

Oracle Data Warehousing

Guided Tour



ORACLE

#1 DM Review Survey 2001

Data Warehousing & Business Intelligence Vendor Recognition

*Jennifer Leung
Technical Team Leader
Sales Consulting
Financial Services
Oracle Corporation
jennifer.leung@oracle.com*


ORACLE®

Agenda

- Oracle rdbms datawarehousing marketshare
- Oracle Data Warehousing Process Products and Tools
- Oracle VLDB Capabilities
- Query Performance and Optimization Capabilities

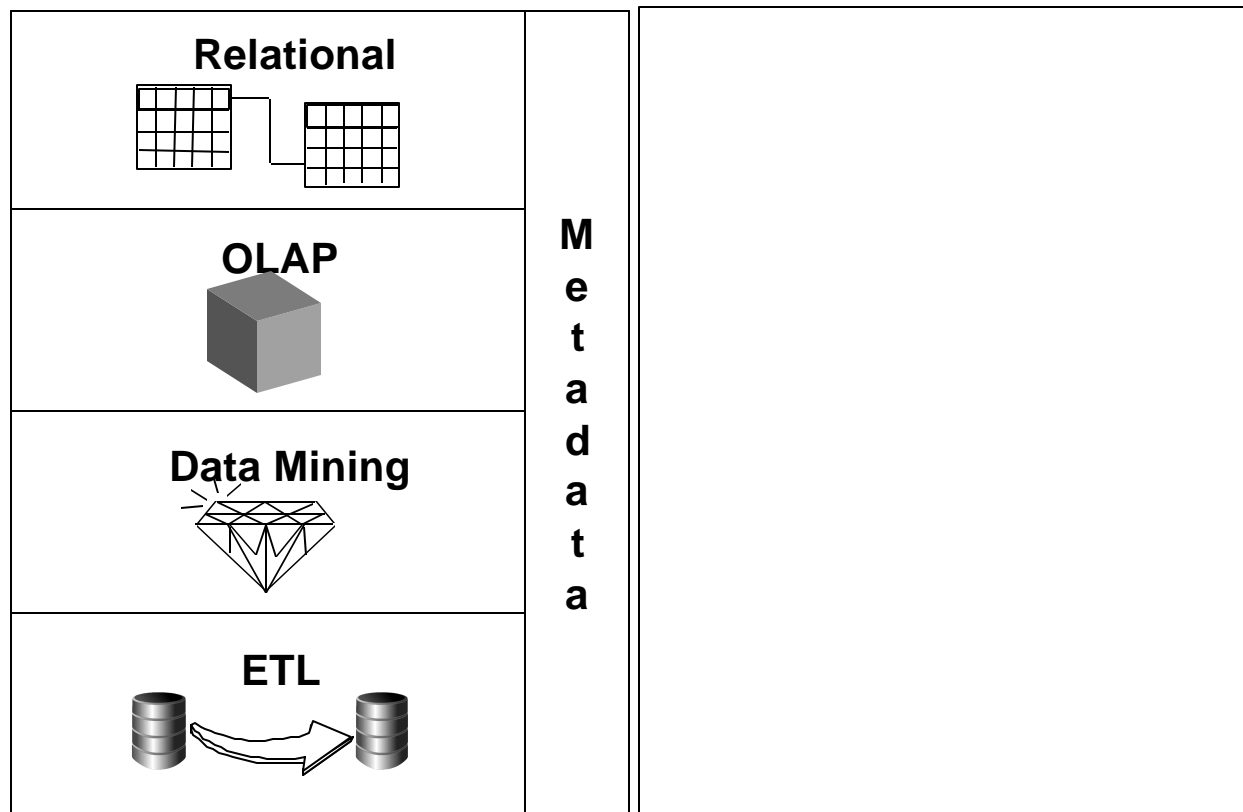


Agenda

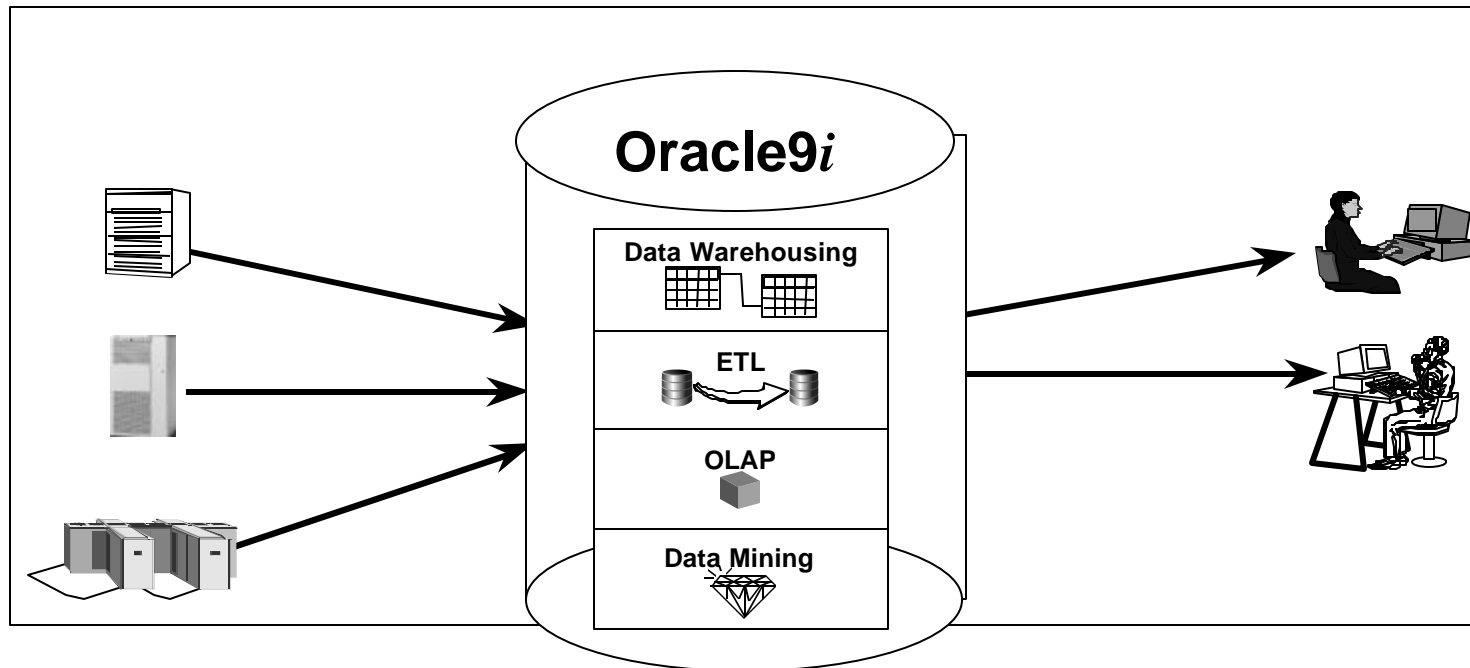
- 
- Oracle rdbms datawarehousing marketshare
 - Recent Benchmarks
 - Oracle Data Warehousing Process Products and Tools
 - Oracle VLDB Capabilities
 - Query Performance and Optimization Capabilities

Oracle9i Database

Single business-intelligence data server



The New Way: Oracle9i

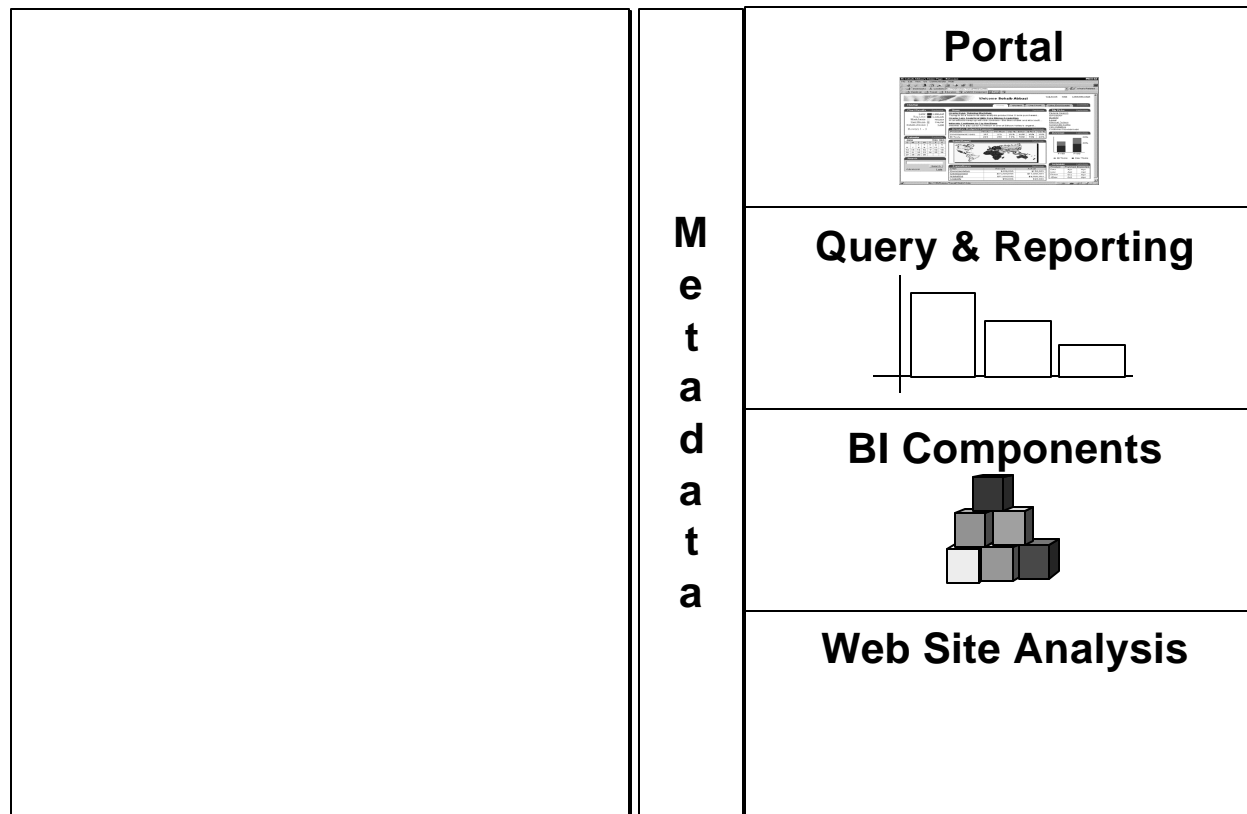


- **Single business intelligence platform**

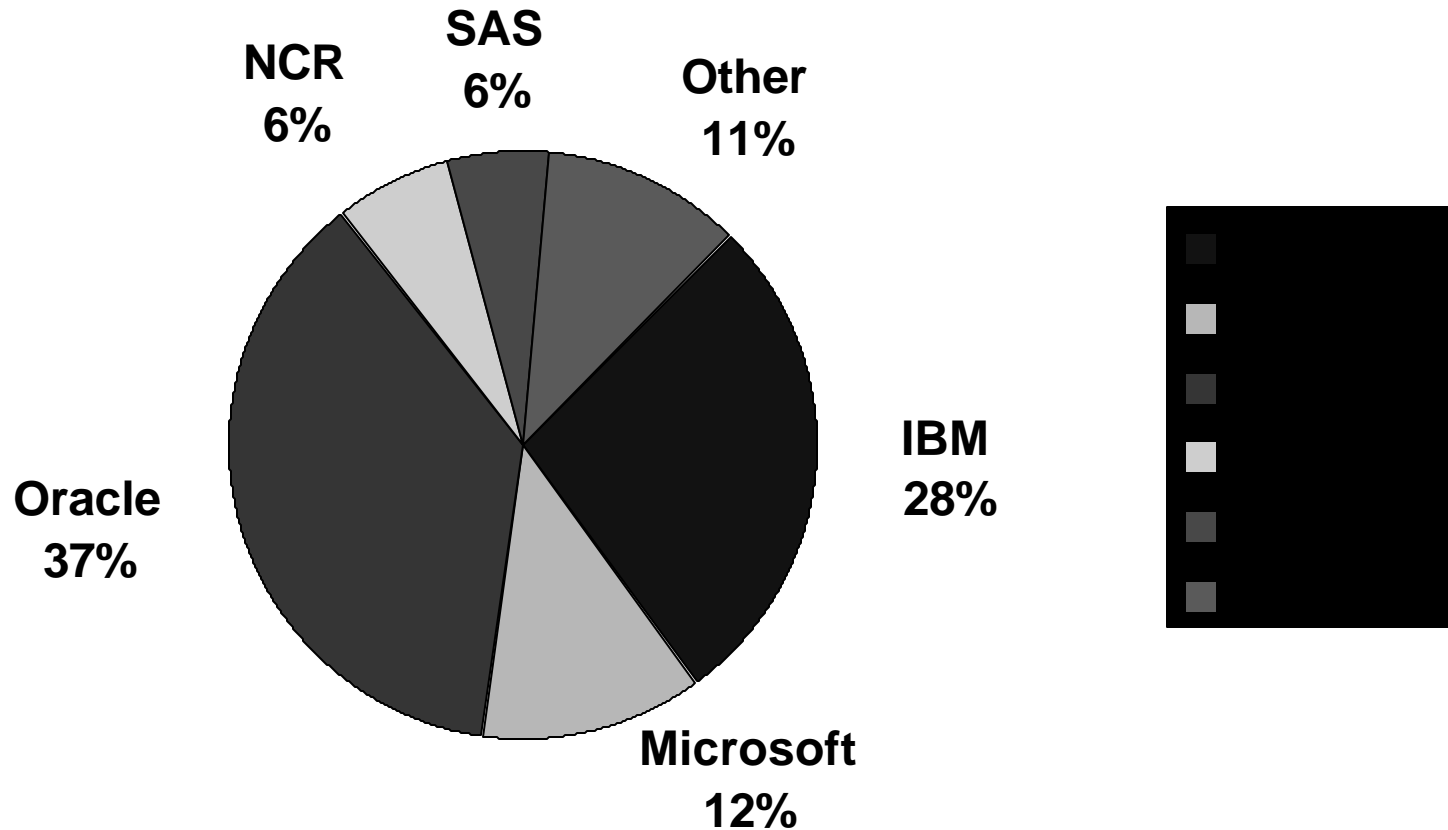
- Reduce administration and implementation costs
- Faster deployment
- Improved scalability and reliability

Oracle9i Application Server

Runs All Your Business Intelligence Applications



Data Warehouse DBMS Market Share



Source: "Data Warehousing Tools: Market Forecast and Analysis, 2001-2005", IDC, 2001

