From GeoSpatial to BioSpatial:
Managing 3D Structure Data

Xavier R. Lopez
Director, Location Services
Oracle Corp.

ORACLE

Overview

- Market & Technology Trends
- Spatial Database Technology
- GeoSpatial DBMS in GeoSciences
- Life Sciences Data Management Challenges
- BioSpatial DBMS in Life Sciences

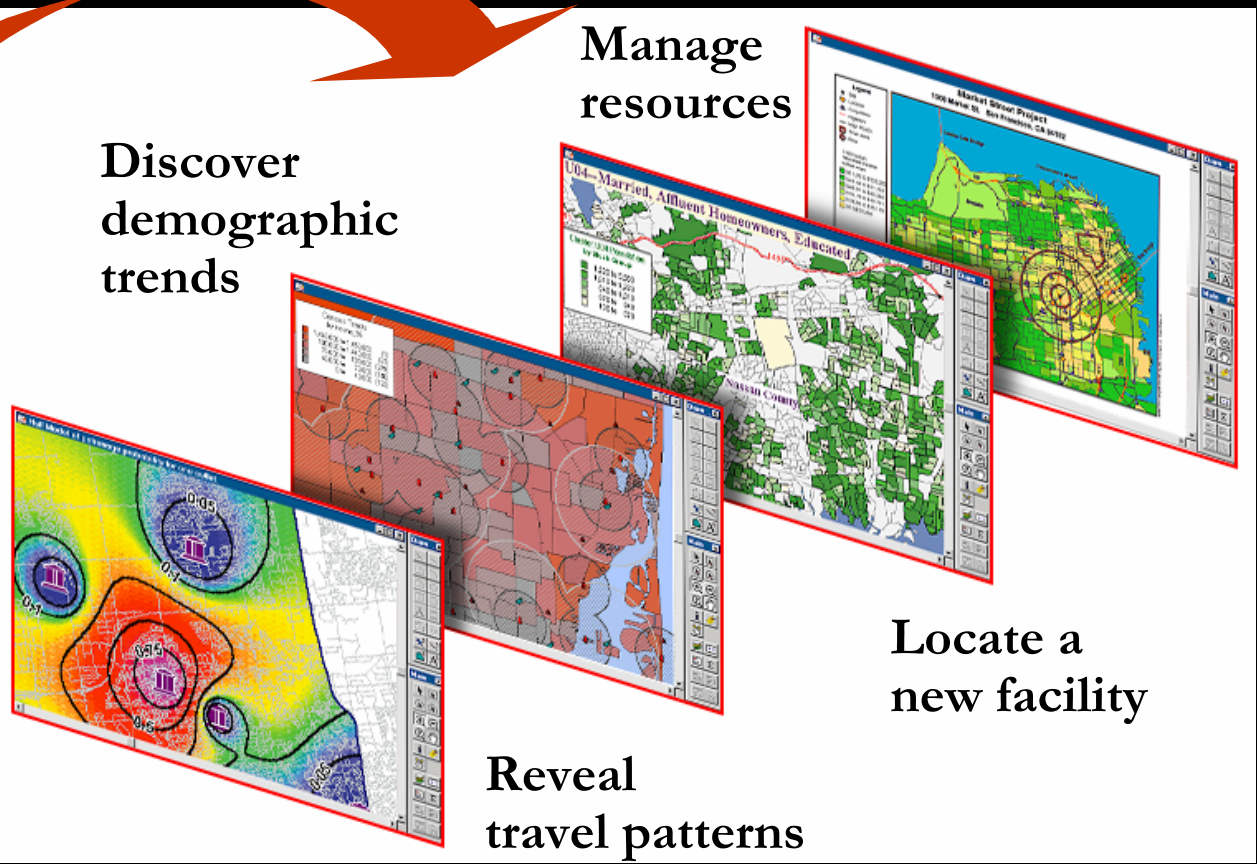
Spatial data becoming ubiquitous

- Location Aware and Enabled Infrastructure
  - Defense, Logistics, Mobile devices
- Internet Portals: MapQuest, Yahoo, MapPoint.NET
- Automobiles: by 2006, 80% of new cars will have some telematics navigation access (eyeforauto 2001)
- Structure Databases: Proteomics, Materials Science

Spatial Analysis

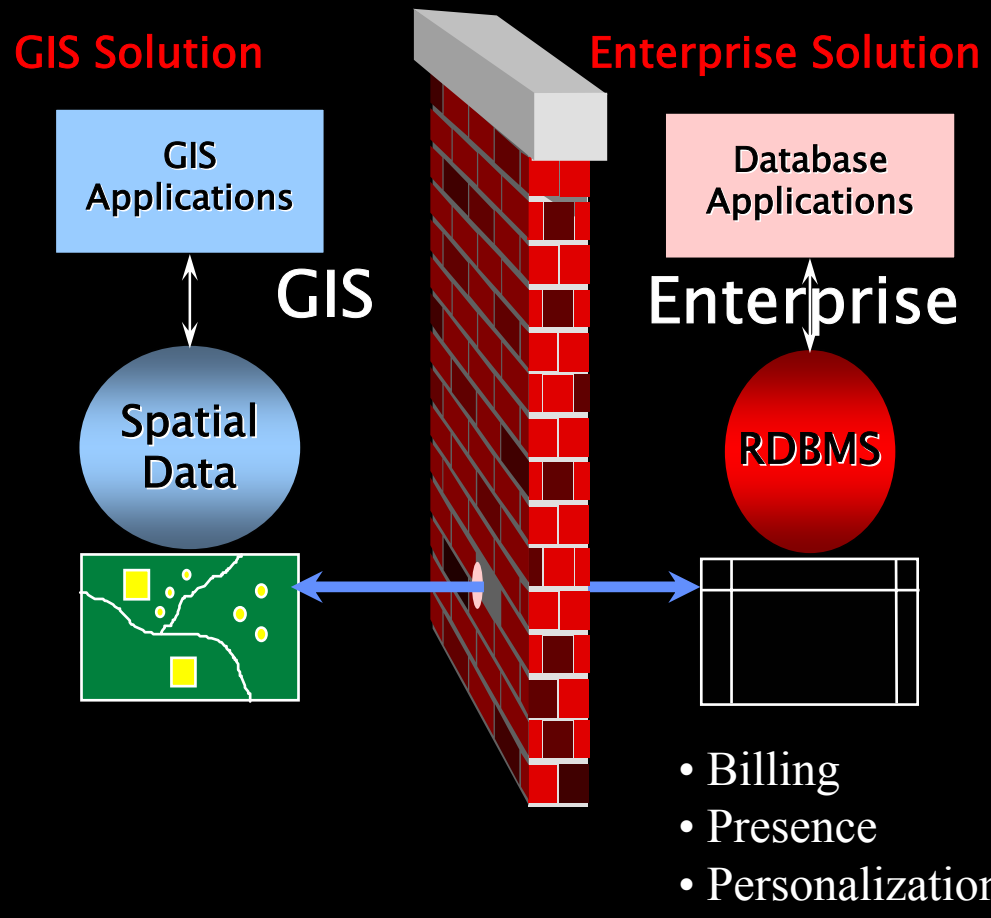
Revealing patterns, relationships & trends

