

# JBoss Enterprise Data Services Platform in the Enterprise

Mani Subramanyam, Genentech  
Michael Walker, Red Hat

# Agenda

Introduction

About Genentech

Challenges

Integration Services

DSP Selection Criteria

EII Architecture at Genentech

Use Cases

- Single view of Employee

- BI/OI Reporting against Cloud

Future Work

Q&A

# About Genentech

Leading bio-technology company, founded 1976

Wholly-owned subsidiary of Roche as of March 2009

Develops drugs for significant, unmet medical needs

Oncology

Immunology

Tissue growth and repair

Merger with Roche brings 31 new products to portfolio

Annual revenue of >\$11B

40% annual revenue growth

>10,000 employees

Fortune “100 Best Companies to Work for” for the 9<sup>th</sup> consecutive year

# Challenges

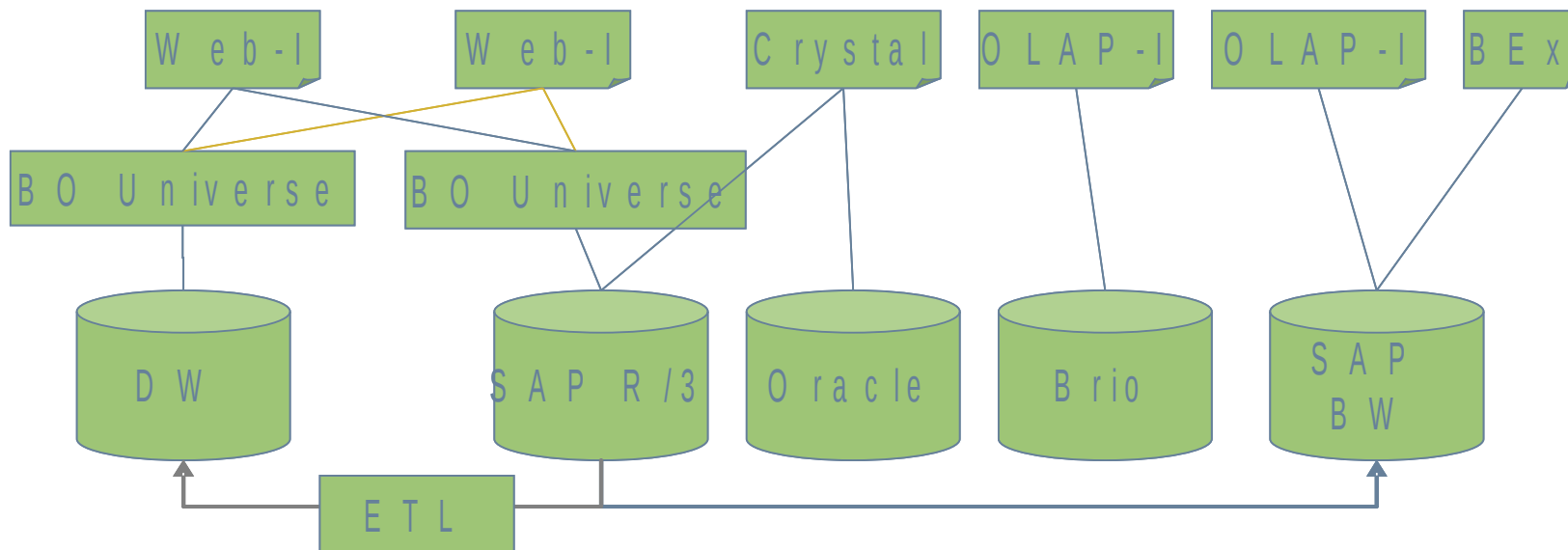
Rapid growth - Annual revenue growth @ ~ 40%

Large number of business software applications and enterprise data sources

No single system of record

Roche integration

Ad-hoc and point-to-point application integrations were implemented to meet short term requirements



# IT Strategy: Integration Services

Cross-functional group that drives integration technology adoption and reuse using:

- Enterprise Information Integration (EII)

- Enterprise Service Bus (ESB)

- Business-to-Business (B2B)

- Extract-Transform-Load (ETL)

- Enterprise Data Mashup (EDM)

SMEs from each field drive decision-making

Leverage existing technology investments and past successes

# Enterprise Information Integration via MetaMatrix Data Services Platform

Provides a single point of access to combine and query data from one or more data sources in real time.

Speeds application development by simplifying access to distributed data.

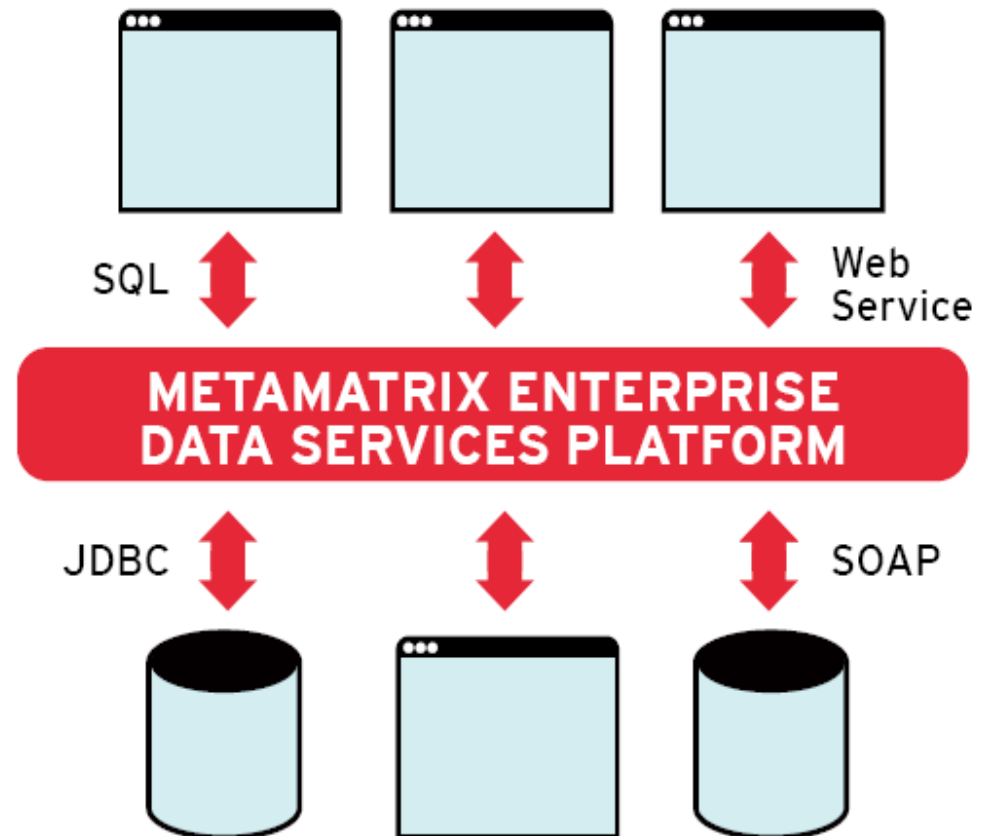
Transforms data structure and semantics through data virtualization.

Consolidates data into a “single view” without the need for more data.

Centralized access control, auditing through robust security infrastructure.

Creates services that provision data to business process in an SOA.

