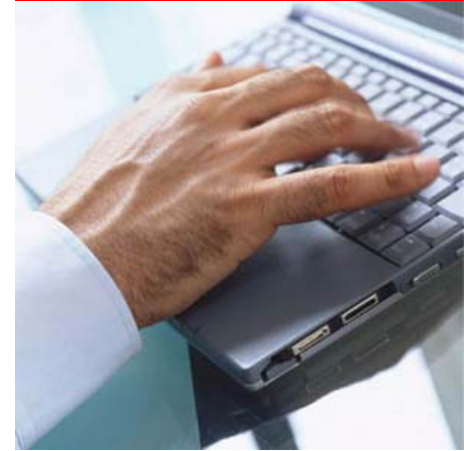
- Oracle Data Integrator Supplies Standard, Inline Data Quality and Data Profiling Capabilities with Every ETL and E-LT Job

■ Continuous

- Oracle Data Profiling and Oracle Data Quality for Data Integrator support Integrated Workflow, Recycling, and Steward GUIs

Data Integration: **Agenda**

- Obstacles to Unifying Enterprise Information
- Oracle's Data Integration Strategy
- What's new in ODI-EE 11g?
- ODI-EE Architectural Advantages
- Why Data Quality
- **Oracle DQ offerings in DI**
- DQ Lifecycle management
 - Investigate
 - Improve
 - Govern
- Oracle Data Integrator in Action
- How to get started



Oracle Data Quality Offering

- Joint Development with Trillium Software
 - Recognized Leader in Data Quality (1600+ customers)
 - More than 30 years experience
 - Full Breadth of Data Quality functionalities
 - Proven Scalability and Performance

Data Quality Process	Product
Advanced Data Profiling	Oracle Data Integrator + Oracle Data Profiling
Standard Profiling and Integrity	Oracle Data Integrator
Advanced Data Quality - Standardize, Match, Enrich	Oracle Data Integrator + Oracle Data Quality for ODI



Best of Breed Data Quality & Profiling

Introducing Two New Oracle Products

Data Governance [Auditing, Cleansing, Matching, Parsing and Standardization] for Enterprise Data

BENEFITS

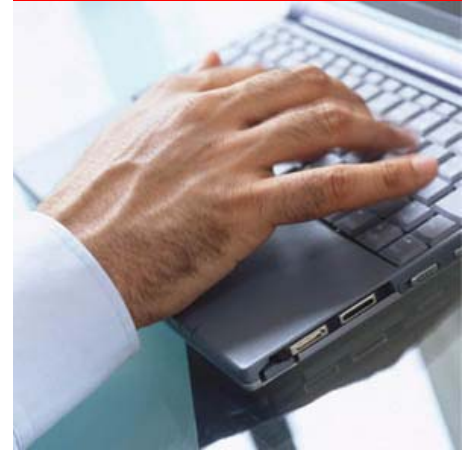
1. *Governance*
2. *Global Reach*
3. *Performance*
4. *Flexibility*
5. *Track Record*

KEY DIFFERENTIATED FEATURES

Core Cleansing & Matching
Internationalization
Single-Pass Architecture
Wide Platform Support
Proven Technology

Data Integration: **Agenda**

- Obstacles to Unifying Enterprise Information
- Oracle's Data Integration Strategy
- What's new in ODI-EE 11g?
- ODI-EE Architectural Advantages
- Why Data Quality
- Oracle DQ offerings in DI
- **DQ Lifecycle management**
 - **Investigate**
 - **Improve**
 - **Govern**
- Oracle Data Integrator in Action
- How to get started





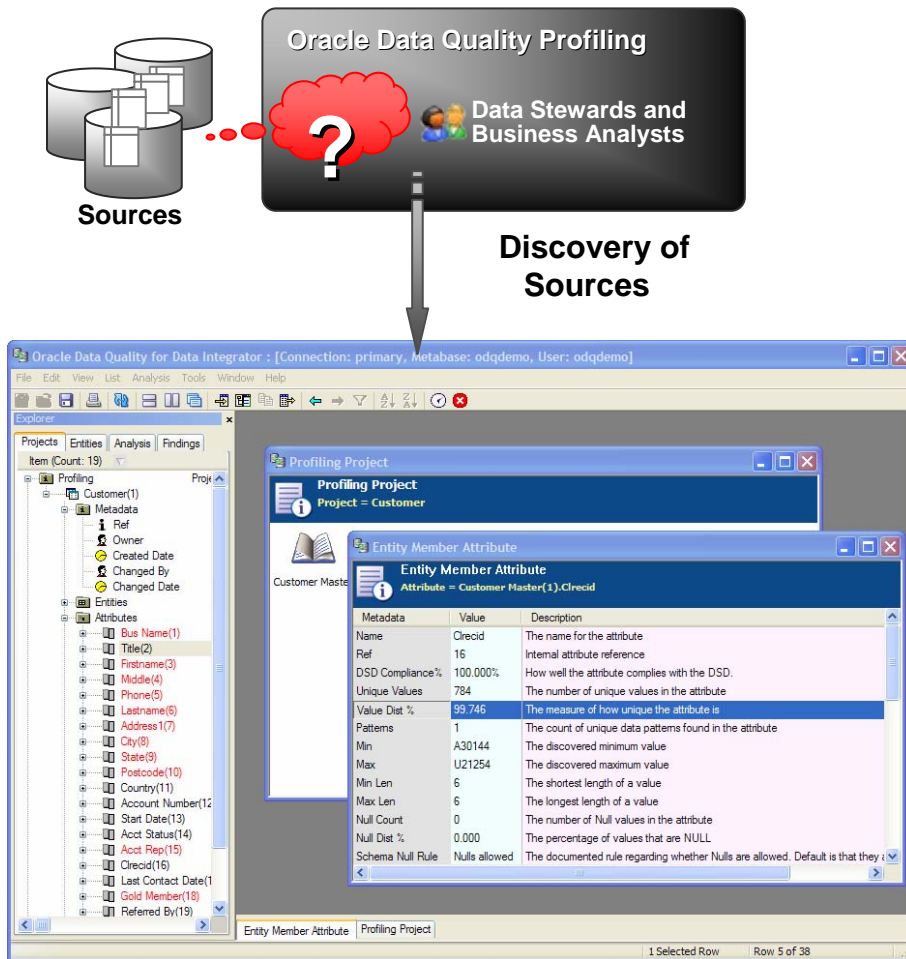
First, investigate...