Location	Client Name	Usage
AUSTRIA	**Hallein Municipality	Local authority
AUSTRIA	**Ludesch	Local Government
AUSTRIA	ARG Verrmessung, Dornbirn	Survey and mapping
AUSTRIA	ILF-Dornbirn -8	
AUSTRIA	ILF-Innsbrueck - 2	
AUSTRIA	ILF-Prague - 2	
AUSTRIA	ILF-Vienna - 2	
AUSTRIA	ILF-Villah - 1	
AUSTRIA	Ingenieurgemeinschaft Laesser-Fezlmayr (ILF),	Engineering company
AUSTRIA	Lochau Municipality, Vorarlberg	Local government
AUSTRIA	Manahl, Feldkirch	Engineering company
AUSTRIA	Vorarlberg Erdgas, Dornbirn	Gas distribution
BOSNIA	City of Zageb(CV)	Local government
BOSNIA	Computech (CV)	Reseller
BRAZIL	Systenge	Reseller
CANADA	City of Edmonton	Local government
CANADA	City of Luduc	Local government
CANADA	District of Oak Bay	Local government
CANADA	Energy & Mines (Ottawa)	
CANADA	Energy & Mines (Quebec)	
CANADA	Geopower Technologies, Inc.	Reseller
CANADA	H.H. Pillar Corp.	
CANADA	University of Toronto	Education
CHINA	Beihai Urban Construction	
CHINA	Beijing Urban Archive	Local government
FINLAND	Pohjois-Satakunnan paikkatietopalvelu OY	GIS systems house
FINLAND	Tampere muncipality (PCX 100 USER LICENCE)	Local government
FRANCE	Cabinet Dulac	Survey and mapping
FRANCE	District Bayonne - Anglet - Biarritz	Local government consortium
FRANCE	EPA Cergy-Pontoise	New town development
FRANCE	France Telecom	Telecommunic. company
FRANCE	Gaz de France	Gas distribtuion
FRANCE	Institut Geographique National (IGN)	National mapping agency
FRANCE	ITMI	Software developer/integrator
FRANCE	Municipality of Dijon	Local government
FRANCE	Nancy District	Local government
FRANCE	School of IGN	IGNs training school
FRANCE	University of Caen	Educational



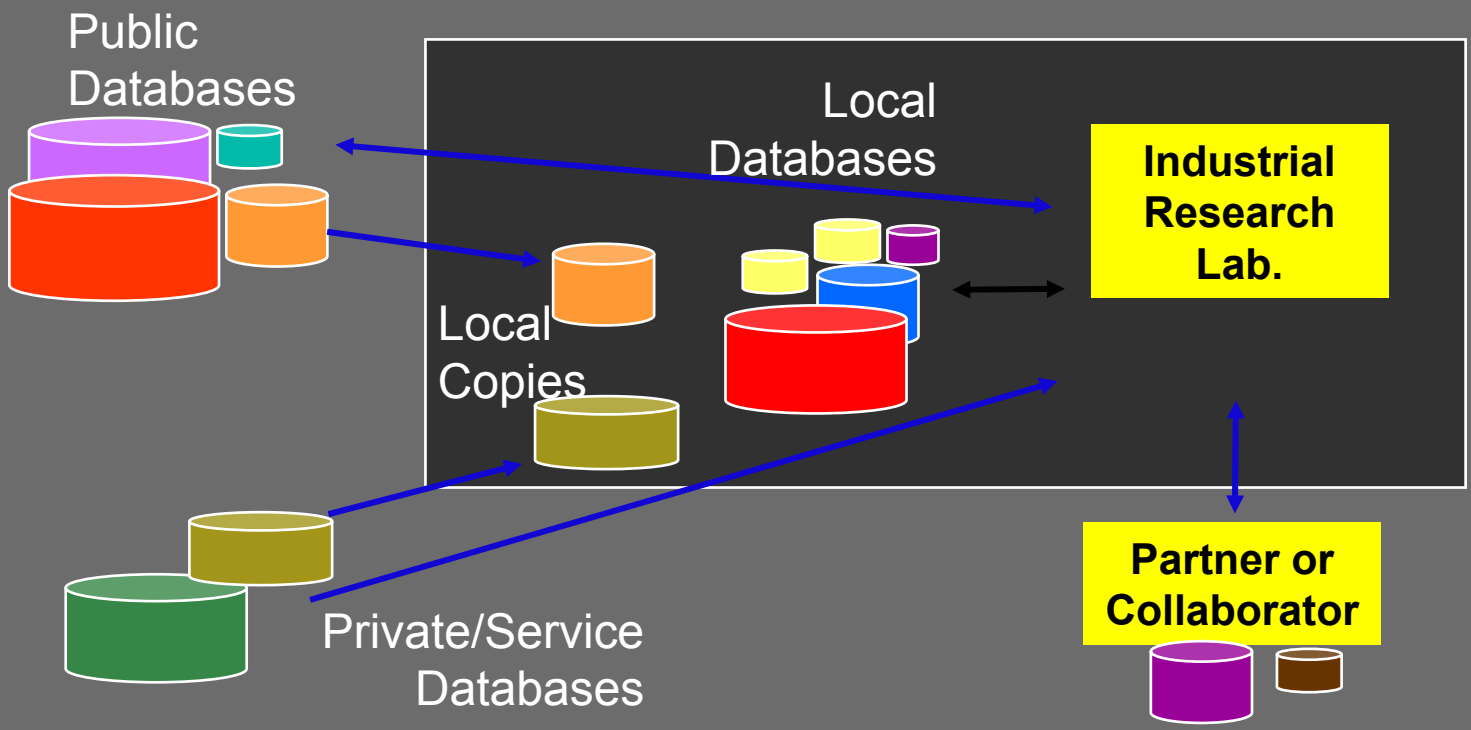
Overcoming Application "Stovepipes"