Agenda

- Oracle rdbms datawarehousing marketshare
- Oracle Data Warehousing Process Products and Tools
- Oracle VLDB Capabilities
- Query Performance and Optimization Capabilities

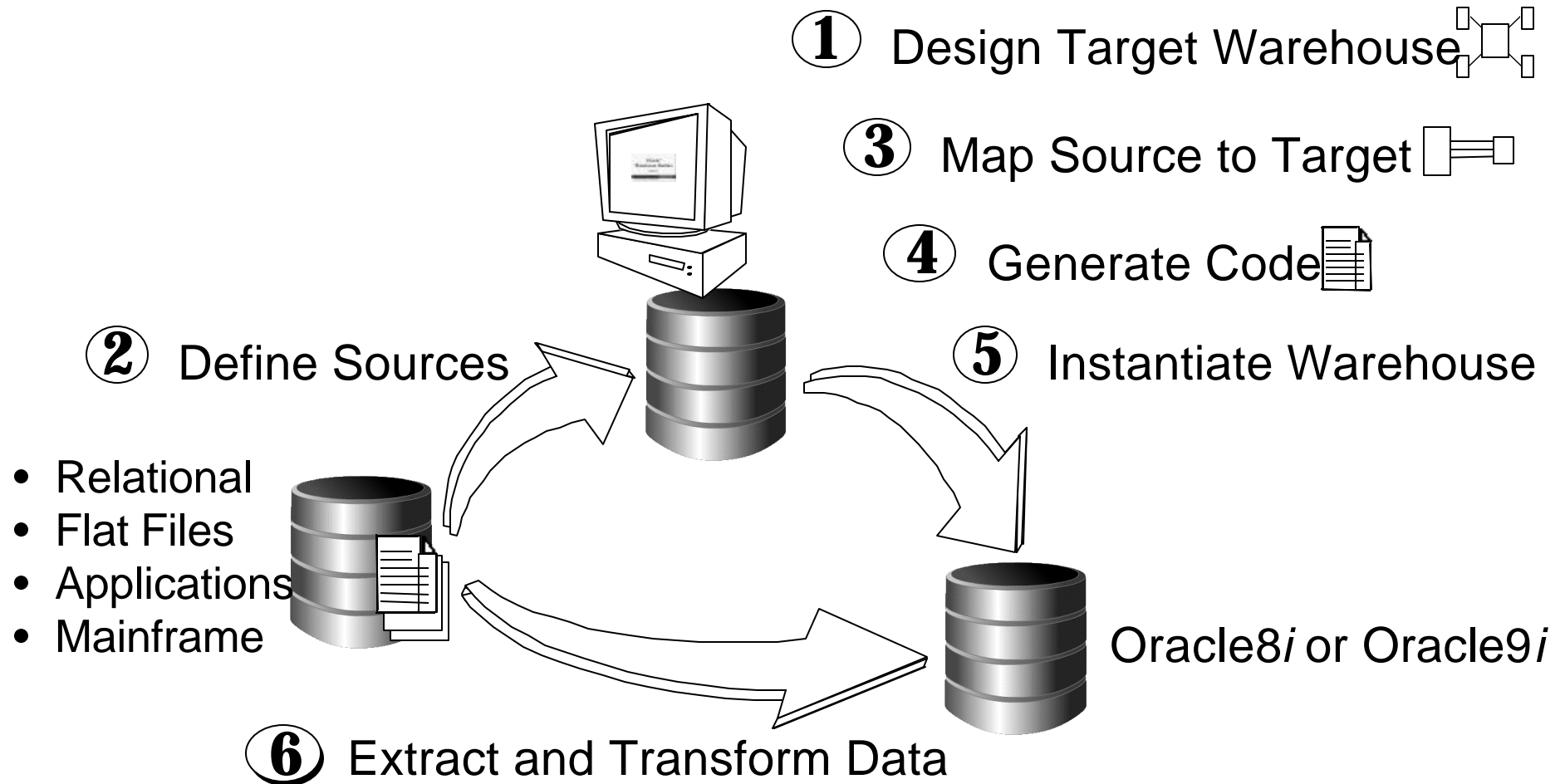


What Is Oracle Warehouse Builder?

- An extensible common warehouse metadata (cwm) based framework for designing, deploying and managing enterprise data warehouses, data marts and e-Business intelligence applications
- Completely CWM based
- Included with Oracle9i/Developer Suite

Oracle Warehouse Builder

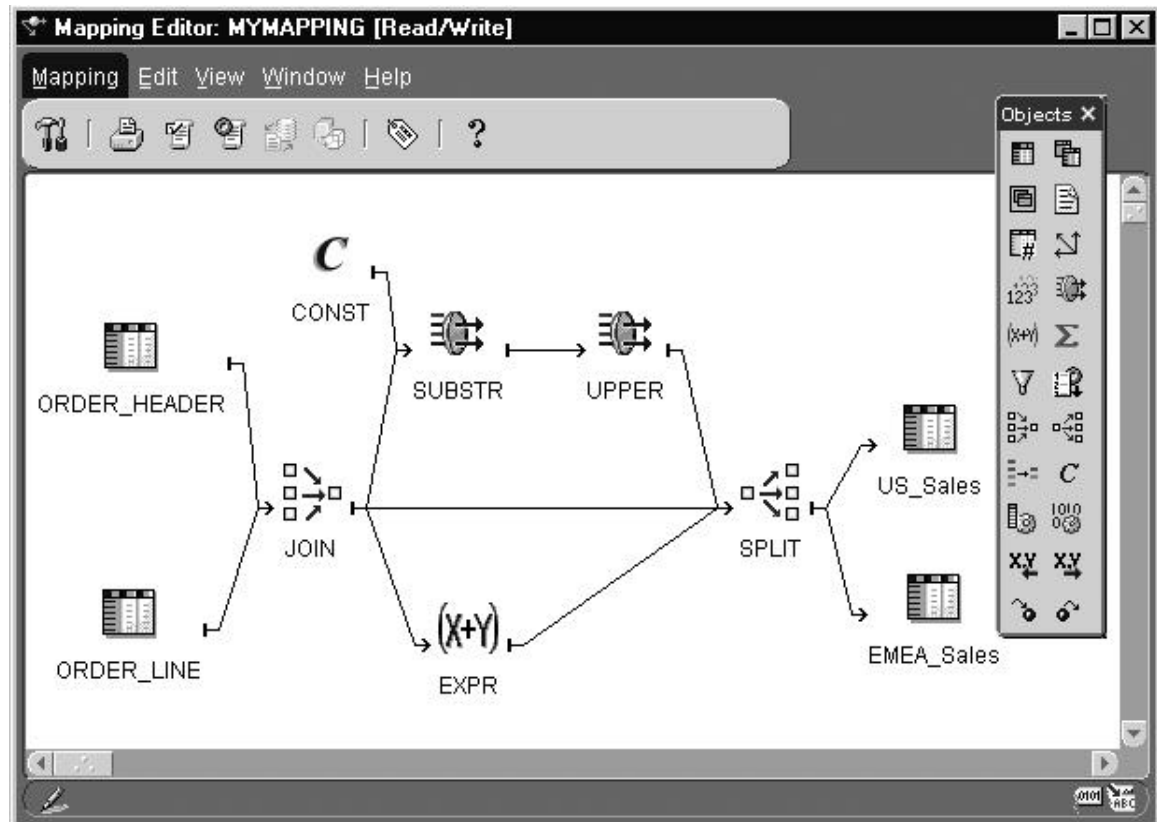
Design, Deploy, Manage



OWB9i Mapping Sources to Targets

Mapping Editor

- Joiner
- Filter
- Aggregator
- Deduplicator
- Sorter
- Splitter
- Sequence
- Inline Expressions
- Transformations
- Pre- and Post-Mapping Processes



OWB9i Incremental Code Generation

The screenshot displays two windows from the Oracle WebLogic 9i environment:

- Mapping Editor: MYMAPPING [Read/Write]**: Shows a mapping diagram with a table icon labeled "ORDER_LINE" connected to a "JOIN" node. The "Generate" menu is open, with a sub-menu showing "Mapping...", "Intermediate Result...", and "UPP".
- Code Viewer: MYMAPPING**: Shows the generated SQL code. The "Operating mode" dropdown is set to "Set based". The code includes an "INSERT" statement and a "SELECT" statement with a "WHERE" clause.

```
INSERT
/*+ APPEND PARALLEL("EMEA_SAL
INTO
"EMEA_SALES"
("RESULT",
"MARGIN")
SELECT
/*+ DRIVING_SITE("ORDER_HEADER_OLTPSOURCE")*/
(UPPER( (SUBSTR("ORDER_LINE_OLTPSOURCE"."ITEM_ID", (MYMAPPING."CON
("ORDER_LINE_OLTPSOURCE"."SALES" - "ORDER_LINE_OLTPSOURCE"."COST
FROM
"ORDER_HEADER"@OLTPSOURCE" "ORDER_HEADER_OLTPSOURCE",
"ORDER_LINE"@OLTPSOURCE" "ORDER_LINE_OLTPSOURCE"
WHERE
"ORDER_LINE_OLTPSOURCE"."ORDER_ID" = "ORDER_HEADER_OLTPSOURCE"."
"ORDER_HEADER_OLTPSOURCE"."REGION" = 'EMEA';
COMMIT;
```

Ln 1, Col 1 | INS

The OWB Repository

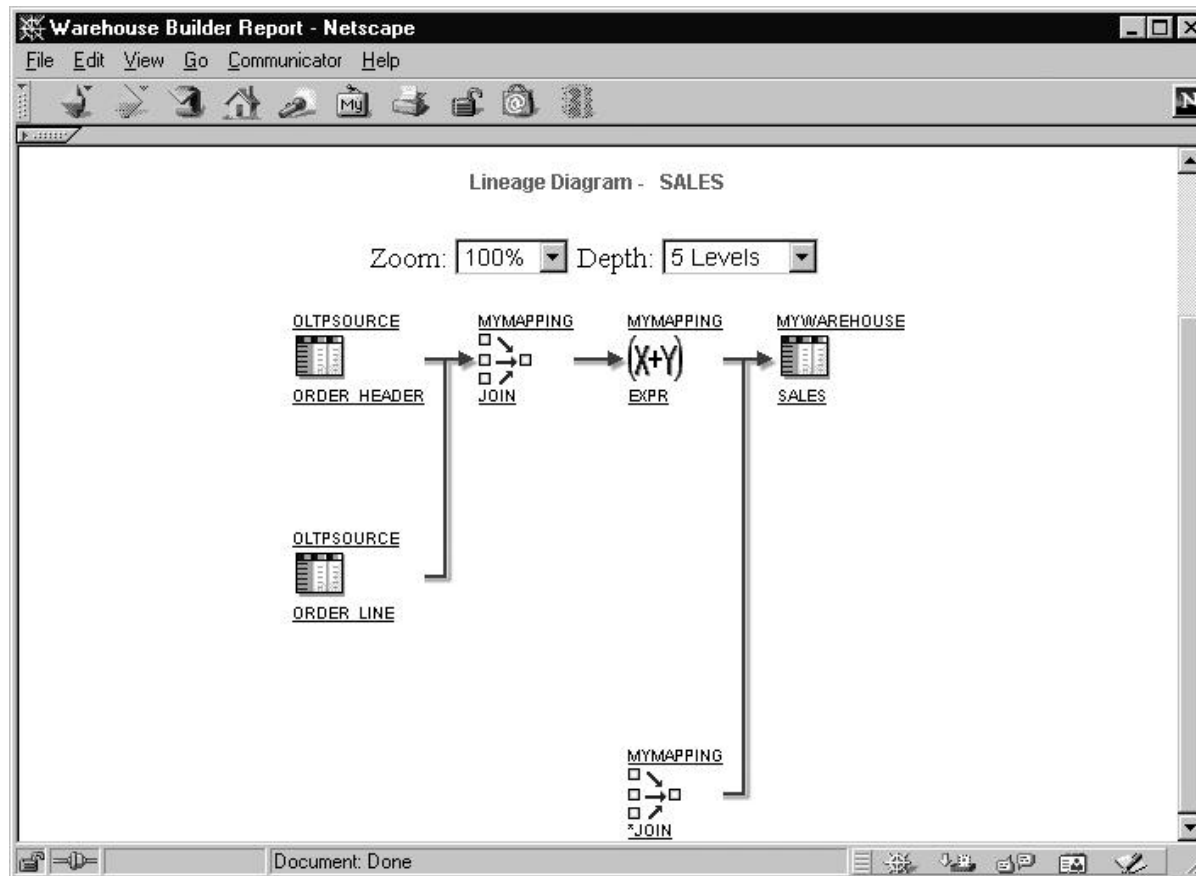
- Multi-user repository
 - Single update, multiple read
 - Object locking
 - Synchronize
- Metadata Import/Export
- Archive/Restore
 - Leverage 3rd party source control system

OWB Metadata Integration

- **OMG CWM**
 - Open Standard, XML Metadata Interchange (XMI)
 - Powerful object model
 - Spans metadata related to ETL, analysis
- **Oracle Warehouse Builder Bridges**
 - Oracle Discoverer, Express
 - CA ERwin
 - Sybase PowerDesigner
 - OMG CWM
- **Metadata exchange from Designer, Pure tools**

OWB Metadata Browser

Reporting, Data Lineage, Impact Analysis



OWB 9i Toolkit

- Oracle9i ETL Toolkit includes
 - Change data capture
 - External tables
 - Table functions
 - Merge (insert / update)
 - Multi-table insert
 - Resumable statements
 - Transportable Tablespaces with different block sizes

Metadata Reporting

The screenshot shows a Netscape browser window titled "Warehouse Builder Navigation - Netscape". The address bar contains the URL: `avigation_page?p_repository_id=1&p_role=wmp_developer&p_sheet=2&p_type_name=InformationSystem&p_id=1478`. The page header features the Oracle logo and the text "Warehouse Builder Navigation" with a "Home" link. Below the header, the date "June 28, 2001" is displayed, along with links for "Add to My Favorites", "Browse My Favorites", "Customize", and "Logout". The main content area shows the breadcrumb path: "Warehouse Builder : Repository > OWB_DEMO_PROJECT : Project > RELATIONAL_SOURCE : Module". A question mark icon is visible next to the path. Below the path, the text "(no description available)" is shown. At the bottom of the main content area, there are four tabs: "Properties", "Contents", "Related", and "Reports". The "Contents" tab is currently selected. Below the tabs, the text "Browse content objects." is displayed. A table with three columns: "Name", "Type", and "Actions" is shown. The table lists three tables: "GEOGRAPHIES", "ORGANIZATIONS", and "PARTY_ADDRESS". Each table entry includes a small table icon and the text "Table". The "Actions" column for each table contains links for "properties", "contents", "related", and "reports".

