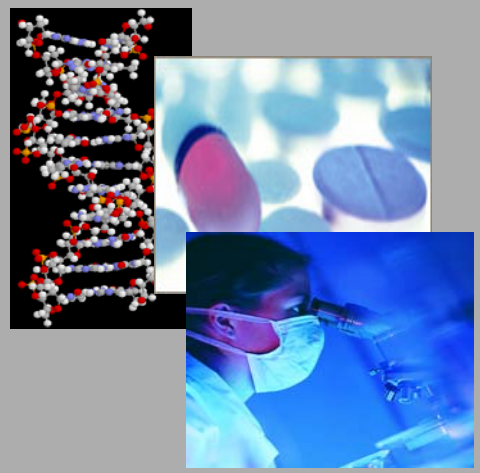
The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decision. The development,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

THE INFORMATION COMPANY



Oracle Life & Health Sciences Platform and 11g Overview

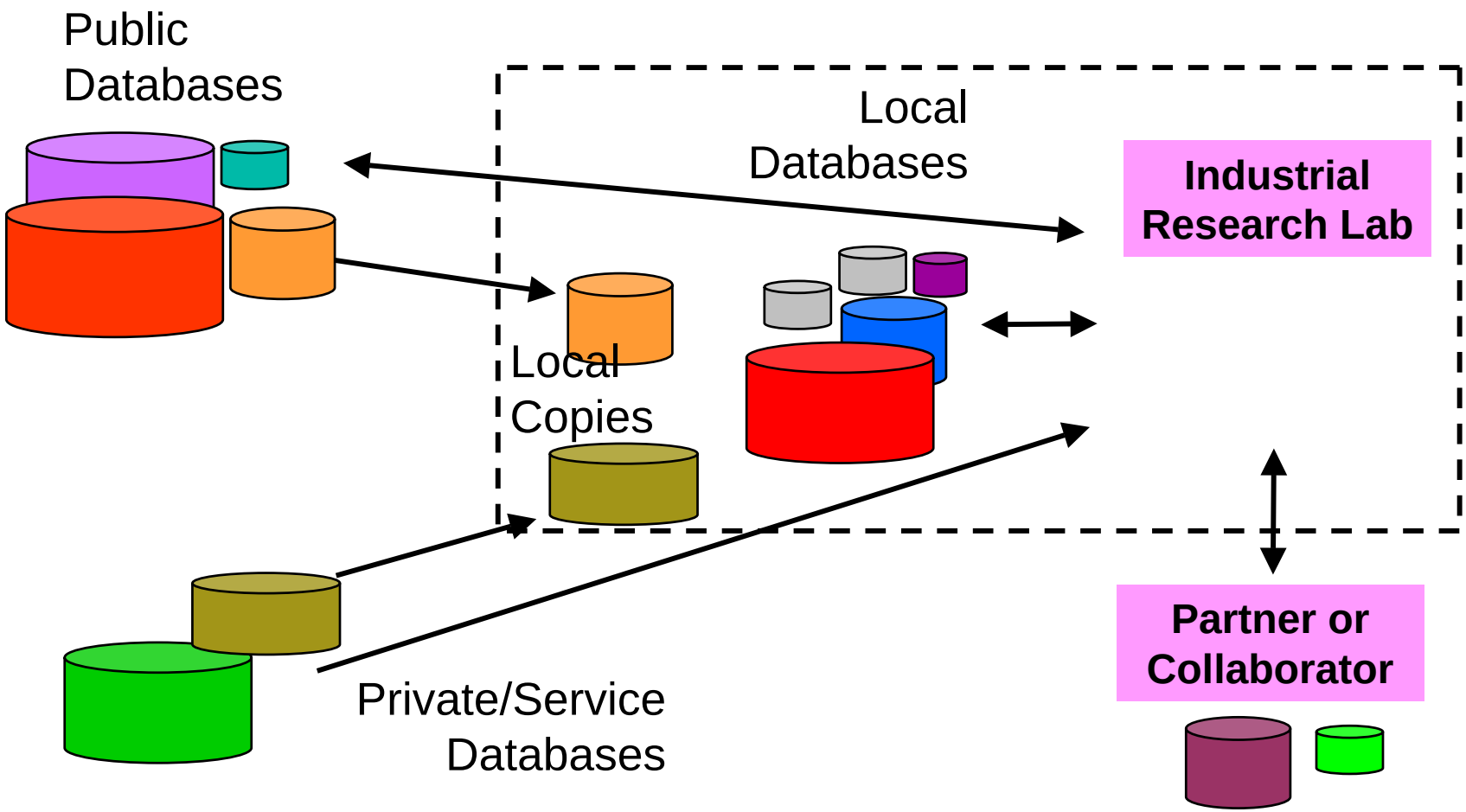
ORACLE

Charlie Berger
Sr. Dir. Product Mgmt
Life Sciences & Health Sciences Industries & Data Mining Technologies
Oracle Corporation
charlie.berger@oracle.com

Oracle's Solutions for Life & Health Sciences

Life Science Challenge

Typical Research Environment



Oracle Life & Health Science Platform

Access distributed data

Gateways, External Tables, SQL Loader, Streams, Transparent Gateways, etc.

Integrate a variety of data types

RDF, XML DB, InterMedia, Text, Semantic, etc.

Manage vast quantities of data

RAC, ASM, Partitioning, Grid, etc.

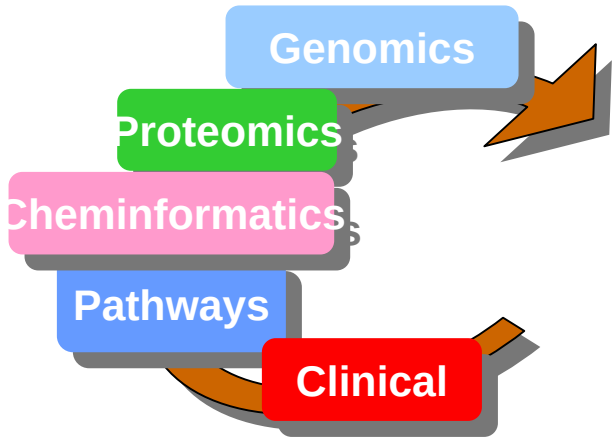
Collaborate securely

Collaboration Suite, Oracle FilesOnline, Portal, Security, etc.

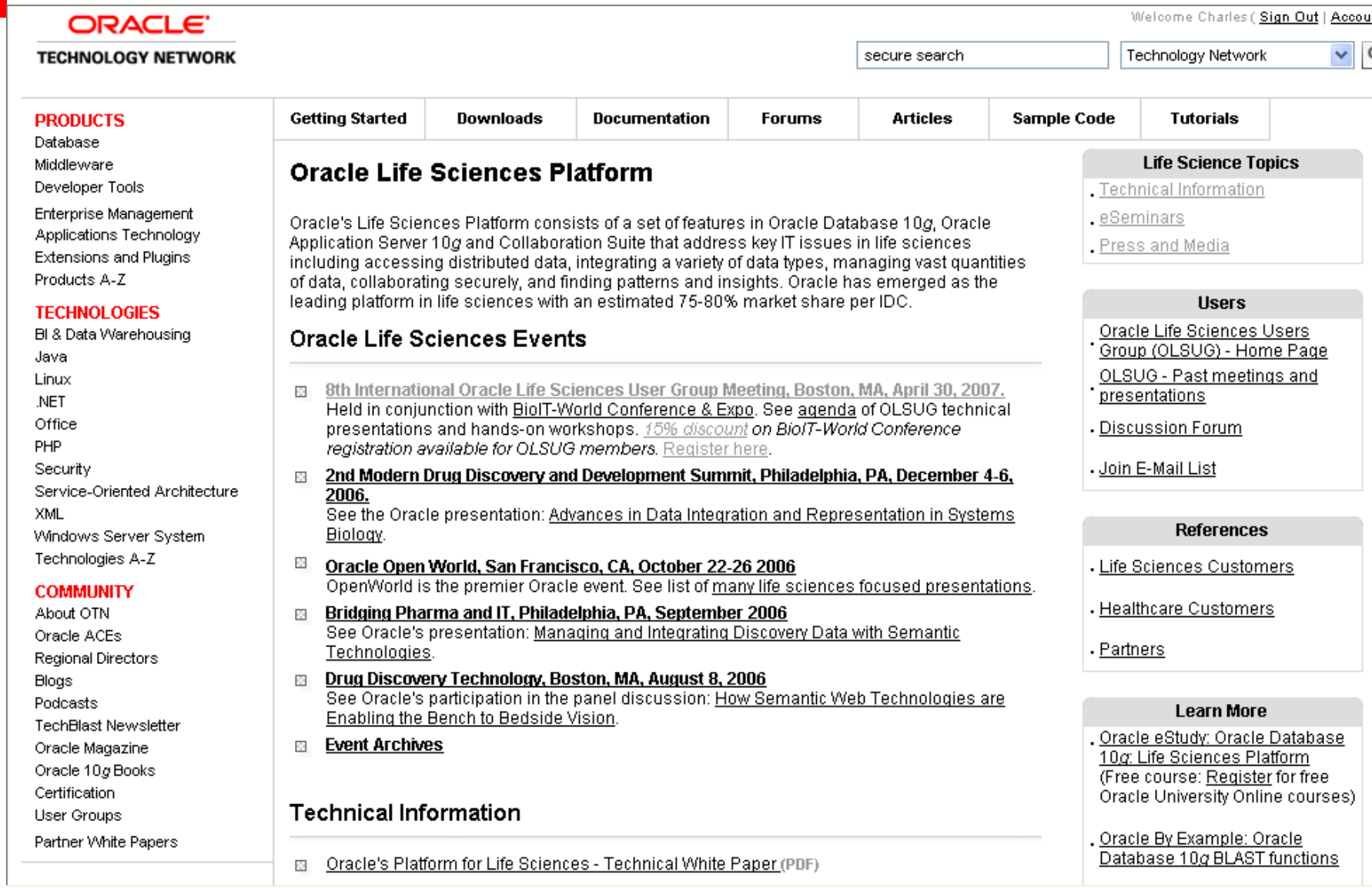
Find patterns and insights

Data Mining, Text Mining, Statistics, BLAST, Regular Expression Searches, etc.

ORACLE 11g DATABASE



Oracle Life & Health Sciences User Community



OLSUG

IOUG independent oracle users group IOUG CLICK here to go to wwww.ioug.org

Welcome to the Oracle Life Sciences User Group Wiki

[edit]

Become a Member **JOIN OLSUG !!!** *Send an email to OLSUG to become a member*

Existing Members Login *Existing Members Login*

8th International Oracle Life Sciences User Group Meeting

It was held April 30, 2007 and was co-located with Bio-IT World, at the World Trade Center, Boston, MA, USA

The **OLSUG Meeting** consisted of exciting keynote presentations from industry luminaries and technical ex AstraZeneca, NIH, Yale, The Broad Institute, SAIC and Oracle.

There where hands-on tutorials, and Oracle partners including Spotfire, ChemAxon, Tom Sawyer, Cambridge described their latest offerings.

New OLSUG Board of Directors

Current List of board members.

Wiki Competition

Congratulations to our most recent most Wiki contributor winner!

John Morris won with his excellent entry **Scoring Text Relevance: A case study**

The **new wiki competition** runs from 1st March to 30 June 2007.

Post a relevant contribution to the wiki and you could win a $300 Amazon voucher.

Past Winners

Rules for Entry

Contribute to the OLSUG Wiki

A new OLSUG Wiki account creation procedure has been introduced, which enables OLSUG members to post content and edit our Wiki pages. This simply involves mailing the OLSUG board and we will set up an account for you.

Previous OLSUG Meetings

See the agenda and slides from past meetings

Main Oracle Technology Categories

- Oracle Life Sciences User Group, Inc.
- 5+ years
- Goal: Share best practices in applying Oracle technology in life sciences & healthcare use cases
- Active User Group
  - 1-2 User Group Meetings per each
  - Wiki web site

Oracle Life & Health Sciences User Community

Bio-IT World
Meeting Sponsor

8th International Oracle Life Sciences User Group Meeting

World Trade Center, Boston, MA, USA

April 30, 2007

$199 commercial. $99 Academic

OLSUG08 Topics & Speakers: (click on links to download presentation)

- *Extracting Biological Meaning from High-Dimensional Datasets* John Quackenbush, Professor of Computational Biology and Bioinformatics, Dana-Farber Cancer Institute
- *Oracle's Platform for Life & Health Sciences* Charlie Berger, Senior Director of Product Management, Oracle
- *MedXminer: mining MEDLINE using Oracle's XMLDB* Uma Mudunuri, Programmer Analyst, SAIC
- *Designing a Powerful Research Data Warehouse that is Intuitive to the Scientists: Our strategies and Solutions* Robert Cain, Principal Scientist, Allergan, Jeff Pierick, CEO, The Pierick Group *Movie*
- *Data Integration for Systems Biology* Giles Day, Director, Research Informatics Site Head, Research Technology Center, Pfizer
- *An Architecture for Drug Discovery Research* Marcus Collins, Enterprise Architect, Novartis
- *Unsupervised Discovery, Supervised Interpretation in Systems Pharmacology* Shunguang Wang, Director of BioIntegration, BG Medicine Inc.
- *Experience Of Using the Oracle RDF Data Model to Integrate Neurodegenerative Data* Hugo Lam, Researcher, Yale University
- *Experience in Integrating Large RDF-based Biomedical Knowledge Resources with Oracle* Kelly Zeng, Senior Technical Consultant, National Library of Medicine, NIH
- *Support for Building Semantics-Driven Life Science Applications in Oracle* Souri Das, Consultant Member Technical Staff, Oracle
- *RDF Friendly Chemical Taxonomies for Semantic Web Powered by Oracle* Bhat Narayana, Project Leader, Bioinformatics, NIST

Technical Workshops (20 PCs available) (click on links to download workshop materials)

- *Introduction to SQL Developer and Application Express* Mark Forster, Chemical Informatics Team Leader, Syngenta R&D Information Systems
- *R-ODM: An R-environment interface for Oracle Data Mining* Pablo Tamayo, Sr. Computational Biologist, Broad Institute; Consulting Member of Technical Staff, Data Mining Technologies, Oracle
- *Using Oracle interMedia to Build an Image Archive* Melliyal Annamalai, Principal Member of Technical Staff, Oracle
- *Using Oracle Data Mining for Life Sciences Problems* Charlie Berger, Senior Director of Product Management, Data Mining Technologies and Life & Healthcare Sciences Industries, Oracle presentation QuickStart using Lymphoma data Data Oracle Data Mining Hands-on Tutorial & Related info.
- *Using Oracle 10g's In-Database Statistical Functions* Henri Tuthill, Principal Technical Architect, Discovery Information, AstraZeneca R & D Supplemental slides
- *Berkeley DB* Greg Burd, Berkeley DB Senior Product Manager, Oracle

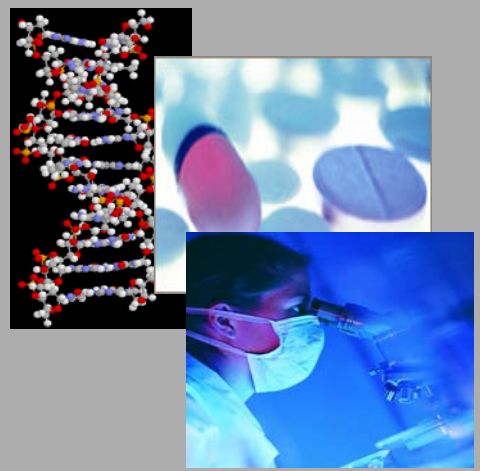
ISV Lightning Rounds

- Applied Biosystems
- ChemAxon
- Cambridge Technology Enterprises
- Spotfire
- SPSS
- Tom Sawyer

OLSUG Partners

Spotfire · Tom Sawyer Software · ChemAxon · CTE Cambridge Technology Enterprises · Maverick Solutions · SPSS · Applied Biosystems

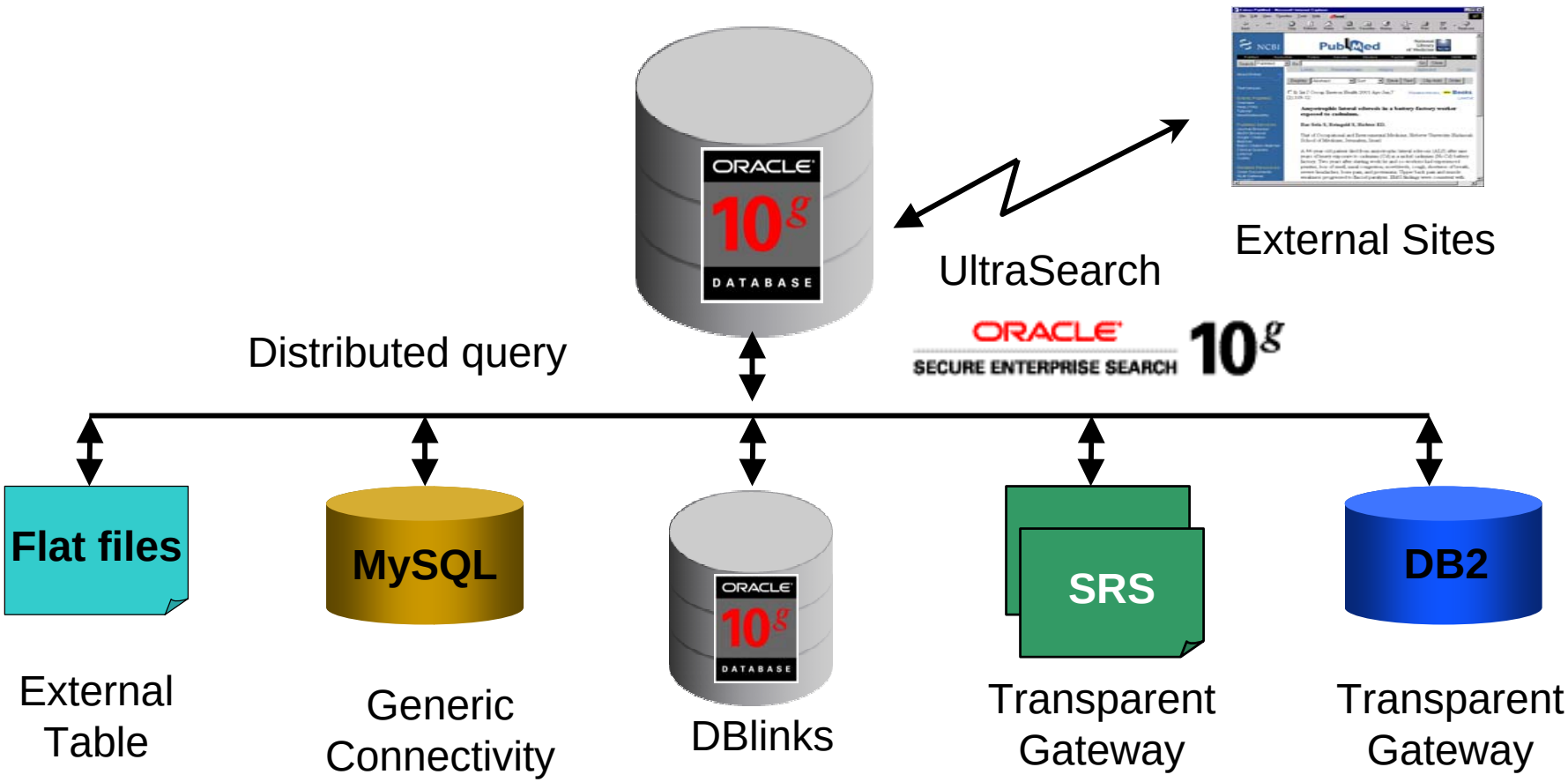
ORACLE



Access Distributed Data

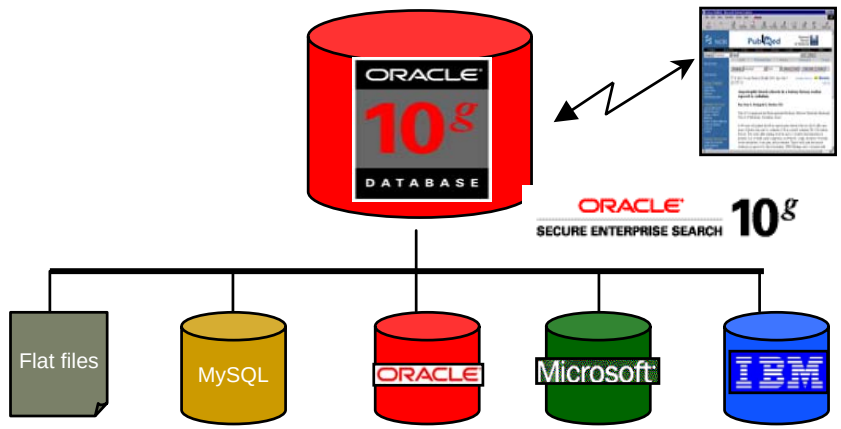
ORACLE

1. Access Distributed Data



1. Access Distributed Data

- SQL*Loader
- Heterogeneous Transportable Tablespaces
- Oracle Warehouse Builder
- Merge Statement
- Oracle Streams
- Migration Toolkits
- High Speed Import/Export
- SRS Gateway
- Migration Toolkit
- Secure Enterprise Search