Enterprise-proven – flexible, scalable, high-performance.



# Business Benefits of EII

Analytical Reporting across multiple data sources

Integrated view from heterogeneous sources

Combine and use data in various formats

- XML and relational

  - Application and relational

  - Expose to users as relational, or web services

  - Flat Files

## Application Retirement

- Insulate Commercial reports from location of data in Cloud

- Insulate applications from location of Employee data in Locator or LDAP

Reduced development and maintenance of mapping rules

# Evaluation Criteria

## Architectural Fit

Can the vendor retrieve data from majority of our sources

## Data Security

Interface with LDAP for authentication and role management

Supports data entitlements authorization at table/column/row granularity

## Maintainability

Repository versioning and concurrent access control

Ease of use

Ability to provide data dictionary like descriptions to end users

## Performance

Minimal overhead over direct access to data sources

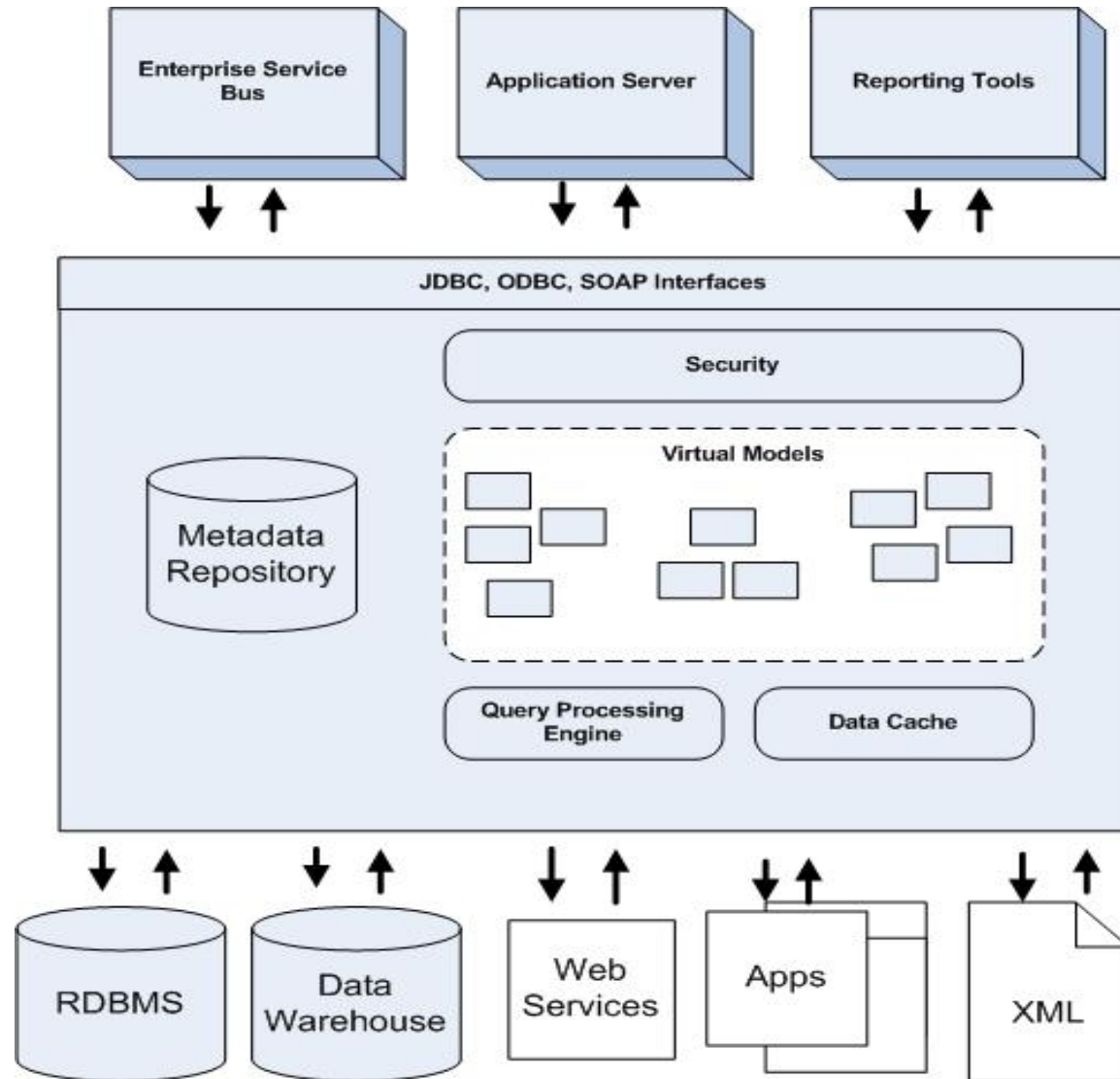
Tuning and administrative capability

Data caching

# ICC Technology Decision Matrix

ESB	ETL	B2B
<ul style="list-style-type: none"><li>• Real-time integration using message bus</li><li>• Publish/Subscribe messaging model</li><li>• Batch integration with small data packets</li><li>• Used for data movement in more than pub/sub</li></ul>	<ul style="list-style-type: none"><li>• Operational data store extraction to analytics repository</li><li>• is more than for operational data stores</li></ul>	<ul style="list-style-type: none"><li>• External partner integration</li></ul>
EII	EDM	
<ul style="list-style-type: none"><li>• Real-time data access</li><li>• Virtual data aggregation from multiple systems</li><li>• Synchronous request/response</li><li>• Access to external sources</li><li>• SQL or Web Services client interface</li></ul>	<ul style="list-style-type: none"><li>• Web Automation: Data extractions from websites to optimize B2B relationships</li><li>• Index and Search: powerful crawling, indexing and search of unstructured contents</li><li>• Unstructured content data source (documents, email...)</li></ul>	