- Specialty GIS servers
  - Data isolation
  - High systems admin and management costs
  - Scalability problems
  - High training costs
  - Complex support problems
- Information not aligned with Business Processes
- Applications can't leverage brute force of large servers

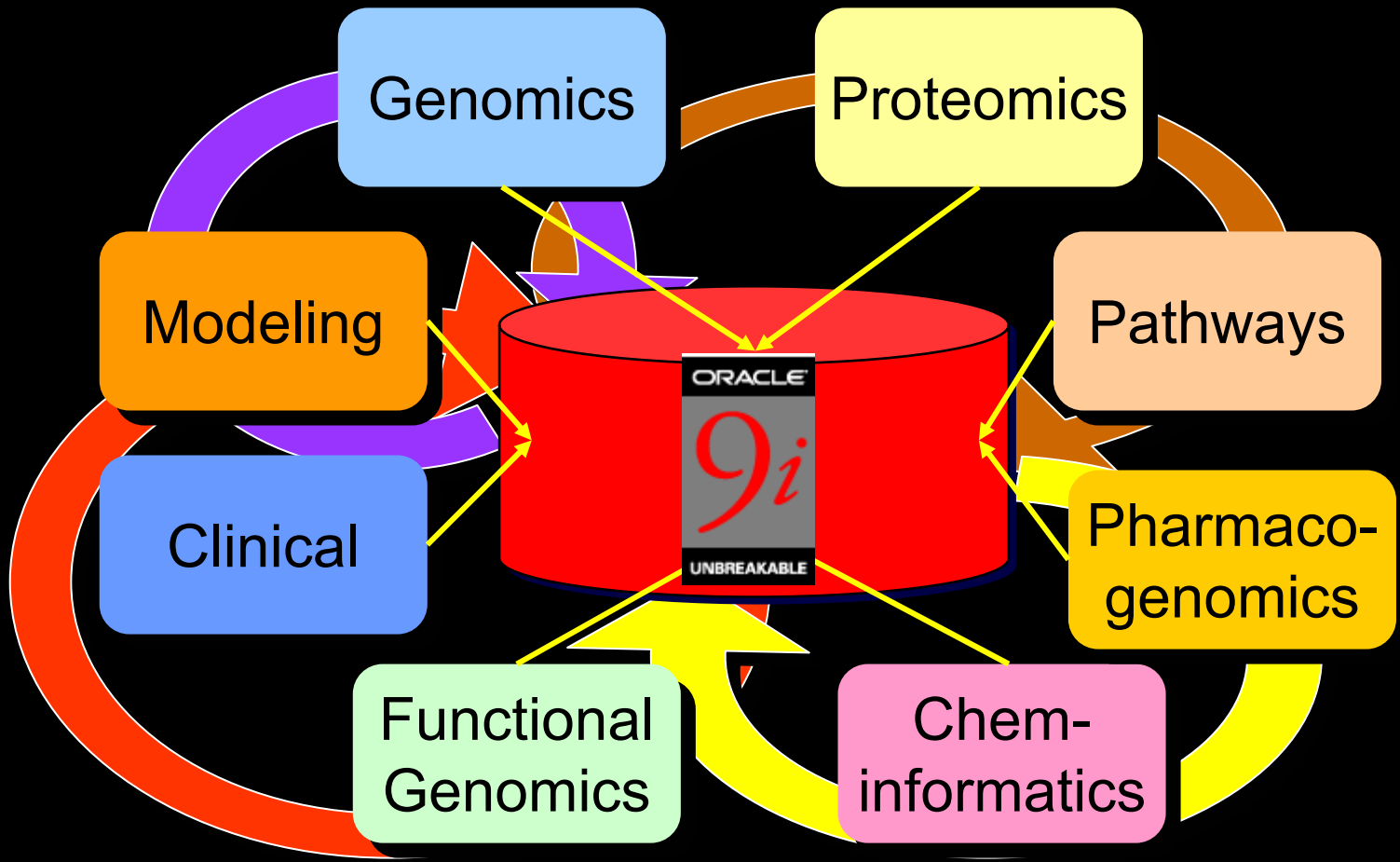


Life Sciences: Drug Discovery