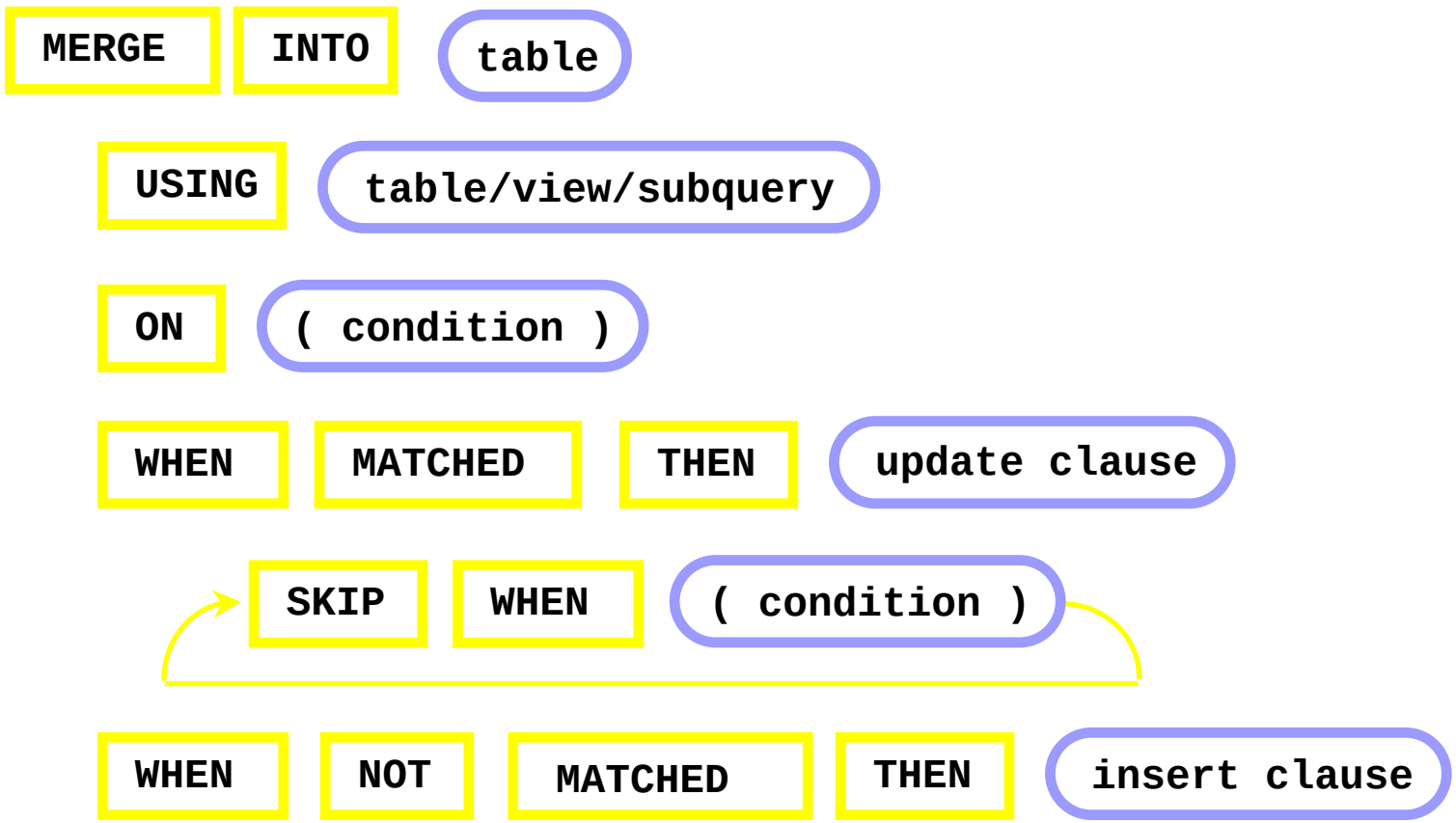


SQL*Loader

- High-speed data loading utility
  - Loads data from external files into tables in an Oracle database.
  - Accepts input data in a variety of formats
  - Performs filtering
  - Loads into multiple tables during the same load session
- Three methods for loading data:
  - Conventional Path Load
  - Direct Path Load
  - External Table Load

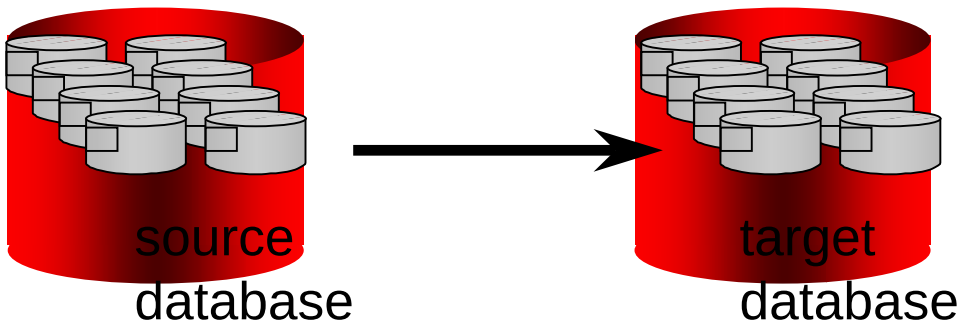
Merge Statement

- Fast insert, update or conditional update/insert of records



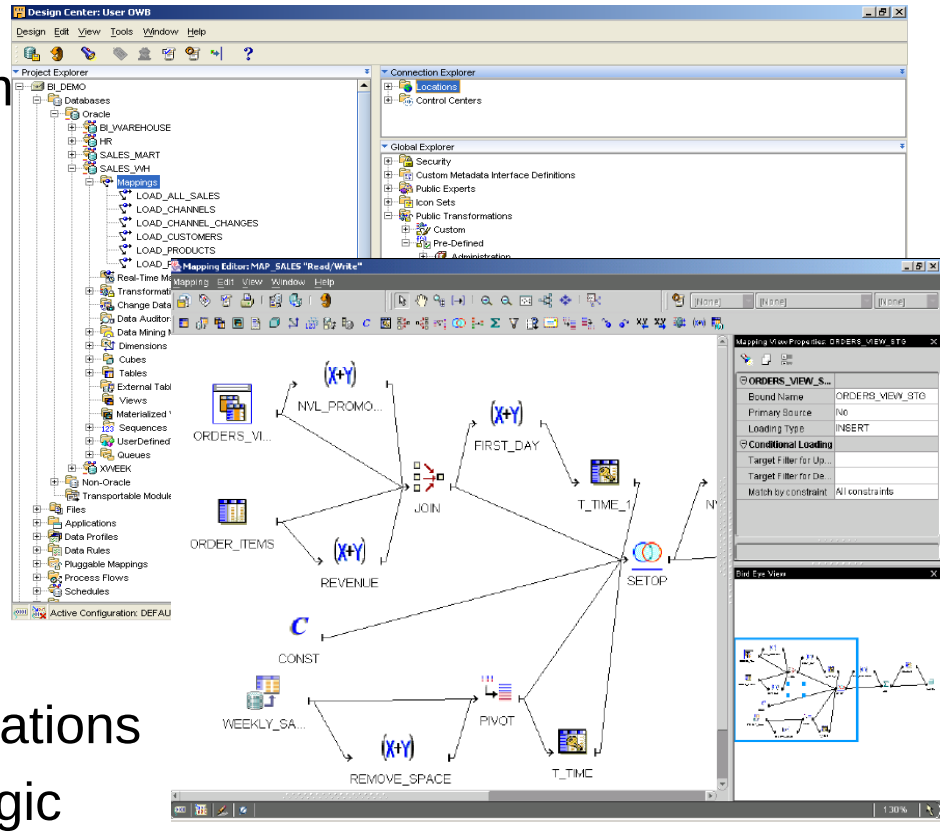
Transportable Tablespaces

- Mechanism to quickly move a tablespace between Oracle databases
- Most efficient means to move bulk data between databases
- Enhanced to support different hardware platforms & operating systems



Oracle Warehouse Builder (OWB)

- Enables the extraction, transform loading of data
- Graphical declarative modeling of data flows
- Generates SQL & PL/SQL
  - Merge, transportable tablespaces, sqlloader, table functions*, streams, xml data types*, BLOBS/CLOBS*
- Leverage custom data transformations
- Nested maps for reusability of logic



Oracle Data Pump

- High speed bulk data and metadata movement (Import/Export) between Oracle databases

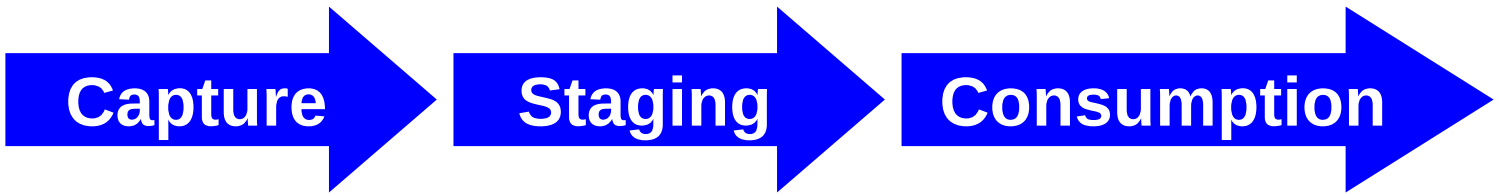
- Speedup of 10x for import and 2x for export for serial execution
- Automatically scales using parallel execution
- Accessible via
  - expdp and impdp utilities
  - PL/SQL API
  - Enterprise Manager

Distributed Query Optimization

- Enhanced cost based optimizer
- Capture complete statistics for remote tables
- Consider network bandwidth & latency in deciding what parts of query plan should be remotely mapped
- Support different execution cost at different nodes (e.g. based on node ownership)

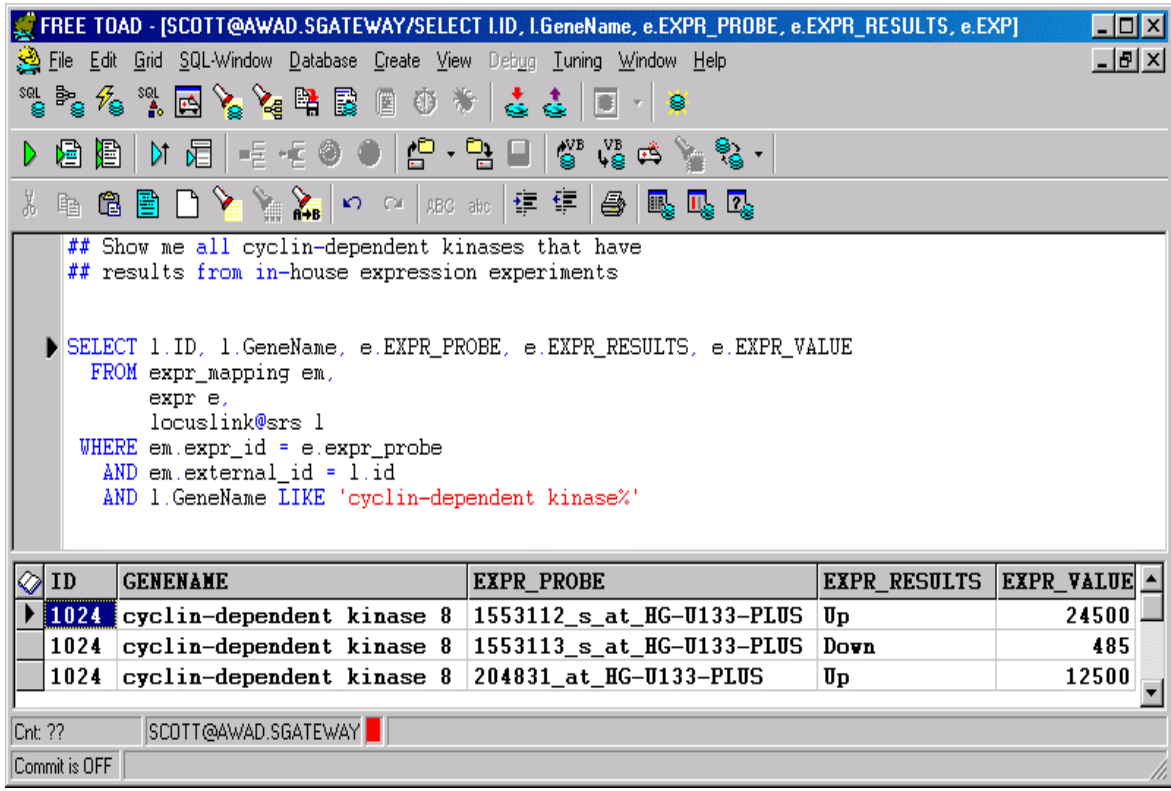
Oracle Streams

- Enables rule-based information sharing among multiple systems
  - Captures and manages events
  - Shares events with other databases and applications
  - Routes published information to subscribed destinations
  - Integrated with new job scheduler



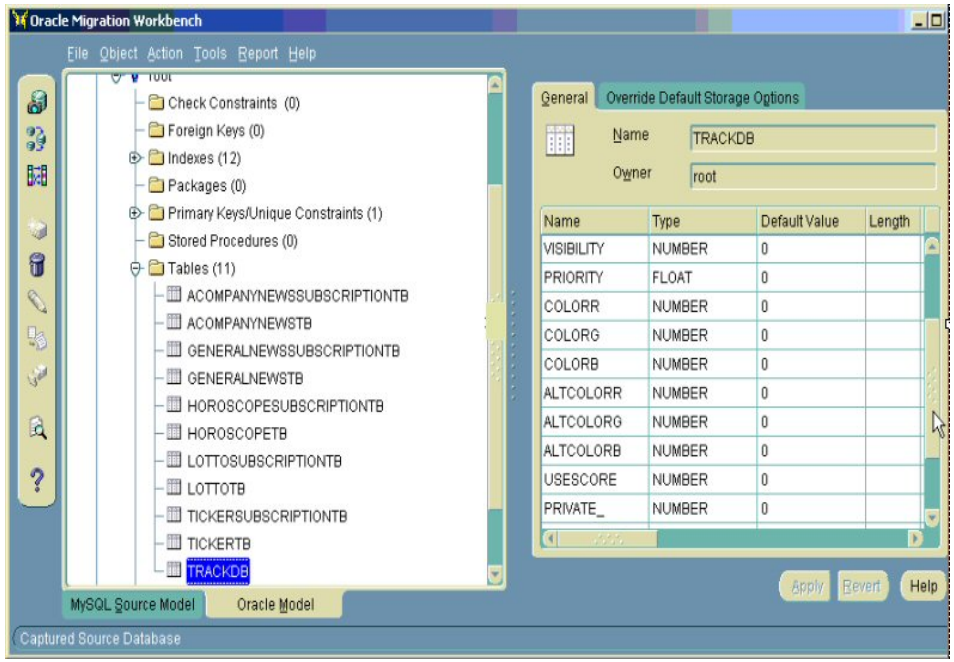
SRS Transparent Gateway for Oracle

- Data behaves as if they are in Oracle
- Oracle re-writes user's SQL query into syntax understood by SRS, using capability table & index of Gateway
- The query is executed in SRS
- If mapping entire query to SRS syntax is not possible, after fetching the data, Oracle will do some functions/joins locally



Migration Toolkits

- Oracle has a series of migration toolkits that can be used to rapidly migrate data in a non-Oracle database into an Oracle database e.g.
  - MySQL to Oracle

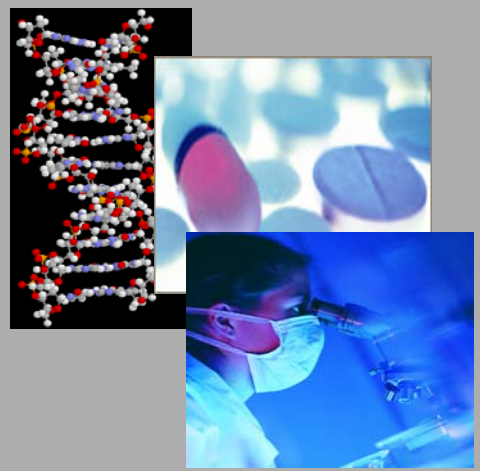


Caprion

- Discover & develop innovative products for the diagnosis & treatment of diseases
  - Scalability for a multi-TB system
  - Integration of all components with existing computing environment
  - Security & protection of data integrity
- Oracle Environment
  - Oracle Database
  - Oracle9i Application Server
  - Oracle9i Developer Suite
  - Oracle9i AS Discoverer
  - Oracle Warehouse Builder
- Key Advantages of Oracle
  - Easy access & management of integrated information
  - Rapid deployment of new ad hoc query
  - Scalability necessary to accommodate growth
- "The Oracle Data Warehouse is a key component of our IT platform for proteomics analysis. The massive amount of information we produce every day requires a system with proven performance to effectively capture our biological data". - Bernard Gagnon, IT Director

Oracle Secure Enterprise Seach

- Oracle Secure Enterprise Search 11g, a standalone product from Oracle, enables a secure, high quality, easy-to-use search across all enterprise information assets.
- Key features include:
  - Seach and locate public, private and shared content across Intranet web-servers, databases, files on local disk or on file-servers, IMAP email, document management systems, applications, and portals
    - Search for protocols, lab notes, research papers, emails, etc.
  - Highly secure crawling, indexing, and searching
  - A simple, intuitive search interface
  - Analytics on search results and understanding of usage patterns
  - Sub-second query performance
  - Ease of administration and maintenance leveraging your existing IT expertise

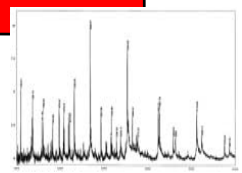


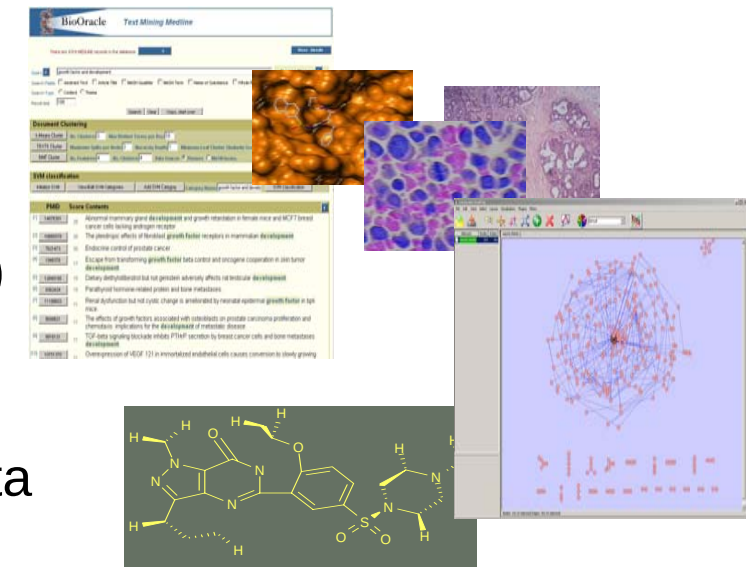
Integrate a Variety of Data Types

ORACLE

2. Integrate a Variety of Data Types

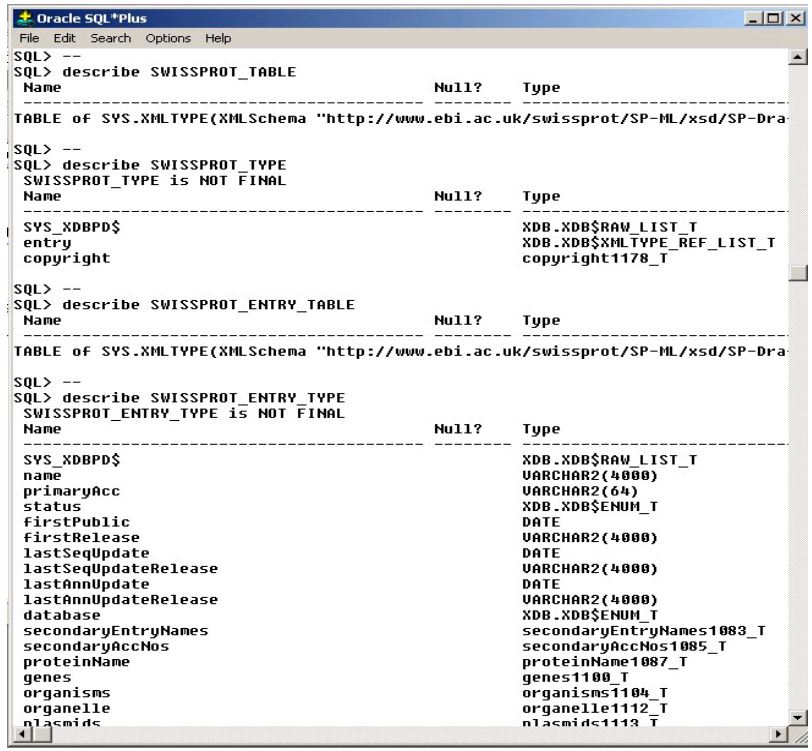
- XML DB
  - Unite XML content & SQL/relational data
- LOBs
  - Manage unstructured data e.g. BFILES, BLOBs, CLOBs, URIs
- Files
  - Central repository for structured & unstructured data
- Text
  - Index & fast query of text content
- *inter*Media
  - Manage audio, video & image data
- Network Data Model (Oracle Spatial)
  - Graph (arc node) relationships
- Extensible indexing
  - Manage & index complex scientific data