Investigate

- Run automated profiling
- Find targeted conditions
- Standards compliance for:
 - Completeness
 - Values and frequencies
 - Patterns, shapes, formats
 - See entity relationships
 - Confirm referential integrity

Oracle Data Profiling Function



- Investigate the Data
 - Data Structure
 - Data Content
 - values, statistics, frequencies, ranges
 - Data Relationships
 - dependencies, keys, joins
- Assess Data Compliance
 - Create and validate data rules
- Report & Alerts
- Monitor Data Quality over time

Data Profiling and Analysis Metrics

Data Browser for Business Users to Identify Problems

Characteristic	Sample Profiling Metrics
Completeness	<ul style="list-style-type: none"> ▪ Null counts ▪ Minimum & maximum lengths
Conformity	<ul style="list-style-type: none"> ▪ Field structure ▪ Data types ▪ Patterns/ masks
Validity	<ul style="list-style-type: none"> ▪ Unique values ▪ Specific business rules
Consistency	<ul style="list-style-type: none"> ▪ Soundex ▪ Metaphones
Integrity	<ul style="list-style-type: none"> ▪ Dependencies ▪ Keys ▪ Joins

Unique Patterns				
Attribute = Customer Master(1).Phone				
Pattern	Value Length	Value Count	Frequency	Dist %
d3_d3pd4	12	184	273	34.733
d5_d6	12	102	190	24.173
d2pd2pd6	12	56	95	12.087
d3pd3pd4	12	50	75	9.542
d10	10	52	54	6.870
d3_d8	12	11	22	2.799
d4_d7	12	11	22	2.799
d5pd6	12	14	14	1.781
d4pd7	12	2	2	0.254

Unique Values					
Attribute = Customer Master(1).City					
Value	Frequency	Dist %	Length	Soundex	Metaphone
BRISTOL	42	5.344	7	B623	BRSTL
London	33	4.198	6	L535	LNTN
BRITSOL	18	2.290	7	B632	BRTSL
HAMBURG	15	1.908	7	H516	HMRK
Chicago	12	1.527	7	C220	XKK
CLEVEDON	11	1.399	8	C413	KLFTN
MUNICH	11	1.399	6	M520	MNX

Data Profiling and Analysis **Drill-Down**

Intelligent Drilling Guides Business Users in Context

1

Unique Patterns				
Attribute = Customer Master(1).Phone				
Pattern	Value Length	Value Count	Frequency	Dist %
d3_d3pd4	12	184	273	34.733
d5_d6	12	102	190	24.173
d2pd2pd6				12.087
d3pd3pd4				9.542
d10				6.870
d3_d8				2.799
d4_d7				2.799
d5pd6				1.781
d4pd7				0.254

Drill down to Matching Values

- Filter...
- Bookmark
- Convert view to Entity...
- Back
- Forward
- Change View
- Export
- Export to Server
- Copy



2

Unique Values					
Attribute = Customer Master(1).Phone					
Value	Frequency	Dist %	Length	Soundex	Metaphone
99-11-332231	3	0.382	12		
23-11-848484	2	0.254	12		
23-11-878787	2	0.254	12		
32-11-2					
32-11-5					
33-11-3					

Drill down to Matching Rows

- Recode Values...
- Export as Recode Tables...

2. Drill down to values list to see all rows containing a particular value



3

Data Rows [Filter by Selected Values in 'Phone']								
Entity = Customer Master(1)								
Row	Bus Name	Title	Firstname	Middle	Phone	Lastname	Address1	
346	KLEIN PHARMAZIE	DR	BOB	H	32-11-234234	KLEIN	FUNDLANDSTR 18	E:
347	KLEIN PHARMAZIE	DR	ROBERT		32-11-234234	KLEIN	FUNDLANDST 18	E:

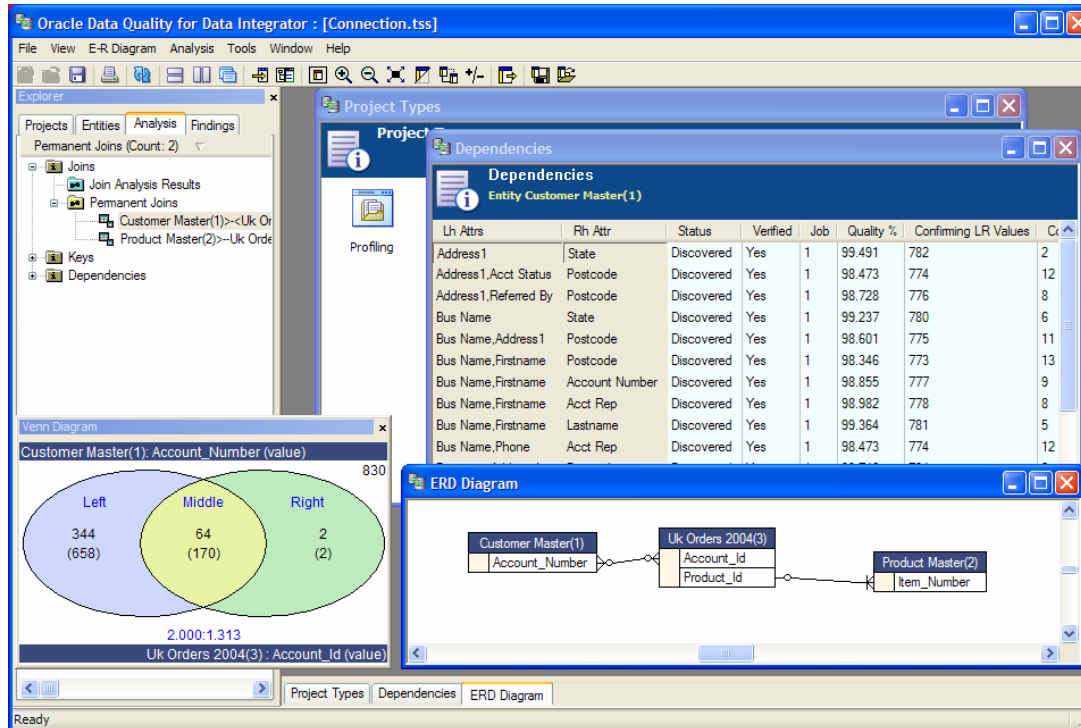
Drill example (using Patterns):

1. Drill down to patterns list to see all pattern values

3. Complete understanding and data examples allow you to quickly take action

Data Profiling and Analysis **Visualization**

See How your Data Really Is, Not Just by Schema Constraints



Beyond Profiling

- Discover joins, keys and dependencies
- Generate entity-relationship (ER) diagrams
- Detailed phrase analysis
- Check specific business rules
- Search and filter data and metadata
- Create Venn diagrams to understand outlier records

...then, **improve...**

Investigate

- Run automated profiling
- Find targeted conditions
- Standards compliance for:
 - Completeness
 - Values and frequencies
 - Patterns, shapes, formats
 - See entity relationships
 - Confirm referential integrity

Improve Standardize Link Enrich

- Automate improvement
- Cleanse and correct
- Apply/enforce standards
- Standardize unstructured and free form text data
- Identify related records
- Consolidate duplicates
- Apply survivorship logic
- Append content

Country-Specific Standardization

How do I make sense of data provided?

Name1	Flugtaggen GMBH
Name2	rhamer strasse 20
Address	dus
City/Town	40489
Post Code	
Country	



Business Name	Flugtaggen GMBH
Personal Name	
Street Name	rhamer
Street Type	strasse
Street Number	20
City/Town	dus
Post Code	40489
Country	DE

Value Add to Customers...