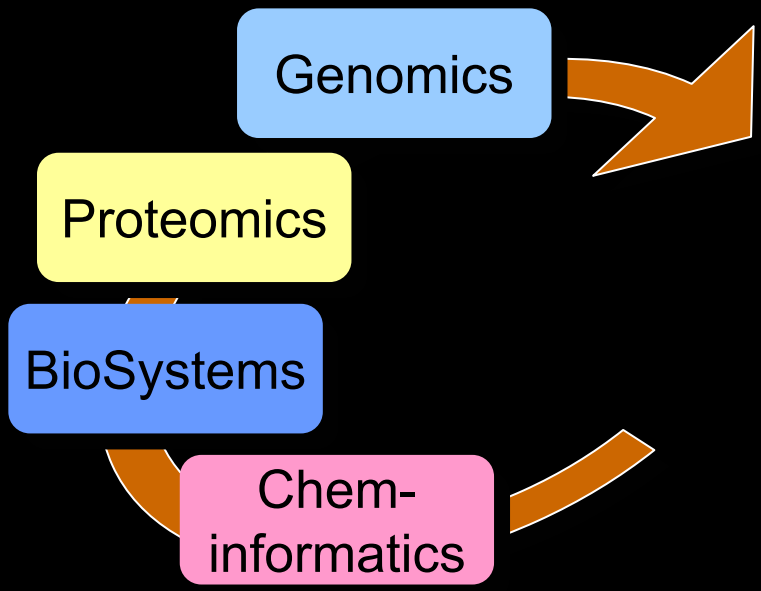
- The Process

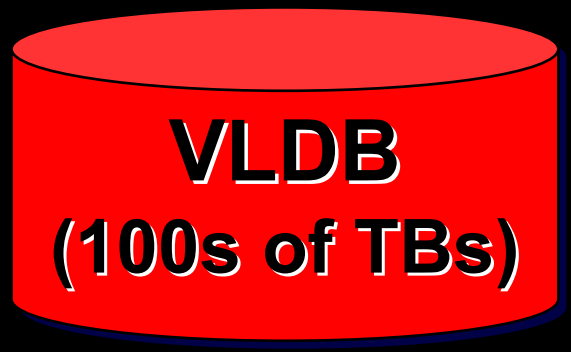


Many Different Kinds Data



IT Challenges

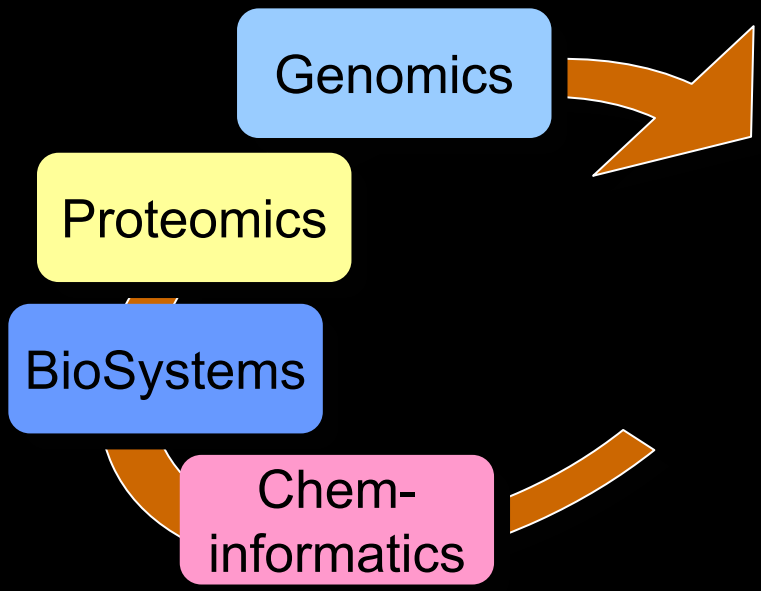


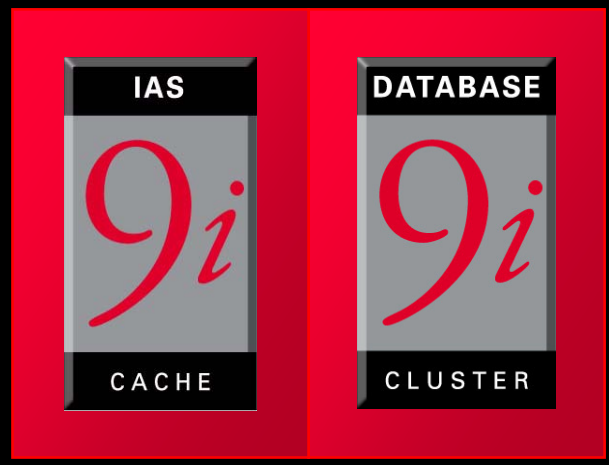


Load
Aggregate
Collaborate

Store
Search
Match
Mine
Visualize

Oracle Platform

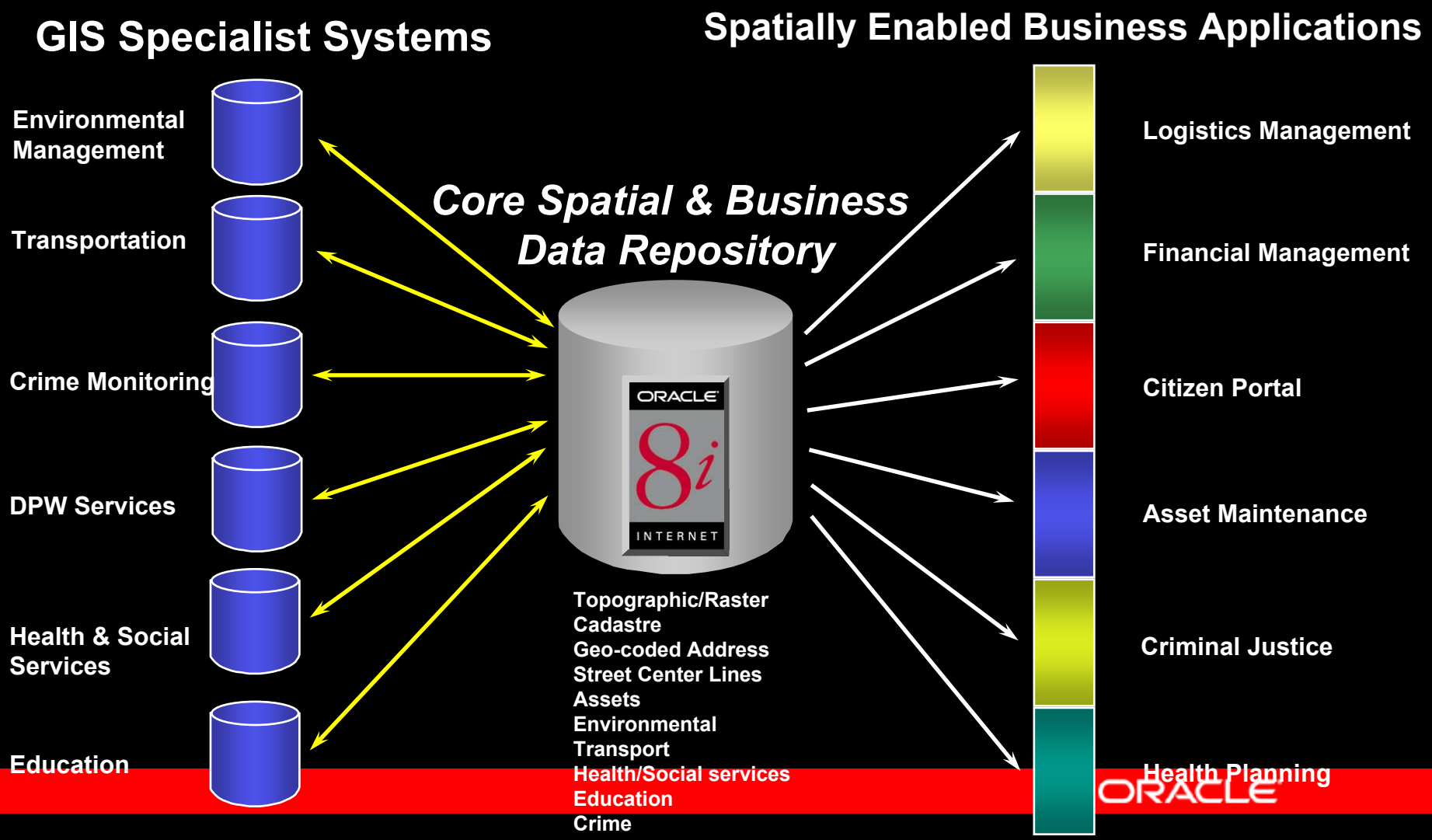