ORACLE

Warehouse Builder
Navigation

Home

June 28, 2001 [Add to My Favorites](#) [Browse My Favorites](#) [Customize](#) [Logout](#)

RELATIONAL_SOURCE ?

Path: [Warehouse Builder : Repository](#) > [OWB_DEMO_PROJECT : Project](#) > [RELATIONAL_SOURCE : Module](#) Warehouse Engineer

(no description available)

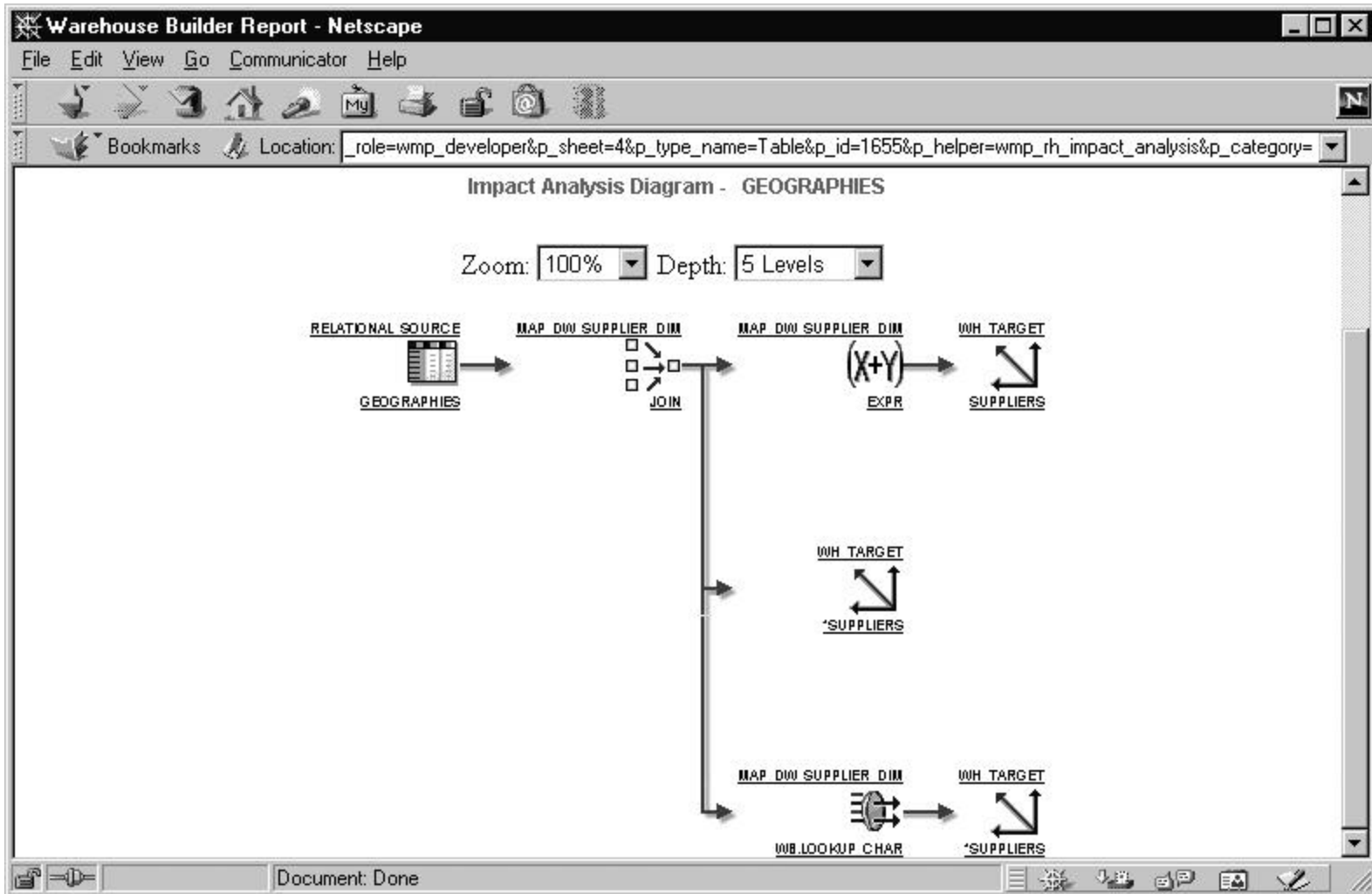
Properties Contents Related Reports

Browse content objects.

Name	Type	Actions	?
GEOGRAPHIES	Table	properties contents related reports	
ORGANIZATIONS	Table	properties contents related reports	
PARTY_ADDRESS	Table	properties contents related reports	

Document: Done

Metadata Reporting – Impact Analysis



Types of Analysis

Ad-Hoc &
Reporting

OLAP

Data Mining

**Extraction of
detailed and
summary data**

**Summaries,
trends and
forecasts**

**Knowledge discovery
of hidden patterns
and insights**

Information

Analysis

Insight and Prediction

**Who charged
most on credit
cards in the last
3 years?**

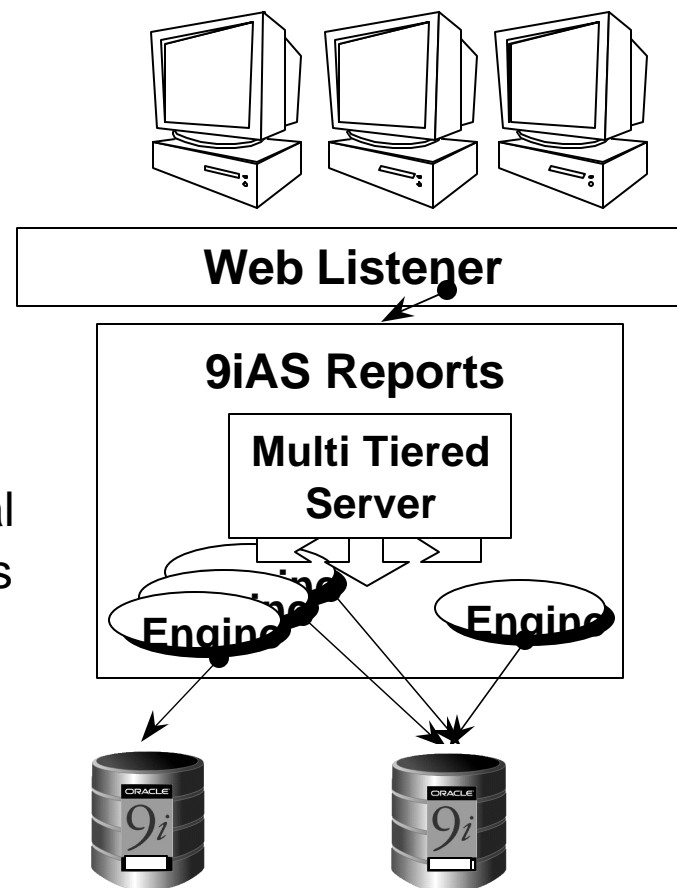
**Usage
distribution
of credit
card users?**

**Who will switch
corporate cards
in the next 6
months
and why?**

Oracle9iAS Reports

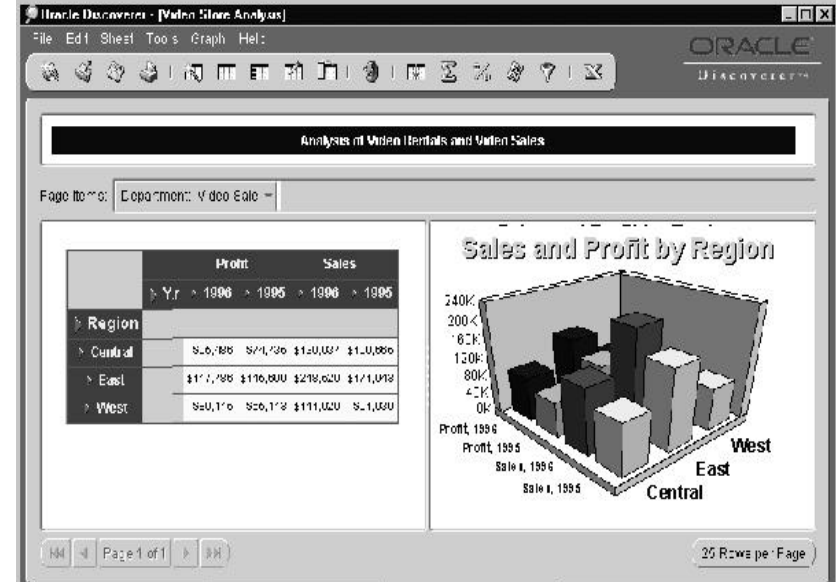
Application Server based Reporting

- **Reports dynamically generated on the server**
- **Batch scheduling, output caching**
- **Load balancing & clustering**
- **New in Oracle9iAS rel. 2**
 - JSP-Based runtime for high quality HTML publishing
 - Seamless integration with Oracle Portal
 - Pluggable Data Sources & Destinations
 - Event Based Reporting
 - Reports Services Enhancements
 - Enhanced Reports Bursting



Oracle9iAS Discoverer

- **Interactive web queries**
 - Power user
 - Build, modify, format queries and graphs
 - Export to popular applications
- **Recently added features**
 - Dynamic drill, pivot, graph
 - Analysis using SQL analytic functions
 - Materialized views support
 - Firewall support



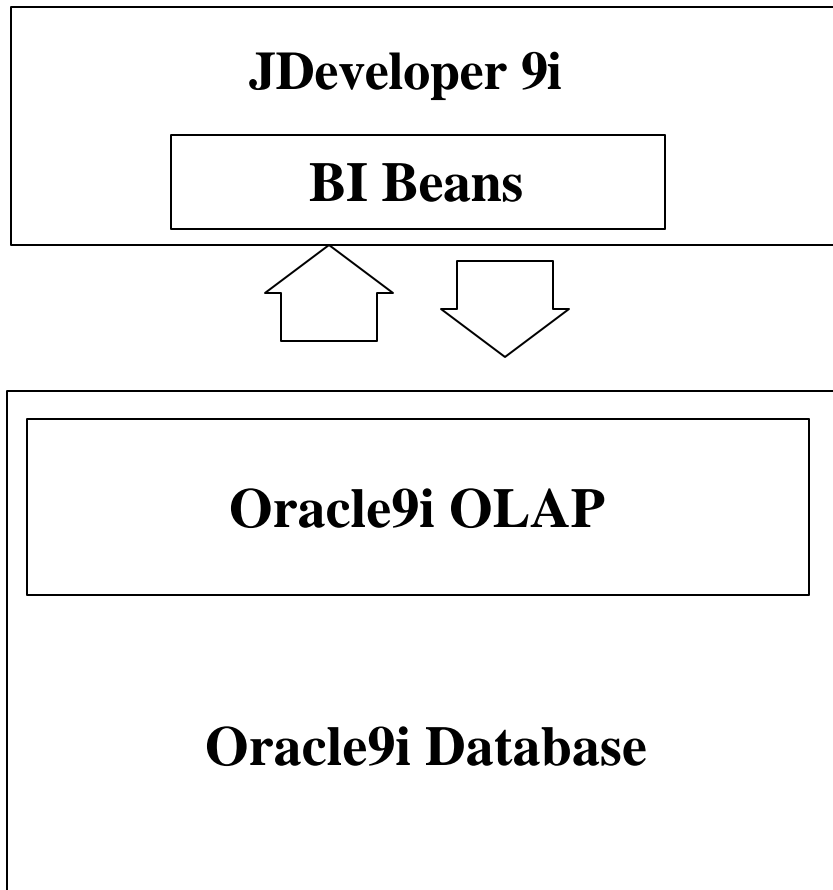
Pure Java, Firewall support

9i OLAP Functionality Overview

Key points:

- Move the OLAP engine into the database
- Achieve scalability
- Achieve security
- Achieve de-mystifying of OLAP by
 - Allowing SQL access
 - Allowing mix-and-match of OLAP and Relational