- Manage global transactional data
- Apply country intelligence (names geographic, etc.)
- Standardize critical data elements
- Context-sensitive data interpretation

Address Validation

How do I ensure data accuracy?

Data from Standardization

Business Name	Flugtaggen GMBH
Personal Name	
Street Name	rhamer
Street Type	strasse
Street Number	20
City/Town	dus
Post Code	40489
Country	DE



Data - Post Address Validation

Business Name	Flugtaggen GMBH
Personal Name	
Street Name	Rhamer
Street Type	Str.
Street Number	20
City/Town	Düsseldorf
Post Code	40489
Country	DE

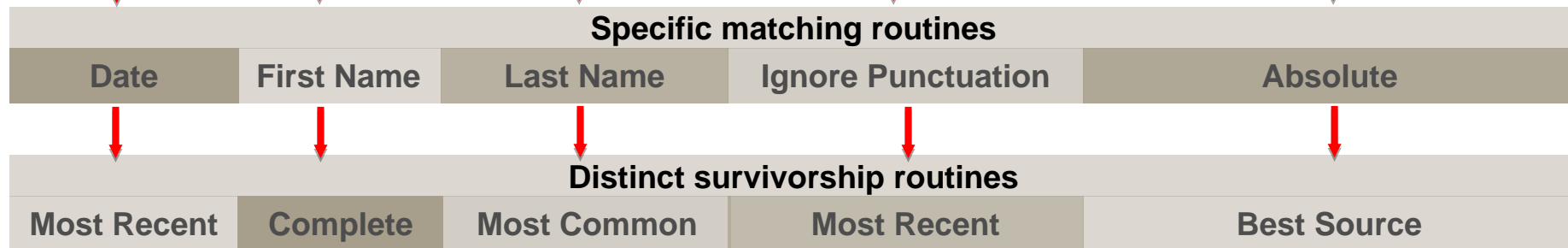
Value Add to Customers...

- Ensure data accuracy
- Enrich data (geocoding)
- Increased accuracy = better business processes and better matching

Matching

Intelligently identify links and relationships, consolidate data with precision

Date	First	Last	Phone	Email	Fail Code
08/02/00	Art	Barrios		bigwheels@hotmail.com	0
12/02/2003	A.	Barros	908-845-1234	abarrios@accen.com	1
	Arthur	Barrios	(902)-845-4417	abarrios@accen.com	8



12/2/2003	Arthur	Barrios	908-845-1234	abarrios@accen.com
-----------	--------	---------	--------------	--------------------

Ultimate flexibility for creating cleansed, standardized, consolidated views from multiple sources

Data Quality Context-Aware Cleansing

Key Words and Data are Standardized Contextually

Original Record

Mr. Bob James St. John III
St. John Chemicil Corp.
 101 S. Main St.
St. Louis, MO 63118
 www.b_stjohn@stjohnchem.com
 Code - Spcl. Deliv. - Hazd MatrIs, Flam

Title: Mr.
 First Name: Bob (Robert)
 Middle Name(s): James
 Last Name: St. John
 Generation: III
 Gender: M
 Business Name: St. John Chemical Corporation
 House Number: 101
 Directional: S (South)
 Street Name: Main
 Street Type: St.
 City: St. Louis (Saint Louis)
 State: MO
 Zip Code: 63118
 Email: www.b_stjohn@stjohnchem.com
 Delivery Code: Z99 - Special Handling
 Delivery Comments: Hazardous Materials, Flammable



1. Key word identification
2. Pattern recognition
3. Data Standardization
4. Postal Matching



ORIGINAL DATA						APPENDED DATA									
Name	Bus. Name	St. Add.	City	State	ZIP	Pr. FN	PR MN	PR LN	PR BN	PR HsNo	PR St N	PR St T	PR City	PR State	PR Zip

Data Quality Domain Agnostic Cleansing

Context Aware Cleansing Operates Regardless of Data Domain

P/N	DESCRIPTION
1774-5674	TUBE, CENTRIFUGE POLY S 15ML (CS/500)CONICAL-BOTTOM
1774-5675	TUBE, CENTRIFUGE PPL 15ML (CS/500)CONICAL-BOTTOM
1774-4532	TUBE, CENTRIFUGE PPL 50ML (CS/500)CONICAL-BTTPCK 25/RACK
1774-4538	TUBE, CENTRIFUGE POLY S 50ML (CS/500)CONICAL-BTMPK 25/RACK
645-4556	PIPET, CLEAR SEROLOGICAL 2ML (CASE/500)
195-7934	NUT, LOCK RH,11"
3324-7955	VIAL, WHEATON 33* CLEAR 4ML (CS/144)



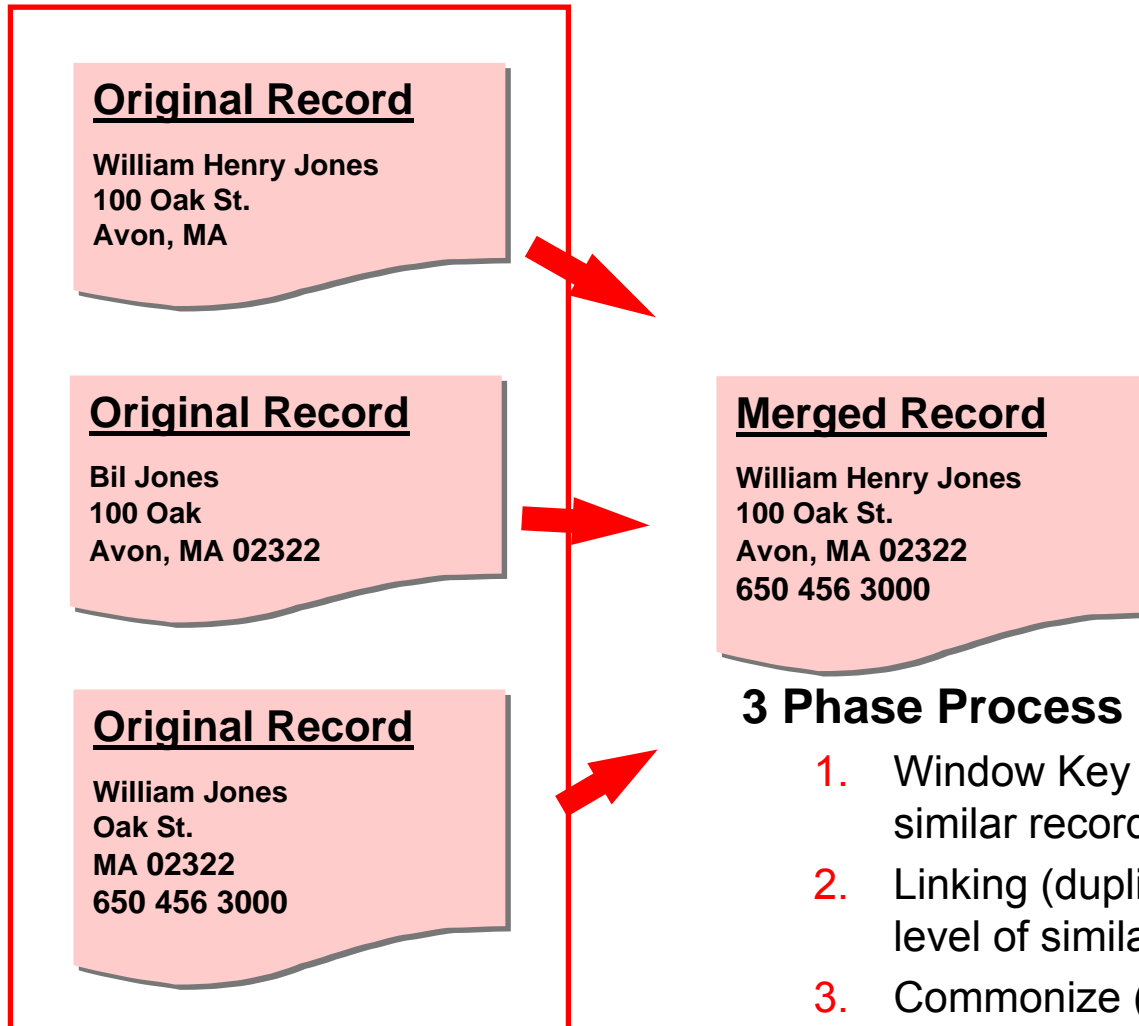
P/N	ITEM NAME	MATERIAL	SIZE	UOM	DESCRIPTOR	PACKAGE	PACK METHOD
1774-5674	CENTRIFUGE TUBE	POLYSTERENE	15	ML	CONICAL	CASE/500	BOTTOM PACKED
1774-5675	CENTRIFUGE TUBE	POLYPROPYLENE	15	ML	CONICAL	CASE/500	BOTTOM PACKED
1774-4532	CENTRIFUGE TUBE	POLYPROPYLENE	50	ML	CONICAL	CASE/500	BOTTOM PACKED 25/RACK
1774-4538	CENTRIFUGE TUBE	POLYSTERENE	50	ML	CONICAL	CASE/500	BOTTOM PACKED 25/RACK
0645-4556	SEROLOGICAL PIPET		2	ML	CLEAR	CASE/500	
0195-7934	LOCK NUT		11	IN	RIGHT HAND		
3324-7955	WHEATON VIAL		4	ML	CLEAR	CASE/144	