# Data Services Platform: Where It Fits

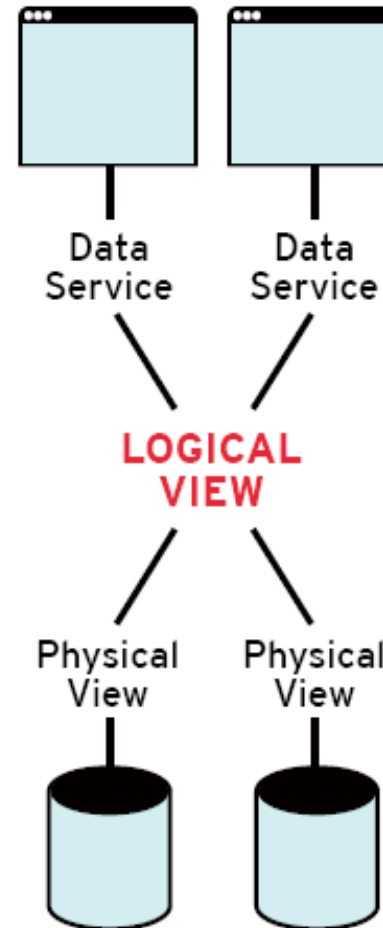


# How It Works

Create logical views of physical data sources using GUI

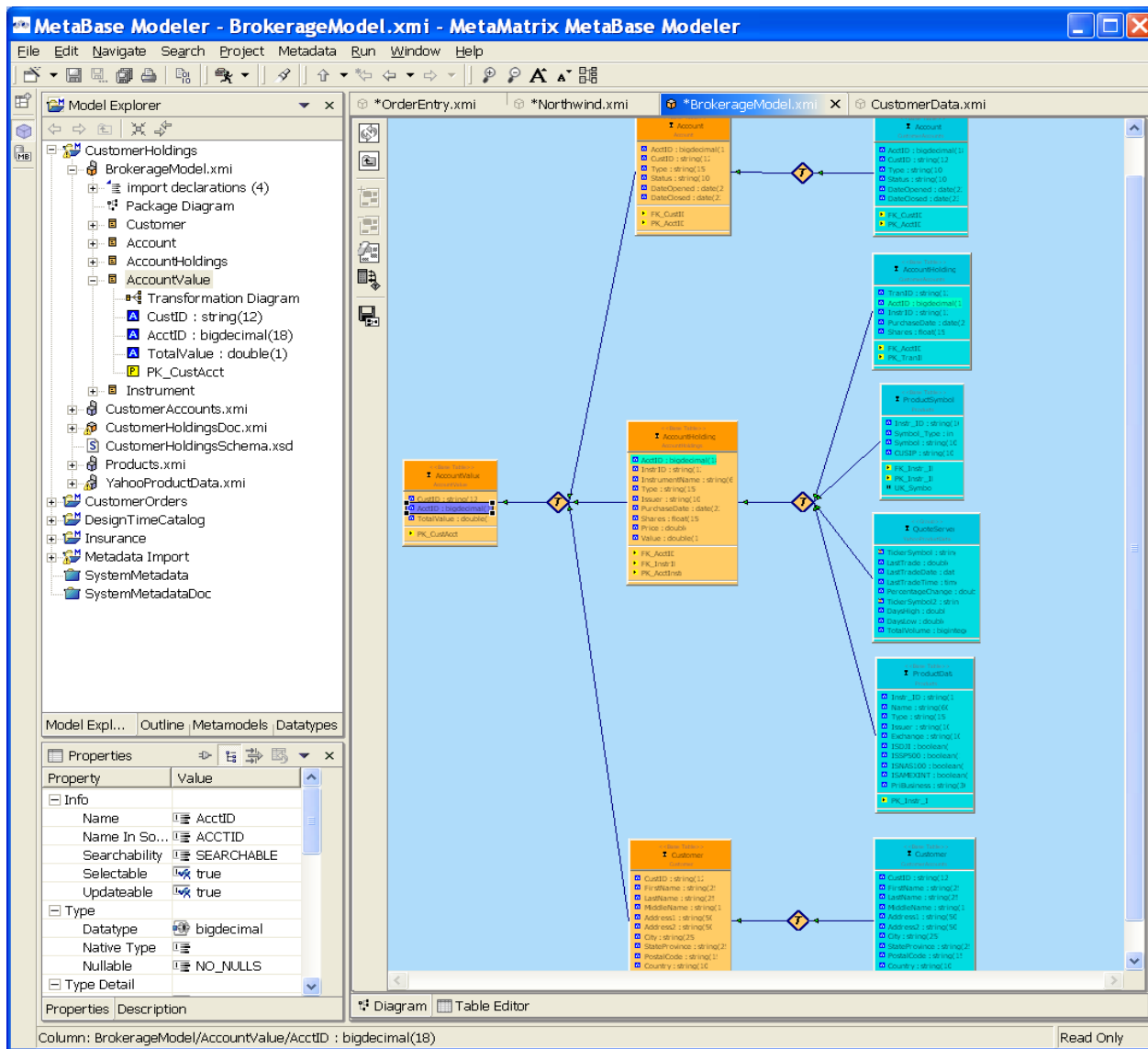
Application-specific services are defined, and mapped to logical view

No programming required



**MetaMatrix Data Abstraction.** Starting with physical views of existing data sources, create logical views and application-specific views without programming.

# Enterprise Designer: How It Works



Defined by Models,  
Not Code

Transformations

Select

Join/Aggregate

Filter

Functions

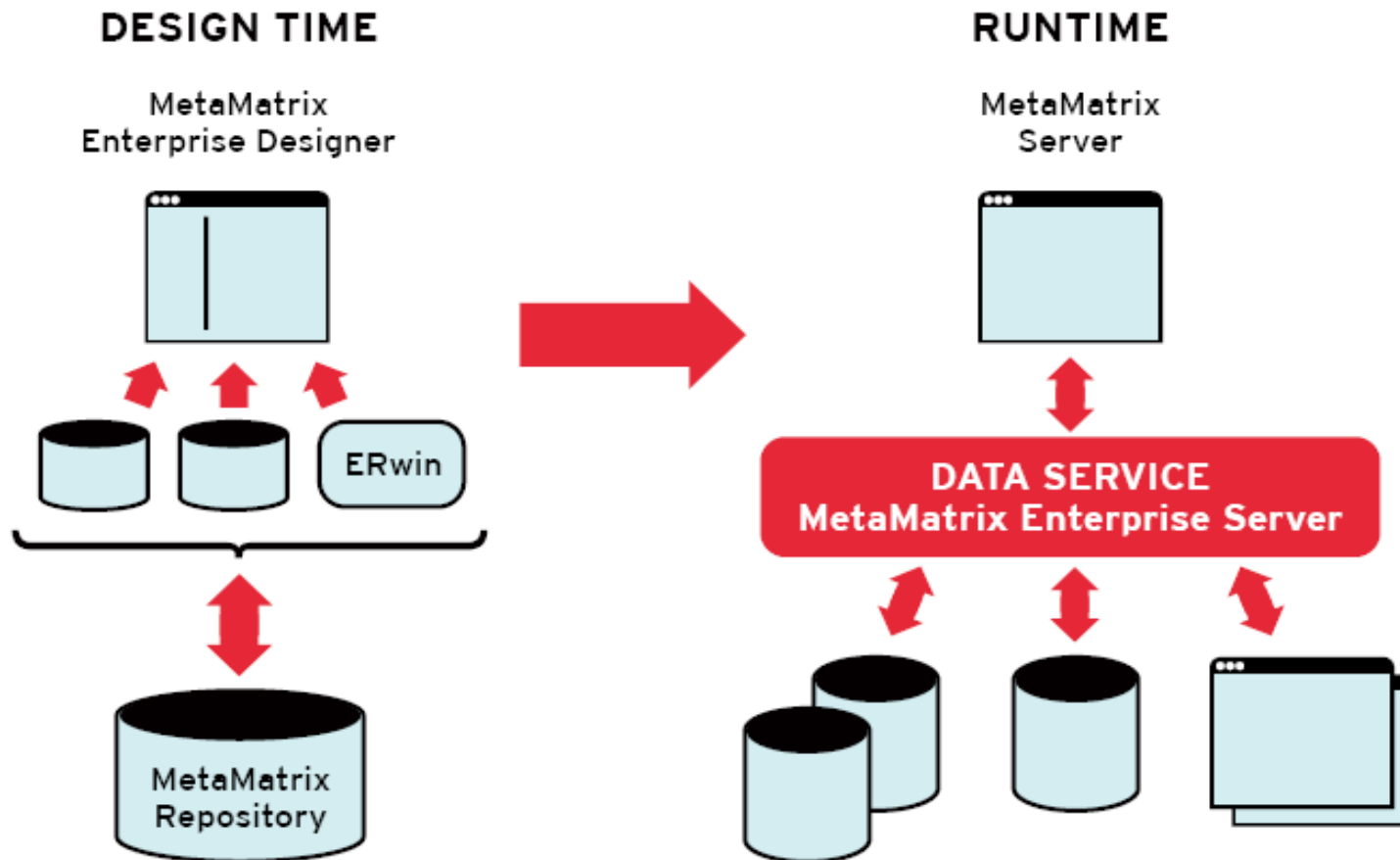
Text/String

Numeric

Decode

User-Defined

# From Design Time to Runtime



MetaMatrix Enterprise Components. Use MetaMatrix Enterprise to create, deploy, execute, and manage data services.