OLAP Application Platform – OLAP Services

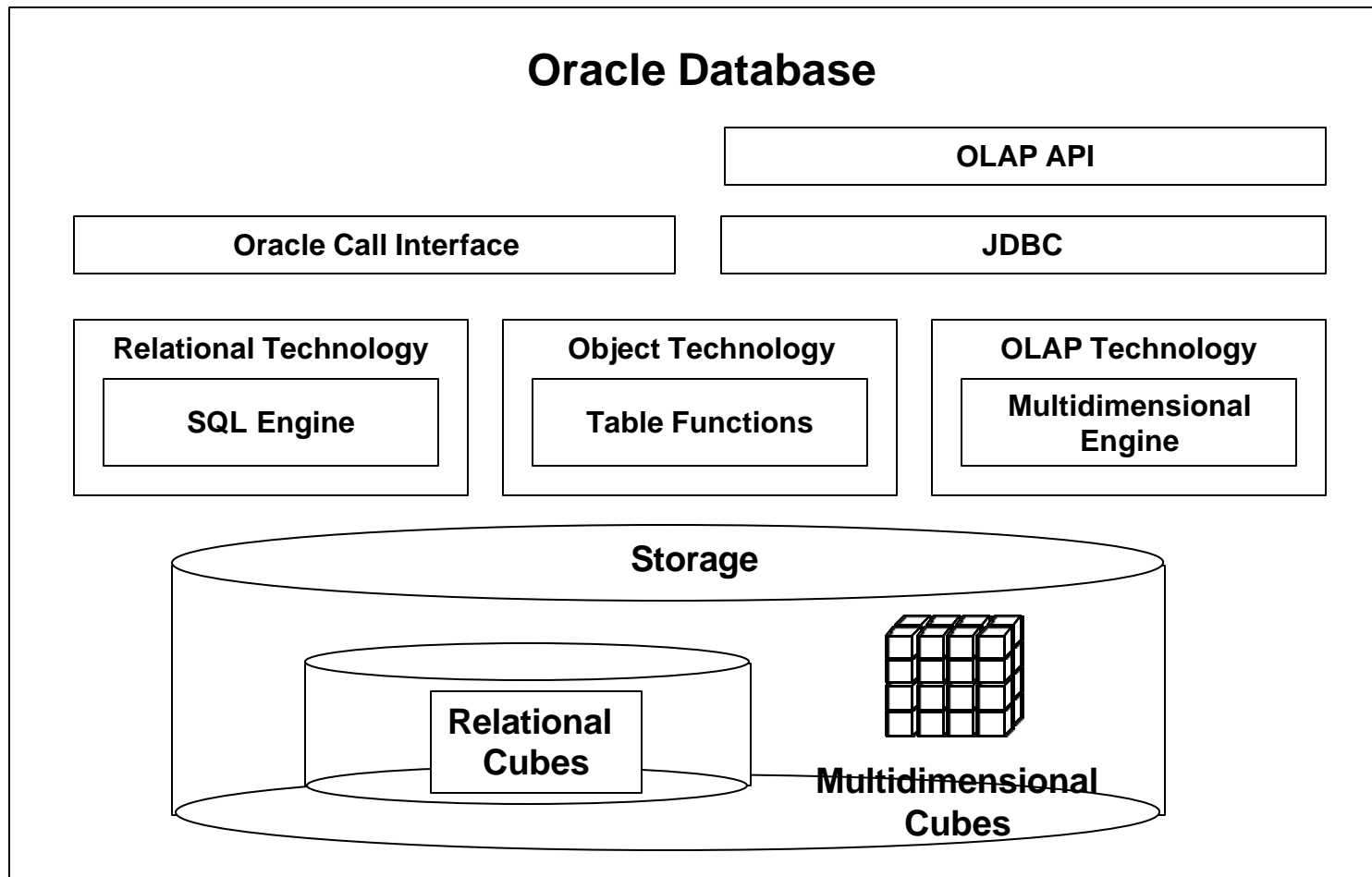


**Rapid application development
OLAP ready**

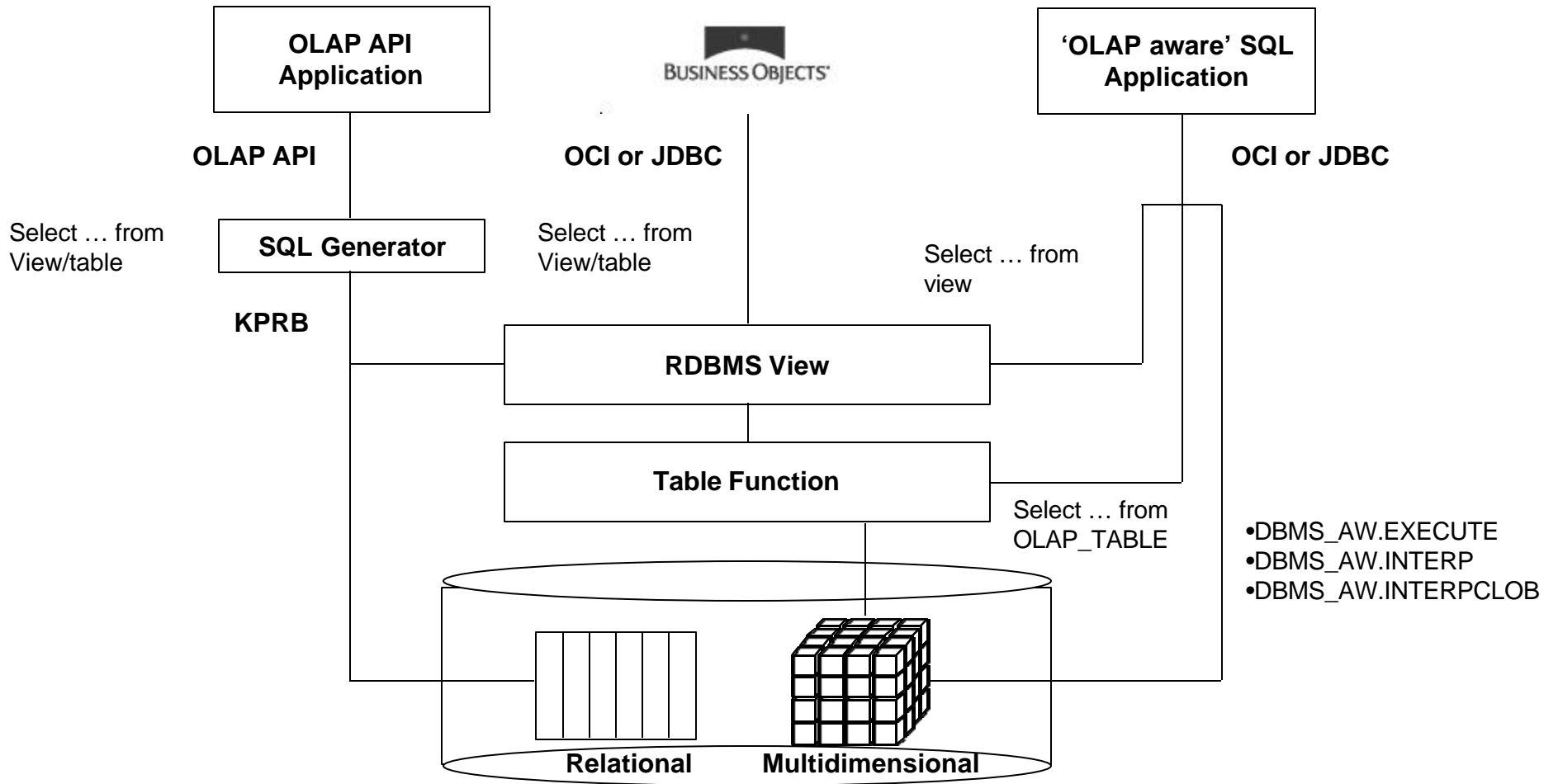
**Java OLAP API
Predictive OLAP functions**

**Scaleable data store
Integrated metadata
Summary management
SQL analytic functions**

9iOLAP Integrated RDBMS-MDDS



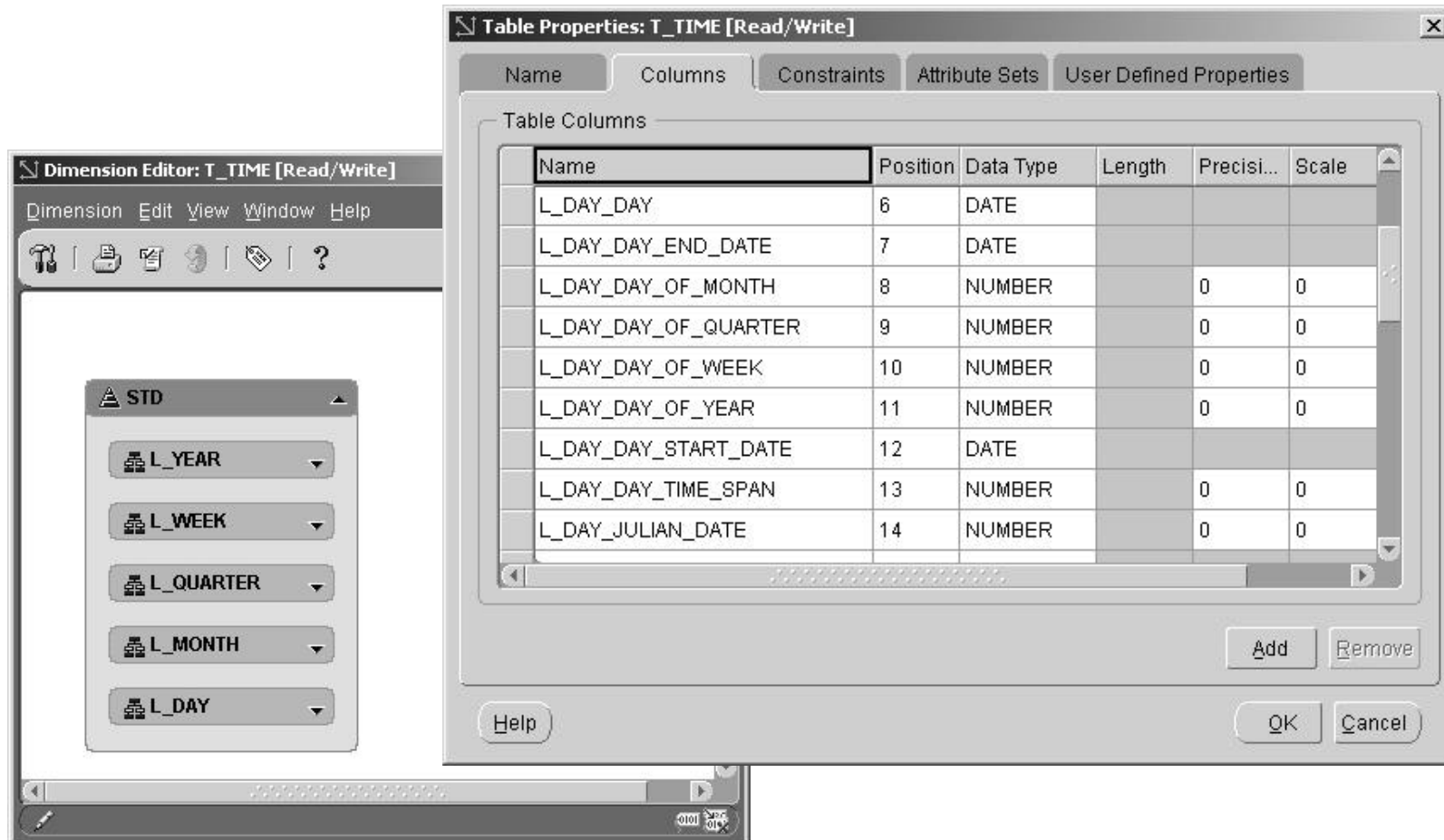
Query Methods



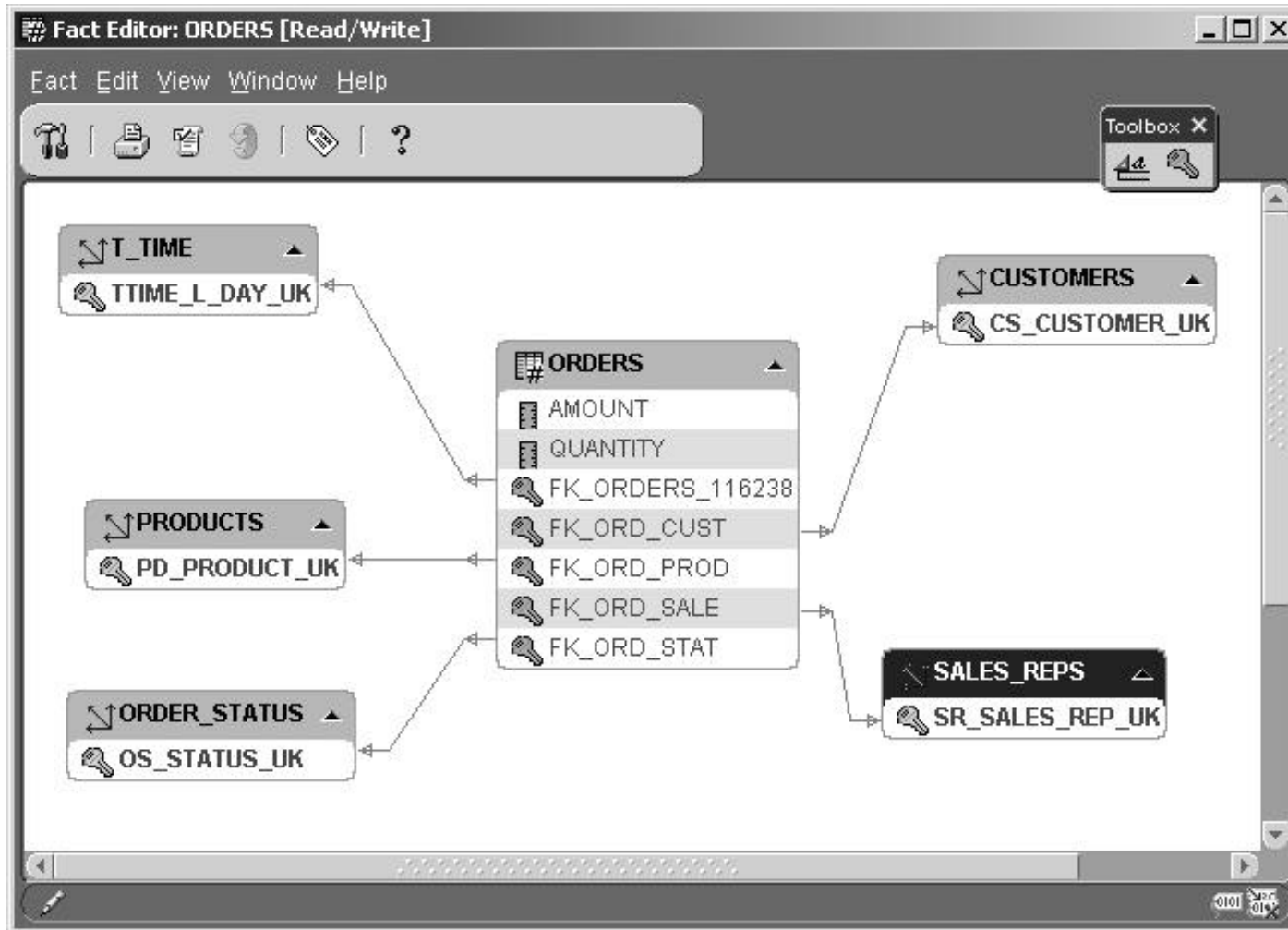
OLAP integration with OWB

- Design the Warehouse in Warehouse Builder
 - Dimensions
 - Cubes
- Design the Data Flow for Relational Objects
- Add the Data Flow for Multi-Dimensional Objects
- Deploy the Design to the Database
- Run the Loads
- Query the Results