Increased accuracy = better business processes & better matching

Data Quality **Matching & Merging**

De-duplicate records

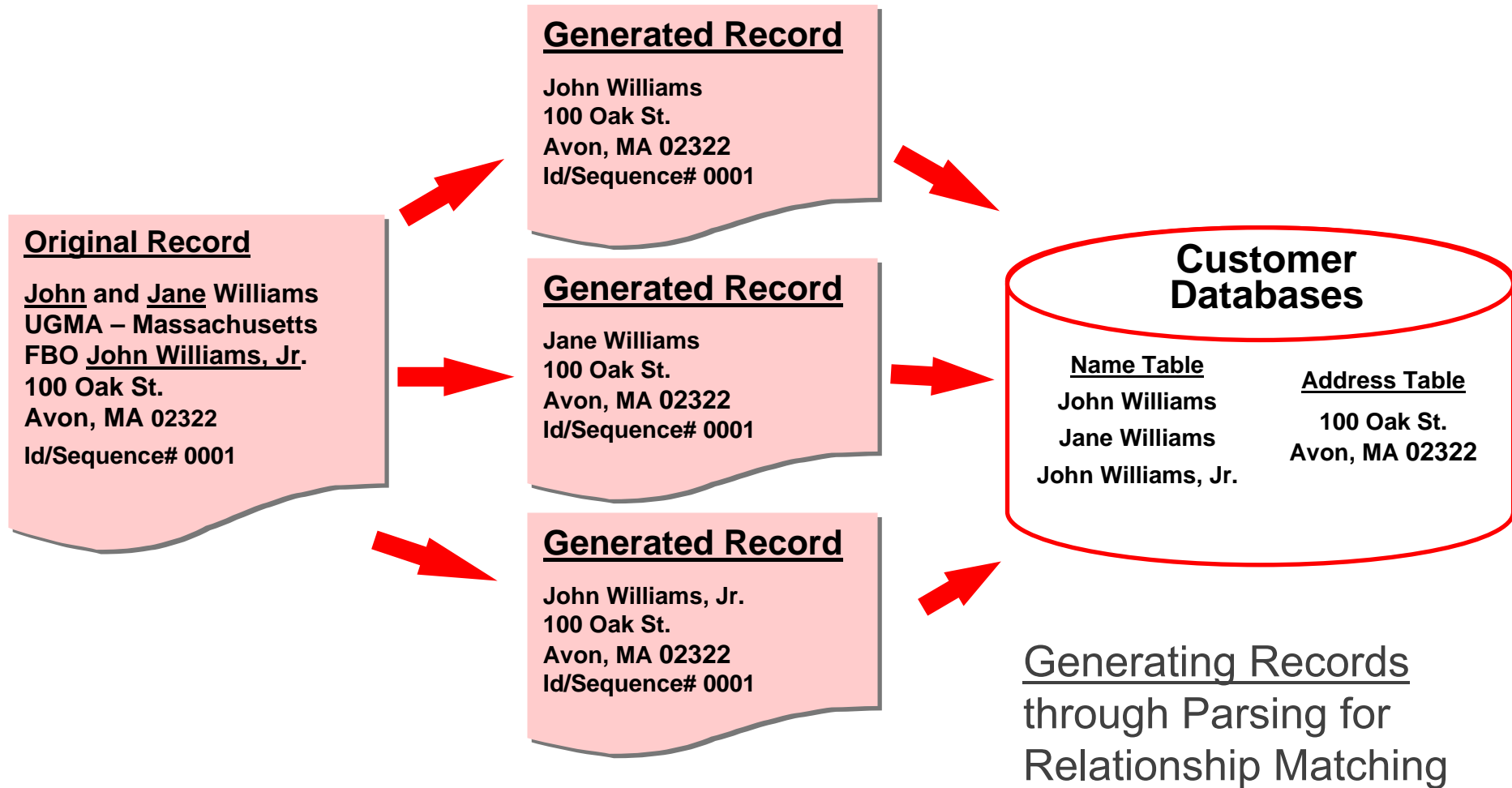


3 Phase Process

1. Window Key Generation (grouping similar records)
2. Linking (duplicates identification with level of similarity)
3. Commonize (create best record)

Data Quality Entity Identification

Parsing Operations Can Extract Multiple Entities Per-Record



...and, **govern.**

Investigate

- Run automated profiling
- Find targeted conditions
- Standards compliance for:
 - Completeness
 - Values and frequencies
 - Patterns, shapes, formats
 - See entity relationships
 - Confirm referential integrity

Improve Standardize Link Enrich

- Automate improvement
- Cleanse and correct
- Apply/enforce standards
- Standardize unstructured and free form text data
- Identify related records
- Consolidate duplicates
- Apply survivorship logic
- Append content