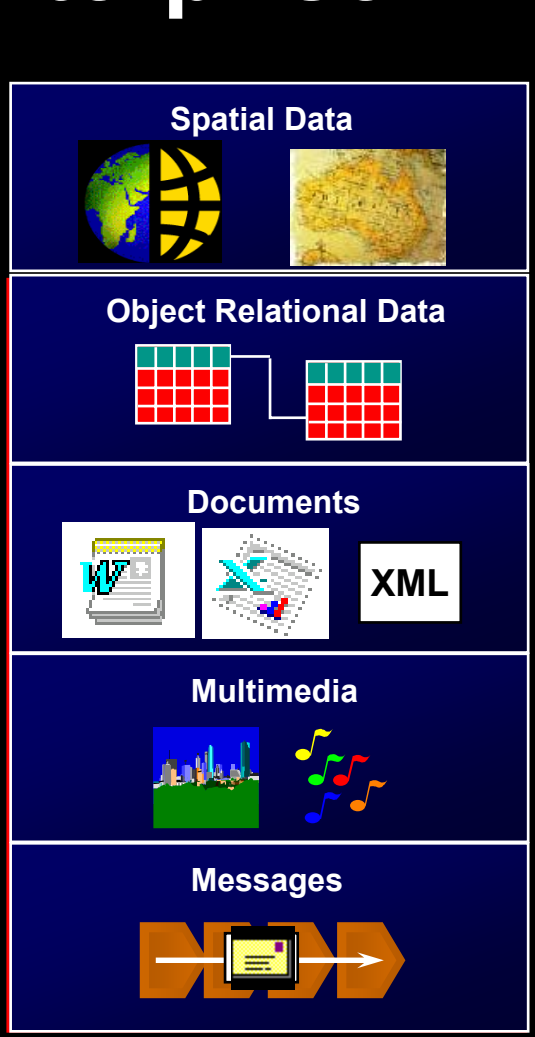
- Distributed Queries
- Incremental Updates
- XML Data Types/Searches
- iFS/collaboration
- Data Mining
- Extensible Indexing
- Partitioning & parallel computing

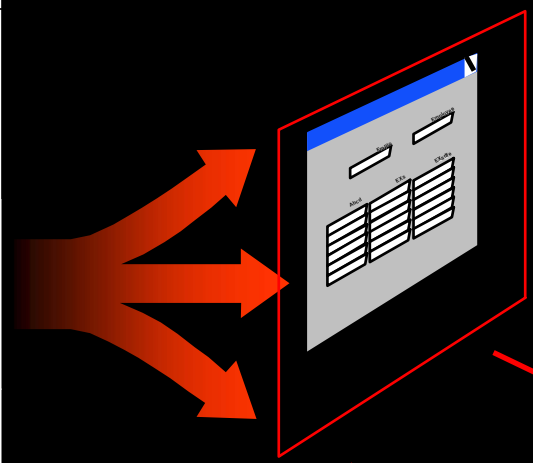
- Unlimited Scalability
- Reliability (RAC)
- Security
- Workflow
- Text searches
- Portal
- Images &Video