Metadata Design - Dimensions



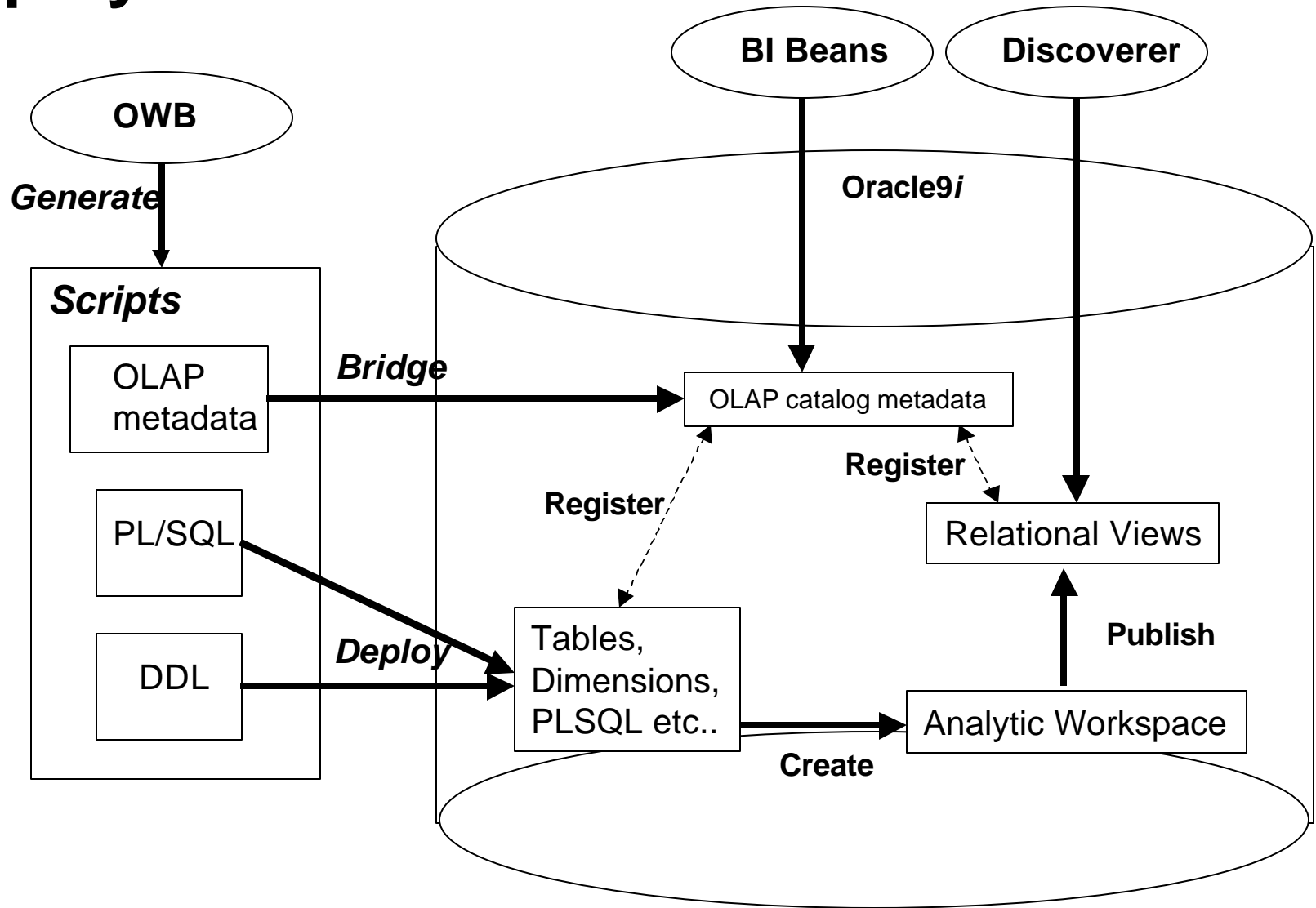
Metadata Design - Cubes



Deployment

- Deploy the regular scripts using the deployment Manager
- Deploy OLAP metadata via the new/improved OLAP bridge, this:
 - Creates all skeleton objects (empty)
 - Registered the objects in the OLAP catalog
 - Binds the OLAP objects to the relational objects
 - BI Beans enabled environment
- Creates a ROLAP environment

Deployment



Deployment - Bridge

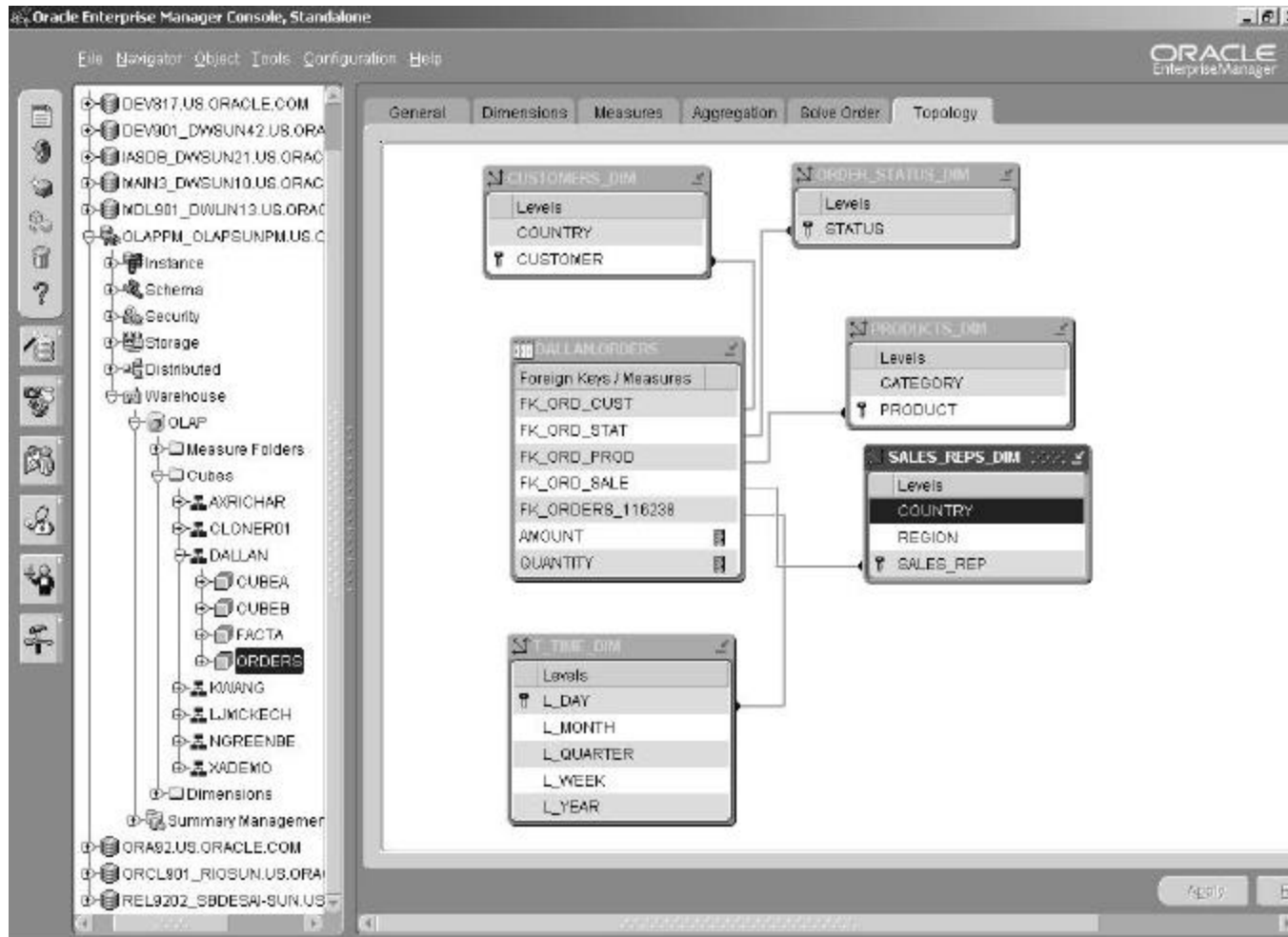
The screenshot shows the Oracle Warehouse Builder Transfer Wizard, step 2 of 3: Transfer Parameter Identification. The wizard is overlaid on the Warehouse Module Editor, which displays a logical tree of the warehouse structure. The wizard's main window is titled "Oracle Warehouse Builder Transfer Wizard, 2 of 3: Transfer Parameter Identification" and contains a section titled "Metadata Object/Filenames and Detailed Transfer Parameters". Below this title, it instructs the user to "Enter or change the transfer parameter values." A table lists various transfer parameters and their values. The parameters include OWB Exported Business Areas (set to ORDERS), OWB Translated Language (English), Username (dallan), Password (masked with asterisks), Hostname (olapsunpm), Port (1521), SID (olappm), PL/SQL Output File (c:\olap_ordersx.sql), and Deploy PL/SQL in database (No). The wizard has navigation buttons for Cancel, Help, Back, Next, and Finish.

Transfer Parameter Name	Transfer Parameter Value
OWB Exported Business Areas	ORDERS
OWB Translated Language	English
Username	dallan
Password	*****
Hostname	olapsunpm
Port	1521
SID	olappm
PL/SQL Output File	c:\olap_ordersx.sql
Deploy PL/SQL in database	No

Deployment – Bridge

Collection Name	-	<i>Collection to export</i>
OWB Translated Language	-	<i>MLS Language</i>
Deploy to AW	-	<i>Do you want to create an AW definition</i>
AW Name	-	<i>Name for the AW</i>
Generate View Definitions	-	<i>Do you want to generate views for this AW</i>
Generated View Prefix	-	<i>Prefix for the views</i>
Access Type	-	<i>OLAPI, DISCO (currently ignored)</i>
Generated View Directory	-	<i>Directory on <u>server</u> for generated view script</i>
Deploy PLSQL in Database	-	<i>Do you want to deploy the PLSQL in the db?</i>
Username	-	
Password	-	
Hostname	-	
Port	-	
SID	-	
PLSQL Output File	-	<i>Resultant PLSQL generated</i>
Log Level	-	<i>Information / Trace / Error</i>