Govern

- Define data metrics
- Monitor results
- Quantify problem areas
- Quantify improvements
- Relate metrics to business impact
- Communicate results
- Prioritize next steps

Monitor Data Quality over time

- Time Series controls data quality over time
 - Set right timeframe, right business thresholds
 - Alert if DQ metrics and business rules are not met

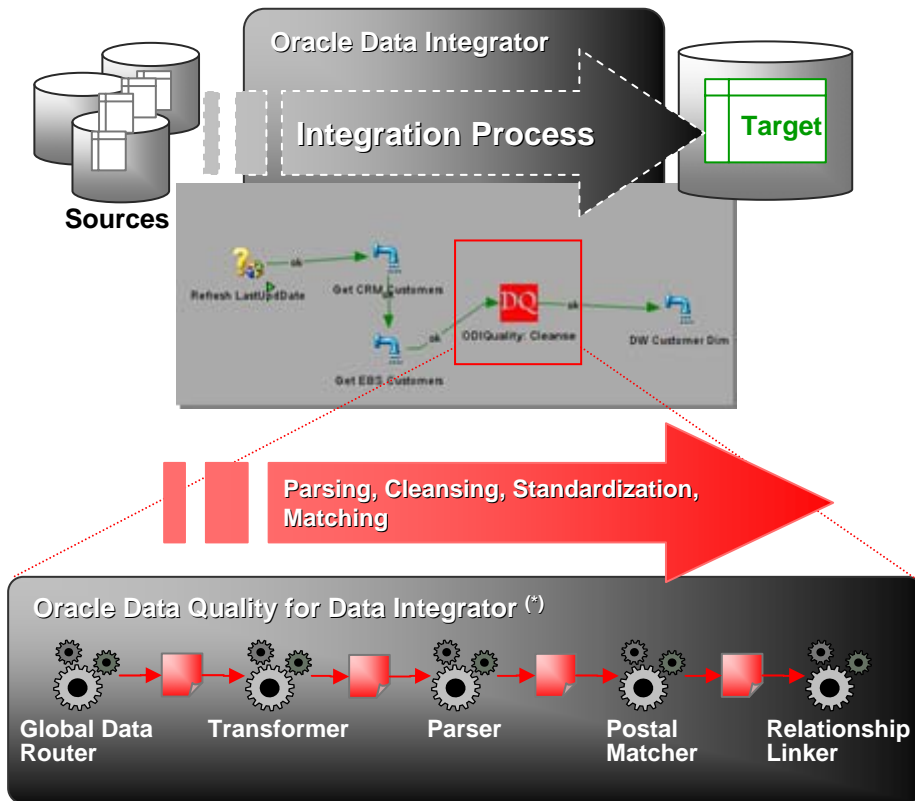
The image illustrates the configuration of a Time Series project for monitoring data quality. It consists of four main components:

- Edit Time Series Project Details:** Shows the project name 'Customer Master', description, and data source 'customer_master.bt'. The automation interval is set to 'Every 1 Days' starting from '29/09/2009'.
- Entity Business Rule:** A rule named 'Check order amount' is defined with the expression '[Total Orders] >= 100'. The description states: 'Check Total orders amount for each customer. Customer must have Total Orders greater than 100, at least 93% of all customers'. The threshold is set to 93%.
- Attribute History:** A tree view showing the history of attributes: Total Orders(1), Name(2), Phone(3), Address1(4), City(5), State(6), Postcode(7), and Country(8). The 'History' folder is expanded to show 'Values' and 'Strings'.
- Series Attributes Table:** A table showing the results of the data quality checks for the series 'All'.

Entity Id	Entity Loaded Date	DSD Compliance %	Unique Values	Value Dist %	Patterns	Min	Max	Min Len	Max Len	Null Count	Null Dist %	Space Count	Space Dist %	Strings	Strings Dis
1	2009/09/29 12:34:08	100.000%	4	0.509	2	CAN	USA	2	3	0	0.000	0	0	4	100.000
2	2009/09/29 12:41:54	100.000%	4	0.509	2	CAN	USA	2	3	0	0.000	0	0	4	100.000
3	2009/09/29 12:43:06	100.000%	4	0.509	2	CAN	USA	2	3	0	0.000	0	0	4	100.000
4	2009/09/29 12:58:15	100.000%	4	0.509	2	CAN	USA	2	3	0	0.000	0	0	4	100.000

Integrate DQ in ODI flows

Oracle Data Quality for Data Integrator

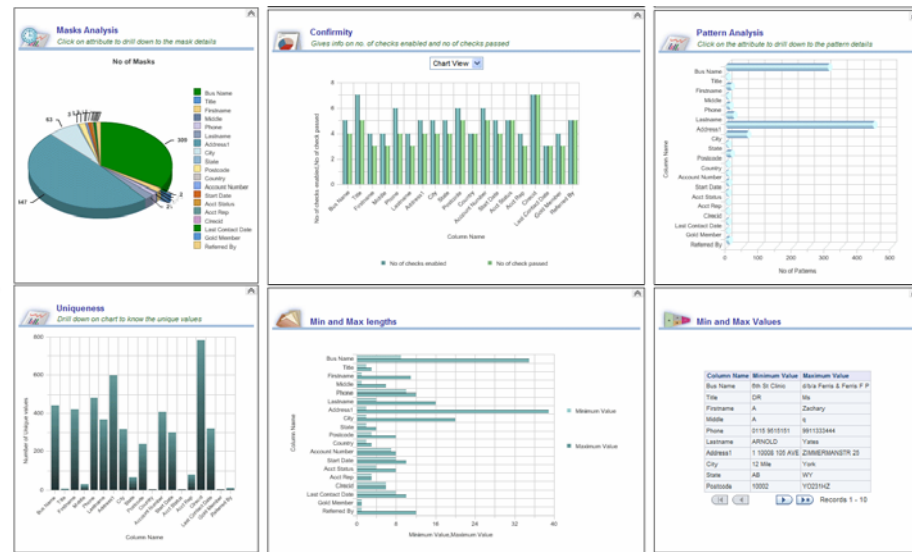
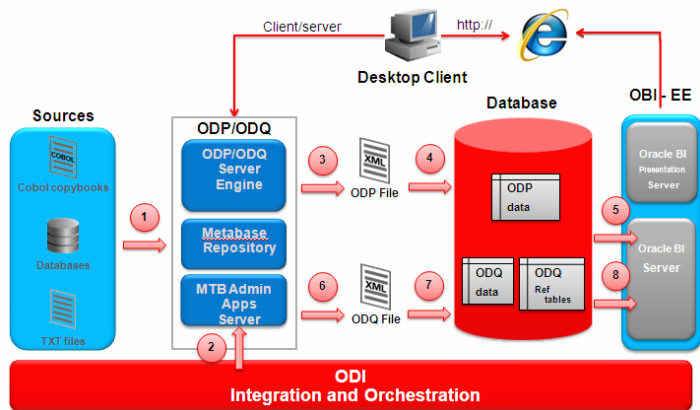