Integrated NYC Spatial Architecture

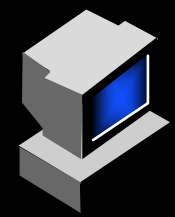


Managing All the Data in an e-Enterprise

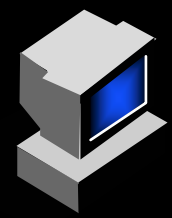




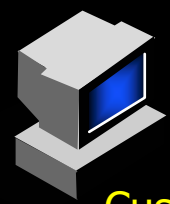
Field



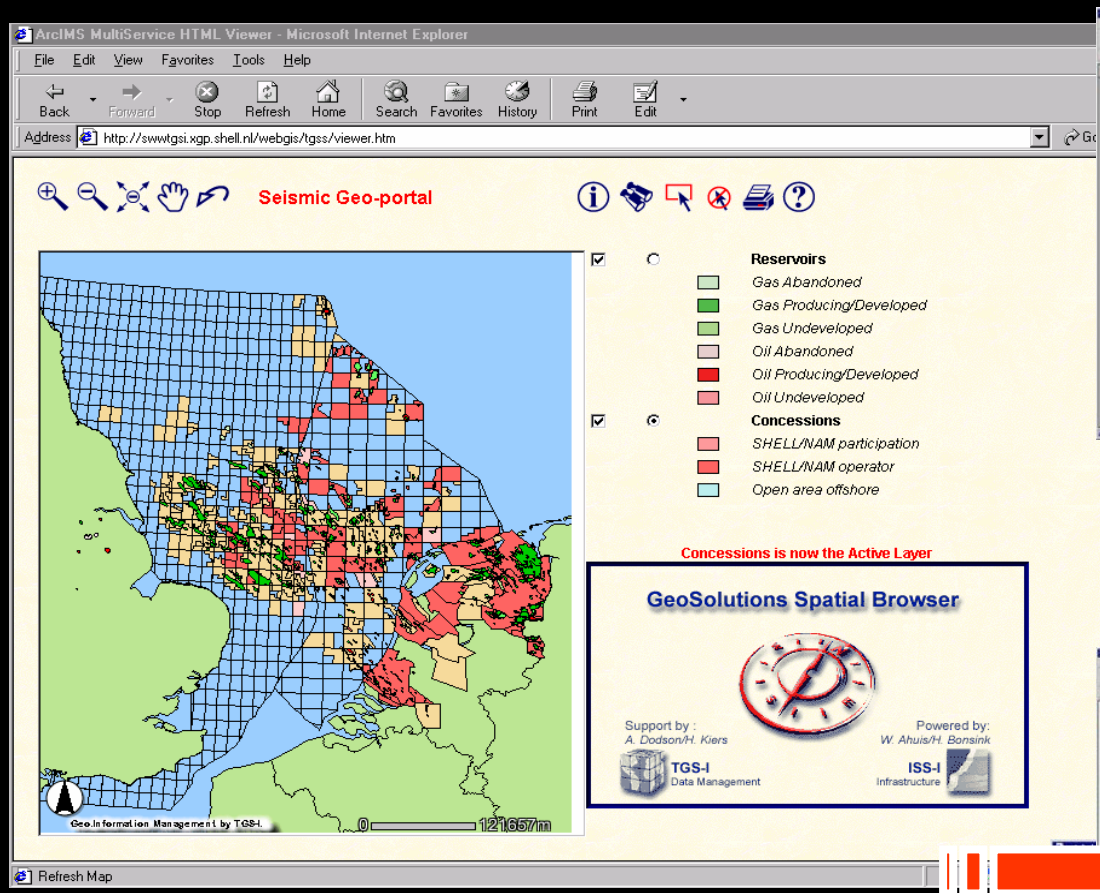
Prospects

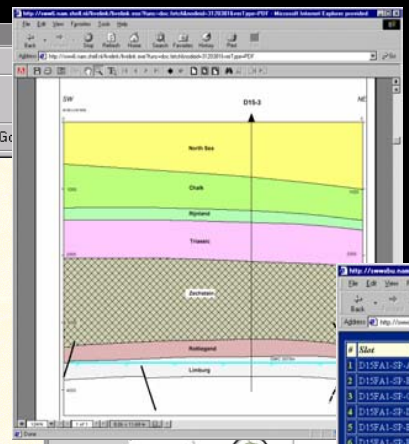


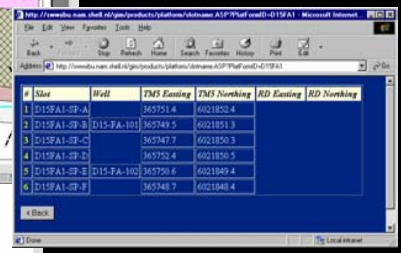
Infrastructure

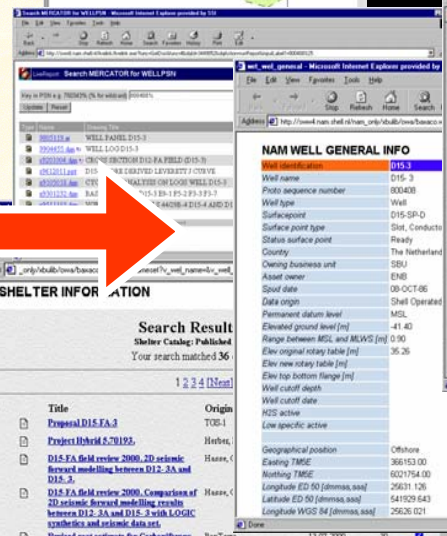


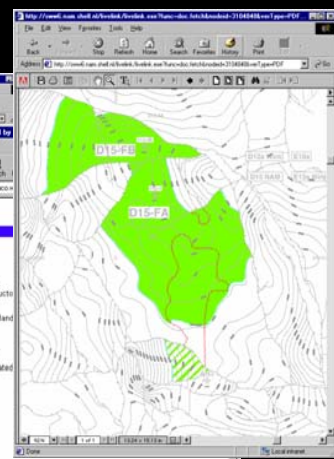
Customers











Shell International:

Web enabled GIS provides browser based access to users of corporate and geo-spatial data from the Oracle RDBMS and Spatial databases in one integrated window