XML Support

- Oracle Database supports XML data model
  - XMLType, XMLSchema, DOM Fidelity, Xpath, …
- Query Language: SQL/XML and XML Query
- Transparent storage optimizations
- A new XML Content Repository
  - Hierarchical organization of the data
  - WebDAV compliant with indexing for fast access
- Copy-based Schema Evolution for XMLType
- SQLX standards compliance



XDK Advances XML APIs

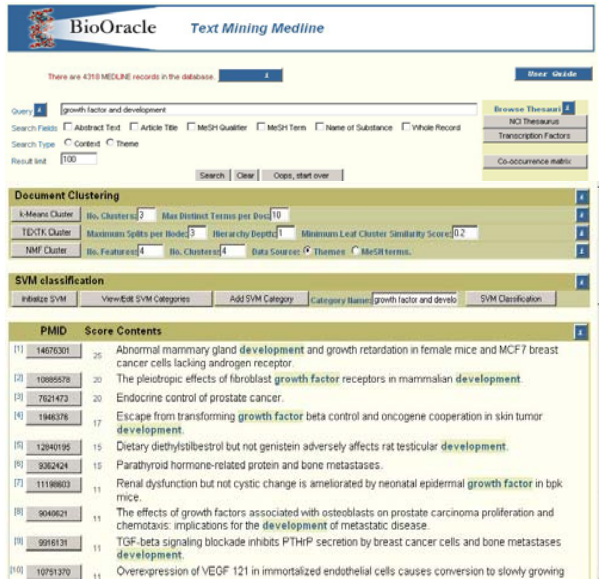
- XDK unifies XML APIs in/outside Database
  - Simplifies XML Application development in the Database, Mid-tier & Clients
  - Eliminates multi-step processing by operating directly on XMLType
  - Improves application performance in Java, C, and C++
- XSLT performance increase up to 100%
- Additional XML Standards Support
  - DOM 3, XSLT 2, XPath 2
  - XML Pipeline, XPointer, JAXB

Reed Elsevier

- Largest technical publishing conglomerate $8B annual revenue
- More than 1700 scientific, technical & medical peer-reviewed journals
- Over 59 million abstracts
- Over two million full-text scientific journal articles , another one million full-text articles via CrossRef (http://www.crossref.org/) to other publishers' platforms
- Oracle XML DB chosen as Repository Database

Oracle Text

- Powerful text search and intelligent text management capabilities
- Fully integrated with the database
- Text can be ASCII, HTML, XML, or formatted (150+ formats supported)
- Offers premier text search quality
- Document Services such as themes, gist, term highlighting and markup
- Classification and clustering capabilities
- Simply text applications development via JDeveloper Wizards



European Bioinformatics Institute

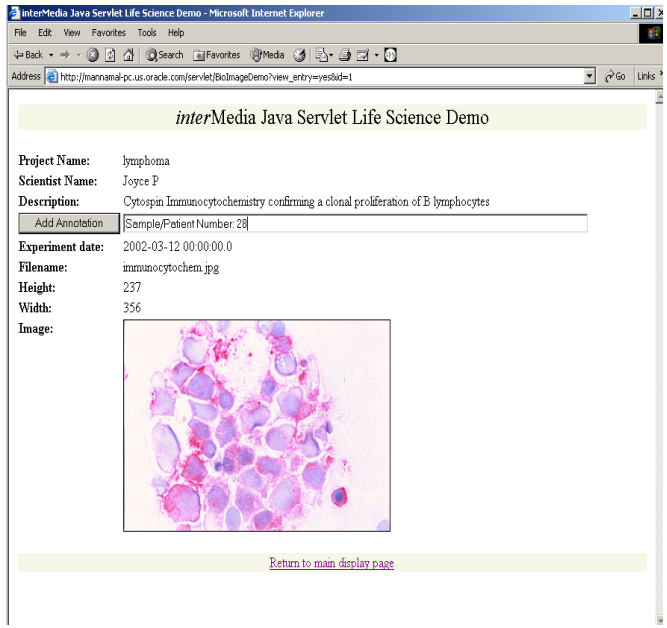
- Manages major public databases (e.g. SwissProt, EMBL Nucleotide Sequence Database, Medline) in Oracle. (Total: > 5 TB)
- Uses Oracle XML DB and Oracle Text for Medline – in development.
  - Size: 11 million records, 200 GB
- Uses Oracle Database and Application Server

Large Objects (LOBs)

- Enables storage and management of large blocks of unstructured data inside or outside the database
- There are three types of LOBs:
  - Binary LOB (BLOB) – Stored in DB
  - Character LOB (CLOB) – Stored in DB
  - Binary File (BFILE) – Stored in OS files
- LOBs enable users to manage unstructured data in the same table that contains the structured data
- In 11g LOB columns are unlimited in size

*inter*Media

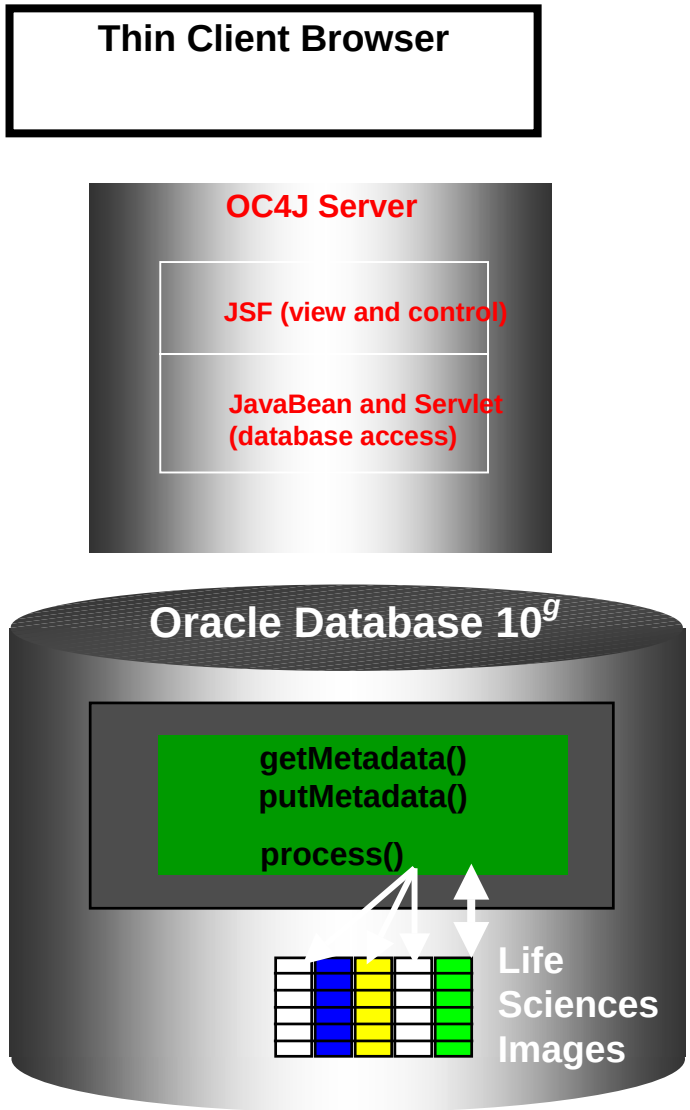
- Ability to store wide range of image types
  - Processing functionality
    - Rotate/flip, brighten/darken using gamma processing, adjust contrast, change bit depth
- Access through SQL, Java & Web interfaces
- Restrict access via security roles
- Conform to SQL/MM still image standard
- Store images as columns
  - Tight integration with annotations
  - Ability to annotate a region of an image (10gR2)



*inter*Media

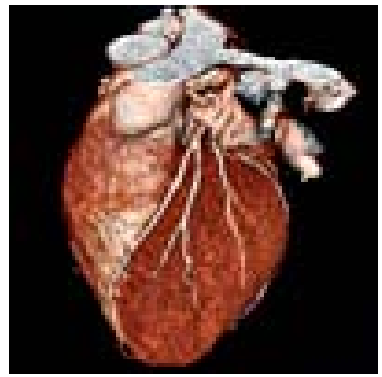
Support for DICOM

- Reads a subset of DICOM image metadata
- Creates XML Schema: patient info, study, series, properties, unique IDs
- Metadata managed as an XML document that can be stored persistently in an XMLType column or handed to an application
- DICOM Image stored in OrdImage

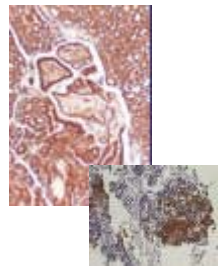


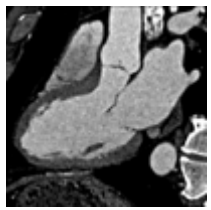
*Inter*Media - DICOM Support

- Support for DICOM version 3
- *inter*Media JAVA and PL/SQL APIs to extract metadata about entities
- Standard way to represent the image metadata
- The metadata can be stored in the database, indexed, and searched
- The APIs for retrieving this metadata return it in the form of an array of XMLType



*inter*Media
DICOM Support





- *inter*Media now supports the most common medical imaging format, DICOM version 3
- *inter*Media JAVA and PL/SQL APIs to extract metadata about patients, physicians, diagnoses, treatments, tests and procedures, and other relevant information included in the DICOM format
  - Standard way to represent the metadata when it is separate from the image file
  - All of the metadata can be stored in an Oracle database, indexed, searched and made available to applications using the standard mechanisms of the Oracle database
  - Since image files can contain many instances of metadata, the APIs for retrieving this metadata return it in the form of an array of XMLType

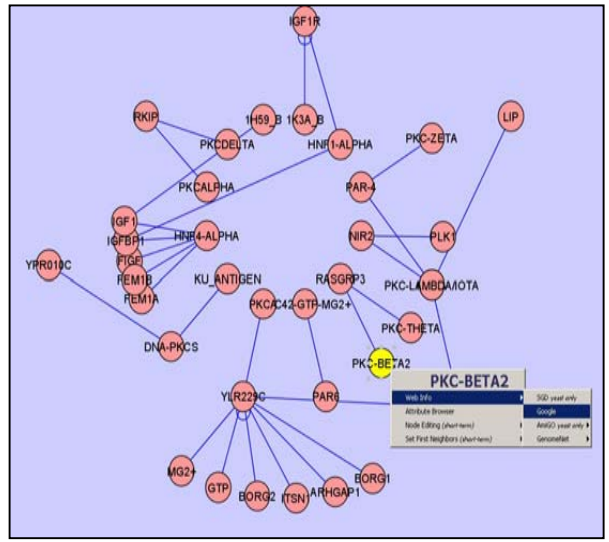
Customer Success

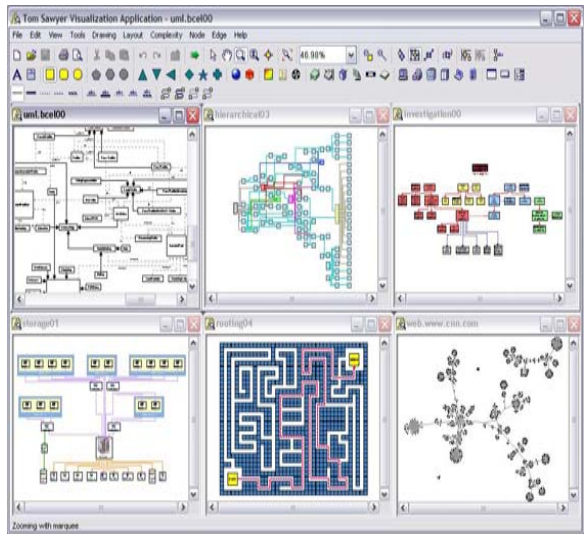
- Clinical Image Repository Helps AstraZeneca Integrate, Manage and Protect Images to Improve Data Integrity and Advance Regulatory Compliance
  - …"The company, in recent years, found that it was using and storing a growing number of images -- from x-rays to microscope slides -- as part of its clinical trial processes. In early 2006, AstraZeneca launched an initiative to create a centralized clinical image repository -- built using the Oracle *inter*Media features of Oracle Database 10g -- to enable rapid access for approved users, ensure data integrity, and streamline compliance. AstraZeneca expects its image repository will exceed 100 terabytes of data within the next year. …
  - … "Our desire was to build a scalable system for managing images that is as robust as our system for managing clinical data," said Dr. Goutham Edula, Business Lead for Clinical Imaging Informatics, AstraZeneca. "Oracle *inter*Media helped us plan for the future by creating a centralized repository to serve our needs today and moving forward. It also allowed us to create a singular back end for data storage, giving us the flexibility to support multiple workflows in AstraZeneca's various practice areas."

Source: http://biz.yahoo.com/prnews/070305/sfm074.html?.v=84 Monday March 5, 2007

Network Data Model

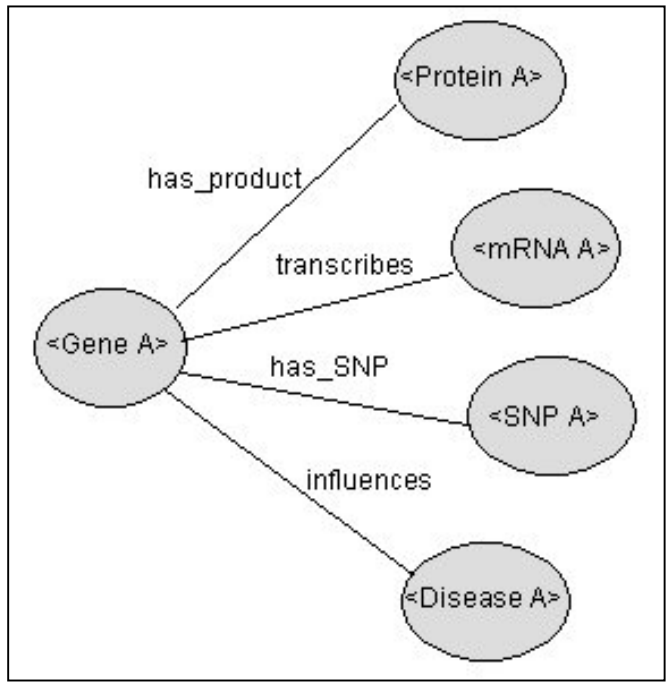
- Model, store, manage & analyze generic connectivity relationships
  - i.e. represent data as nodes & links
- Can model hierarchies, logical or spatial information, directionality
- Network analysis at client or application level, e.g. shortest-path, tracing, within-distance analysis, minimum cost spanning tree, nearest neighbor
- Network management, e.g. add, delete, modify, load





Resource Description Framework (RDF) Data Model Storage

- W3C standard
  - Object-relational implementation
  - Based on triples (subject–predicate–object)
  - Everything has a URI
  - Ontologies used to label the RDF tagged elements
  - Set of triples form an RDF/OWL graph (model)
  - Optimized storage structure: repeated values stored only once (uses normalization)
- Scales to very large datasets
  - Incremental load and bulk load
  - Can handle multiple lexical forms of the same value - Eg: “0010”^^xsd:decimal and “010”^^xsd:decimal



Working with RDF Data

- Query RDF Data
  - SPARQL-like graph pattern embedded in SQL query
  - Can use SQL operators/functions to process results
  - Avoids staging when combined with queries on relational data
  - Millisecond query times for data sets of 10M+ triples
  - Native inferencing for RDF, RDFS & user-defined rules
  - Graph analytics available through Java API
- Native Inferencing with OWL
  - Basics: class, subclass, property, subproperty, domain, range, type
  - Property Characteristics: transitive, symmetric, functional, inverse functional, inverse
  - Class comparisons: equivalence, disjointness
  - Property comparisons: equivalence
  - Individual comparisons: same, different
  - Class expressions: complement

Query with Semantic Operators

Find <id, diagnosis> info for all patients who have been diagnosed as afflicted with diseases of type Immunodeficiency_Syndrome that are within a specified distance from it.

```
SELECT id, diagnosis
FROM Patients_Data
WHERE SEM_RELATED ( diagnosis,
                    'rdfs:subClassOf',
                    'Immunodeficiency_Syndrome',
                    'NCI', 1) = 1
    AND SEM_DISTANCE (1) <= 2;
```

Network Data Model Reference

"Oracle 10g's Network Data Model feature is great for building a semantic work infrastructure. Oracle 10g's graphical representation is an excellent tool for planning our Y2H protein interaction data storage needs and for building a signaling network from our Nature-AfCS Molecule Pages Database." - Joshua Li, Sr. Computational Scientist, San Diego Supercomputer Center / UCSD

"Beyond Genomics, Inc., as a leading systems biology company, believes that Oracle 10g's network data model will significantly advance the integration of metabolomic, proteomic, transcriptomic, and clinical data sets and the applications that derive value from these data." – Eric Neumann, Vice President Strategic Informatics, Beyond Genomics, Inc.

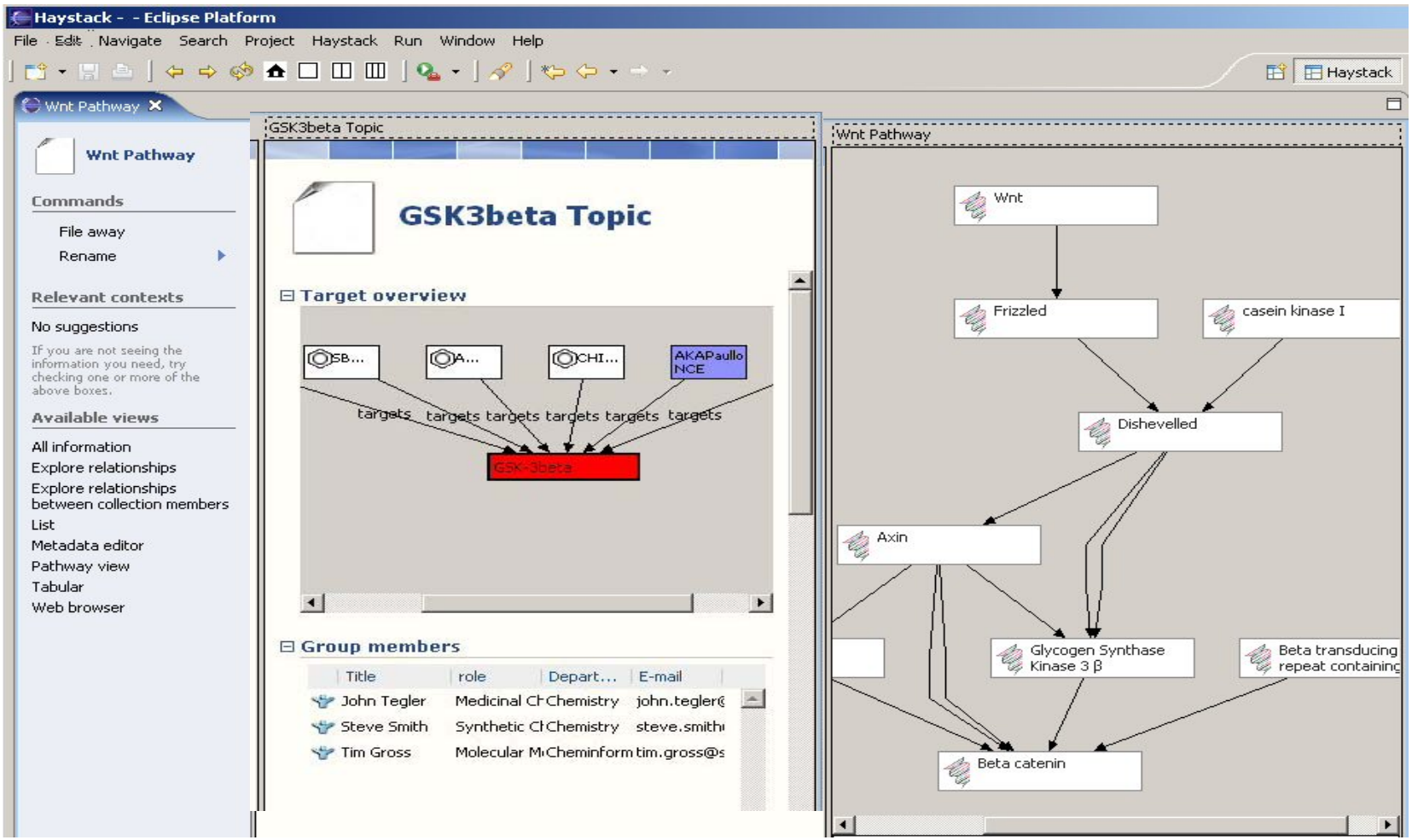
Oracle Spatial
RDF Data Model

- Resource Description Framework (RDF) is a language for representing information about resources in the WWW
  - Statements are essentially broken into triples: {subject/resource, predicate/property, object/value}
- Each triple is a complete and unique fact, in a specific domain, and is represented by a link in a directed “graph”
- RDF triples in the Oracle database as a logical network (using Oracle Spatial Network Data Model)
  - Each RDF triple: {subject, property, object} is treated as one unique database object. As a result, a single RDF document comprising a number of triples will result in multiple database objects.Supports reification
  - Java Ntriple2NDM converter for loading existing RDF data
  - An RDF_MATCH function which can be used in SQL to find graph patterns in RDF (similar to SPARQL)