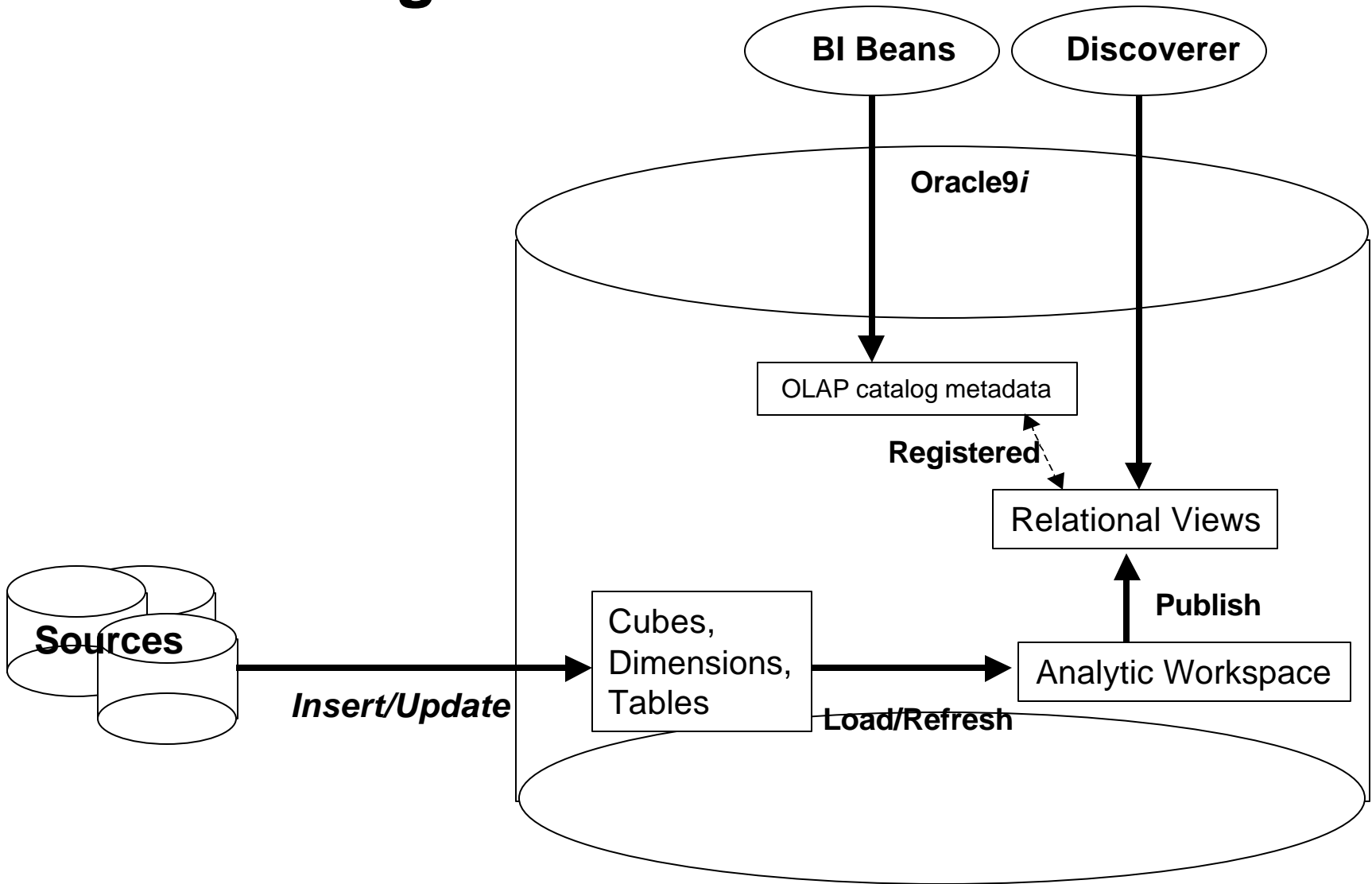
Metadata - OEM



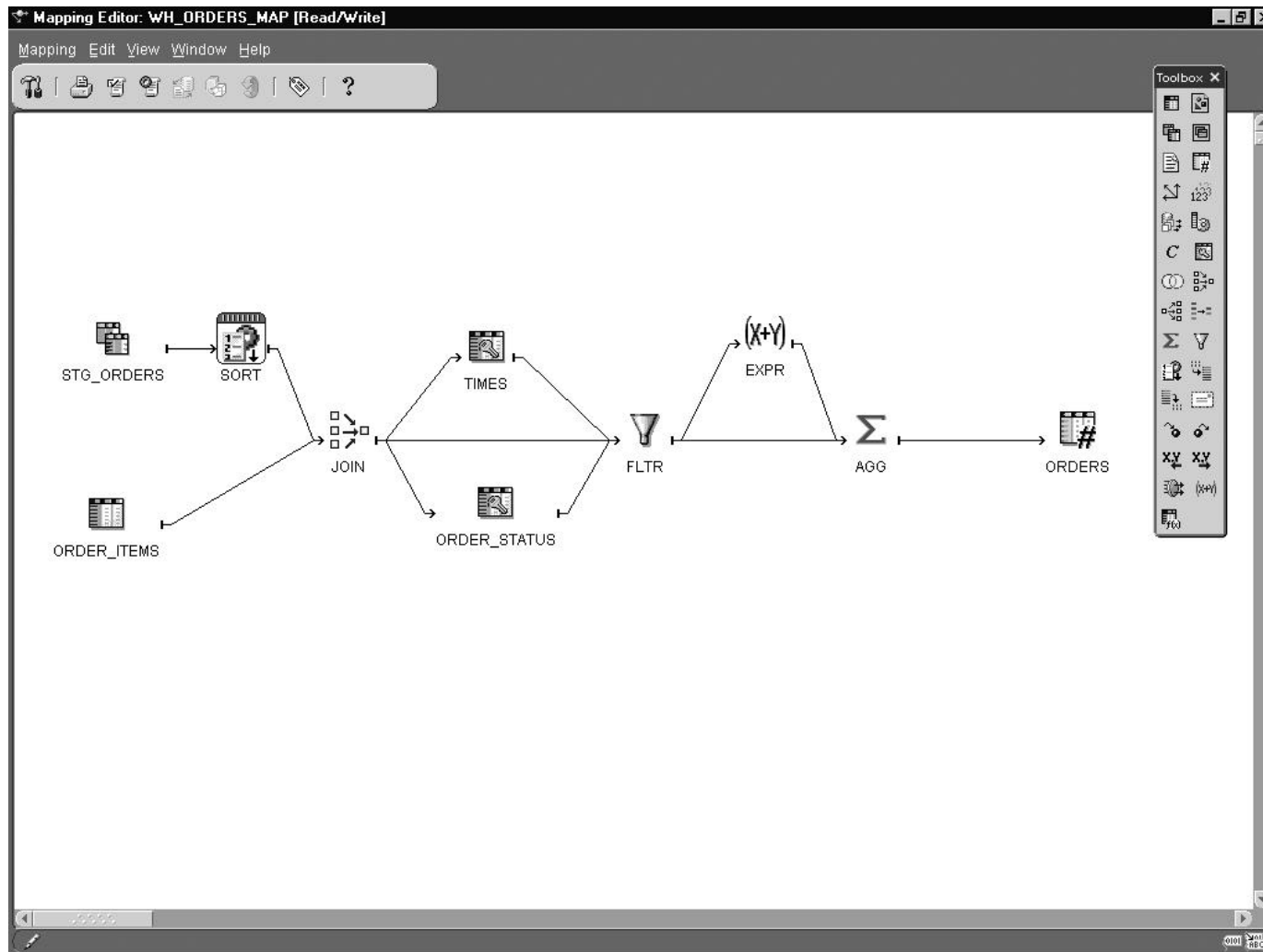
Data Loading

- Load the relational objects via a normal mapping
- Load the OLAP AWs via:
 - Mapping – post mapping process
 - Process Flow activity
- Load the OLAP AWs:
 - Refresh or Insert into Dimensions
 - Refresh or Insert into Cubes
- Using an OWB wrapper procedure on top of the RDBMS pl/sql

Data Loading



Data Movement – Relational

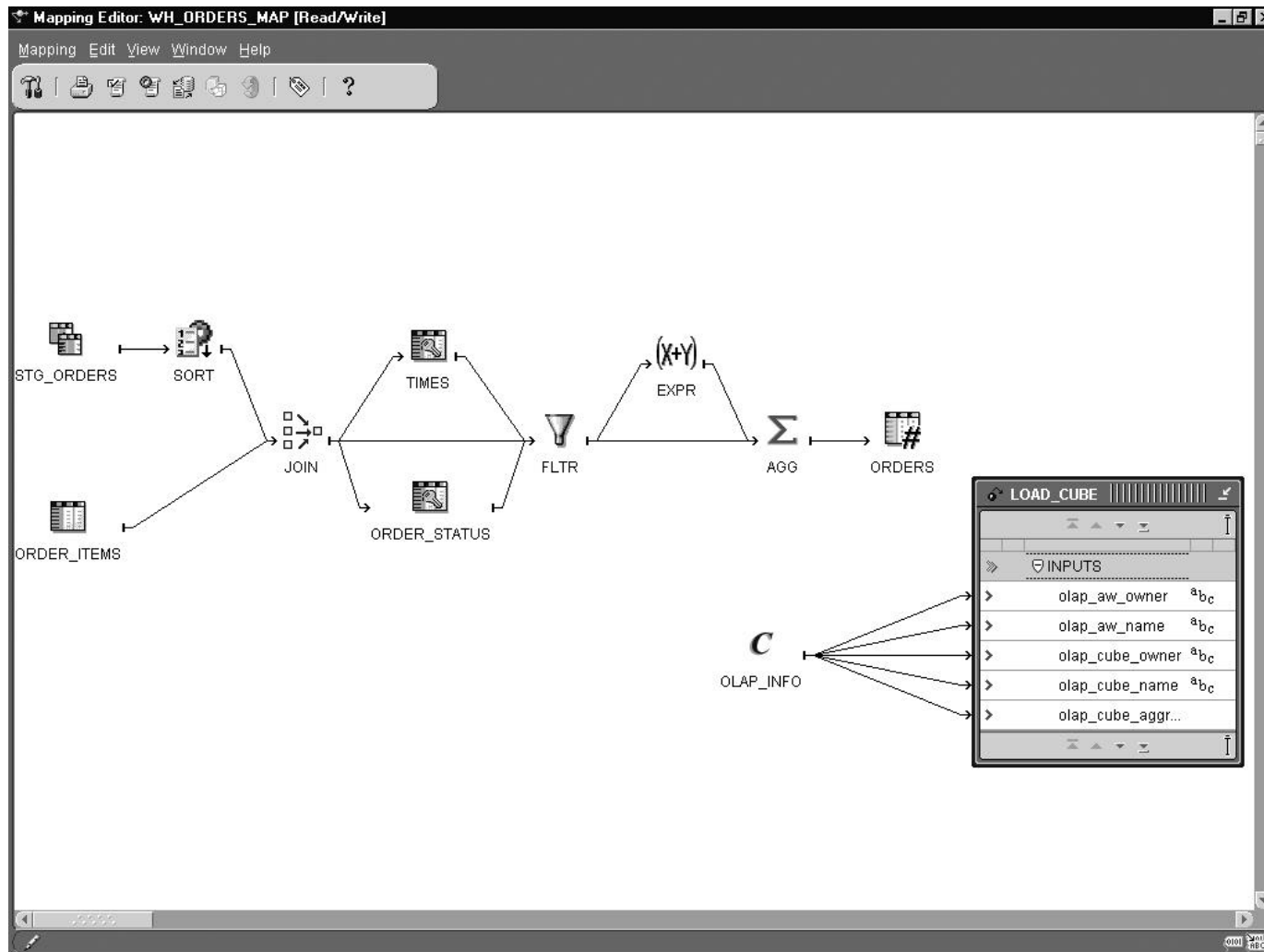


Data Movement – Loading OLAP



Refresh routines for AW
are published as OWB
OLAP procedures

Data Movement – Adding OLAP



Data Movement – Using PFE

Process Editor: LOAD_DIMENSIONS [Read/Write]

Process Flow Edit View Window Help

Graph View Code View

Toolbox

WH_PRODUCTS_MAP
WH_CUSTOMERS_MAP
WH_SUPPLIERS_MAP
AND_ACTIVITY
WB_OLAP_LOAD_DIMENSION_1
WB_OLAP_LOAD_DIMENSION_2
AND_ACTIVITY_1
FORK_1

Activity View

ACTIVITY	PARAMETER	DAT...	DI...	VALUE	DESCRIPTI
WB_OLAP_	OLAP_AW_OWNER	STR...	IN	WH_AW	
	OLAP_DIMENSION_NAME	STR...	IN	'CUSTOMERS'	
	OLAP_DIMENSION_OWN...	STR...	IN	WH_TGT	
	OLAP_AW_NAME	STR...	IN	'ORDERS_AW'	

Transition View

FROM	TO	CONDITION	DESCRIP...
WB_OLAP...	AND_ACTI...		
FORK_1	WB_OLAP...		

Graph Overview

85%

Data Movement – Using PFE

Process Editor: LOAD_FACT_OLAP [Read/Write]

Process Flow Edit View Window Help

Graph View Code View

Toolbox

Activity View

ACTIVITY	PARAMETER	DATATY...	DI...	VALUE	DESCRIPT
WB_OLAP_					
	OLAP_AW_OWNER	STRING	IN	WH_AW	
	OLAP_CUBE_AGGRE...	STRING	IN	SUM	
	OLAP_CUBE_NAME	STRING	IN	ORDERS	
	OLAP_CUBE_OWNER	STRING	IN	WH_TGT	
	OLAP_AW_NAME	STRING	IN	ORDERS_AW	

Transition View

FROM	TO	CONDITION	DESCRIP...
WB_OLAP...	EMAIL		
WH_ORD...	WB_OLAP...		

Add Remove

99%

Viewing the Results

Currently Access is provided through:

- BI Beans
 - The Java query components to enable OLAP
- Discoverer
 - The Ad-Hoc query tool now utilizing OLAP
 - Warehouse Builder is capable of generating Business Areas for Discoverer
 - Impact Analysis and Lineage is available from Discoverer through the ETL process

Viewing the Results – OLAP DML

The screenshot shows the Oracle OLAP Worksheet interface. The main window contains a query editor with the following SQL commands:

```
--> limit t_time_dim to 'YEAR.1845'  
--> limit sales_rep_dim to 'REGION.1'  
--> limit order_status_dim to 'STATUS.141'  
--> report orders_amount
```

Below the query editor, the dimensions are listed:

```
T_TIME_DIM.1.YEAR.1845  
SALES_REPS_DIM.REGION.1  
PRODUCTS_DIM.CATEGORY.19
```

The results are displayed in a table format:

ORDER_STATUS_D	CUSTOMER.1	CUSTOMER.1	CUSTOMER.1	CUSTOMER.1	CUSTOMER.1	CUSTOMER.1	CUSTOMER.1	CUSTOMER.1	CUSTOMER.1	CUSTOMER.1	CUSTOMER.1	CUSTOMER.1	CUSTOMER.1	CUSTOMER.1	CUSTOMER.1	
IM	COUNTRY.CA	COUNTRY.CH	COUNTRY.CN	COUNTRY.DE	COUNTRY.IN	COUNTRY.IT	COUNTRY.JP	COUNTRY.US	01	02	03	04	05	06	07	
STATUS.141	NA	NA	NA	NA	NA	NA	NA	NA	189,100.90	12,738.00	51,236.30	774.00	5,584.00	1,887.00	NA	4,270.30

The status bar at the bottom indicates "Ready DML Mode".

Generated
AW with
fully solved
cube in
OLAP
worksheet

Benefits

Reduced complexity

- User Interface reduces complexity of OLAP environment
- Deployment is made simple from Warehouse Builder
- ETL now covers the OLAP data

Reduced Development time

- One design environment saves:

- Time and effort in building

- Time and effort in training

- Easy to extend existing warehouse with OLAP (up sell possibility)

Reduced Maintenance

- One design environment means one management environment
- Part of the regular ETL flows and processes
- Part of the regular DBA activities

Agenda

- Oracle rdbms datawarehousing marketshare
- Oracle Data Warehousing Process Products and Tools
- Oracle VLDB Capabilities
- Query Performance and Optimization Capabilities



Oracle9i for Data Warehousing

Continuous Innovation

Oracle 7.3

- Hash Join
- Bitmap Index
- Parallel-Aware Optimizer
- Partition Views
- Instance Affinity Function
- Shipping
- Parallel Unique
- Asynchronous Read-Ahead
- Histograms
- Anti-Join

Oracle 8.0

- Partitioned Tables
- Partition Pruning
- Parallel Index
- Parallel Insert, Update, Delete
- Parallel Bitmap Index
- Parallel ANALYZE
- Parallel Constraints
- Server Management
- Point-in-Time Recovery

Oracle8i

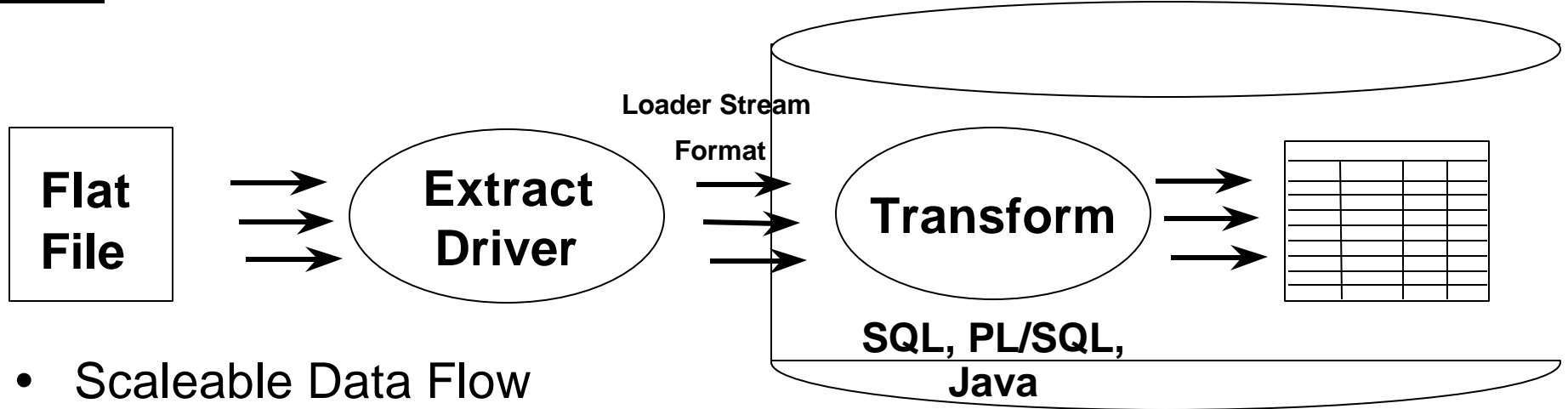
- Hash and Composite
- Resource Management
- Progress Monitoring
- Adaptive Parallelism
- Server-based Parallelism
- Materialized Views
- Transportable Tablespace
- Direct Loader
- Functional Index
- Partition-wise Operations
- Security Enhancements

Oracle9i

- List Partitioning
- Bitmap Join Index
- Dynamic Aggregation Buffer Size
- Materialized Intermediate Results
- Grouping Sets
- Concatenated Grouping Sets
- Aggregate Pruning
- New Analytic Functions
- Self-Tuning Execution Memory
- System Managed Undo
- Dynamic Resizing of Buffer Pool
- ETL Infrastructure
- and much more ...