# MetaMatrix DSP Physical Architecture

6 MMX Environments

SND, DEV, SQA, UAT, TRN,  
PRD

2 x 2-CPU Virtual Machines in  
Cluster

Failover and High Availability

Disaster Recovery

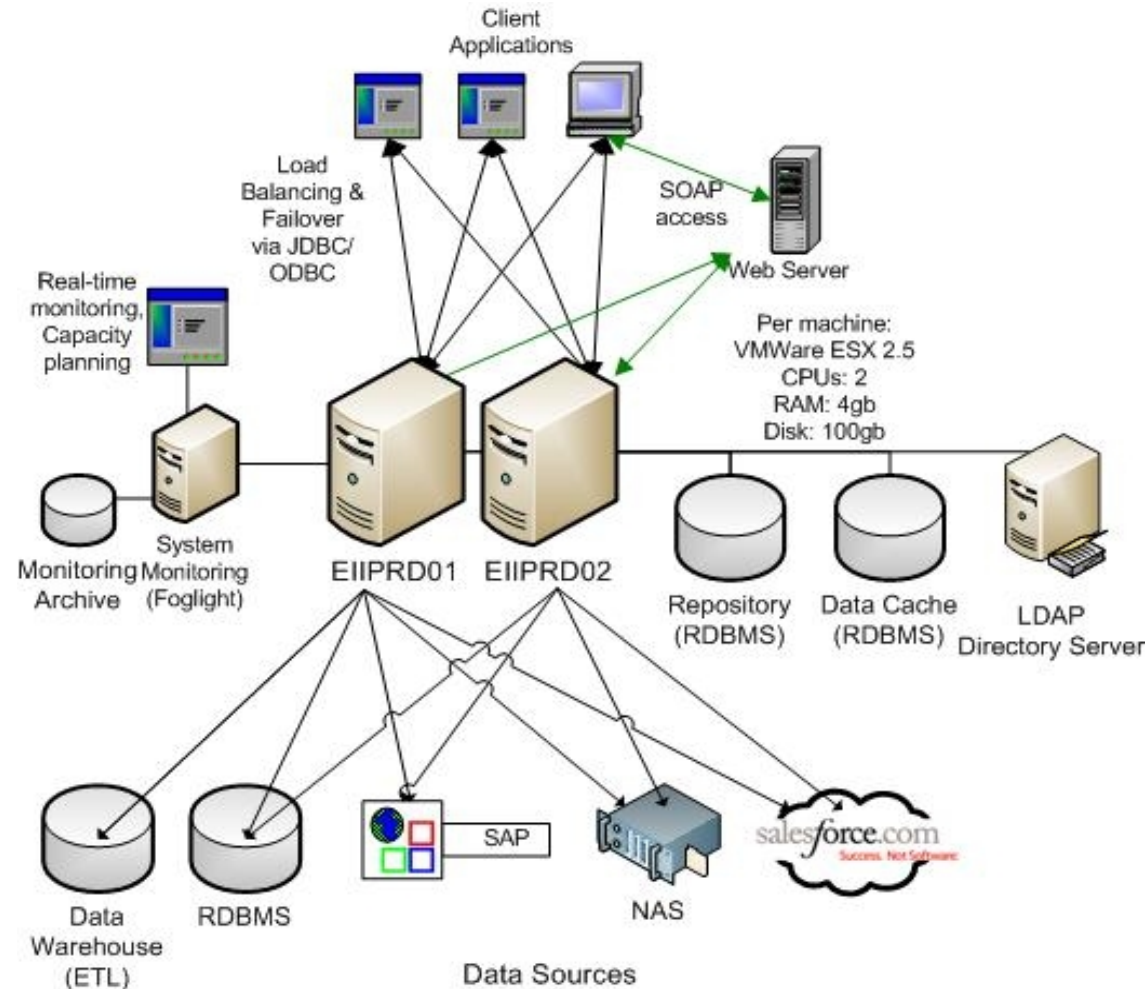
Uniform Security Infrastructure

LDAP Authentication

Data Roles

Proactive/Reactive Monitoring

Capacity and Regression testing



# MetaMatrix DSP Use Cases at Genentech

# MetaMatrix DSP Use Cases at Genentech

## Master Data Management

- 360° View of Employee (GI)

- 360° View of Customers

## Reporting and Analytics

- Operational Intelligence Reporting (Commercial)

- Business Intelligence Reporting (Commercial)

## Security Integration

- Secure Access Layer to Data Warehouses (Commercial)

# MetaMatrix DSP Data Sources at Genentech

## Relational Databases

Data Warehouses

Operational Data Stores

Oracle, SQLServer, PeopleSoft, FileMaker Pro

LAN and WAN

LDAP Directory Server

Salesforce.com Cloud Platform

OSISoft PI System Historian DB

Flat Files

# MetaMatrix DSP Clients at Genentech

## Web Applications

- MetaMatrix Hibernate Dialect

- Perl/CGI scripts

## ESB

- iPhone

## OI/BI Reporting

- Business Objects

- Crystal Reports

## Analytics Tools

- Spotfire

- BRIO

# Use Case: gPeople

Oracle database containing outdated employee information at Genentech

Enterprise LDAP is the actual system of record for employee information

Real-time data access requirement

Many applications use the outdated Locator database (total # unknown)

Audit requirement

Additional data source requirements:

Cost Center Hierarchy (SAP)