Semantic Web offers Life Sciences

- Heterogeneous data integration using explicit semantics
- Expression of well-defined & rich models of biological systems
- Annotating & sharing findings with others
- Embedding models & semantics within papers
- Applying logic to infer additional insights

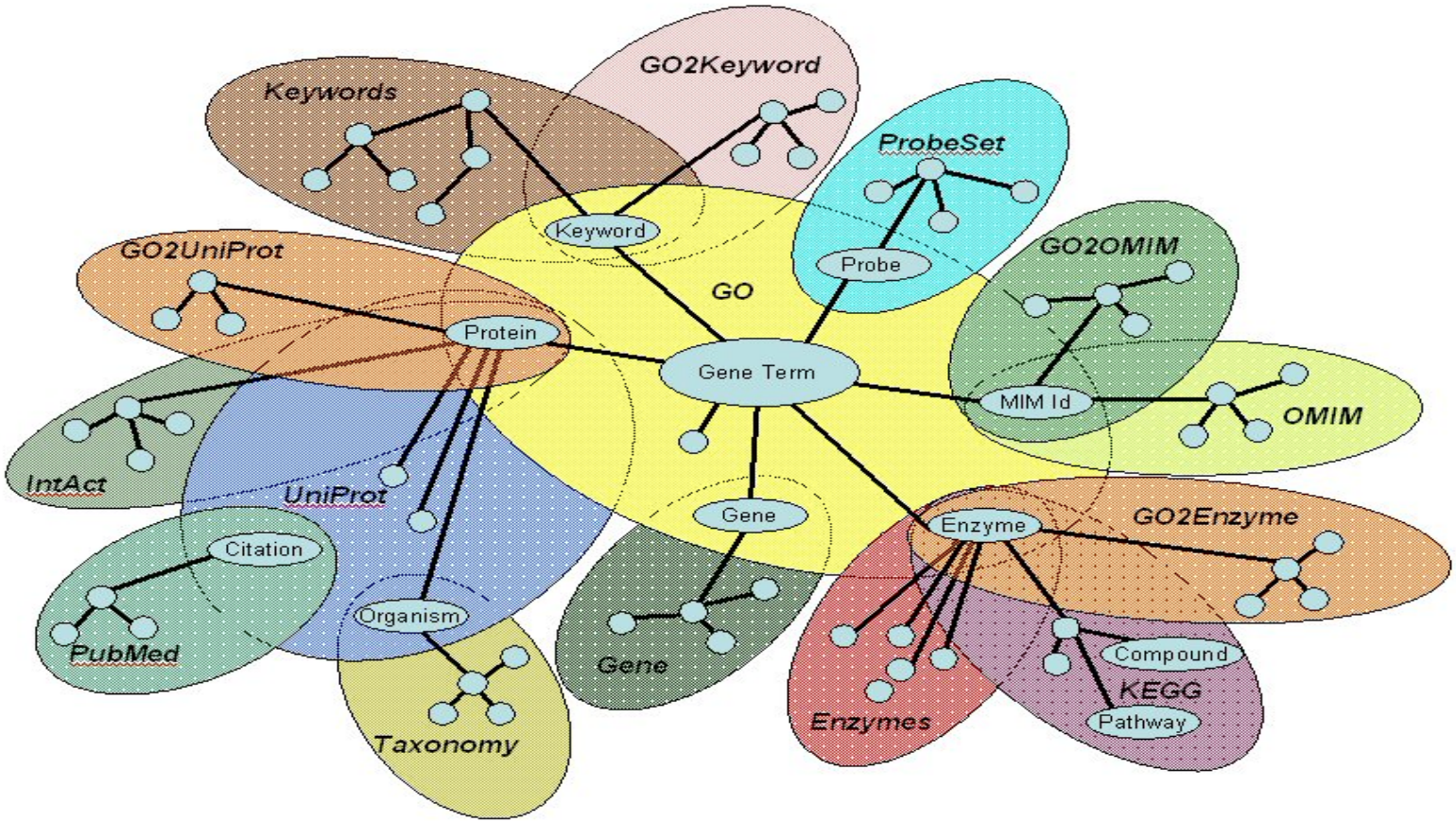
BioDASH



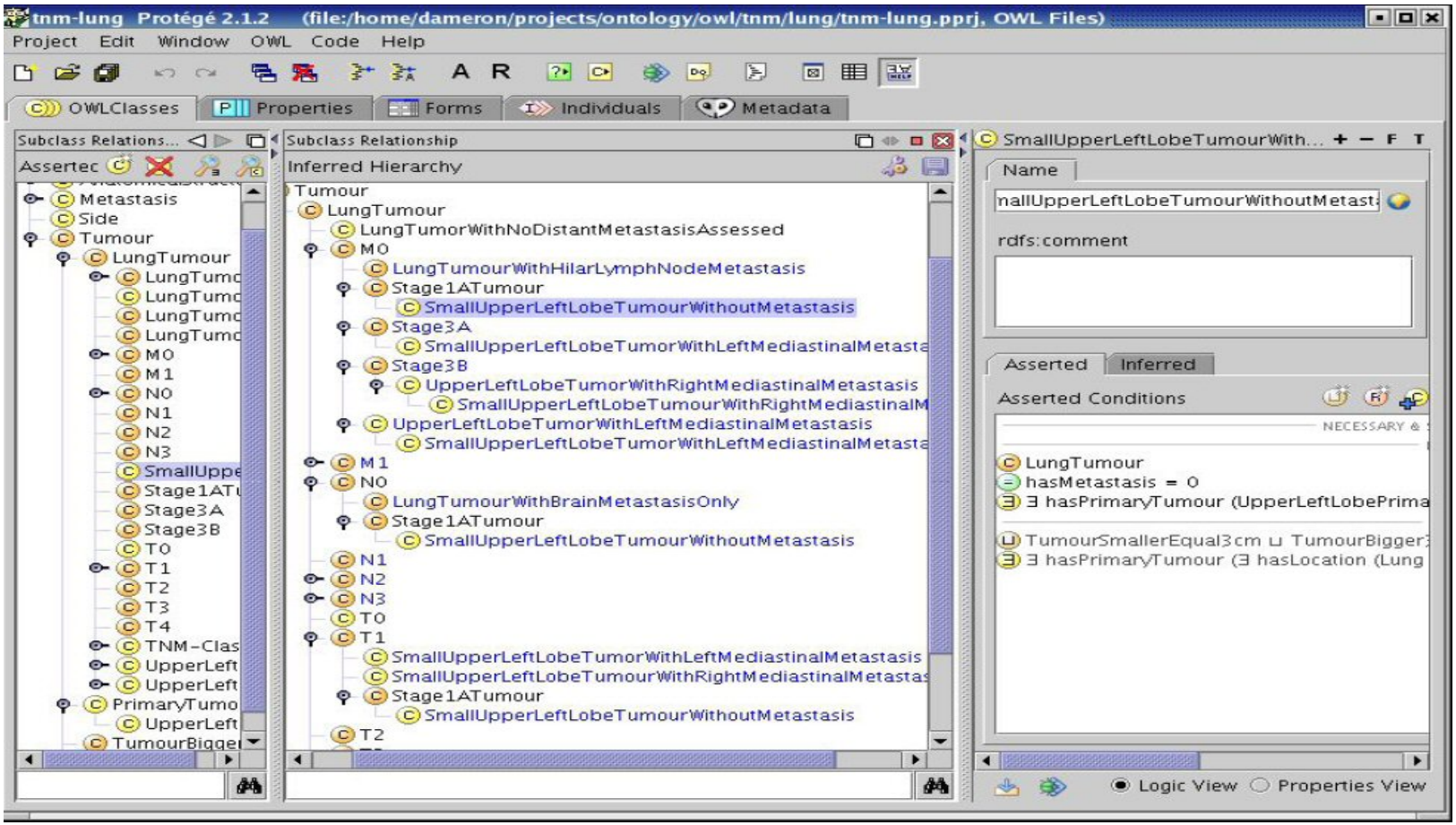
http://www.w3.org/2005/4/swls/BioDash/Demo

Image Source: BioDASH

Integrated Bioinformatics Data

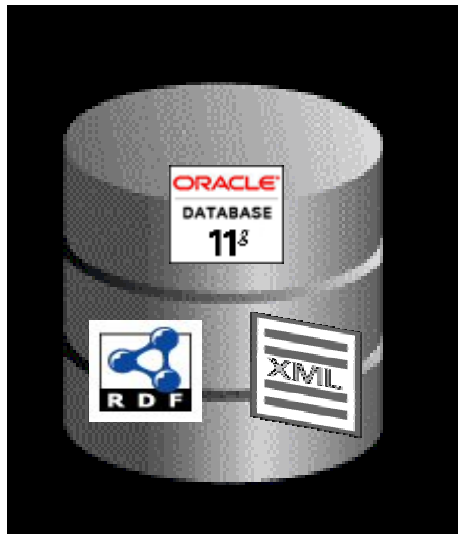


Protégé Ontology Development Tool



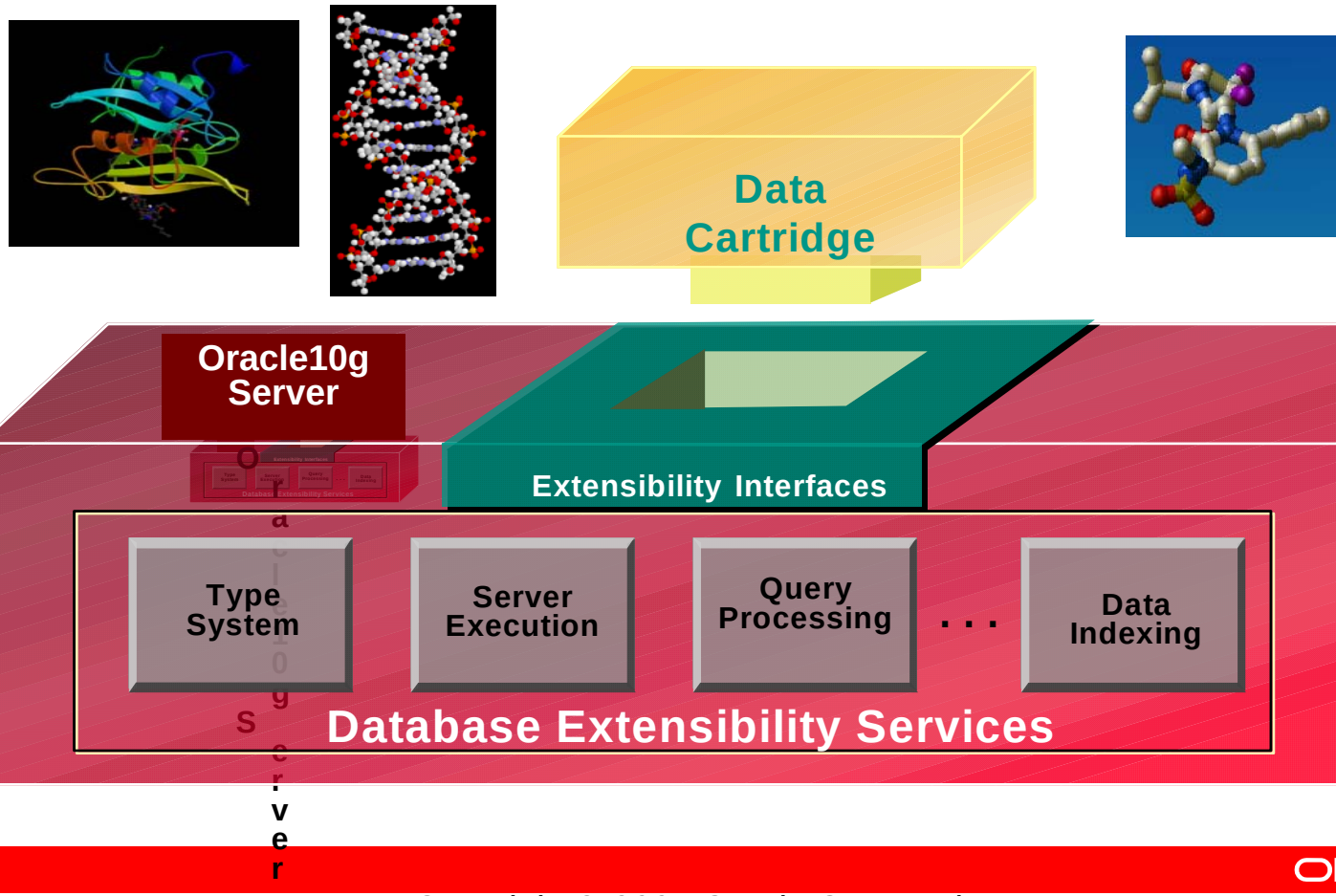
Data Integration

- SQL / RDBMS
  - Concise, efficient transactions
  - Transaction metadata is embedded or implicit in the application or database schema
- XQuery / XML
  - Transaction across organizational boundaries
  - XML wraps the metadata about the transaction around the data
- SPARQL / RDF
  - Information sharing with ultimate flexibility
  - Enables semantics as well as syntax to be embedded in documents



Extensibility Framework

- Data Cartridges
  - Manage complex scientific data



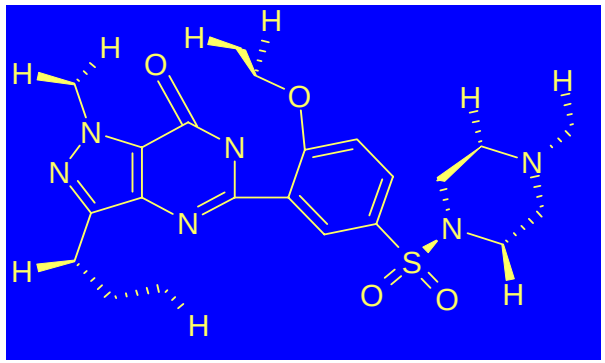
Chemical Searching

Chemistry searching requires special techniques

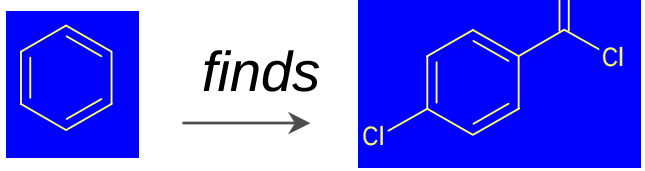
- Chemical name is not unique
- Chemists think graphically

"Viagra®"

"sildenafil citrate"



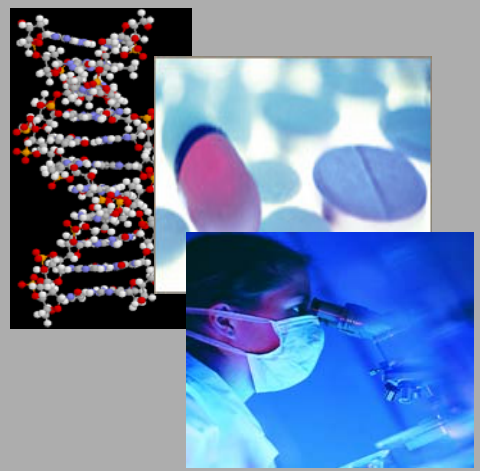
- The solution:
  - A graphical user interface
  - Specialized operators such as substructure search ("sss") = a chemical "contains"



Enhanced Support for Perl

- 10g Release 2 provides support for Perl expressions.
- Perl REGEXP builds on the POSIX standard and has evolved over the years to introduce many proprietary extensions, due to the fact that POSIX sets aside the notation “backslash followed by a character” for tool-specific extensions
- Biologists and life scientists commonly use Perl to rapidly build useful software applications

Character	Description
\d	Match a digit character
\D	Match a non-digit character
\w	Match a word character
\W	Match a non-word character
\s	Match a white space character
\S	Match a non-white space character
\A	Match only at beginning of string
\z	Match only at end of string
\Z	Match only at end of string, or before new line at the end
*?	Match 0 or more times (non-greedy)
+?	Match 1 or more times (non-greedy)
??	Match 0 or 1 time (non-greedy)
{n}?	Match exactly n times (non-greedy)
{n,}?	Match at least n times (non-greedy)
{n,m}?	Match at least n but not more than m times (non-greedy)

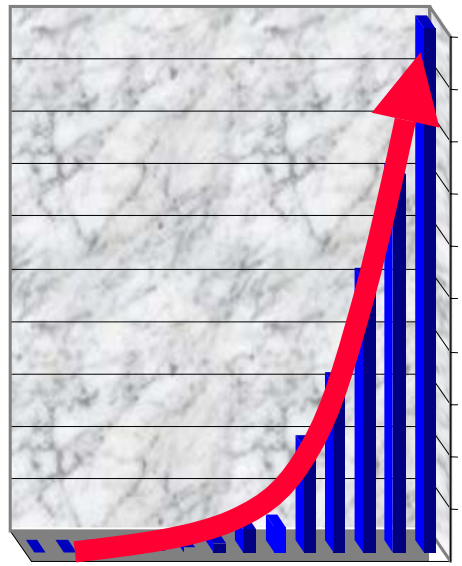


Manage Vast Quantities of Data

ORACLE

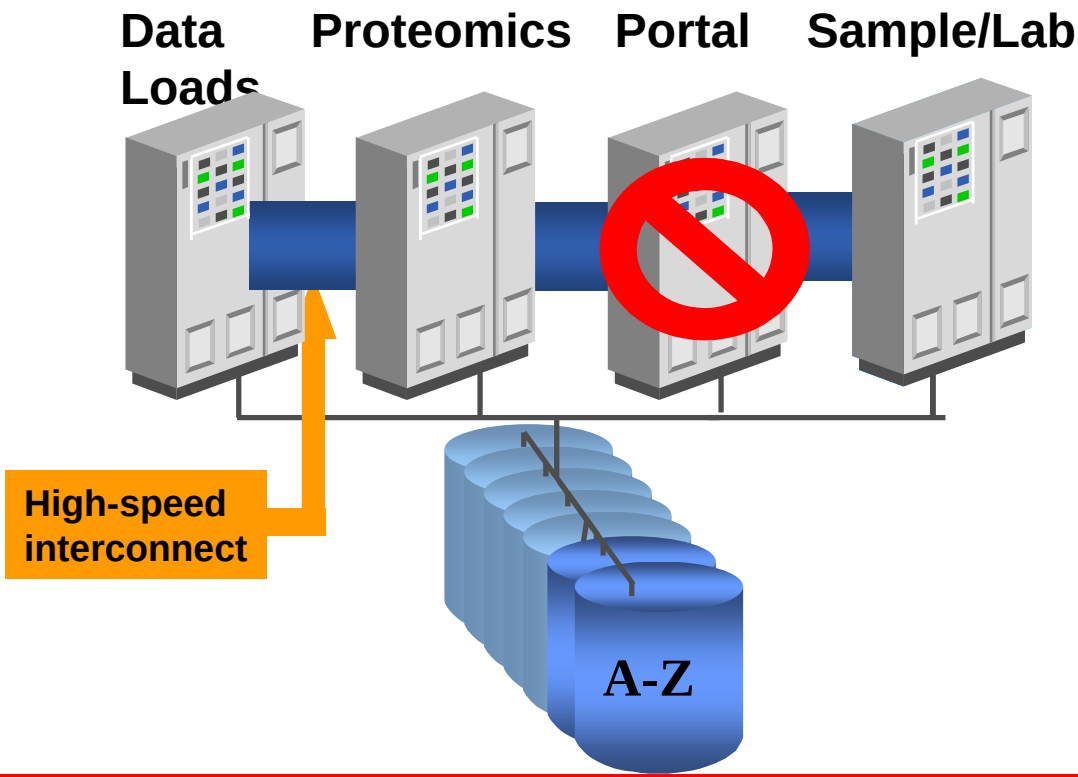
3. Manage Vast Quantities of Data

- Real Application Clusters (RAC)
  - Provides high availability, performance and ease of scalability
- Grid Computing
  - Automated data and computational provisioning
- Automated Storage Management
- Scheduler
- Partitioning
  - Divide and conquer
- Oracle Data Guard
  - Protect data from human or system failures
- Oracle 11g Application Server
  - Provide scalability for middle tier



Real Application Clusters (RAC)