ETL Infrastructure External Tables & Table Functions



- Scalable Data Flow
- Pipelined, Parallel Transformations
 - SQL, PL/SQL, Java (load, insert, update, delete)
- Intra-file Parallelism
 - Eliminates need for manual split of input file
- Currently does not support indexes
- Useful when entire file must be accessed for join or load



Oracle9i -External Table example:

```
CREATE TABLE sales_trxn_ext
(prod_id number(6),
cust_id number,
unit_cost number (10,2),
unit_price number (10,2),
time_id date
)
ORGANIZATION external
(
TYPE oracle_loader
DEFAULT DIRECTORY data_file_dir
ACCESS PARAMETERS
(
RECORDS DELIMITED BY NEWLINE CHARACTERSET US7ASCII
BADFILE log_file_dir: 'sh_sales.bad_xt'
LOGFILE log_file_dir: 'sh_sales.log_ext'
FIELDS TERMINATE BY "|" LDTRIM
)
location
(
'sh_sales.dat'
)
) REJECT LIMIT UNLIMITED;
```

```
INSERT /*+ APPEND */ INTO COSTS
( prod_id,
time_id,
unit_cost,
unit_price
)
SELECT
prod_id,
time_id,
sum(unit_cost),
sum(unit_price)
FROM sales_trxn_ext
GROUP BY time_id, prod_id;
```

Oracle9i - ETL Infrastructure

Key Benefits

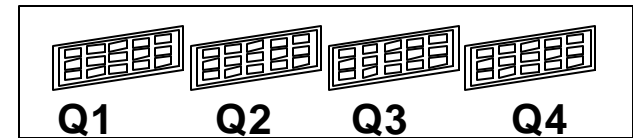
- Increased Load Performance and Scalability
 - Pipelining: Fewer steps/stages
 - Increased Parallelism
 - More powerful transformations
- Improved Manageability
 - Single solution for Data and ETL management
- Reduced Costs
 - Fewer tools to buy and integrate
 - Leverage existing SQL, PL/SQL and Java skills

Oracle8i Partitioning Methods

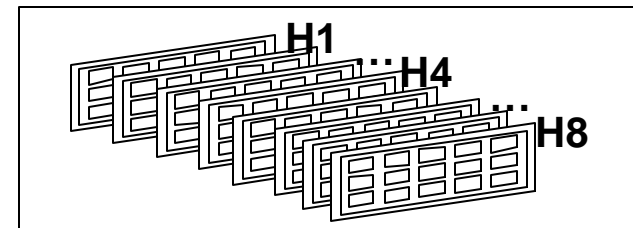
Range, hash and composite partitioning

- Hash partitioning distributes data evenly
 - Improves parallelism
 - Easy striping of data to disk
- Composite partitioning: range + hash
 - Admin and availability benefits of range
 - Parallel benefits of hash
- Parallel maintenance ops support “rolling window”

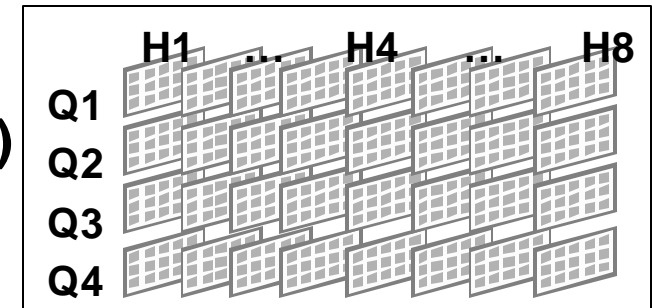
**RANGE
(ORDERS)**



**HASH
(PRODUCTS)**

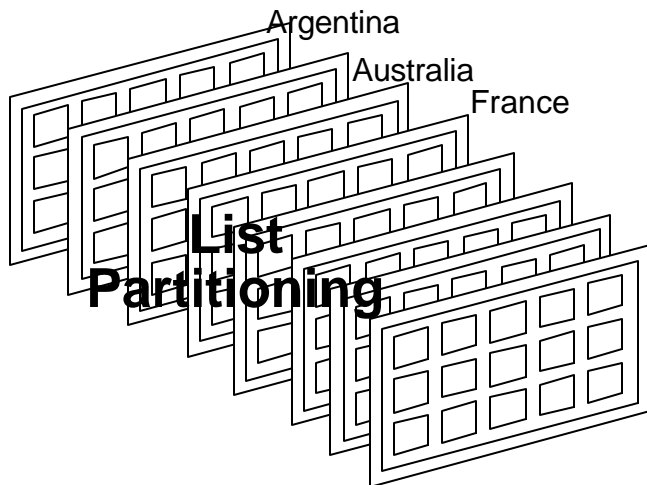


**COMPOSITE
(ORDER ITEMS)**



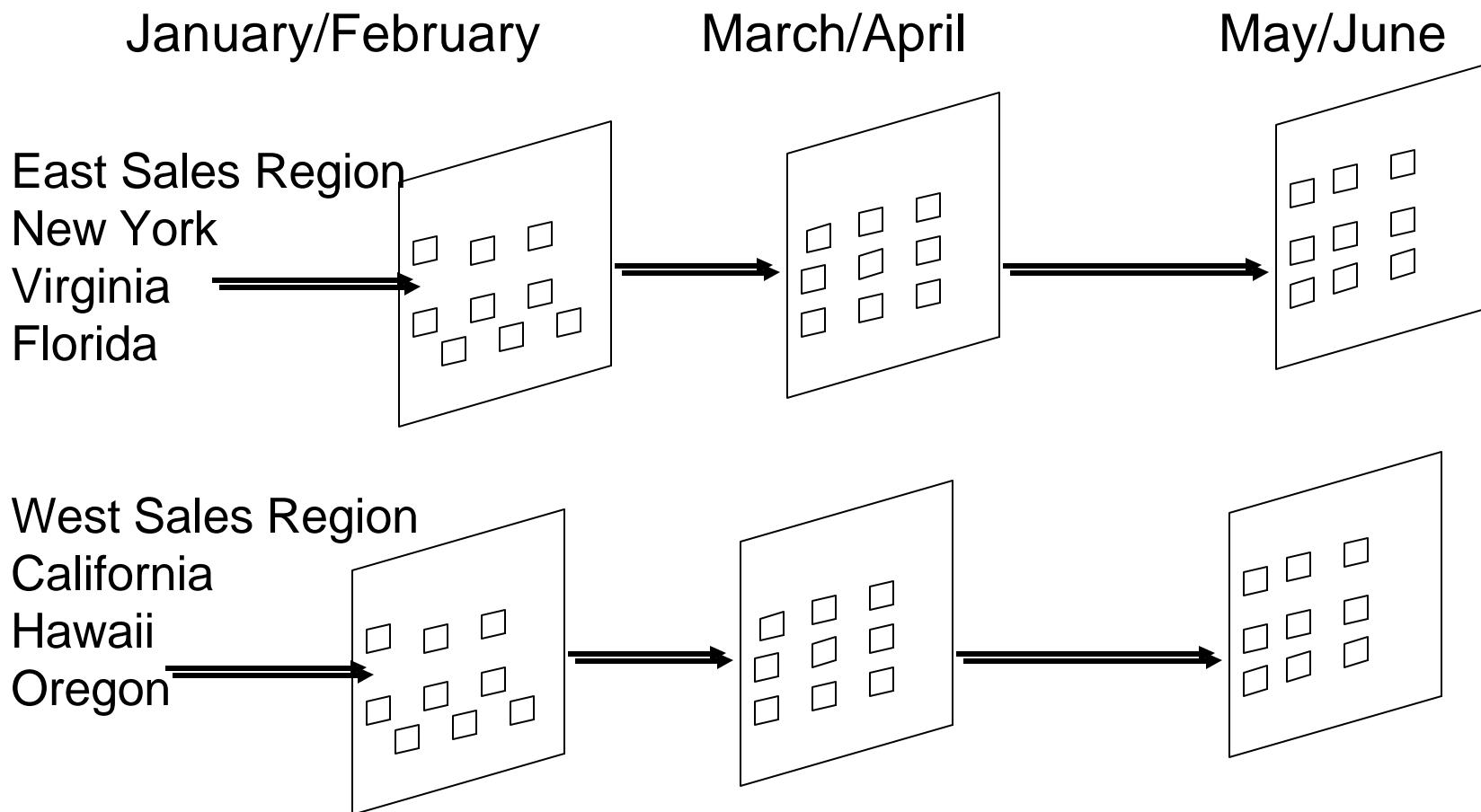
- Physically organizes a table into smaller units for manageability & availability.
- Partition Elimination and partition aware optimization speeds queries.
- Partition level administration improves manageability (export, recover, merge)
- Partitions support in parallel query, truncate, & indexing.

9i - List Partitioning



- Allows to logically designate partitions in-line with business behavior
- e.g. by product groups, geographical regions, by department names etc...
Benefit of list partitioning: More precise control of distributing data among partitions
- Allows a DBA to enumerate the partition-key values for each partition
- Useful for partitioning over discrete domains (e.g. geography, product category)

9i Rel2: Range-List Partitioning

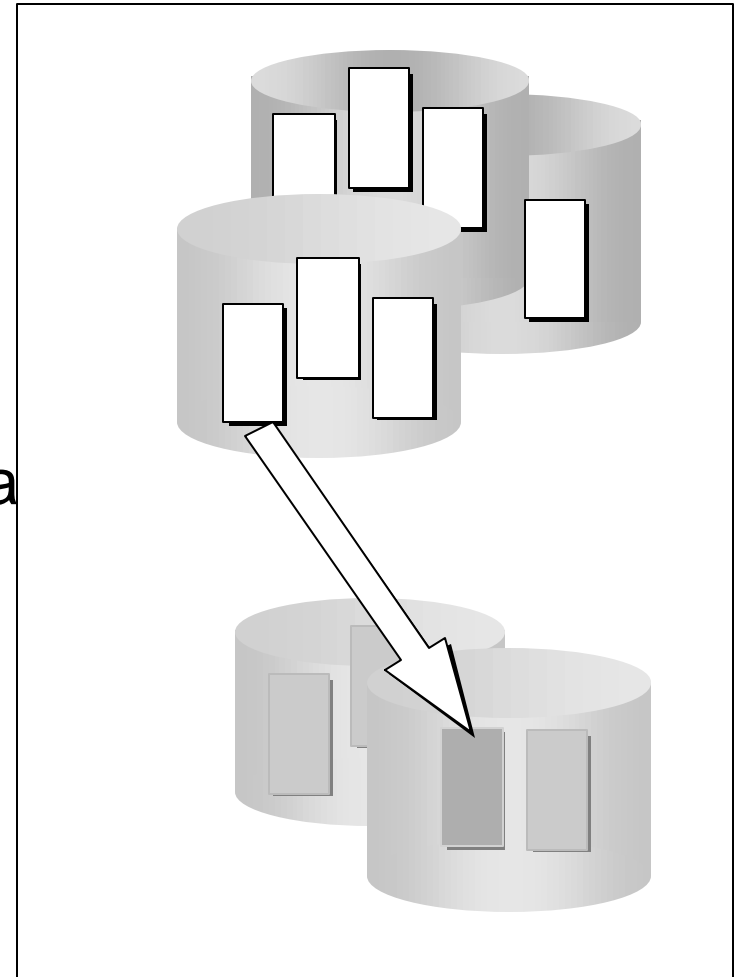


Range, sub-partition by List.

Transportable Tablespaces

- Copy database subsets (tablespaces) between databases
 - Operating system file copy for data
 - Managed transfer of meta data between databases
 - “Plug in and go”

**Result: extremely fast
bulk data transport
between databases**



Analytic Functions

Oracle8i - Release 2

- Ranking functions
 - rank, dense_rank, cume_dist, percent_rank, ntile
- Window Aggregate functions (moving and cume)
 - avg, sum, min, max, count, variance, stddev, first_value, last_value
- LAG/LEAD functions
 - Direct inter-row reference using offsets
- Reporting Aggregate functions
 - sum, avg, min, max, variance, stddev, count, ratio_to_report
- Statistical Aggregates (normal and window)
 - correlation, linear regression, covariance
- These types of queries performed without the OLAP option

Analysis Ready RDBMS Analytic Enhancements

- Oracle8i, Release 2
 - Rank, percentile
 - Window (moving average, cumulative sum)
 - LAG/LEAD
 - Ratio-to-Report
 - Statistical functions (linear regression, correlation)
- Oracle9i
 - Inverse Percentile
 - Hypothetical Rank and Distribution functions
 - Histogram Function
 - First/Last Values

Agenda

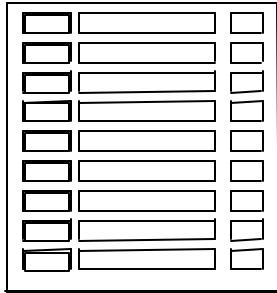
- Oracle rdbms datawarehousing marketshare
- Oracle Data Warehousing Process Products and Tools
- Oracle VLDB Capabilities
- Query Performance and Optimization Capabilities



8i Supported Index Types

Index Organized Tables

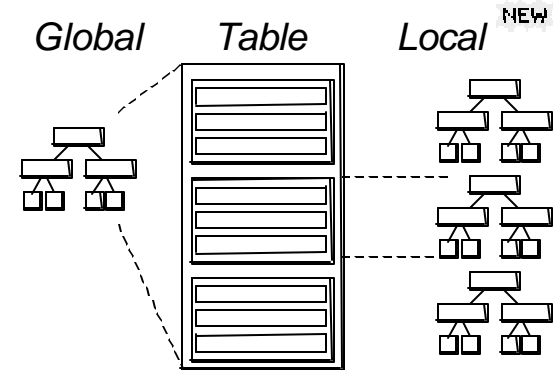
Hash



Bitmapmed Index

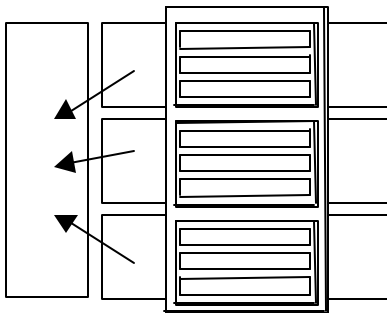
Column	Bitmap
<input type="checkbox"/>	01010 00010 110001
<input type="checkbox"/>	11000 11101 000100
<input type="checkbox"/>	01010 00010 110001
<input type="checkbox"/>	01010 01010 010101
<input type="checkbox"/>	11100 00101 000101