Manager Hierarchy

External Employee LDAP branch

LDAP Change History

# gPeople Project Goals

Shut down legacy DB

Shift many applications to Enterprise LDAP as source for employee information

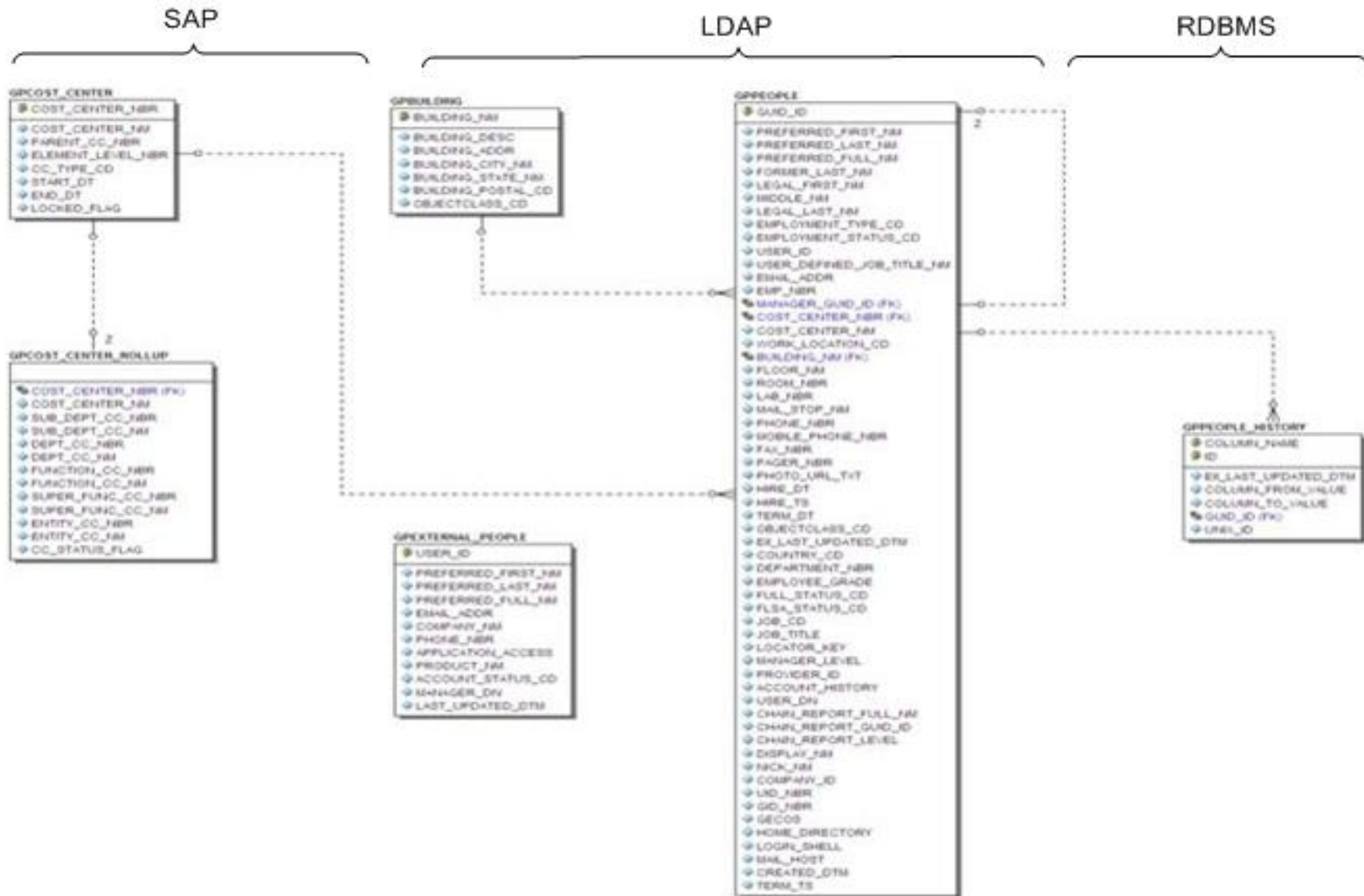
Build a reusable, SQL-based interface to LDAP-based employee data