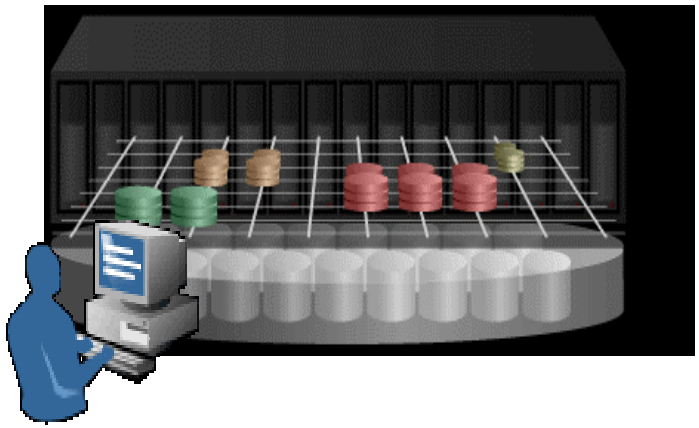
- Start with one server, one database; grow as you grow
- Linear scalability out of the box
- Save on Hardware and Storage costs



- Works with ALL applications
- Fail-over transparent to users
- Easy to administer

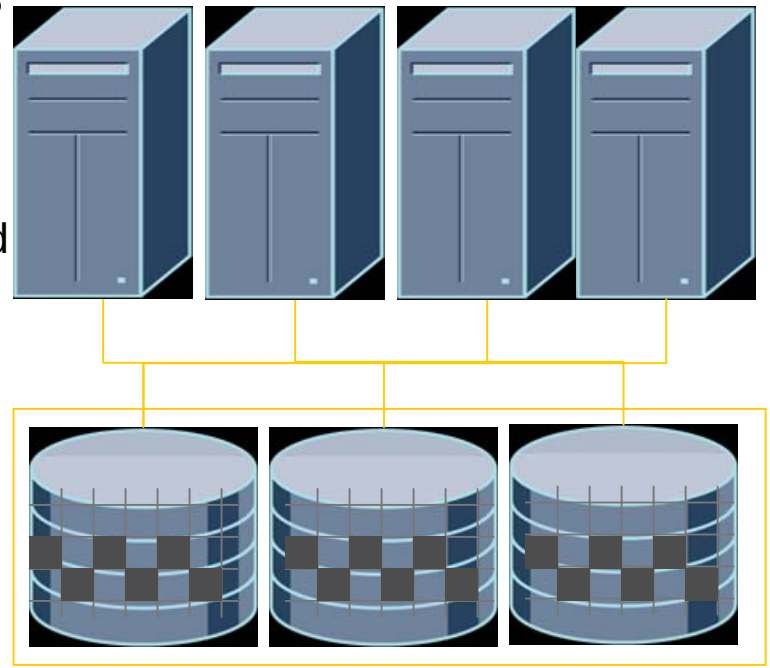
Enterprise Grid Computing

- Mission Critical Quality of Service on Industry Standard, Low Cost Servers
- Integrated clusterware makes RAC easy for everyone
- Grid concepts provided with:
  - Distributed queries, External Tables, Security, RAC, etc.
- Fault tolerant, scales all applications
- Capacity on demand
- Automatic load balancing



Automated Storage Management

- Storage virtualization layer that automates and simplifies the optimal layout of all Oracle database managed disk storage
  - No volumes: just a pool of storage
  - Partitions total disk space into uniform sized megabyte units
  - Efficient, online add/remove of disk with automatic rebalancing
- Configures disk groups to provide data redundancy and optimal layout of all data
- Automatically re-balances and redistributes Oracle Database files to ensure optimal performance across a changed configuration



Automatic Storage Management

Oracle Scheduler

- Provides the ability to schedule a job to run at a particular data and time
- Runs PL/SQL, Java, 3GL, OS Scripts, internal utilities (RMAN)
  - Job classes, priorities, workload windows
  - Integrated with Resource Manager & RAC service framework
- Integrate Platform's JobScheduler with Oracle database
  - Single interface for job scheduling
  - Platform's JobScheduler can create & schedule Oracle database jobs
  - Database jobs can be incorporated into larger job flows
  - Schedule & use resources efficiently for combined database & computational tasks

Partitioning

- Partitioning helps support very large tables and indexes by letting users decompose them into smaller and more manageable pieces called partitions
  - Enables data management and system maintenance at the partition level
  - Improves query performance
  - Implemented without any application modification
- 11g provides following additional support:
  - Hash partitioning of global indexes
  - List partitioning support for index-organized tables (IOTs)
  - Partitioning of IOT's containing large object binaries (LOBs)
  - Automatic global index management

Data Guard

- Protects data from user errors, disasters, storage failures, and planned outages
- Provides an out-of-the box rapid deployment and management interface for a standby database
- Switch instantly to a standby database with no data loss
- Set delay in applying changes to a standby database to allow time to correct human errors
- 11g provides new functionality:
  - Support for rolling upgrades of hardware, operating system, or database version
  - Database authentication prior to shipping or accepting encrypted redo data
  - Compression and check-sum of transmitted data
  - Improved monitoring capabilities

Application Server

- All of Oracle's core middle-tier services are integrated into one product
- Enables customers to build and deploy portals, transactional applications, and business intelligence applications with a single product
- Web Cache stores frequently accessed pages in memory enabling database queries to be processed faster and the database to support more users

"At 22 000 GB the Trace Archive is in the Top Ten UNIX databases in the world. That's not bad for a research organisation of 850 employees in the countryside just outside Cambridge."

"It is possibly the biggest single (acknowledged) scientific RDBMS database in Europe, if not the world."

Martin Widlake, Database Services Manager

Wellcome Trust Sanger Institute

The Winter Corporation database survey 2005 suggests the Trace Archive would rank fifth behind such giants as AT&T, Yahoo and other large international corporations.

Dragon Genomics Center

Dragon Genomics™

TAKARA BIO INC.

- High-Level Project Goals
  - Manage data throughout every step of a complicated process
  - Create a laboratory information management system (LIMS) enabling large scale sequencing
  - Provide reliable back up and recovery of vast amounts of data
- Key Benefits
  - Provided easy access and management for vast amounts of data
  - Ensured scalability needed to accommodate future growth

- Oracle Environment
  - Oracle Database Enterprise Edition
  - Oracle9iAS Enterprise Edition
- "We trust Oracle in its ability to run terabyte-class databases in clustered environments with high availability. And we're pleased to say that Oracle has not disappointed us.“ - Toru Suzuki, Project Manager, Dragon Genomics Center, Takara Bio Inc.

ORACLE

Genentech, Inc.

- Leading biotech company
  - Over 2 TBs of data in Oracle
  - Oracle serves as a centralized information resource for gene searching and database cross-referencing.
  - Oracle used for the entire pipeline from research to clinical data to manufacturing and sales applications.
- Key Advantages of Oracle
  - Improved performance
  - Greater reliability
  - Genentech's corporate goal is 99.999% availability in a 24x7 environment

- Oracle Environment
  - Oracle 9i database
  - Real Application Clusters

- Oracle9i Real Application Clusters provide the foundation for the scalable and highly available database infrastructure we require to meet our growing data demands in all areas of our business.“ -Scooter Morris, Genentech, Inc.

San Diego Supercomputing Center

"In the beginning, we considered using MySQL, Oracle, and another database. But when we evaluated our project needs over the next ten years and realized that our database could grow to terabytes, we decided we needed a scalable database and one that was reliable. We didn't want to be forced to change databases in the middle of the project. …. "We do not need a lot of DBAs to maintain the database."

Joshua Li, Senior Computational Scientist, University of California, San Diego, Supercomputing Center

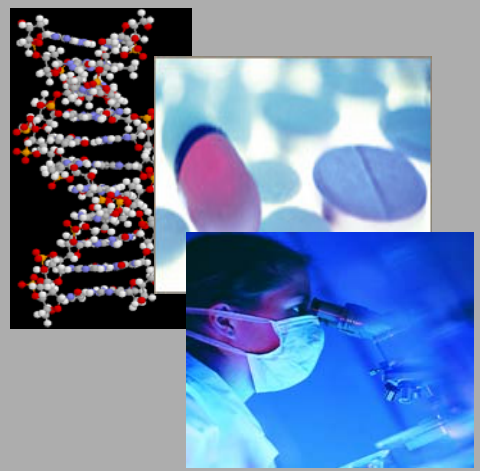
Systemwide, SDSC relies on only three DBAs to run over 40 Oracle databases.

Bioinformatics Center Institute for Chemical Research Kyoto University

The Bioinformatics Center Institute for Chemical Research Kyoto University is leading biotechnology research thanks to its comprehensive studies in various areas, including the life sciences, information sciences, chemistry and physics.

“In order to manage this massive amount of genetic information and to operate efficiently, it is essential to have a platform with paramount stability. Our web site receives accesses from all over the world continuously, 24 hours a day. In order to offer the latest information under such circumstances, performance is also an issue. In this sense, the Oracle Database was the most appropriate since it can handle this enormous amount of data in a fast and stable manner, 24 hours a day.”

– Professor and Director Minoru Kanehisa, Bioinformatics Center Institute for Chemical Research Kyoto University

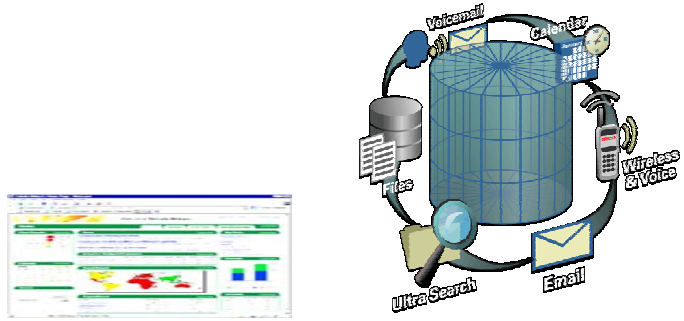


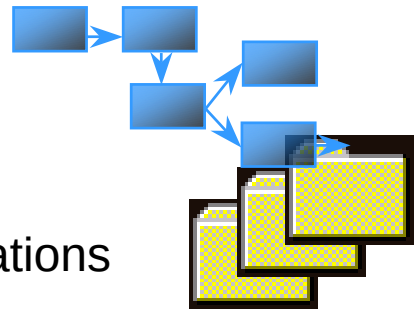
Collaborate Securely

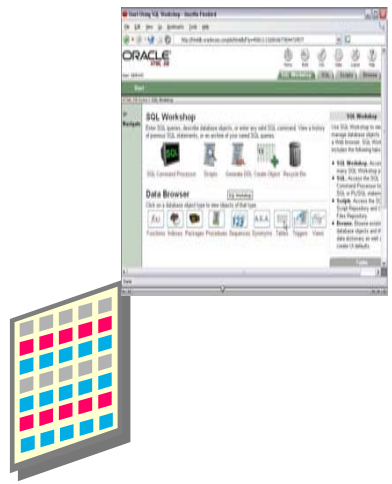
ORACLE

4. Collaborate Securely

- Oracle Collaboration Suite
  - Integrated communications
- Oracle 11gAS Portal
  - Build personalized portals
- Oracle Workflow
  - Automate laboratory and business processes
- Oracle 11gAS Files
  - Enable content management and collaboration
- Applications Express (formerly HTML DB)
  - Develop and deploy database-centric Web applications
- Virtual Private Database
  - Different users have unique access privileges
- Oracle Data Vault
  - Solution for ensuring data is secure
- Oracle Secure Backup
  - Automated encrypted data to tape
- Auditing
  - Create audit trail to facilitate FDA compliance
- Oracle 11gAS Web Services
  - Standard way to collaborate through the Web

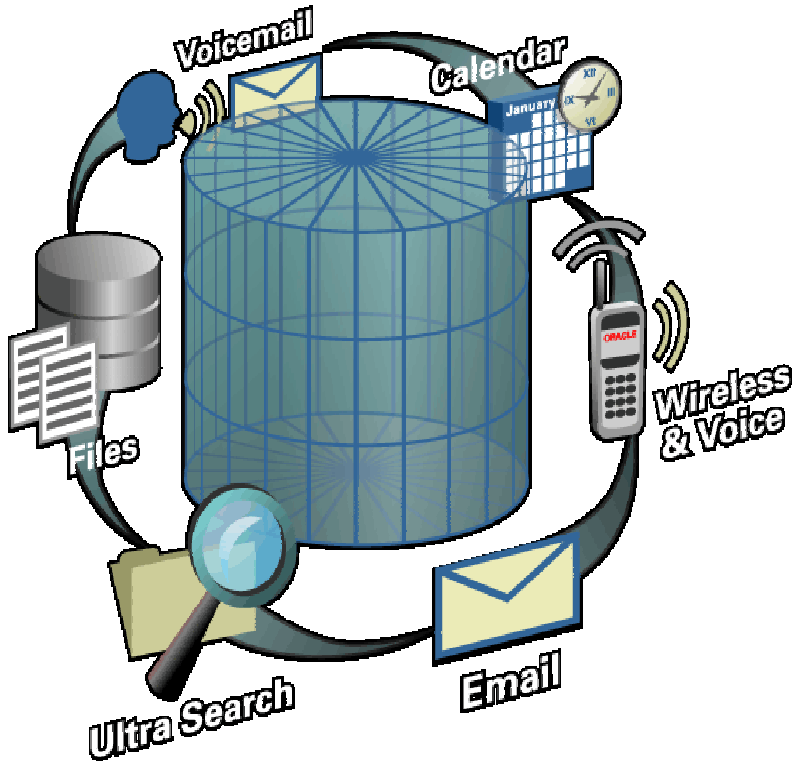






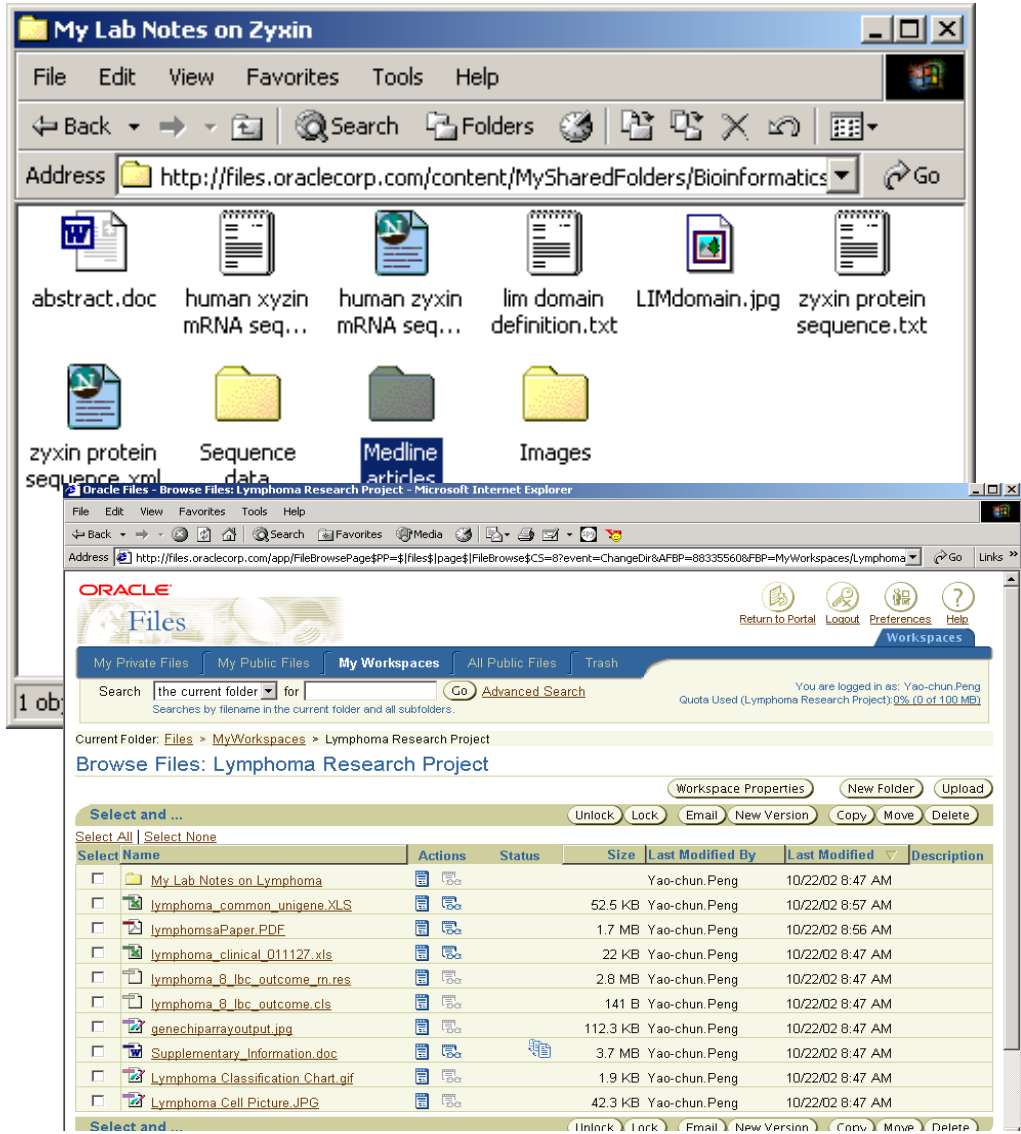
Oracle Collaboration Suite

- Integrated communications
- Single enterprise search across all repositories
- Flexible access



Oracle Files

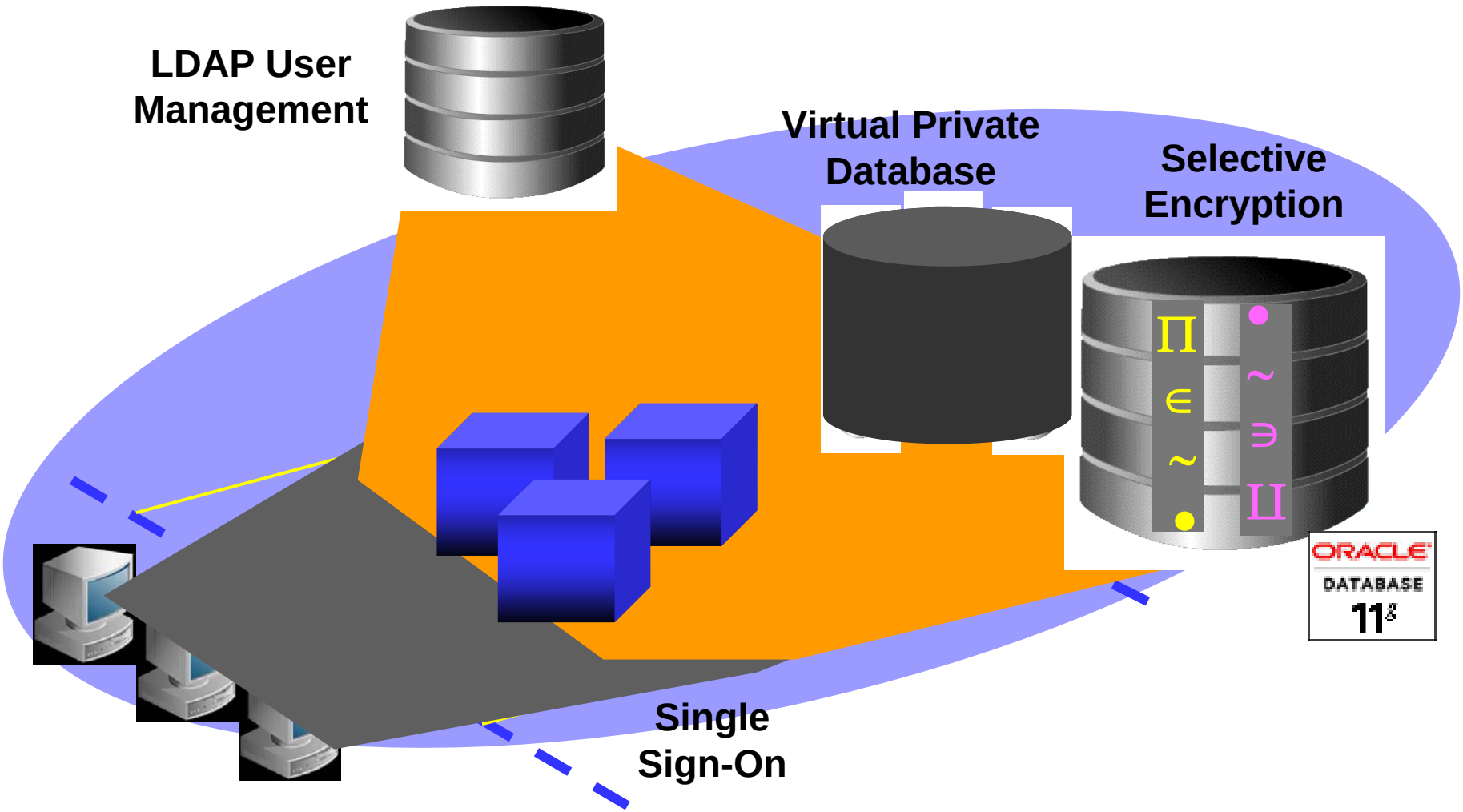
- Collaborate easily and securely via workspaces
  - Groups of users can be created with different project access privileges
- Protect your data from with role-based security
- Oracle Files supports HTTP/WebDAV, FTP, SMB, AFP, and NFS
- Stop sending/receiving email attachments