Spatial Database Technology: Manage Location & Structure Data

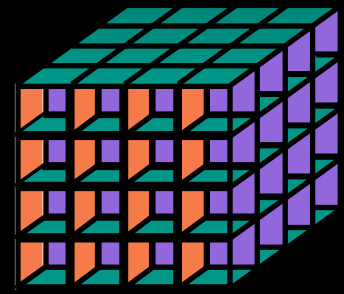
Oracle9*i* Spatial Capabilities

Spatial Data Types

Native Spatial Data Management in the DBMS

Oracle Spatial

Spatial Indexing



Fast Access to Spatial Data

Spatial Access Through SQL

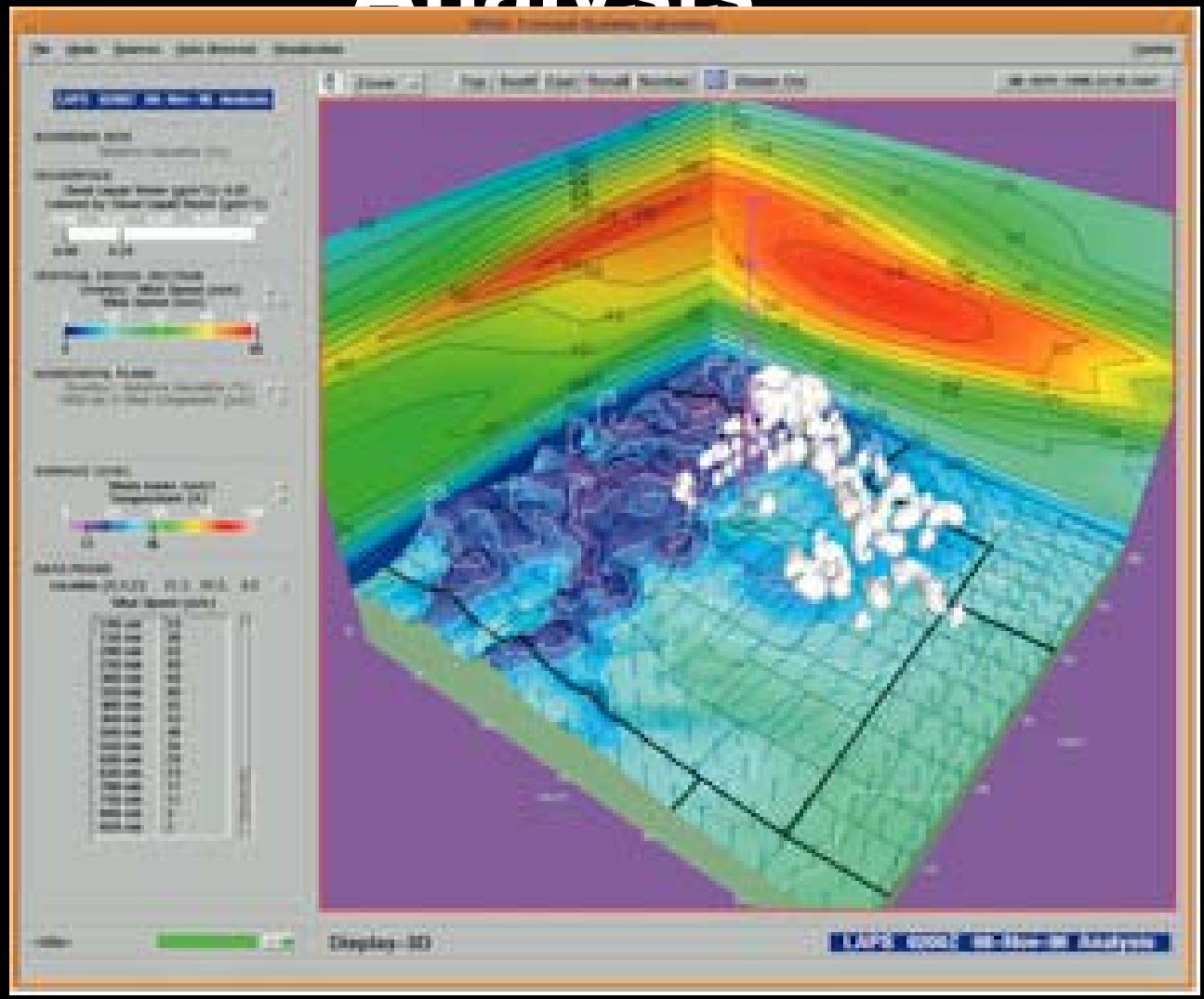
```
SELECT STREET_NAME FROM ROADS, COUNTIES
WHERE SDO_RELATE(road_geom, county_geom,
      'MASK=ANYINTERACT QUERYTYPE=WINDOW') ='TRUE'
AND COUNTYNAME='PASSAIC';
```

Vector Map Data in Oracle Tables

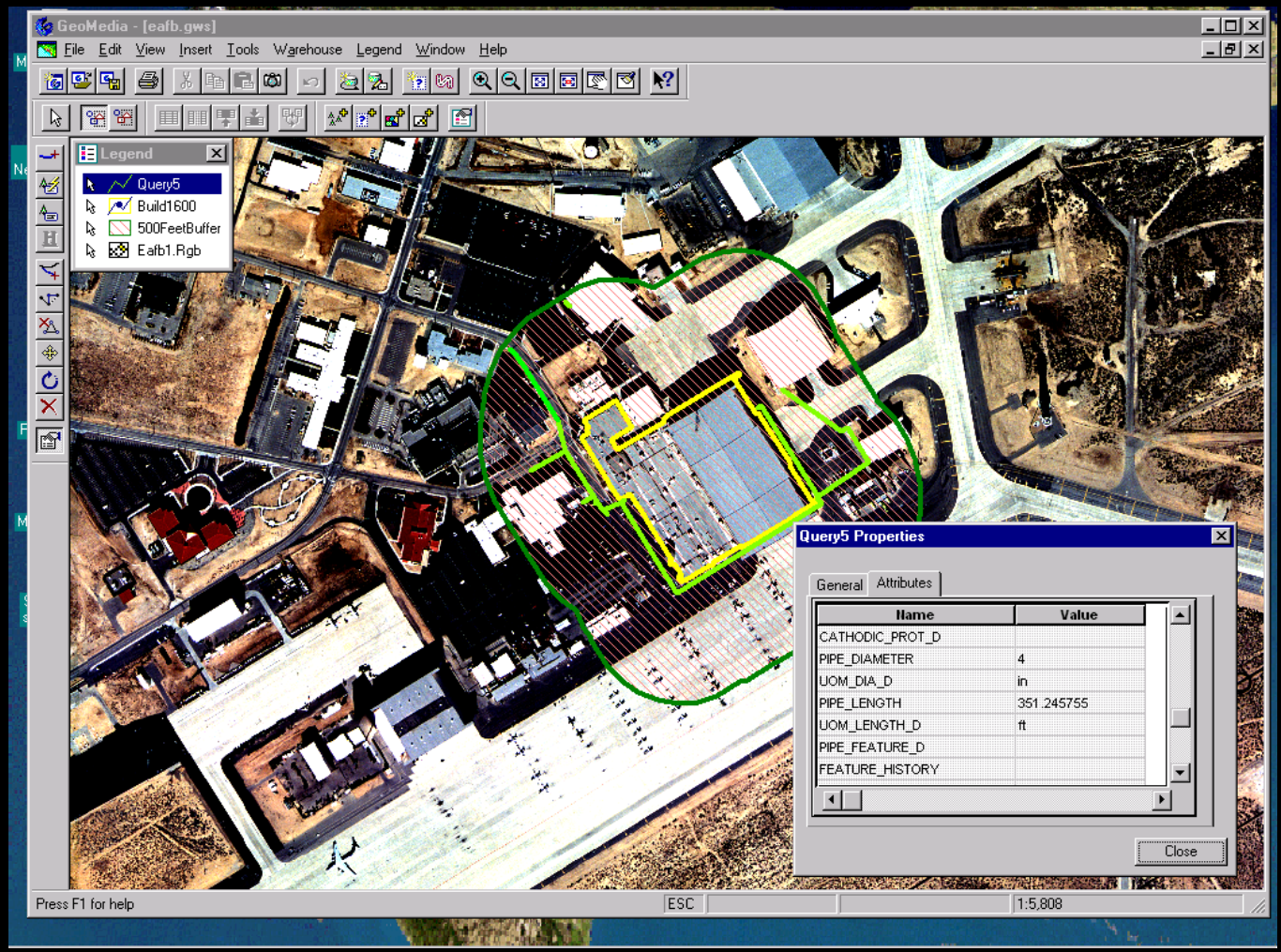
Road

ROAD_ID	NAME	SURFACE	LANES	LOCATION
1	Pine Cir.	Asphalt	4	
2	2nd St.	Asphalt	2	
3	3rd St.	Asphalt	2	