Combine LDAP data with additional sources

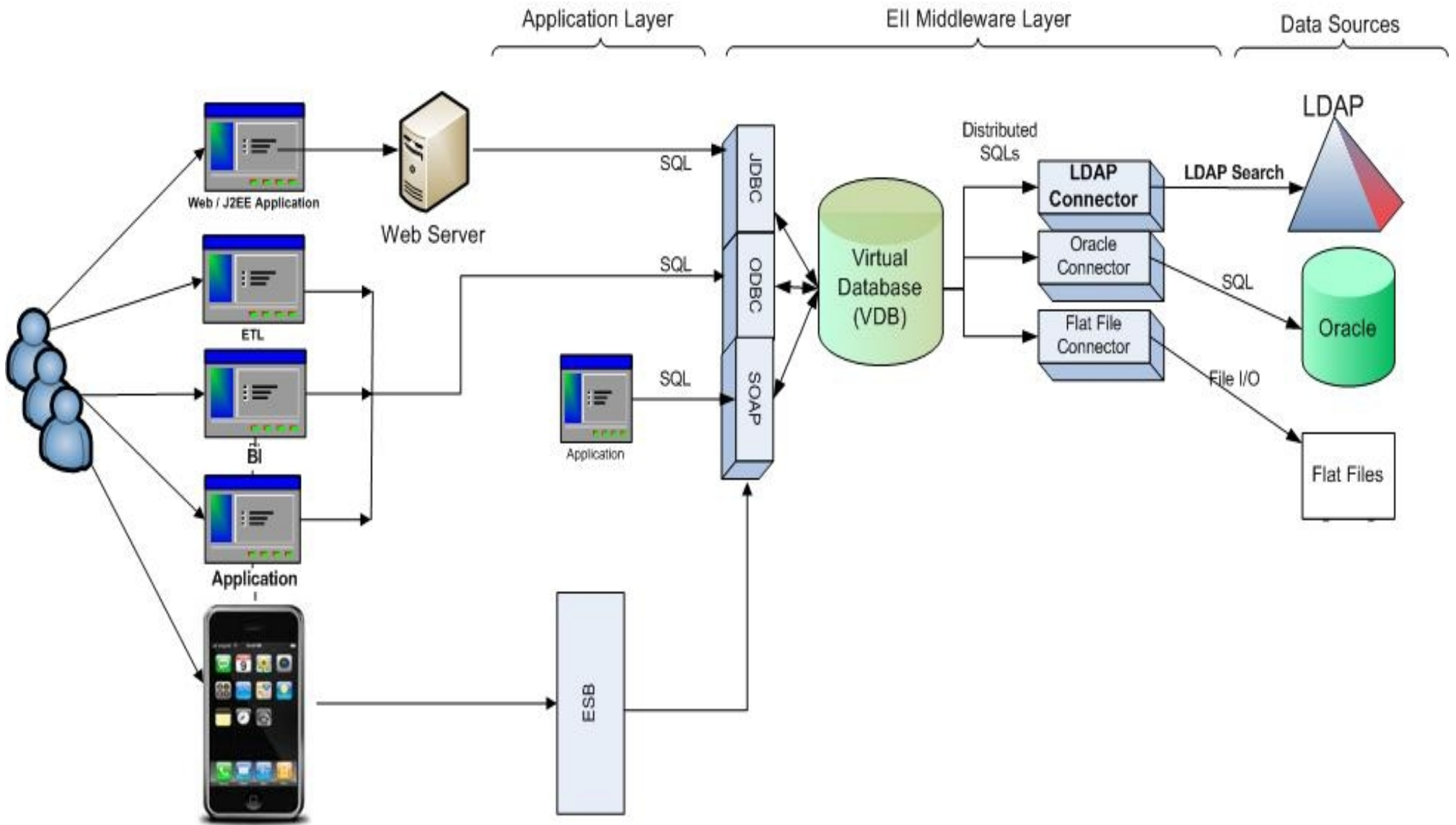
Define reusable Data Access policies

Track history and support for delta load for clients

# gPeople Data Models



# gPeople Physical Architecture



# gPeople in Production

25+ applications using gPeople VDB

- Extensive re-use

- Enhanced as downstream demands increase

## Multiple virtual views

- Public view for more commonly used elements

- Private view is for HR restricted access elements

- More granular security for Private views

- Application specific views

## Support real-time and cached access

- Cache Views to reduce the load on LDAP server

Monitored through Enterprise Foglight monitoring tool

Referenced in other VDBs

# gPeople: Lessons Learned

Best use case for providing SQL interface to non-SQL data

Largest number of clients compared to any other VDBs

Tracking History of changes

Should have modeled all attributes at once, instead of per project

Better approach on defining data access roles upfront

Limitation of using Cluster on some integration technology  
specifically ESB

# gPeople: Future Work

Multi-valued attribute support to extract

Consolidate security model

Publish to JMS Queue

Roche Integration

- Multi-source binding

- Active Directory data source

Caching LDAP Groups

Data Cleansing of gPeople data

gPeople History enhancements

# Use Case: BI/OI Reporting Against the Cloud

Salesforce Cloud Platform migration

Host Enterprise Patient, Case, Sales data

Move off legacy systems, data warehouses

Reduce cost, maintenance

Business Units track various metrics, business processes

200+ reports

Metrics, Brand Scorecard, Operational information

Complex aggregations and process tracking

Ad-hoc analytics tools

# Cloud Reporting: Challenges

SFDC reporting does not meet business need

- Limited support for aggregation, joins, pre-calculated data

- Basic reporting and formatting only

Information must be organized, secured, and filtered using private data in real-time

- Not available in the cloud:

  - Role Management (authorization)

  - Territory alignment data (filtering)

  - LDAP (authentication)

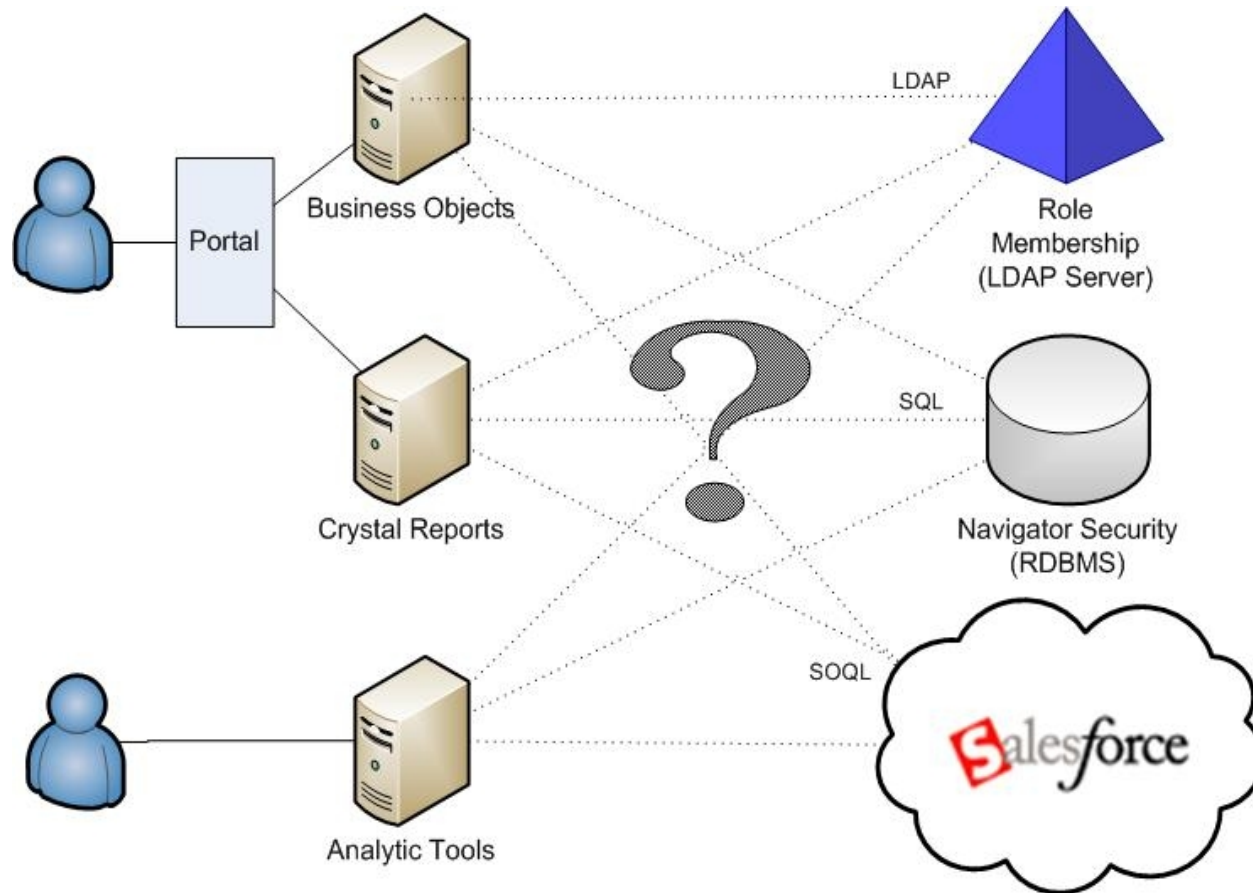
Metrics must be available immediately at system/product launch