Oracle Portal

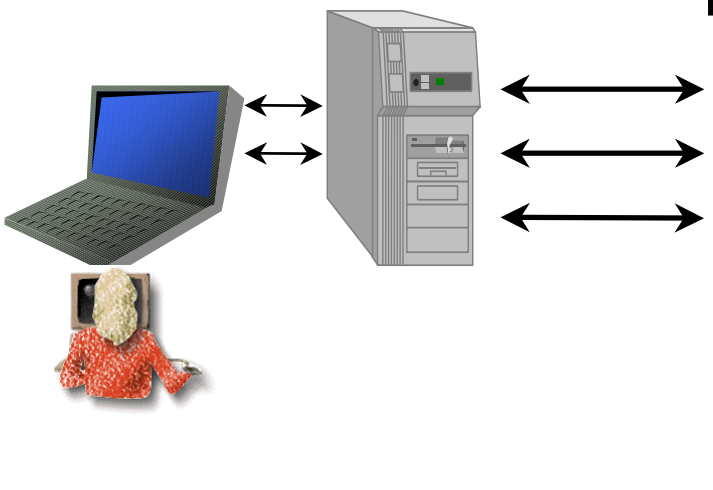
- Rich, declarative environment
  - Create Web interfaces, publish and manage information, access dynamic data, and customize with extensible J2EE framework
- Connect researchers and collaborators with the information they need
- Flexibility to create views tailored to each community

Security



Oracle Label Security Example

User	Label (Level :: Compartment :: Group)
Dr. Murphy	Sensitive :: Orthopedic, Acute :: Active



Row Labels

Identifiable	Ambulatory	Dep
Identifiable	Orthopedic	Active
Sensitive	Radiology	Ret
Confidential	Disease	Active
Sensitive	Orthopedic	Ret
Sensitive	Acute	Active

Levels
Hierarchical

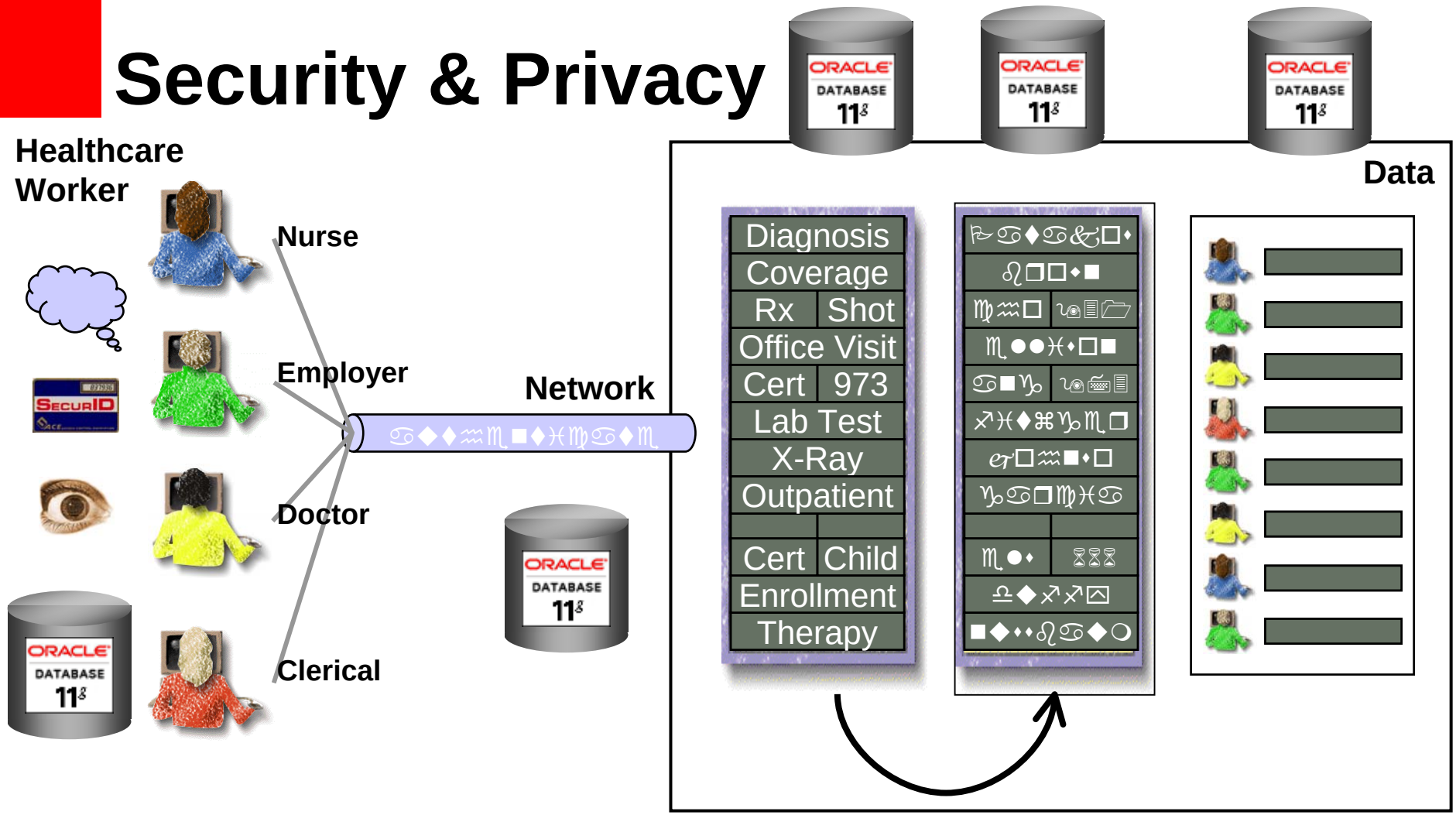
Compartments
Non-Hierarchical

Groups
Hierarchical

Levels : Confidential → Sensitive → Identifiable

Groups : Active → Retired → Departed

Security & Privacy



Oracle11g Unbreakable Security

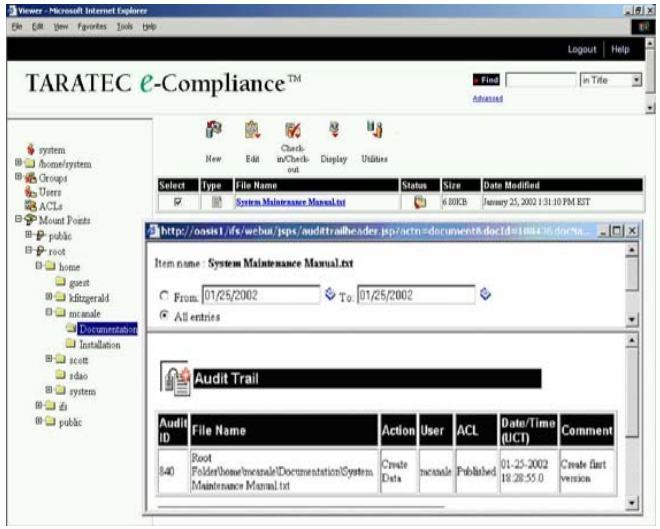
Complete data protection
Manage user access
Detect data misuse with Auditing
Facilitate regulatory compliance (HIPPA, 21 CFR PART 11)

Security Evaluations	Oracle	Microsoft	IBM
US TCSEC, Level B1	1	-	-
US TCSEC, Level C2	1	1	-
UK ITSEC, Levels E3/F-C2	3	-	-
UK ITSEC, Levels E3/F-B1	3	-	-
ISO Common Criteria, EAL-4	4	-	-
Russian Criteria, Levels III, IV	2	-	-
US FIPS 140-1, Level 2	1	Failed	-
TOTAL	**15**	**1**	**0**

Taratec e-Compliance™

TARATEC
Pairing e-Technologies and Regulations in Life Sciences

- Taratec e Compliance™
  - Built specifically to supports FDA 21 CFR Part 11 Compliance
  - Designed for Life Sciences Data & File Management
- Features
  - Versioning, Advance Searching, Check-in/Check-Out
  - Integrated storage of files from any source
  - Universal access through Web browser
  - Complete Audit Trail of File Operations

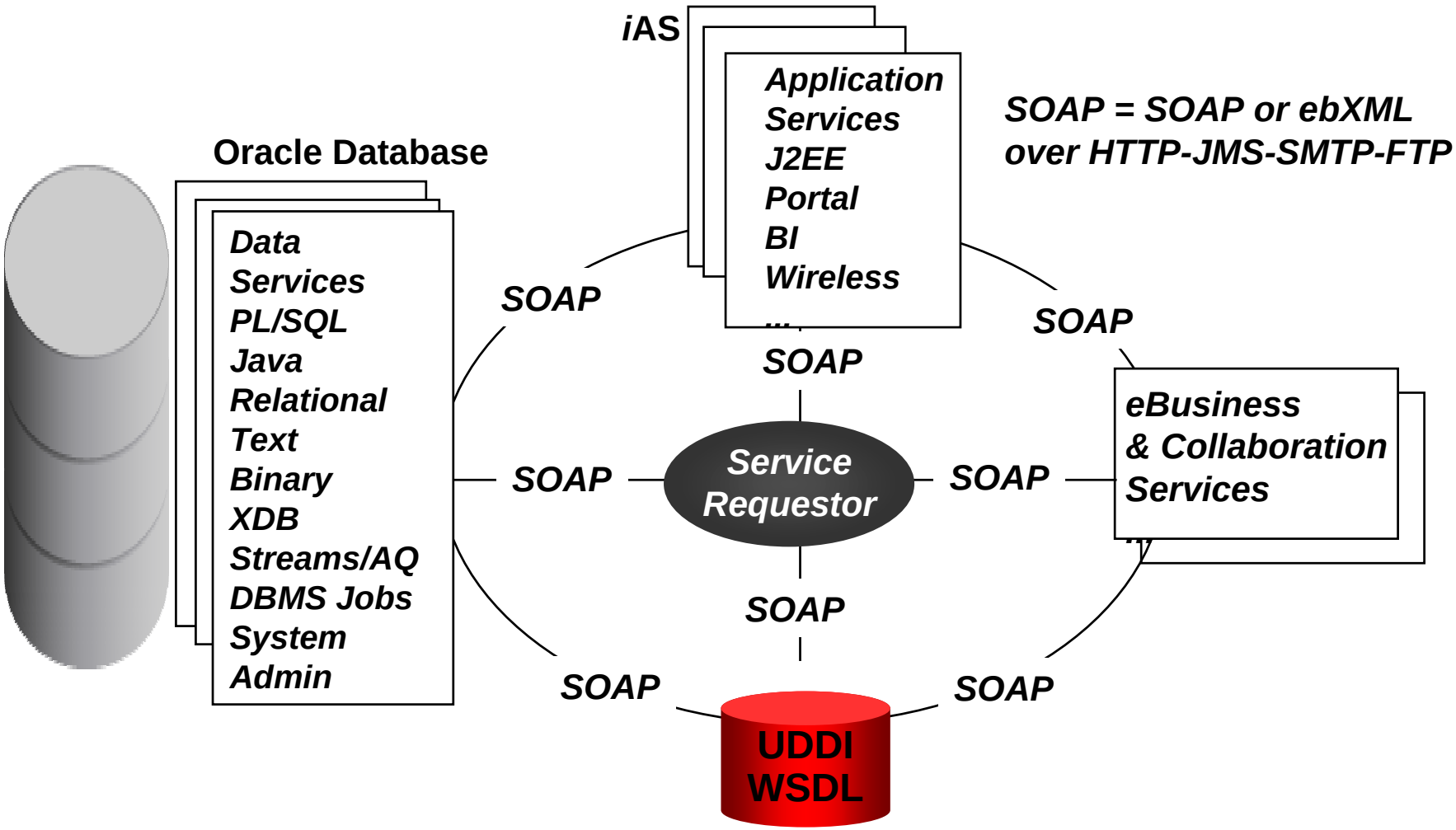


"With Oracle as the foundation, we were able to develop a solution that can secure a vast array of file-based data with vault like security." - Bill Gargano, President and COO Taratec Development Corporation

University of California San Diego School of Medicine

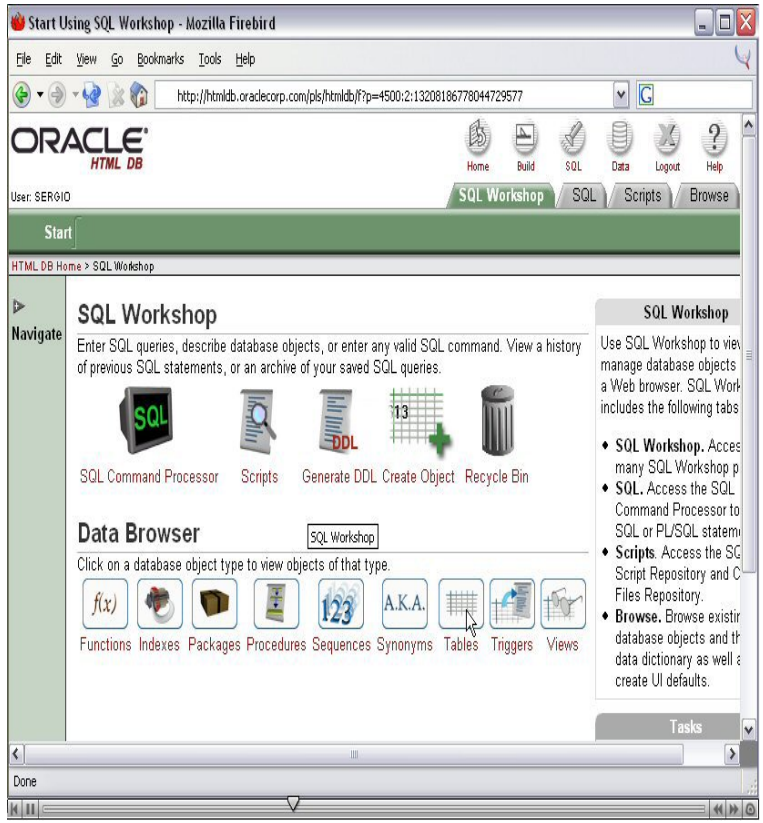
- The Patient Centered Access to Secure Systems Online (PCASSO)
  - 178,000 Medical Records
  - Provides trusted access to a patient's health information from healthcare providers over the Internet
  - Oracle Label Security & Virtual Private Database
    - The security is locked to the data and therefore can't be subverted.
    - No application coding needed to implement security.

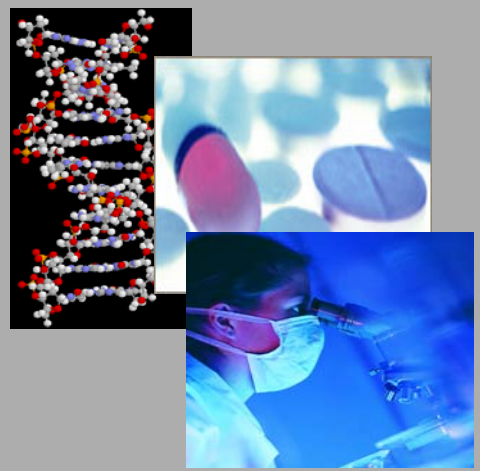
Integrated Data and Web Services Platform



Oracle Applications Express (HTML DB)

- Tool for development and deployment of database-centric Web applications
- Features development with design themes, navigational controls, form handlers and flexible reports
- Using a Web browser, users can quickly build database driven Web application
- Deploys data in spreadsheets and personal databases to the Web



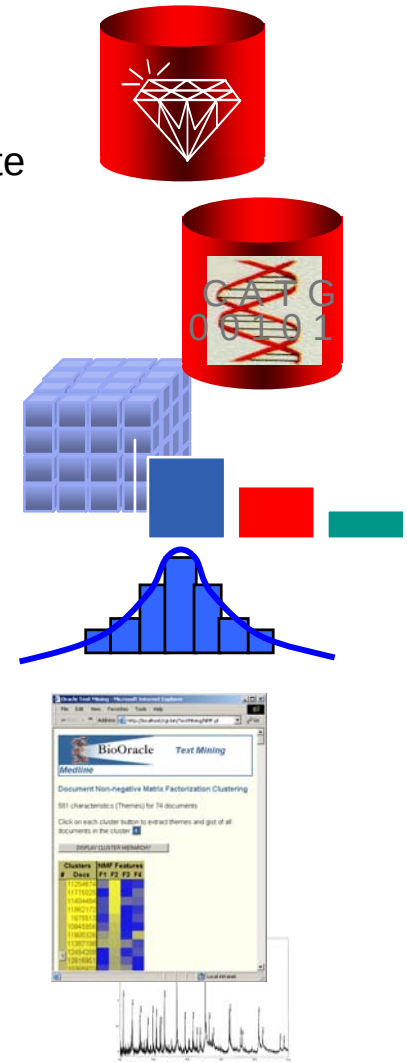


Find Patterns and Insights

ORACLE

5. Discover Patterns and Insights

- Oracle Data Mining
  - Find relationships and clusters
    - Naïve Bayes, Adaptive Bayes Networks, Decision Trees, Attribute Importance, Association Rules, K-Means, O-Cluster, SVM, NMF algorithms
- BLAST—Basic Local Alignment Search Technique
  - SQL queries can pre-filter & post-process BLAST results
- Oracle Discoverer, OLAP, Oracle BI EE
  - Interactive query & drill-down
- Statistics
  - Perform statistics in Oracle
    - For example, summary statistics, hypothesis tests, cross-tab statistics, distribution tests, correlations, linear regression
- Oracle Text
  - Search, index, classify and cluster documents
- IEEE Float support
- Table Functions
  - Implement complex algorithms within the database



5. Discover Patterns and Insights

Life Sciences data

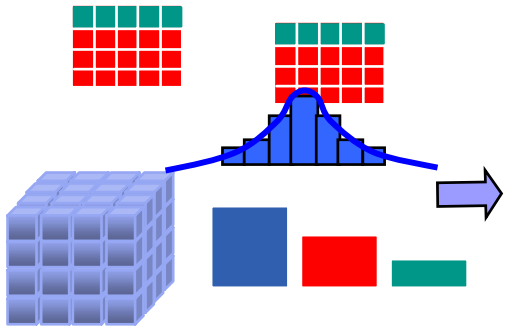
Functional Genomic Databases

Clinical Databases

Proteomics Database

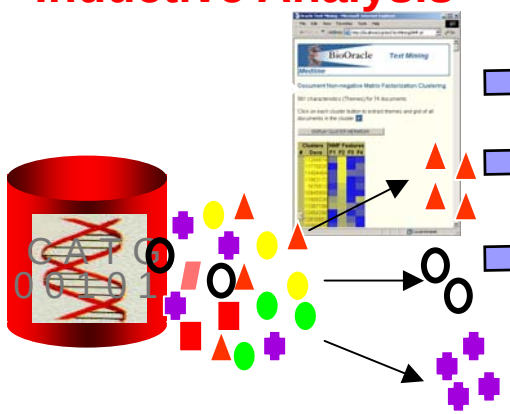
Pharmacological databases

Deductive Analysis



Answer complex questions about the relationships in genomic, clinical and pharmacological data

Inductive Analysis



Finding relationships for classification, class discovery and prediction

Regular Expression Searches

- A powerful method of describing both simple & complex patterns for searching & manipulating
- A multilingual regular expression support for SQL & PL/SQL string types
- Follows POSIX style Regexp syntax
- Support standard Regexp operators
- Includes common extensions such as case-insensitive matching, sub-expression back-references, etc.
- Compatible with popular Regexp implementations like GNU, Perl, Awk

Regular Expression Searches Quote

"Thanks to Oracle 10g's Regular Expressions (RE) query support, it's no longer necessary to export data from the database, process it with a RE enabled tool and then import the data back into the database. Now, RE processing can be handled with a single query." - Marcel Davidson, Head of Database Administration, Myriad Proteomics

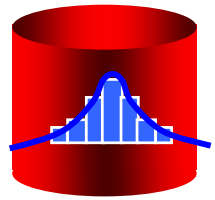
Quotes

- *"Support for regular expressions in SQL and PL/SQL is one of the most exciting features of Oracle Database 10G. Oracle has long supported the ANSI-standard LIKE predicate for rudimentary pattern matching, but regular expressions take pattern matching to a new level. They provide a powerful way to select data that matches a pattern, as well as to manipulate, rearrange, and change that data."*

Oracle Regular Expressions Pocket Reference, O'Reilly Sept. 2003

11g Statistics & SQL Analytics

FREE (Included in Oracle SE & EE)



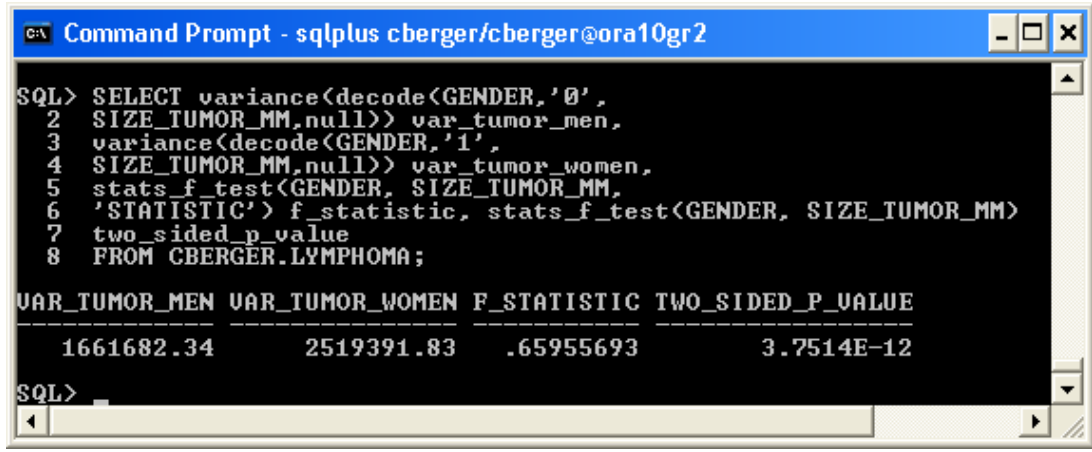
- Ranking functions
  - rank, dense_rank, cume_dist, percent_rank, ntile
- Window Aggregate functions
  (moving and cumulative)
  - Avg, sum, min, max, count, variance, stddev, first_value, last_value
- LAG/LEAD functions
  - Direct inter-row reference using offsets
- Reporting Aggregate functions
  - Sum, avg, min, max, variance, stddev, count, ratio_to_report
- Statistical Aggregates
  - Correlation, linear regression family, covariance
- Linear regression
  - Fitting of an ordinary-least-squares regression line to a set of number pairs.
  - Frequently combined with the COVAR_POP, COVAR_SAMP, and CORR functions.