(*) Joint Development with Trillium

- Out-of-box integration
 - ODI integrates with Quality functions via pre-built Knowledge Modules
 - ODI Model metadata passed to ODQ at design time
 - ODQ source and target Metadata passed to ODI at Design Time
 - Out-of-box ODI Tool for runtime invocation of ODQ processes
- Packaged Quality Rules
 - Delivered Out-of-the-Box by Oracle
 - More than 40 Countries & Domains covered

Leverage DQ with Dashboards and Reports

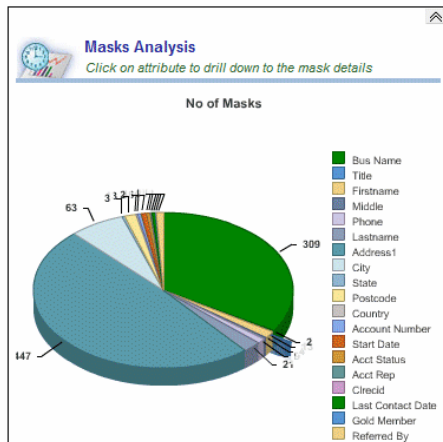
- ODPQ&ODP + OBI-EE: extreme quality control
 - Enable further analysis and report generation
 - Integrate BI features plus DQ results



DQ Dashboards on OBIEE

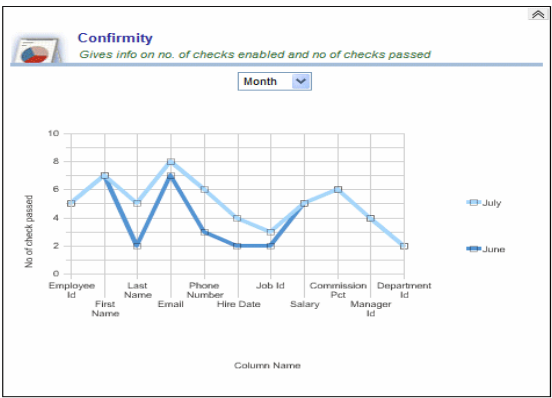
Attribute analysis, historical analysis, DQ stats

Data Profiling Metrics & KPIs, Data Quality Statistics at a glance



Correctness
Gives info on datatype mismatches

Attribute	Target Type	Dsddatatype	Match
Bus Name	string	string	Match
Title	string	string	Match
Firstname	string	string	Match
Middle	string	string	Match
Phone	string	integer	Mismatch
Lastname	string	string	Match
Address1	string	string	Match
City	string	string	Match
State	string	string	Match
Postcode	string	integer	Mismatch



Global Parsing

Description	Frequencycount
Records contain Directory Post Code type 0	1
Records contain Directory Post Code type 4	4
Records contain Directory Post Code type 5	0
Records contain Directory Post Code type 8	219
Records contain Directory Post Code type 9	0
Records contain Directory Post Code type OTHER	76

Review Groups

Code	Description	No of Records
000	No Targeted Conditions Found	218
001	Unidentified Item	0
002	Mixed Name Forms	0
003	Hold Mail	0
004	Foreign Address	0
005	No Names Identified	0
006	No Street Identified	5
007	No Geography Identified	2
008	Unknown Name Pattern	0
009	Derived Genders Conflict	0
010	More Than One Middle Name	5
011	Unknown Street Pattern	0
012	Invalid Directional	0
013	Unusual or Long Address	0
014	No City or County Identified	69
015	Geography Too Long	0
016	Corrected City Name Too Long	0
017	Conflicting Geography Types	0
018	Unable to Verify City Name	0
019	Unidentified Line	0
020	Multiple Street Line Types	1

Compliance
Click on attribute to drill down to the quality checks analysis

Column Name	Compliance	DSD Compliance
Bus Name	100%	80.000%
Title	100%	71.429%
Firstname	100%	75.000%
Middle	100%	75.000%
Phone	100%	66.667%
Lastname	100%	75.000%
Address1	100%	80.000%
City	100%	80.000%
State	100%	80.000%
Postcode	100%	83.333%

Data Quality Lifecycle Management

Investigate

- Run automated profiling
- Find targeted conditions
- Standards compliance for:
 - Completeness
 - Values and frequencies
 - Patterns, shapes, formats
 - See entity relationships
 - Confirm referential integrity

Improve Standardize Link Enrich

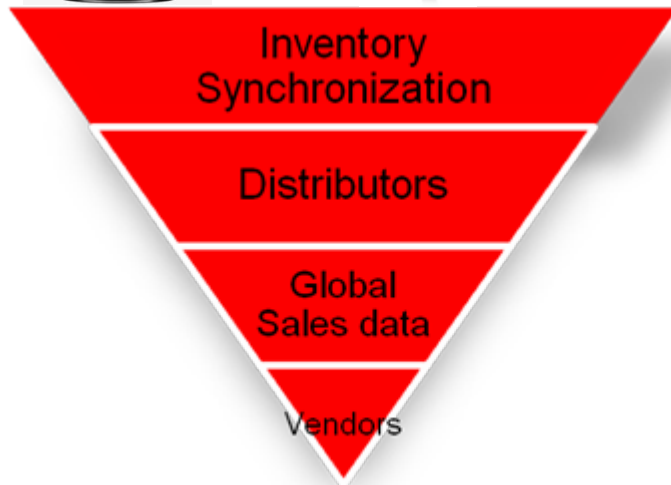
- Automate improvement
- Cleanse and correct
- Apply/enforce standards
- Standardize unstructured and free form text data
- Identify related records
- Consolidate duplicates
- Apply survivorship logic
- Append content

Govern

- Define data metrics
- Monitor results
- Quantify problem areas
- Quantify improvements
- Relate metrics to business impact
- Communicate results
- Prioritize next steps

Product data governance

product data comes in all shapes and sizes...



- **Use:** Quickly profile and standardize product data to:
 - Create consistency across multiple suppliers
 - Perform inventory consolidation
 - Identify variances across sites (plants, warehouses, suppliers)
 - Maintain "Item Masters" across multiple sites
- **Apply:** Leverage a definitive, complete product view for:
 - New system loads
 - Global data synchronization
 - Spend and inventory analysis
 - Regulatory compliance in areas like recall management, material maintenance, lot management and tracking

- **Profit:** Benefit from improved performance and process results such as:
 - Instant identification of item substitutes or alternatives (for sales, manufacturing, or warranties)
 - Rapid and thorough adherence to changes in regulations
 - Improved leverage of global supply contracts

Category	Product Vendor	Product Code	Product Family	Measure	Unit of measure	Technology
monitor	HP	HP-19LCD	Screen	19	inch	LCD
Tv	Samsung	SMSG-234UE23	Screen	23	Inch	LED
cable	voka	VK-15USB	Accessory	1.5	Meter	USB
cable	Belkin	BLKN-FWR121	Accessory	2	Meter	FIREWIRE
Hard disk	Seagate	SGT-2TbSATA	Storage	2048	Gb	S-ATA
DVD-ROM	Verbatim	VRBTM-RWDL	Storage	4096	Gb	DVD-RW