- Data warehouse/ODS initial overhead too high

Information must be encrypted

- HIPAA-compliance

# BI/OI Reporting: Challenges

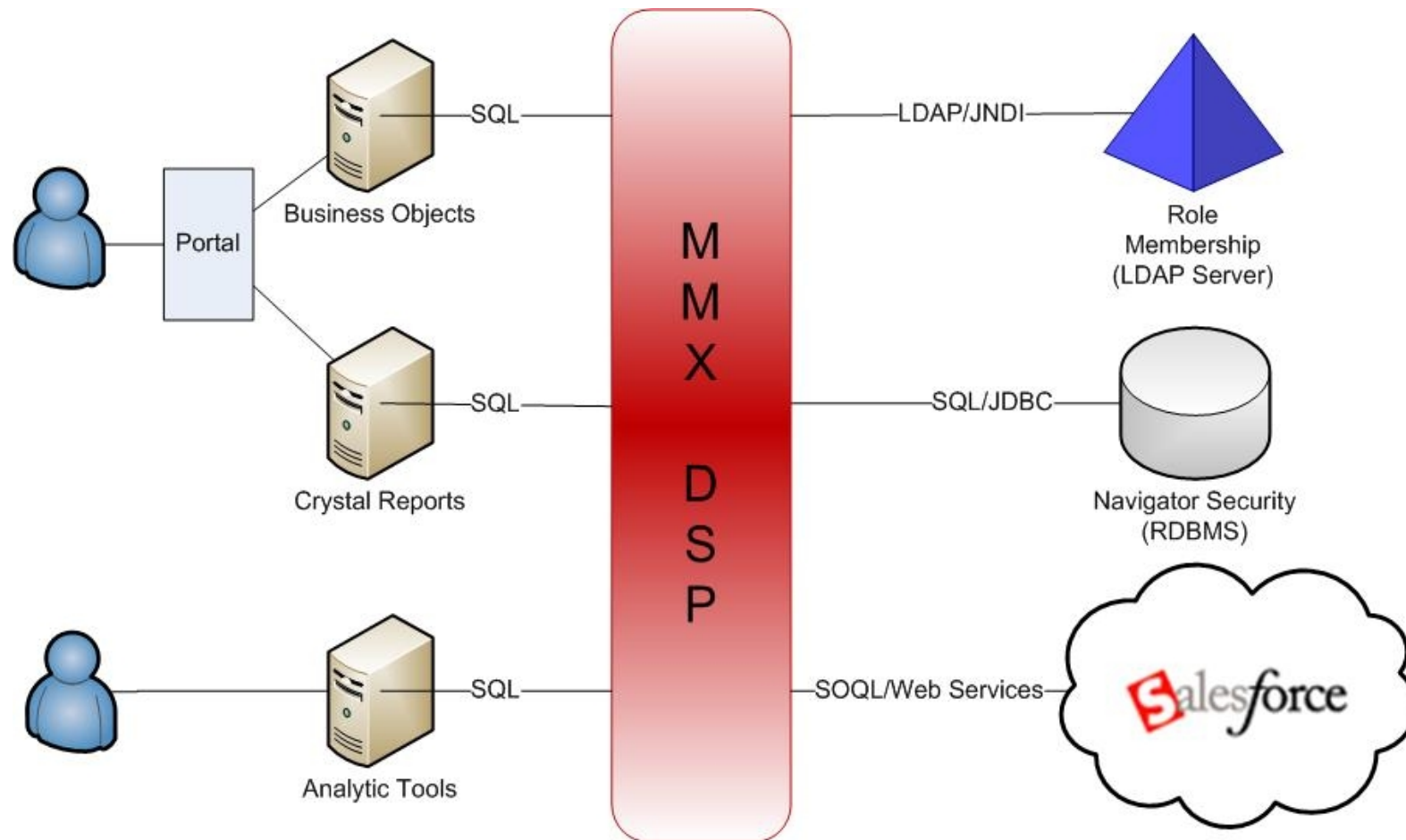


Need to securely integrate public cloud data with multiple private systems

Uniform, reusable SQL interface needed

Need blend of real-time and pre-computed data

# BI/OI Reporting Physical Architecture



Provide uniform, reusable interface to heterogeneous data sources

Mask underlying interfaces to source systems

Combine real-time access with cached aggregate data for reporting

# Salesforce Connector

Creates Relational Models

Imports SFDC object metadata

Tabular representation

Translates SQL to SOQL

Masks underlying XML/Web Services interface

Supports DML

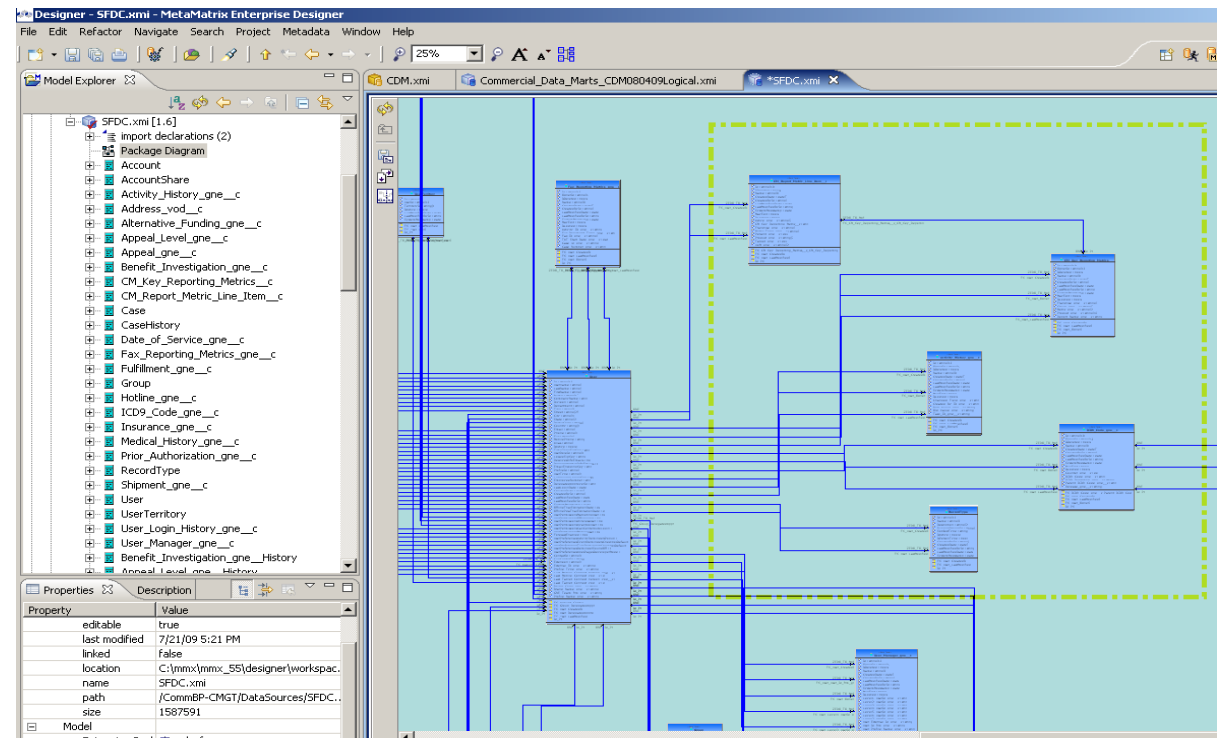
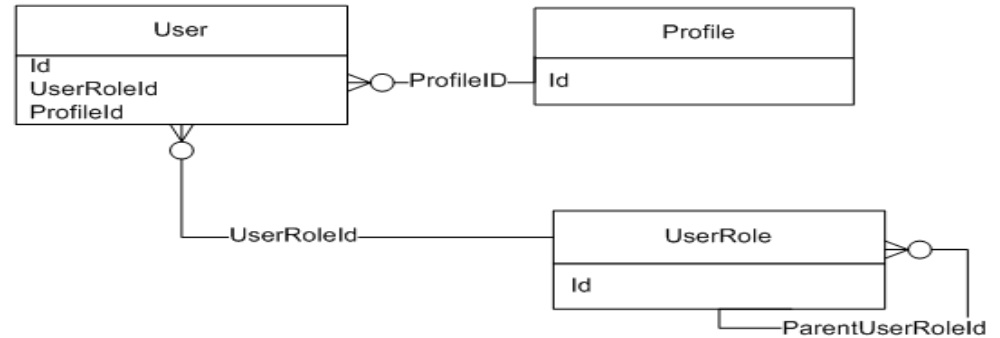
Insert, Update, Delete