Note: Statistics and SQL Analytics are included in Oracle Database Standard Edition

- Descriptive Statistics
  - average, standard deviation, variance, min, max, median (via percentile_count), mode, group-by & roll-up
  - DBMS_STAT_FUNCS: summarizes numerical columns of a table and returns count, min, max, range, mean, stats_mode, variance, standard deviation, median, quantile values, +/- n sigma values, top/bottom 5 values
- Correlations
  - Pearson's correlation coefficients, Spearman's and Kendall's (both nonparametric).
- Cross Tabs
  - Enhanced with % statistics: chi squared, phi coefficient, Cramer's V, contingency coefficient, Cohen's kappa
- Hypothesis Testing
  - Student t-test , F-test, Binomial test, Wilcoxon Signed Ranks test, Chi-square, Mann Whitney test, Kolmogorov-Smirnov test, One-way ANOVA
- Distribution Fitting
  - Kolmogorov-Smirnov Test, Anderson-Darling Test, Chi-Squared Test, Normal, Uniform, Weibull, Exponential
- Pareto Analysis (documented)
  - 80:20 rule, cumulative results table

In-Database Statistics

- Powerful classical statistical functions
- Simpler architecture
- FREE vs. expensive SAS alternative



"Our experience suggests that Oracle 10g Statistics and Data Mining features can reduce development effort of analytical systems by an order of magnitude."

Sumeet Muju

Senior Member of Professional Staff, SRA International

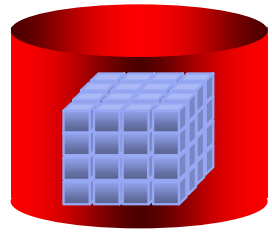
(SRA supports NIH projects)

```
248 rows selected.

SQL> select peak_id peak,avg(decode(E.sample_group,'CNS',s.intensity,null)) avg_CNS, avg(decode(E.sa
mple_group,'ND',s.intensity,null)) avg_ND, stats_ks_test(E.sample_group,s.intensity,'STATISTIC') ks_
stat, stats_ks_test(E.sample_group,s.intensity) ks_p_value,  stats_t_test_indep(E.sample_group,s.int
ensity) t_test_p_value, avg(subs_mass) AVG_MASS from exp_descriptor E, celd_spectrum s where E.exp_i
d = s.exp_id and E.chip_id = s.chip_id and E.spot_number = s.spot_number and (sample_group = 'CNS' o
r sample_group='ND')  Group By peak_id order by  stats_t_test_indep(E.sample_group,s.intensity);

      PEAK    AVG_CNS     AVG_ND    KS_STAT KS_P_VALUE T_TEST_P_VALUE   AVG_MASS
---------- ---------- ---------- ---------- ---------- -------------- ----------
       178  1.3314339 2.17817187 .673333333 7.2556E-16     2.6544E-17 5952.91674
       181  5.0996028 7.89194275 .626666667 8.4480E-14     1.4848E-14  6075.9581
       180 2.27649538 3.47917519 .606666667 5.8453E-13     8.8539E-14 6055.14643
       182 1.82166302 2.70982458 .586666667 3.7986E-12     1.7684E-13 6093.52256
       112 1.43756807 .415726202         .6 1.0984E-12     4.5081E-13 4033.66603
       179 .470304995  .71366692 .546666667 1.3289E-10     6.0678E-13 5976.18528
       162  .32065549 .488111947 .606666667 5.8453E-13     6.3174E-13 5384.91078
       176 1.71447936 2.70554235 .553333333 7.4775E-11     1.7747E-12 5914.53224
       185 .336895407 .472142857         .55 9.9772E-11     1.9222E-12 6260.71013
       186 .401995708 .562915017 .506666667 3.6175E-09     2.1445E-12 6281.69466
       177  2.3623861 3.80160199 .586666667 3.7986E-12     4.1808E-12 5933.83033
```

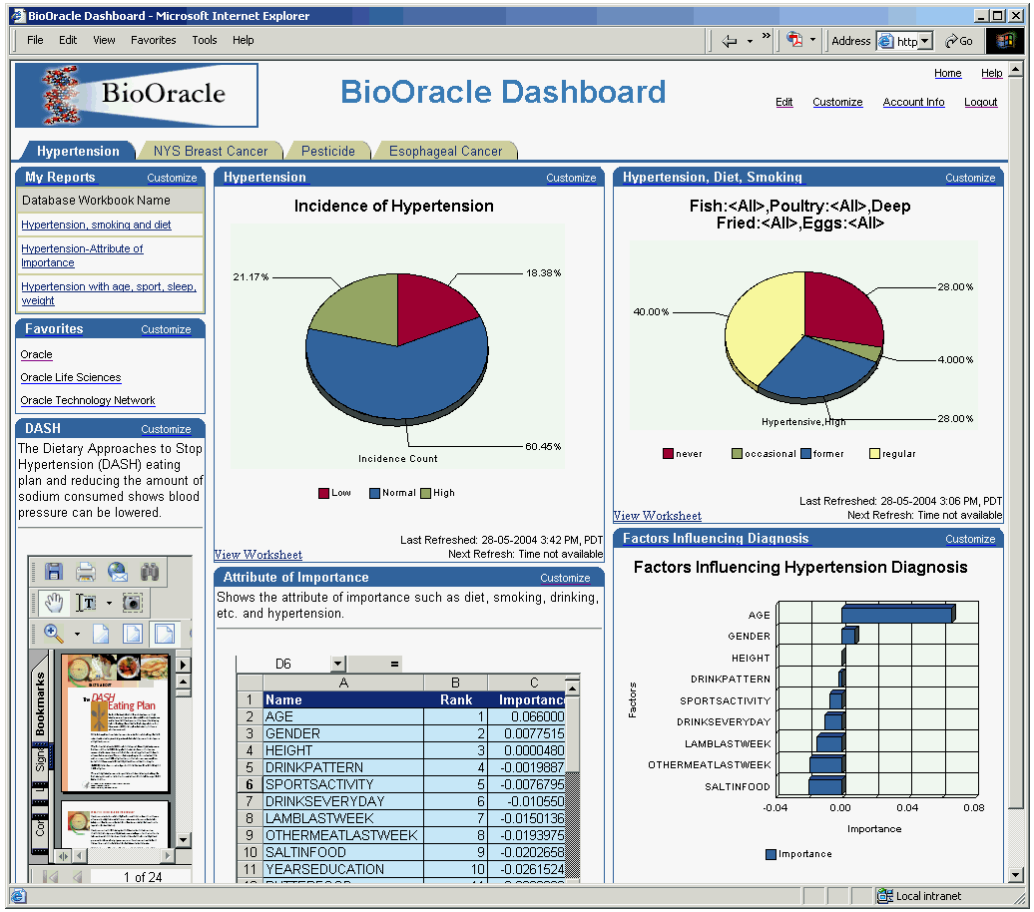
Oracle OLAP



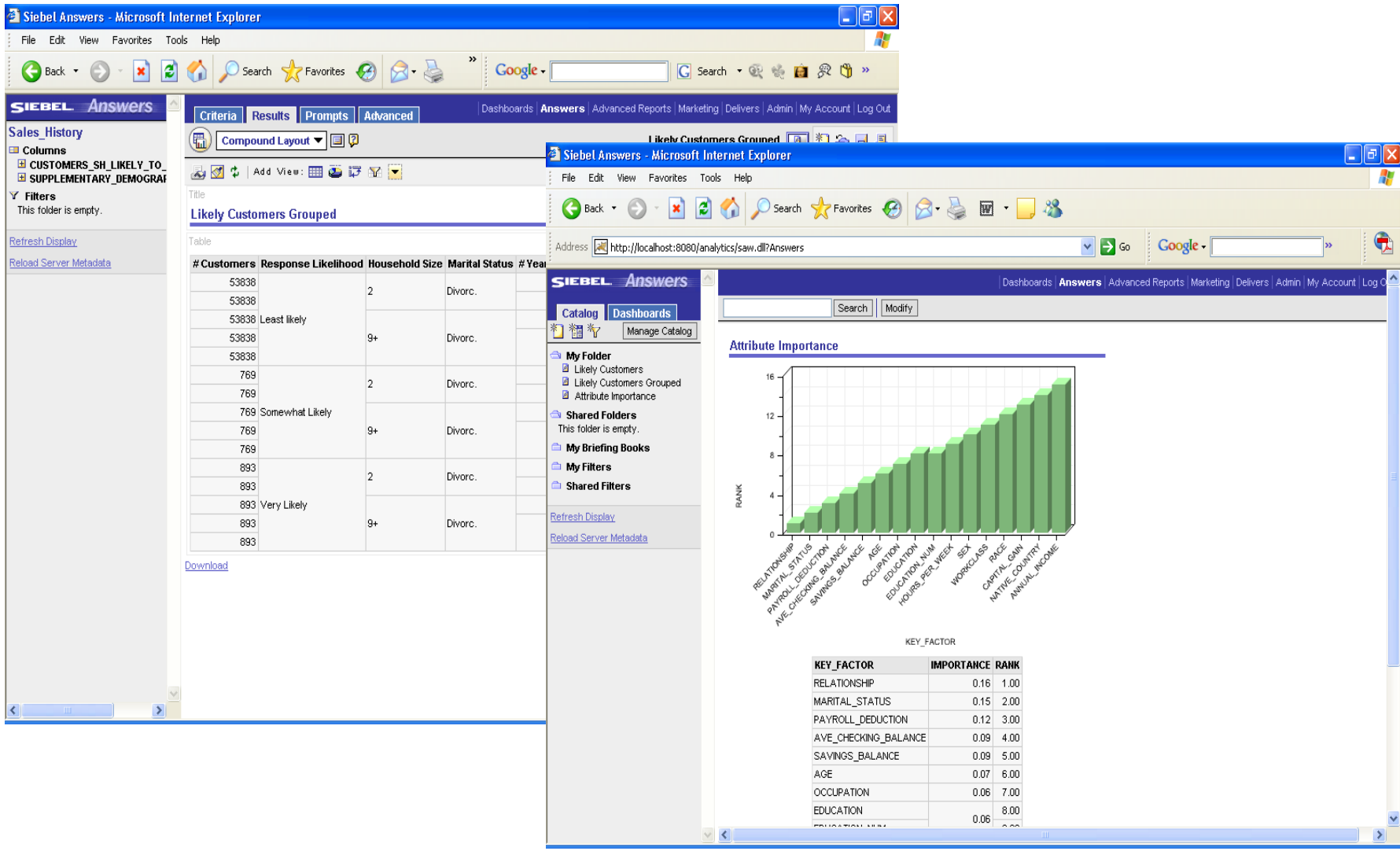
- Build multi-dimensional data cubes to enable slicing and dicing of data
- New 10*g* functionality includes:
  - Enhanced OLAP capabilities using the database's built in analytical workspaces
  - PL/SQL and XML interfaces for creation of workspaces based on cubes and dimensions defined in the OLAP catalog
  - Cross-tabular analysis capabilities support the aggregation of attributes within a dimension
  - Parallel capabilities are provided for AGGREGATE and SQL IMPORT operations, making it faster to load and materialize analytical workspaces from relational data

Oracle Discoverer

- Ad-hoc query & reporting
- Web publishing
- Discoverer is included with Oracle Application Server **Enterprise Edition**



Oracle BI EE

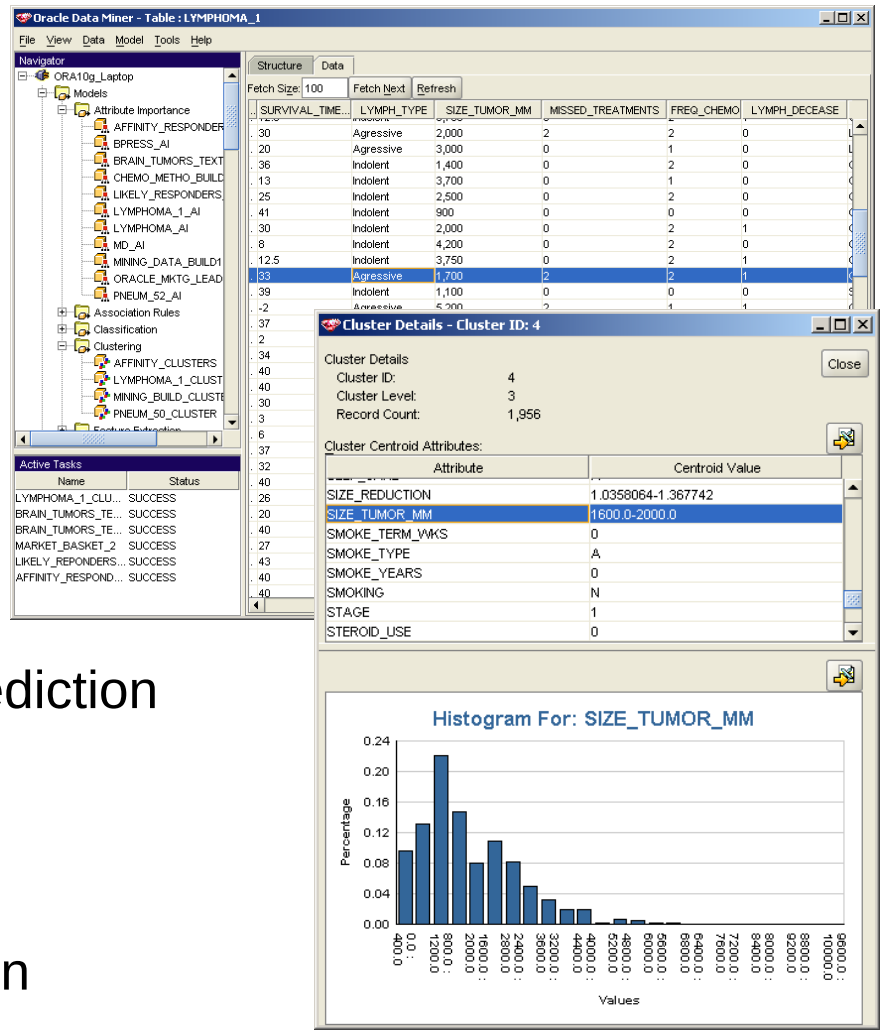


IEEE Floating Point

- Support for industry standard treatment of numbers & precision
- Critical for compute intensive operations
- Faster performance

Oracle Data Mining

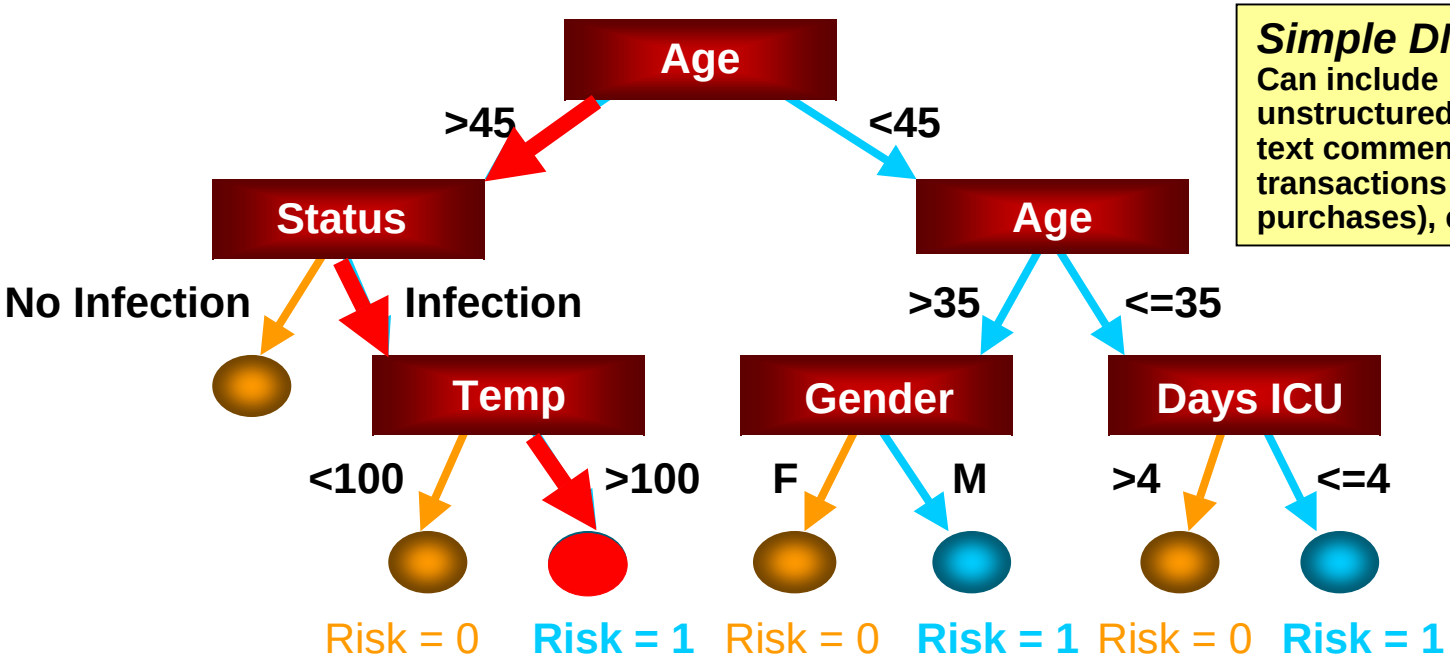
- Oracle mining platform
  - PL/SQL API
  - Java API
  - Oracle Data Miner (GUI)
  - Spreadsheet Add-In for Predictive Analytics
- Range of algorithms
  - Structured & unstructured data
  - Attribute importance
  - Classification, regression & prediction
  - Anomaly detection
  - Association rules
  - Clustering
  - Nonnegative matrix factorization
  - BLAST



Oracle Data Mining 11g

Decision Trees

- Classification, Prediction, Patient “profiling”



Simple DM model:
Can include unstructured data (e.g. text comments), transactions data (e.g. purchases), etc.