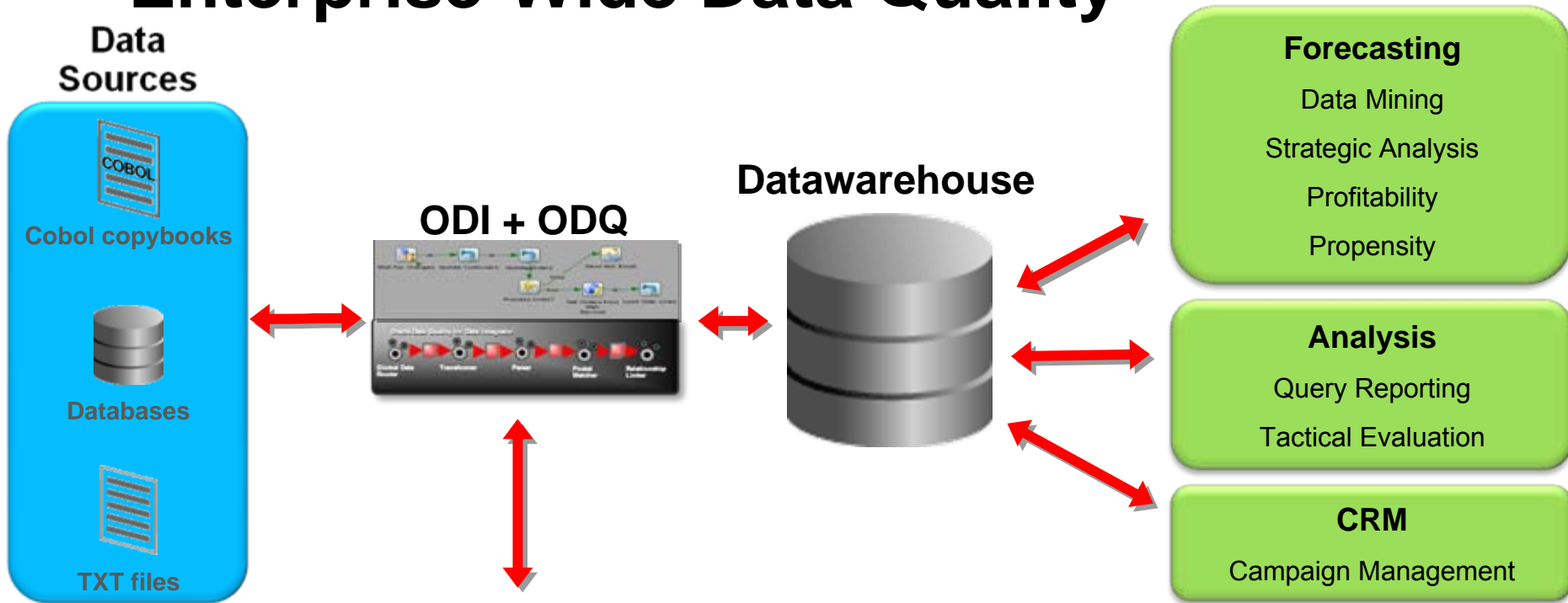
Data Quality Firewall

profile, repair, check, alert, report



- DWH, improving reliability and quality
- New ERP/CRM installation, and legacy data integration
- Master Data Management projects
- Data synchronization projects

A **Data Warehousing** approach to Enterprise-Wide Data Quality



Customer Touchpoints

Call Center

Direct Mail

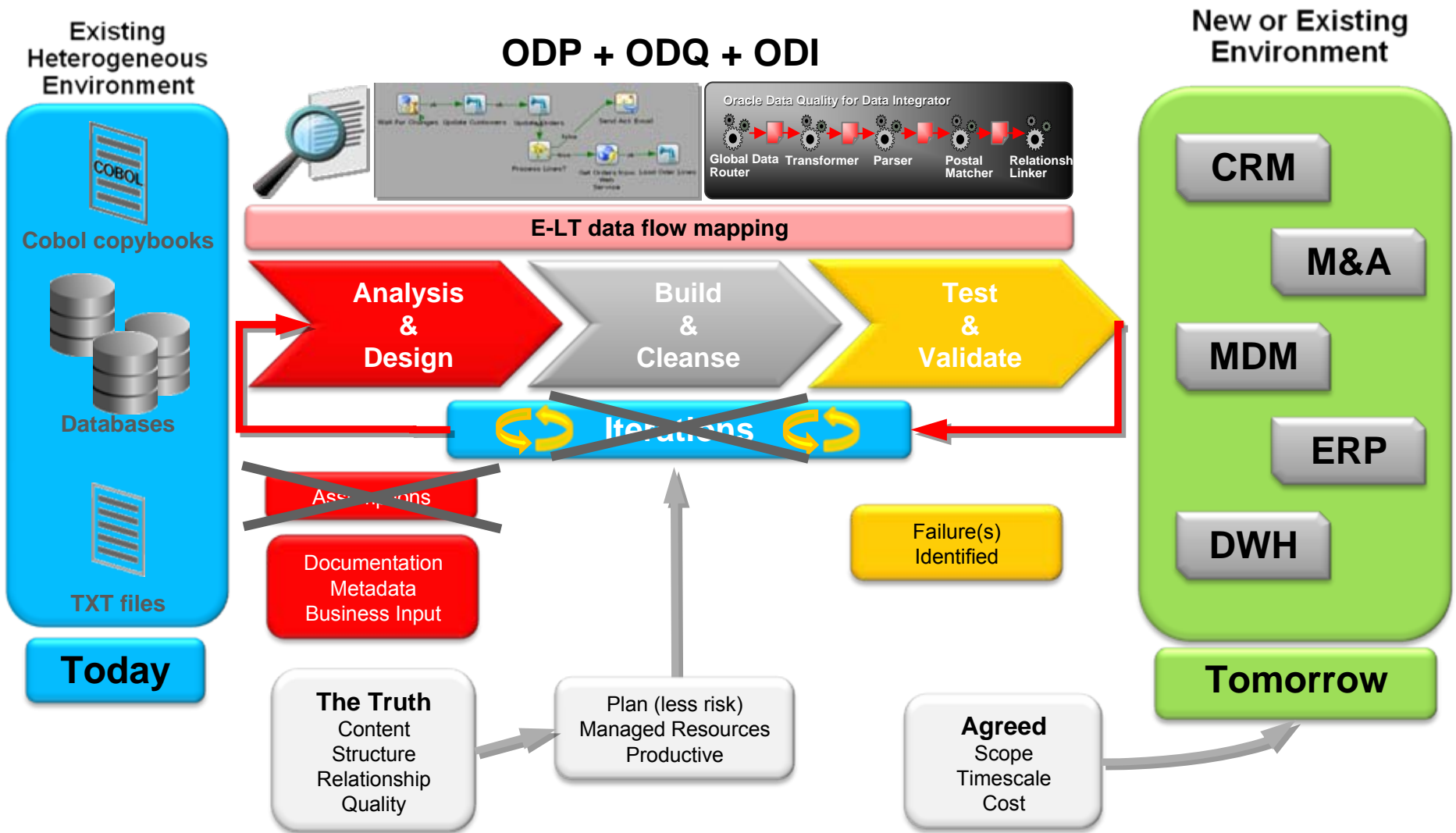
Account Mgmt.