Supports SFDC-specific functions in SQL statements

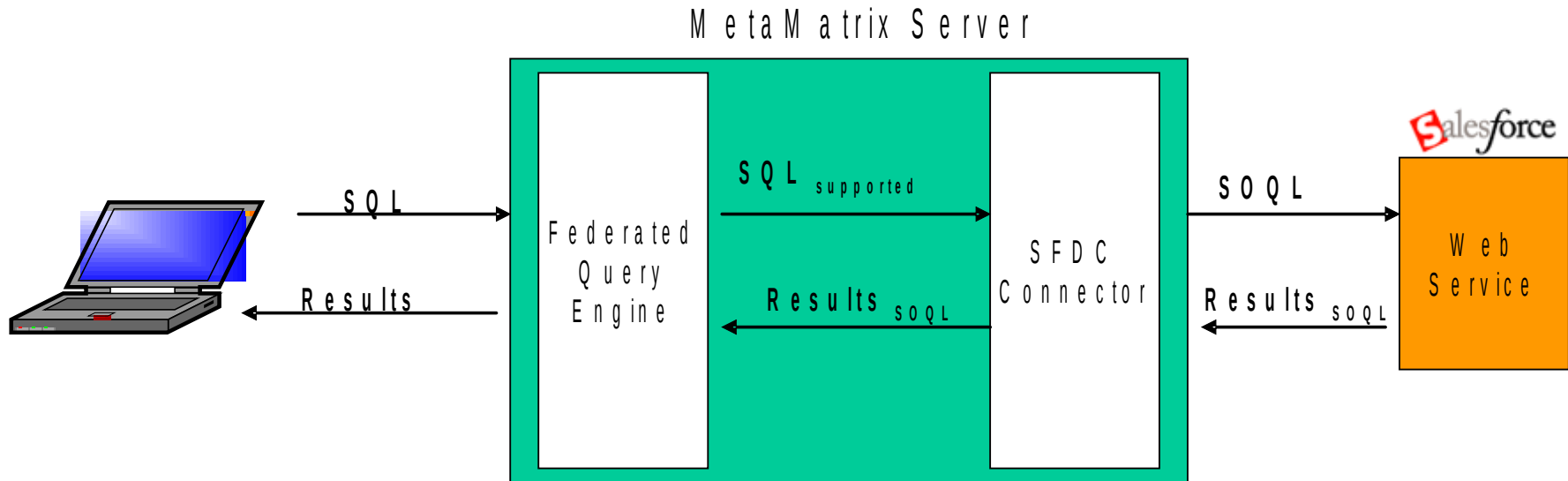
Includes, excludes

Supports custom objects

**Note:** All OwnerId, CreatedById, and LastModifiedById fields on other objects relate to the User object.



# Salesforce Connector



App issues SQL:

```
SELECT SUM(Reports) FROM Sales where Division = 'customer support';
```

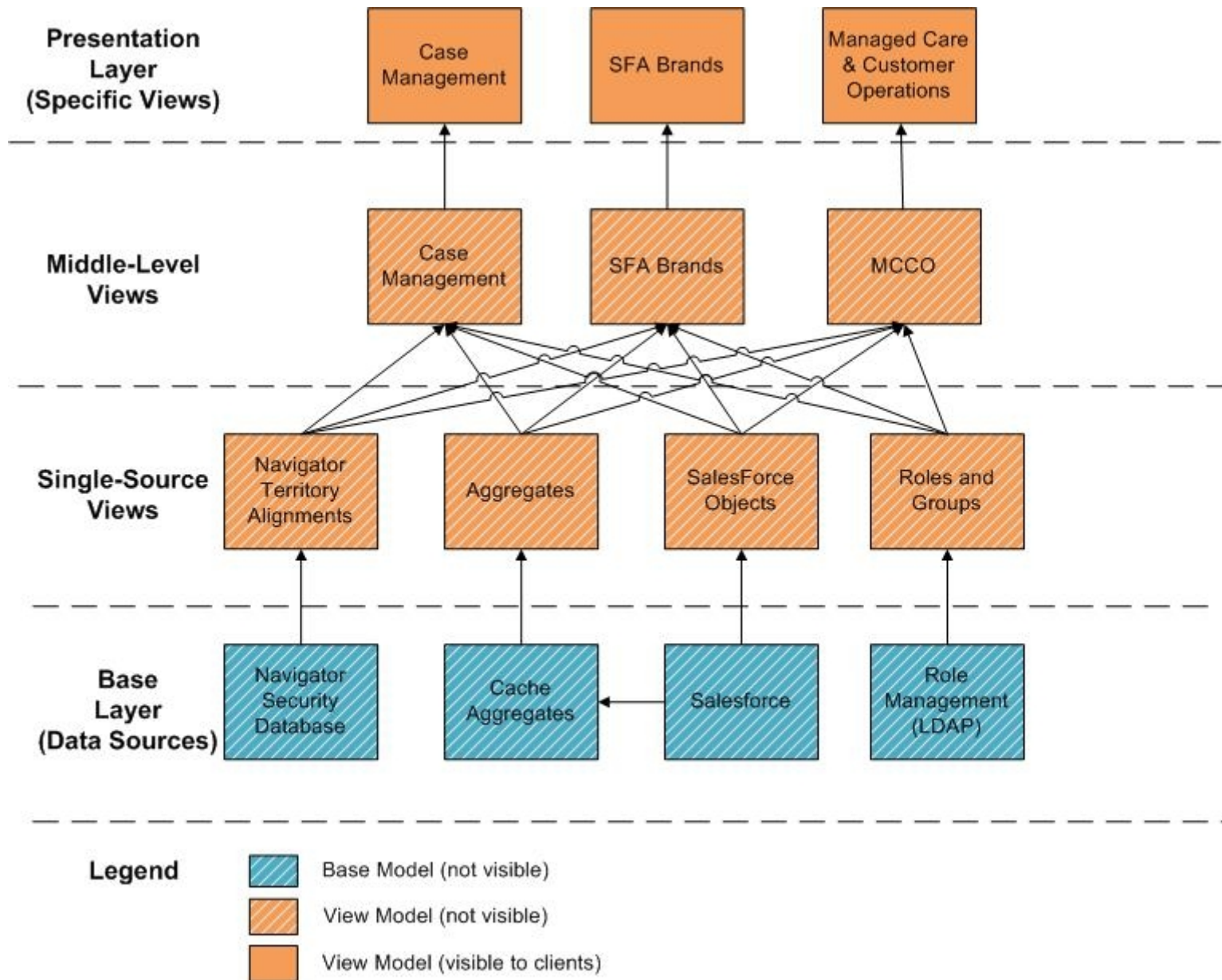
Query Engine translates SQL into **SQL<sub>supported</sub>** based on SOQL-supported capabilities.

```
SELECT SUM(Reports) FROM Supervisor where Division = 'customer support';
```

Salesforce Connector manages SOQL request/response, returns results

Query Engine applies additional SQL to form complete result (SUM)

# BI/OI Reporting Data Models



# Cloud Reporting: Lessons Learned and Future Work

Incremental Refresh option for SFDC data

Notifications, Outbound Messaging

Via ESB integration

GetDeleted(), getUpdated() API calls

Not recommended

Managing schema changes in Salesforce

Release procedure

Delta detection and impact analysis

# MetaMatrix DSP at Genentech: Future Work

MetaMatrix DSP Upgrade to version 6.0

RSS Aggregation

Publish to JMS Queue

Email Connector

google Connector

Delta Detection for Materialization Process

Integration with Active Directory (Roche)

Integrate with ODS

Infrastructure upgrade to support global access

**QUESTIONS?**

**TELL US WHAT YOU THINK:  
[REDHAT.COM/JBOSSWORLD-SURVEY](https://redhat.com/jboss-world-survey)**

# Tips and Tricks

Multi-Source Binding

Materialization table name ending with “ST”

e.g. gppeople\_hist

Text File Descriptors accessing NAS Share

Host controller to run as windows domain account

Not local account

Trailing Space in the where clause filter value

e.g. city = 'South San Francisco '

Query Plan and Access for UNION query

Correlated sub-queries