Partitioned Indexes

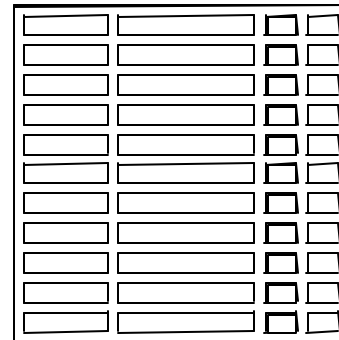


Parallelization

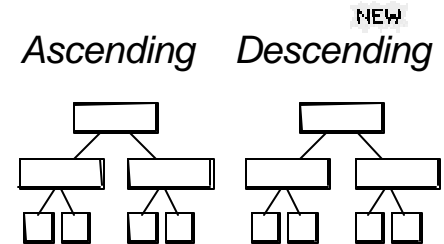
Index



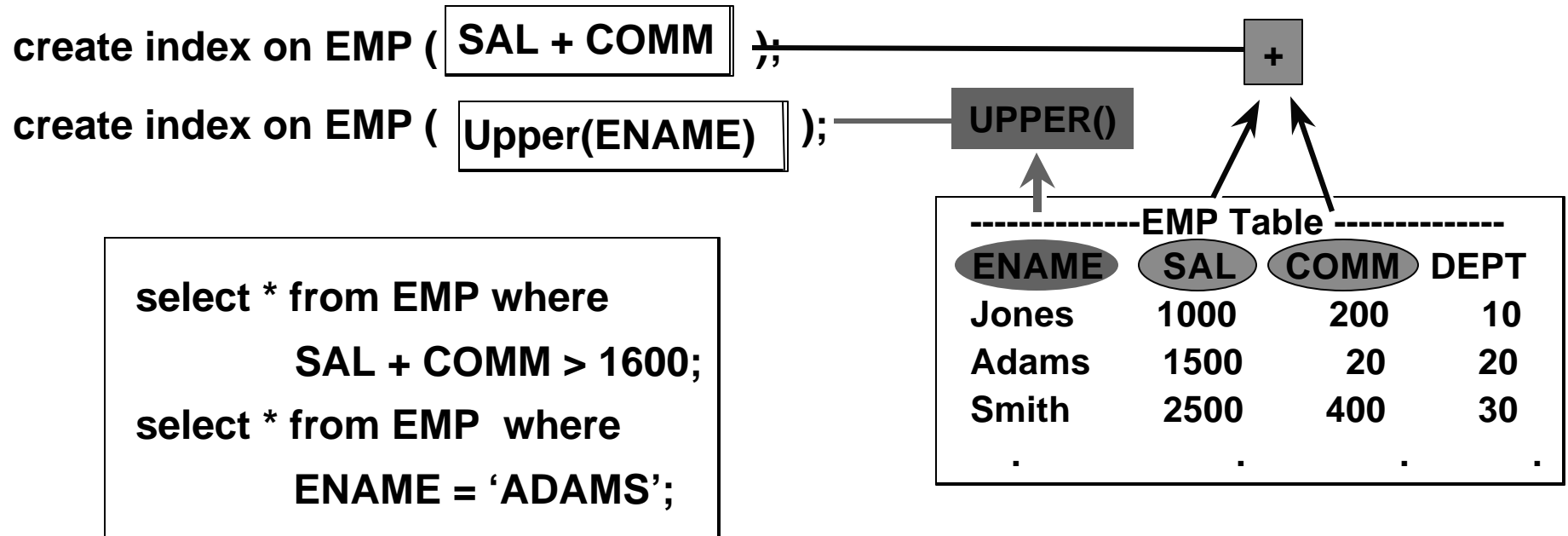
Function Based ^{NEW} $f()$



B-Tree



Function Based Indexes

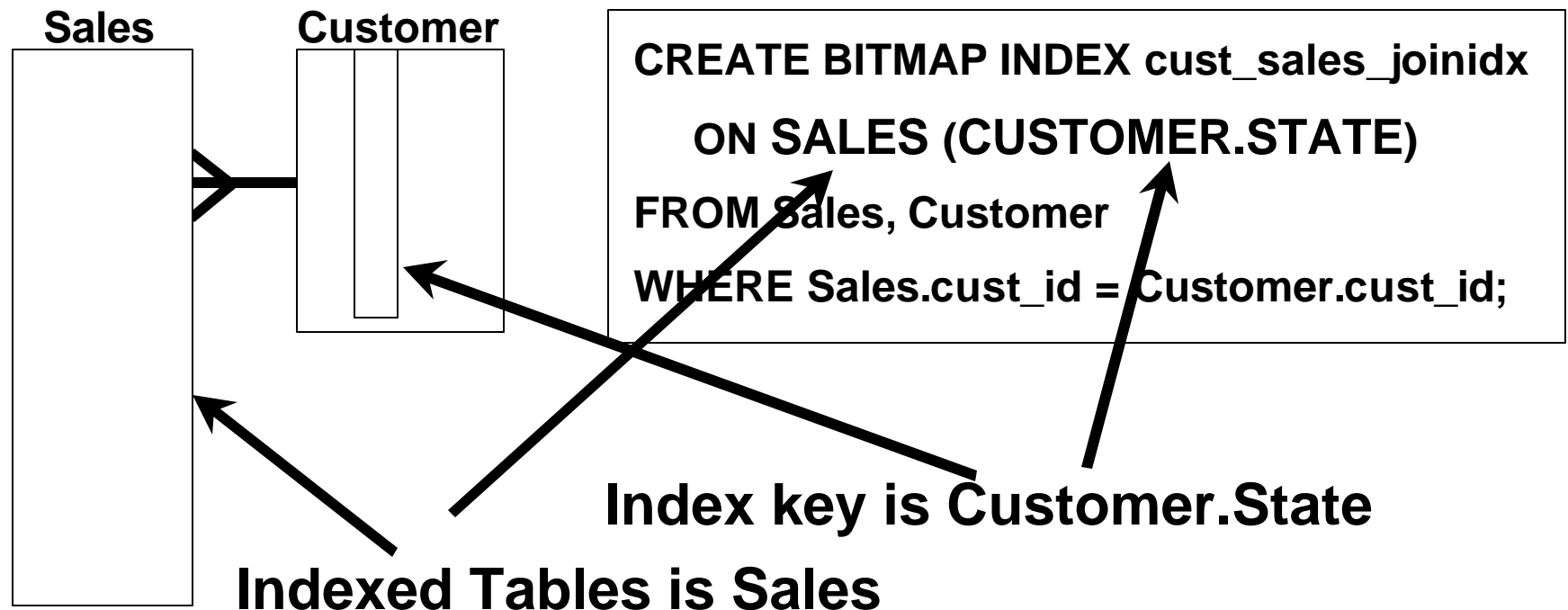


Indexed access replaces full table scan

Fast access to data based on an expression, built-in function or user-defined function

Bitmap Join Indexes

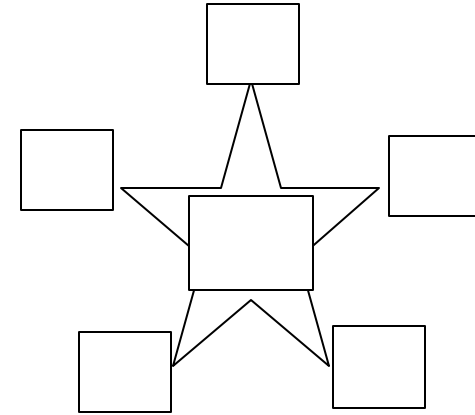
Speed Joins on Two or More Tables



- Especially good for large dimension tables in star & snowflake schemas
- Up to 30x improvement in internal tests; very space-efficient

Join Types

Parallel Bitmap Star Joins



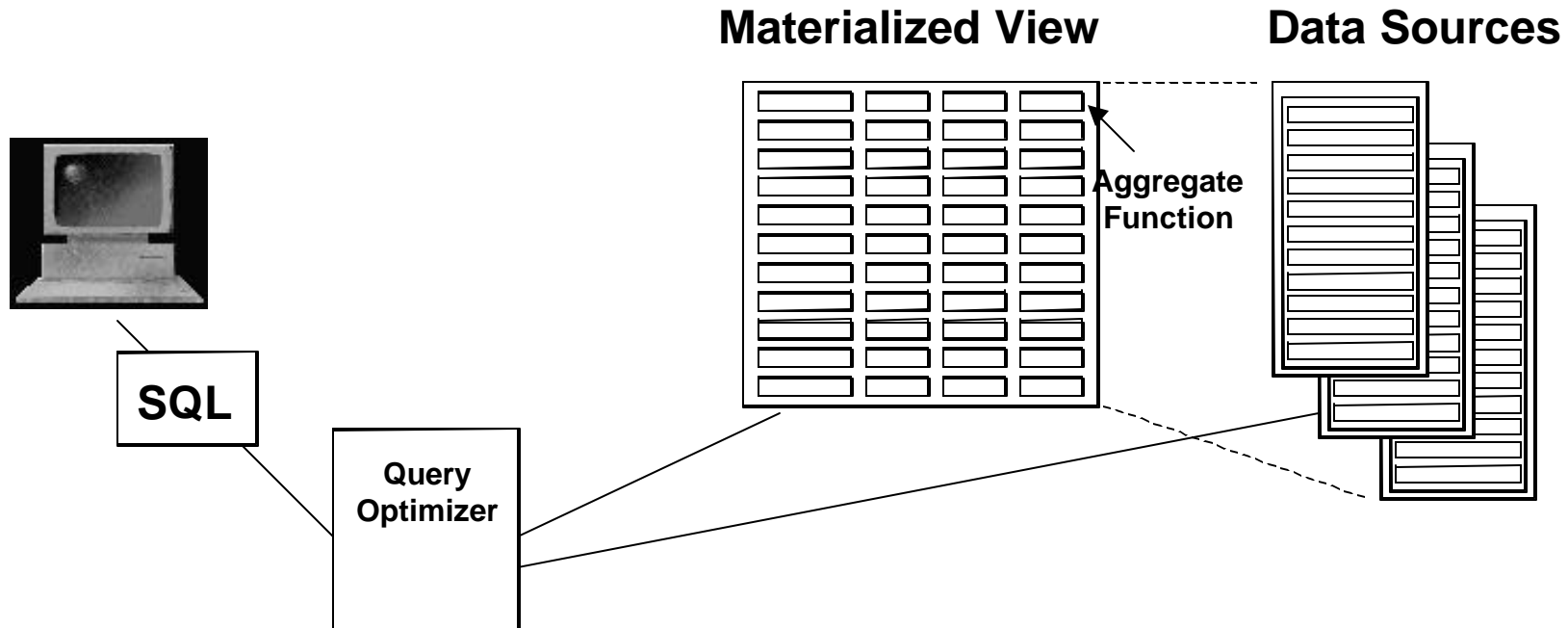
Joins

- Nested-loop join
‘typical’ relational join
- Sort-merge join
- Hash join (7.3)
- Parallel Bitmap Star Join (8.0)
- Hash Anti-Joins

Dramatic Performance Gains

- **Innovative use of bitmap indexes**
- **Complex Star Schemas**
 - multiple fact tables
 - many dimensions
 - unconstrained dimensions
 - large dimension tables
- **Optimized for sparse fact tables**
- **Parallel execution**
- **Query Transformation**

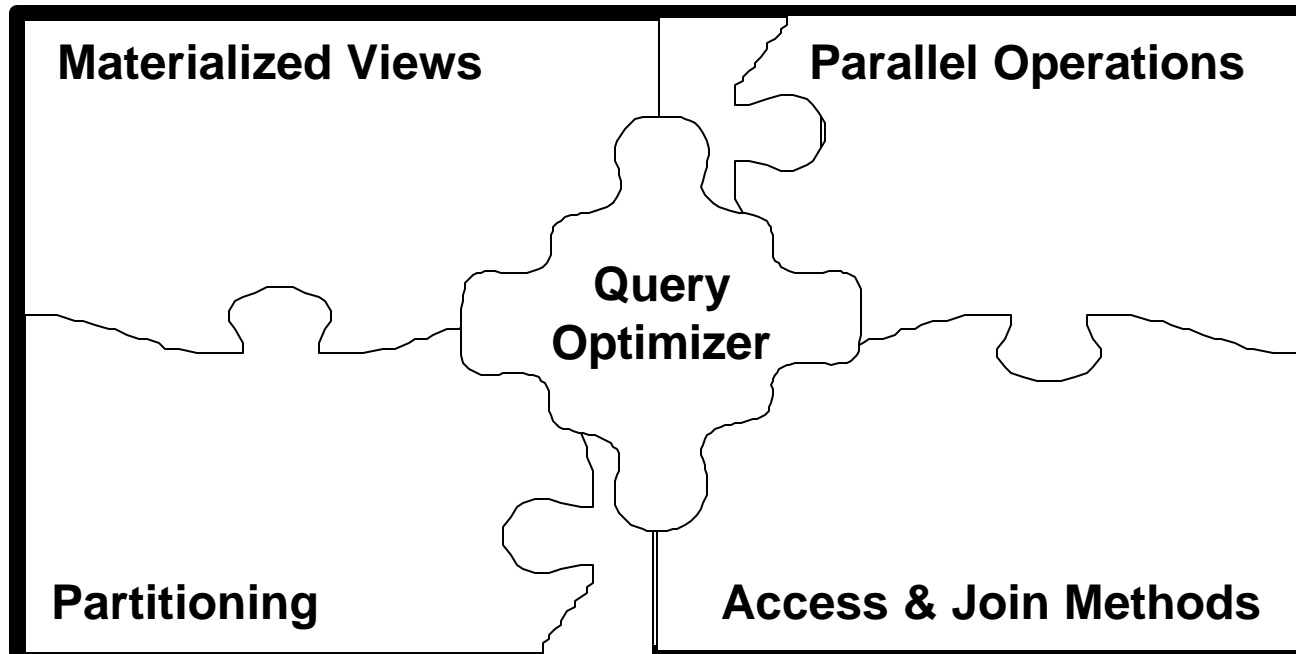
Materialized Views



- Pre-compute “materialized” queries to speed performance
- May Include Aggregates, Joins and Unions
- Query rewrite enables transparent use
- DBMS_OLAP package displays actual use
- Refresh by time, on demand, or in real-time

Fast Query Performance

- The best approach for every query
 - Integrated
 - Comprehensive



Oracle Data Warehousing & Business Intelligence Solution

Analysis Tools & Applications

Discoverer, Express, Business Intelligence Solutions,
Clickstream Intelligence, Personalization, Partners

Design & Development Tools

Warehouse Builder, Designer, JDeveloper, Reports, Portal, Partners

Extract, Transform, Load

Oracle Warehouse Builder, Partners

Database & Server Technology

Oracle8i & Oracle9i, Express Server

**M
e
t
a
d
a
t
a**

CWM
CWMI

**M
a
n
a
g
e
m
e
n
t**

OEM
Packs
Partners

**S
e
r
v
i
c
e
s**

Consult
Educate
Partners

ORACLE

QUESTIONS
ANSWERS

A large, bold, black graphic of the letters 'QA' is centered on the page. The 'Q' is a thick, rounded letter with a small tail that loops back into the 'A'. The 'A' is a simple, blocky letter with a wide base. The 'QA' graphic is superimposed over the text 'QUESTIONS' and 'ANSWERS'.