Internet

Supplier Network

Partner Network

ODP + ODQ + ODI for Data Migration



Data Integration: Agenda

- Obstacles to Unifying Enterprise Information
- Oracle's Data Integration Strategy
- What's new in ODI-EE 11g?
- ODI-EE Architectural Advantages
- Why Data Quality
- Oracle DQ offerings in DI
- DQ Lifecycle management
 - Investigate
 - Improve
 - Govern
- Oracle Data Integrator in Action
- How to get started



Accelerated Business Process Performance by 75%

Ross Stores



Business Challenges

- Connect 900+ store chains together for better managing inventory
- Eliminate data leakage – reduce costs for every missed order
- Standardize the business on a single approach to data integration and process management which would scale, be open and standards-based

Oracle Solution

- 120+ interfaces in production utilizing process orchestration and bulk data integration patterns with **Oracle Data Integrator**
- Oracle SOA Suite footprint spans across merchandising, supply chain and financial applications
- Integration with over 10 B2B trading partners



Return on Investment

- 75% improvement in business process execution times via SET based transformations (E-LT)
- Eliminated errors in inventory replenishment
- Built trusted integration across over 400 retail chains
- Heterogeneous sources helped lower total cost of ownership

Improved Data Quality and Consistency

Verizon



Business Challenges

- Services Operations dept maintains the infrastructure which integrates the various technologies used in the company
- Needed a single integration infrastructure to glue all disparate systems to create a consolidated view of client records
- Manual integrations required between systems led to high TCO

Oracle Solution

- SOA Suite and **Oracle Data Integrator and Oracle Data Quality** provided a shared data services framework
- New integration solution allows the operations team to process large data batch loads automatically



Return on Investment

- Improved data quality and consistency with data correction, monitoring, & audit capabilities
- Saved several millions of dollars by IT consolidation of applications
- 50% performance enhancement on batch processing

Delivered Actionable Business Insight

Nespresso



Business Challenges

- Enable marketing and CRM teams to better target its campaigns and activities across 50 countries
- Scale available for high-value-added actionable data on clients, orders, stock, etc.

Oracle Solution

- Implemented **Oracle Data Integrator** to develop and supply a data warehouse from the consolidated operational transaction database of all countries
- Provided 100 gigabytes of available high-value-added actionable data on clients, orders, stock, etc.
- Developed and supplied a business data mart for marketing and CRM needs through Oracle Data Integrator



Return on Investment

- Replaced weekly insertion with daily recording of change—allowing for more effective client follow-up
- Improved response times for production by transferring reporting processes to the data warehouse
- Improved reactivity for restocking requirements thanks to daily updates on stock quantity and value for each store
- Provided overall view of the status of business intelligence processing, thereby resulting in improved user communication

Increased DW Loads by 40%

Raiffeisen International Bank



Business Challenges

- Improve scalability to populate data warehouses across 12 countries
- Reduce number of hand-coded ETL jobs to minimize complexity and reduce high maintenance costs
- Minimize costs associated with system migrations and upgrades

Oracle Solution

- Implemented **Oracle Data Integrator** to standardize ETL for all processes and jobs across multiple financial chains
- Ensured a high-performance operation via horizontal and vertical scaling options
- Leveraged the flexible architecture of Oracle Data Integrator to develop a converter to automate the migration of ETL jobs



Return on Investment

- Provided efficient access to data sources and significantly improved data warehouse processing performance by up to 40%
- Reduced costs for migrations and upgrades across multiple regions
- Reduced risk of data gaps, data errors and synchronization of data across global regions

Faster Time To Value for Business Intelligence

Finansbank



Business Challenges

- Implement an automatic data management system for business intelligence to reduce man time lost through manual coding
- Streamline systems for ease of use to reduce training time
- Improve data tracking capabilities to solve queries and ambiguities quickly and efficiently

Oracle Solution

- Implemented Oracle Data Integrator to implement all BI/DW loads
- Combined solution for Oracle Database, Oracle Performance Analyzer, Oracle Financial Data Manager to improve performance management



Return on Investment

- Accelerate response delivery to business intelligence requests, reducing time needed from one day to one hour
- Cut learning period for new staff from about six weeks to three days
- Reduced time required to trace business intelligence data to its source from two days to a matter of seconds
- Reduce the time needed to identify the impact of changes to source data from weeks to just minutes



Real Results: Oracle Data Integrator EE

Reduce costs, improve performance, faster time to value

Cost Savings, Efficiencies

- Reduce development costs by at least **30%**
- Improve the speed and accessibility of data by **50%**

Reduce Risk

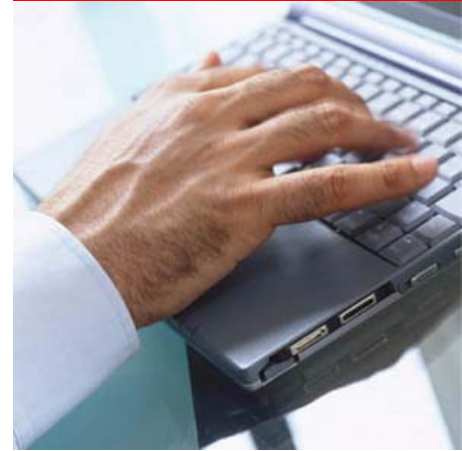
- Improve data quality by integrating cleansing as part of the process
- Eliminate accessibility issues by better connecting data

Business Insight

- End-to-end data lineage
- Ensure pervasive reach of BI/EPM applications to multiple source systems

Data Integration: **Agenda**

- Obstacles to Unifying Enterprise Information
- Oracle's Data Integration Strategy
- What's new in ODI-EE 11g?
- ODI-EE Architectural Advantages
- Oracle Data Integrator in Action
- **How to get started**





Best Practices

1. Leverage E-LT and real-time data replication for real-time BI/DW
2. Commit to a service oriented architecture approach for increased agility
3. Leverage both pre-packaged Oracle solutions, Oracle Partners and Oracle Professional Services


For More Information

Get Started

- Visit the Oracle Fusion Middleware 11g web site at <http://www.oracle.com/goto/fmw11g/index.html>
- Oracle Data Integration on oracle.com www.oracle.com/goto/odi
- Oracle Fusion Middleware on OTN <http://otn.oracle.com/middleware>
- Information on GoldenGate: <http://www.oracle.com/goldengate>

Resources

- Datasheet: <http://www.oracle.com/products/middleware/odi/docs/odiee-datasheet.pdf>
- Blog: <http://blogs.oracle.com/dataintegration>
- Technical information available at: <http://www.oracle.com/technology/products/oracle-data-integrator/index.html>
- Data Integration Events <http://www.oracle.com/events>



The preceding is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.