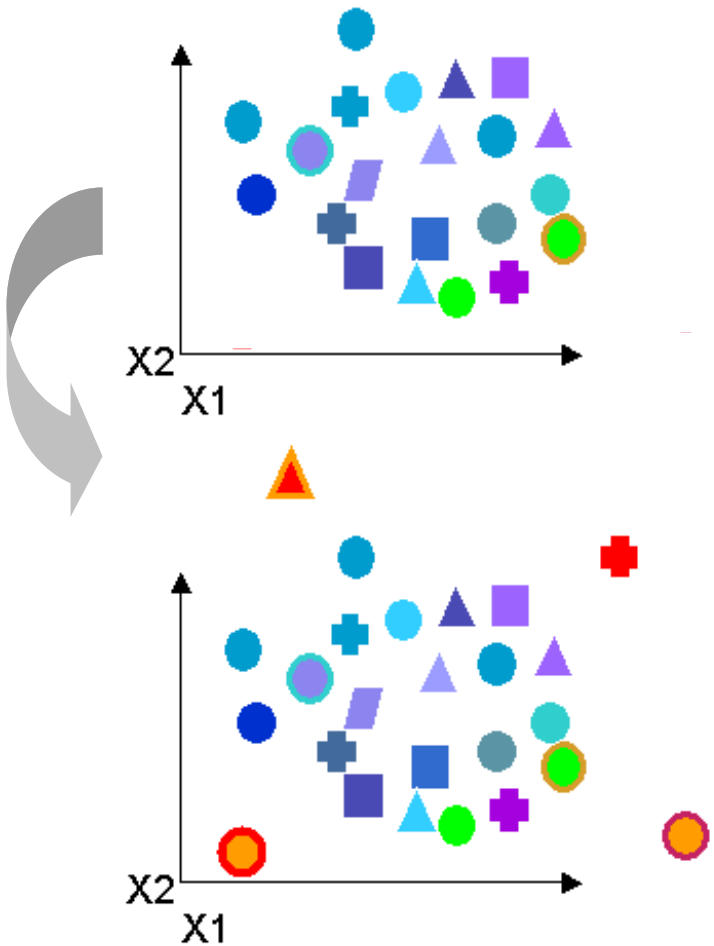
IF (Age > 45 AND Status = Infection AND Temp = >100)
THEN P(High Risk=1) = .77 Support = 250

Oracle Data Mining 11g

Anomaly Detection

Problem: Detect rare cases

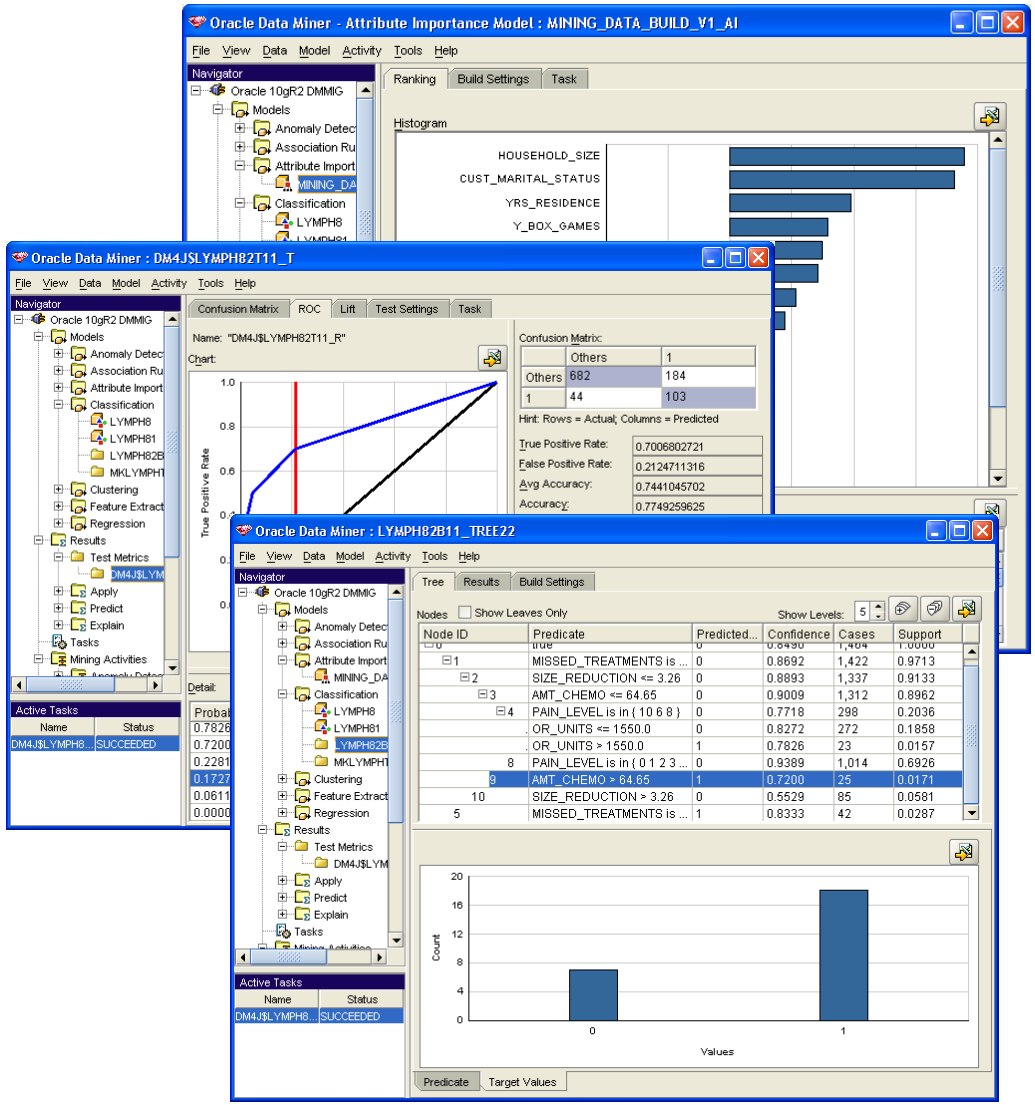
- "One-Class" SVM Models
  - Fraud, noncompliance
  - Outlier detection
  - Network intrusion detection
  - Disease outbreaks
  - Rare events, true novelty



Oracle Data Mining

Easy to Use

- GUI for building, evaluating, and applying ODM models
  - Wizards approach
  - Mining Activity Guides
- Generate SQL & Java code to *"operationalize"* applications
  - Integrate data mining "insights" into other BI tools and applications



Linear Algebra Solvers
BLAS & LAPAK

- PL/SQL interfaces to a set of routines that perform common numerical linear algebra operations on memory-resident vectors and matrices using state-of-the-art algorithms
  - BLAS
  - LAPACK
- Routines used for developing statistics, data analysis, data mining, and life sciences applications

Oracle Data Mining in the Life Sciences

Gene expression analysis

- Problem
  - Given thousands of gene expression values for each patient, can a small subset of the expressions be identified that can be used to distinguish one type of leukemia from another?
- Solution
  - Apply ODM's Attribute Importance algorithm to the data to decrease the size of the problem
  - Build an Adaptive Bayes Network Classification model to predict disease type from the gene expressions

Oracle Data Mining in the Life Sciences

Gene expression analysis

Top Genes (of ~7000) for Classifying Leukemia

Gene Expression	Relative Importance
V00594_s_at	0.298955976210004
D43950_at	0.292217965904811
U34038_at	0.227177556507829
J03827_at	0.227177556507829
U64863_at	0.227177556507829
S85655_at	0.175469338594625
L07758_at	0.17031674247889
U19345_at	0.17031674247889
U89336_cds4_at	0.125995412839
U79295_at	0.125995412839
HG311-HT311_at	0.125995412839
V00599_s_at	0.125995412839

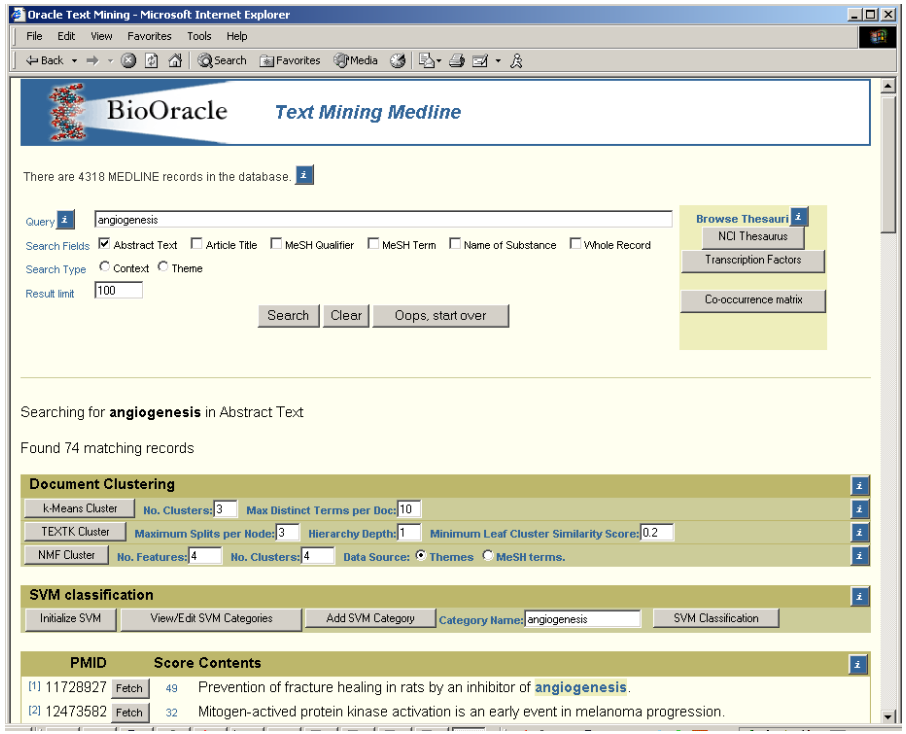
Data Mining Quotes

“Using InforSense discovery workflows built upon the world leading Oracle data mining, text mining and R&D Database functionality, researchers and organizations can now automate large scale and complex knowledge discovery and management activities with performance and reliability.”
- Yike Guo, CEO InforSense

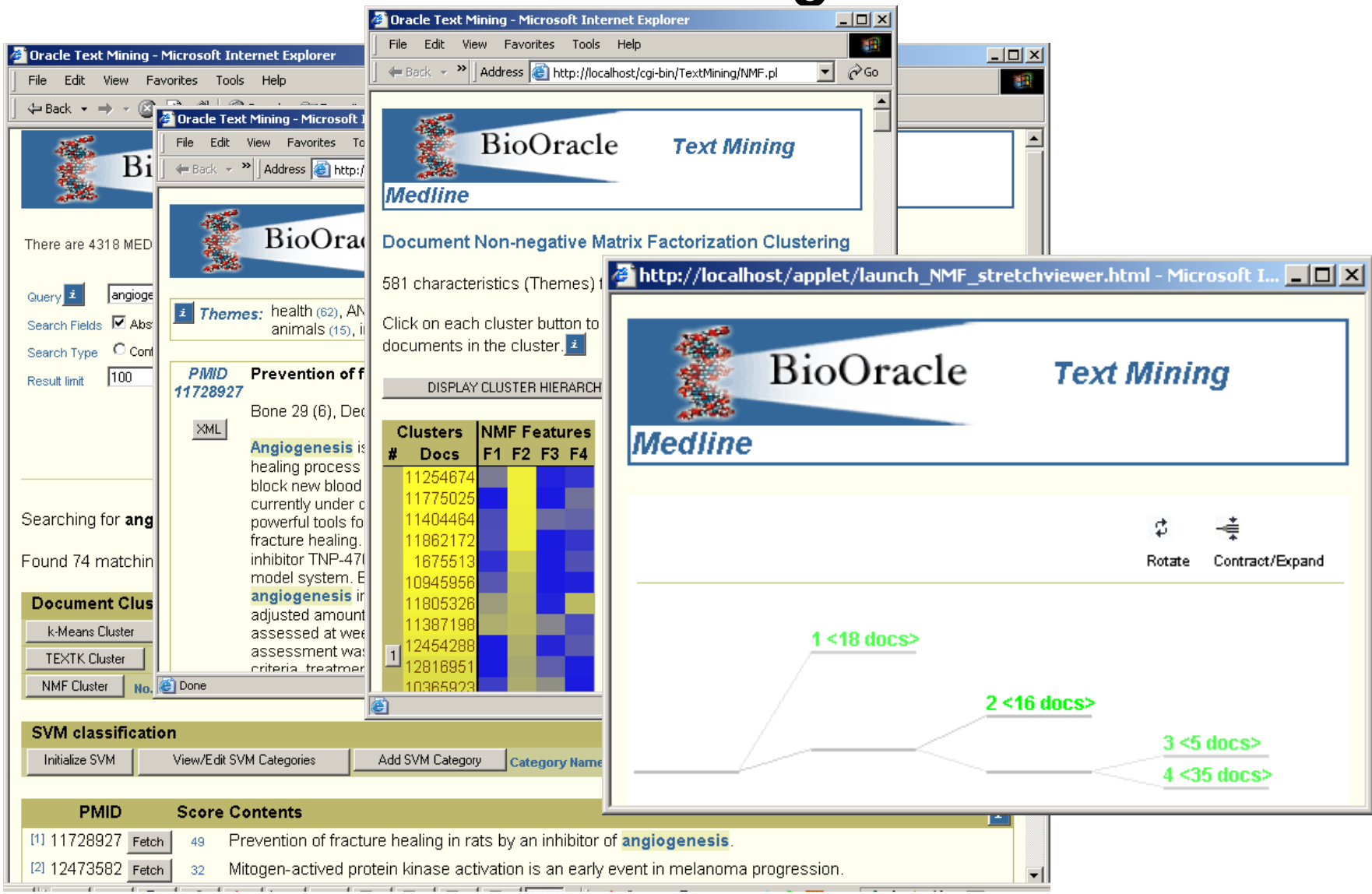
Support Vector Machines gives Oracle Data Mining a very powerful tool for pattern discovery in very wide data sets. Moreover, its ease of use and efficiency, based on the effective parameter tuning and model optimization, enables experienced and inexperienced users to get really great results.“
- Angela Uvarov, Department of Computer Science and Statistics, URI

Oracle Text & Text Mining

- Classify & cluster documents (using data mining algorithms)
  - Find “clusters” of similar documents
  - Develop applications to classify documents likely to be “of interest” based on other example documents



Oracle Text & Text Mining



Walter Reed Medical Center

- Improving clinical outcomes

Decision Support Center Data Warehouse

- Over 80 Gigabytes
- 5 years patient appointments
  - Jan 98 to Apr 04
- 1,820,575 patients
- Outpatient appointments
  - 14,815,520 appointments
  - 18,847,911 diagnoses
  - 9,035,022 procedures
- Inpatient visits
  - 116,173 visits
  - 38,419 diagnoses
- Out of Network visits
  - 6,356,525 visits

Attribute Importance

Diabetic Patients Target = Scorecard

Name	Rank	Importance
PROVIDER_CLASS	1	0.131911
MEPRS_CODE	2	0.060813
DRUG_GROUP	3	0.041141
PCM_MTF	4	0.030109
DMIS_CODE	5	0.028885
VISIT_COUNT9902	6	0.021486
OVERALL_LDL_SUCCESS	7	0.015959
OVERALL_A1C_SUCCESS	8	0.007874
PREVENT_HOSPITAL_COUNT9902	9	-0.00115
PT_SEX	10	-0.00172
ERVISIT_COUNT9902	11	-0.00195
ACV_CODE	12	-0.00221
PCM_CLASS	13	-0.00305
HOSPITAL_COUNT9902	14	-0.00418
MIL_BR	15	-0.00451
MIL_STATUS	16	-0.00476
PT_AGE	17	-0.0097

Table Functions

- Allows researchers to implement their own compute intensive algorithms in PL/SQL in the database or Java, C or C++ outside the database
- Accepts a set of rows as input, provides a set of rows as output, and seamless use with applications
- Benefits include:
  - Integration of additional functionality with the database
  - Making new functionality accessible via SQL
  - Utilization of database functionality, e.g. procedural logic, parallelism and pipelining

Analytical Pipelines

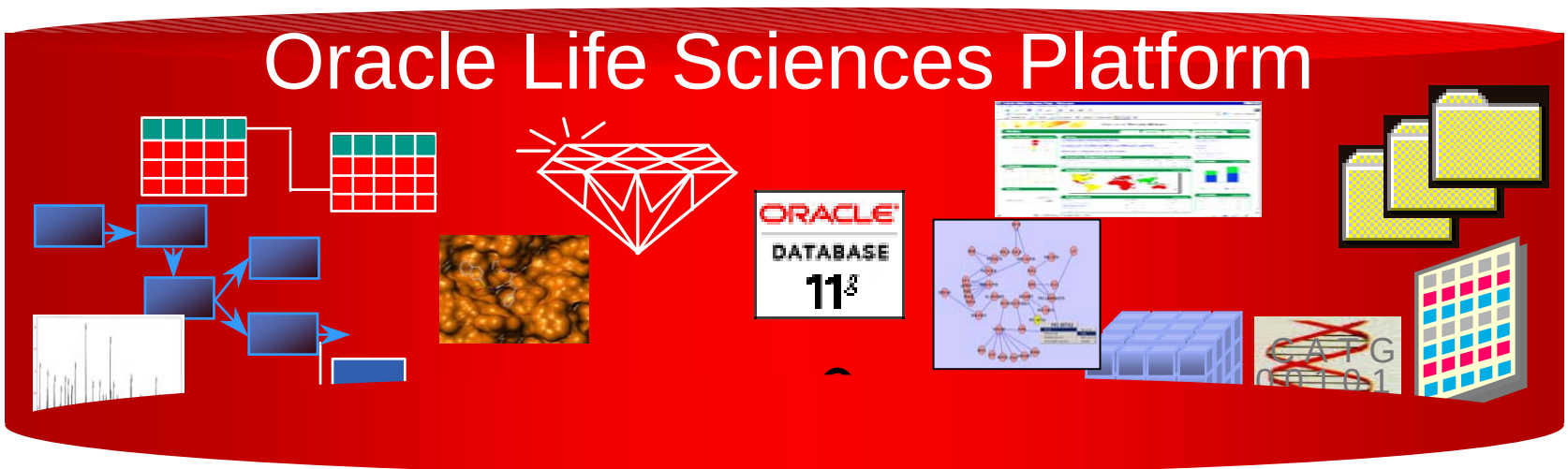
Biological/ Clinical Experiments

Instruments

Data Pre-Processing

Analytical Algorithms

Interpretation of Results



Bio-IT World

Bio-IT World.com
Indispensable Technologies Driving Discovery, Development and Clinical Trials

"At the end of such testimonials, it was very difficult to see whether Oracle has a serious rival in the realm of databases for high-throughput drug discovery. With a well-known 70 percent market share, Oracle is starting to penetrate smaller labs in academia and nonprofit research institutes."

- Mark D. Uehling, Bio-IT World (online) 09/12/03

eWeek

“All are among the features that make Database 10g much more than a large-scale data repository. Old 1960s labels such as "electronic brain" come to mind—Database 10g doesn't just know stuff, it also thinks about it.”

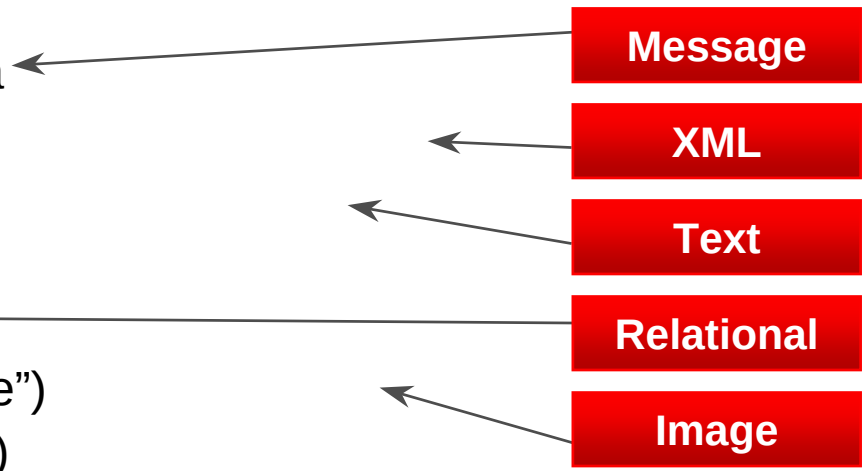
- Peter Coffee, eWeek (online) 05/31/04

Oracle's Contribution to Life Sciences

Find me any compound that looks like my current structure, and that has been tested on any assay in my company where the IC50>200nM, where I know that I have a unique patent position, and hasn't been published in any journal?

Oracle11g

```
select c.id, p.structure,
from compound c, protein p, assay a
where a.compound_id = c.id
and     a.protein_id = p.id
and     a.company = "BIO_SYS"
and     a.IC50 > 200nM
and     similar_to(p.id, "protein kinase")
and     not_published(p.id, "Medline")
and     extract_value(value(p.id), 'Dgene/Protein/Id') = p.id
```



IDC Analysts

"Even IBM's own partners say that DB2 and DiscoveryLink have failed to gain much ground in the life sciences despite IBM's giveaways. According to Hall, **Oracle, the "*de facto standard*," still holds a commanding 75 percent to 80 percent market share in this vertical."**

Mark Hall, Director of Life Sciences, IDC,
quoted in *InfoWeek* 12/12/2002

Roche

"Oracle is an excellent database. It's been around for years, it's been honed and developed, and it's very good at handling large volumes of information—and that's exactly what we need."

Jennifer Allerton, CIO of Pharma division of Roche

quoted in *Oracle Profit magazine* July 2004

Oracle Life & Health Sciences Platform

- Oracle 11g Enables you to:
  - Access distributed data
  - Integrate a variety of data types
  - Manage vast quantities of data
  - Collaborate securely
  - Find patterns and insights
- Oracle 11g is an ideal platform for health & life sciences


QUESTIONS
ANSWERS