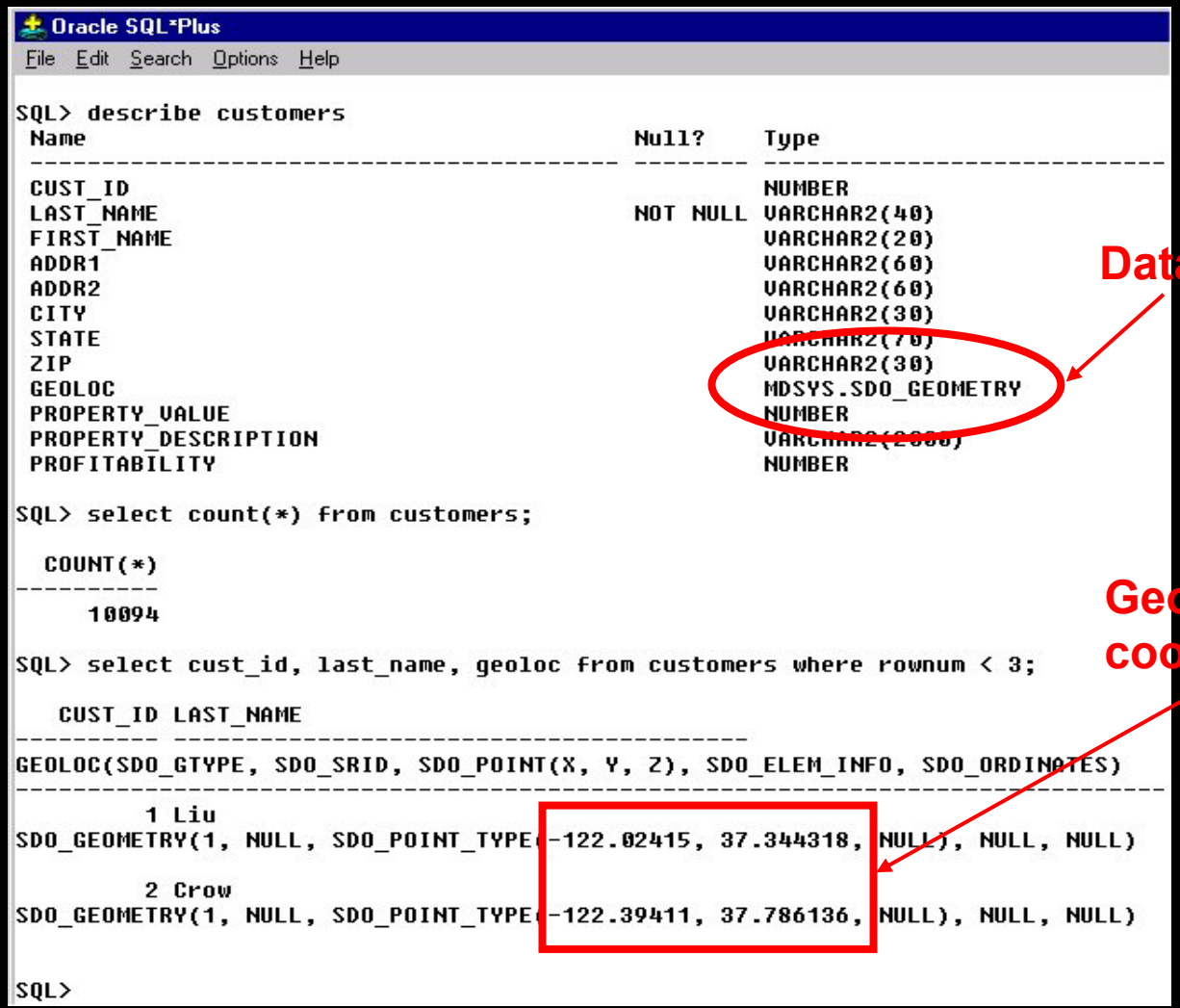
Sub-surface Geological Analysis



Raster/Vector Mapping



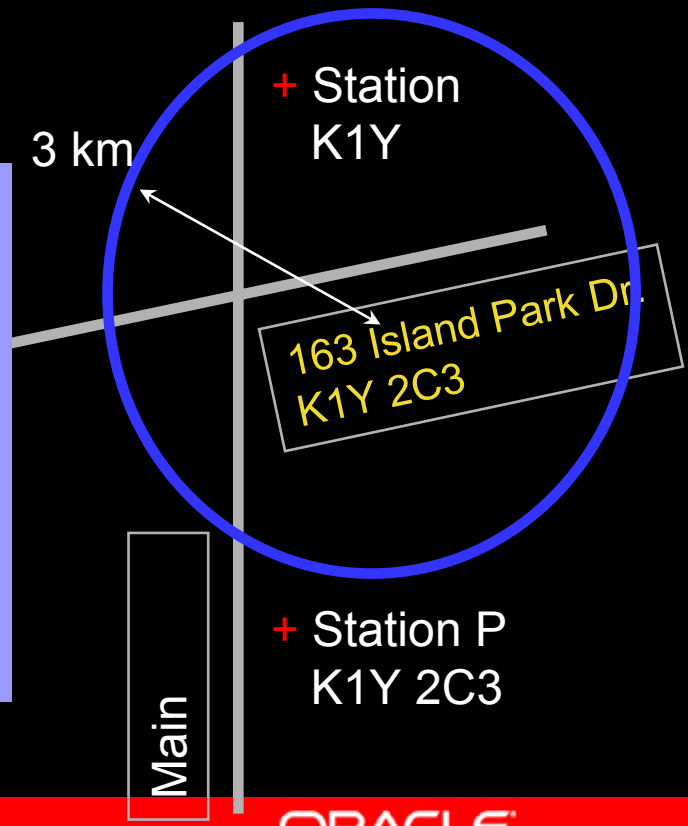
How Spatial Data Is Stored



Performing Location Query in Oracle9i

Example:What are the nearest post offices to my office?

```
SQL> SELECT P.Post_Office_Name, P.Address
  2>   FROM Post_Offices P,
  3>  Address_Master A
  4>  WHERE
  5>  A.St_Address ='163 Island Park Dr.'
  6>   and A.City = 'Ottawa'
  7>    AND MDSYS.SDO_WITHIN_DISTANCE(
  8>         A.Location, P.Location,
  9>         'distance=3') = 'TRUE';
```



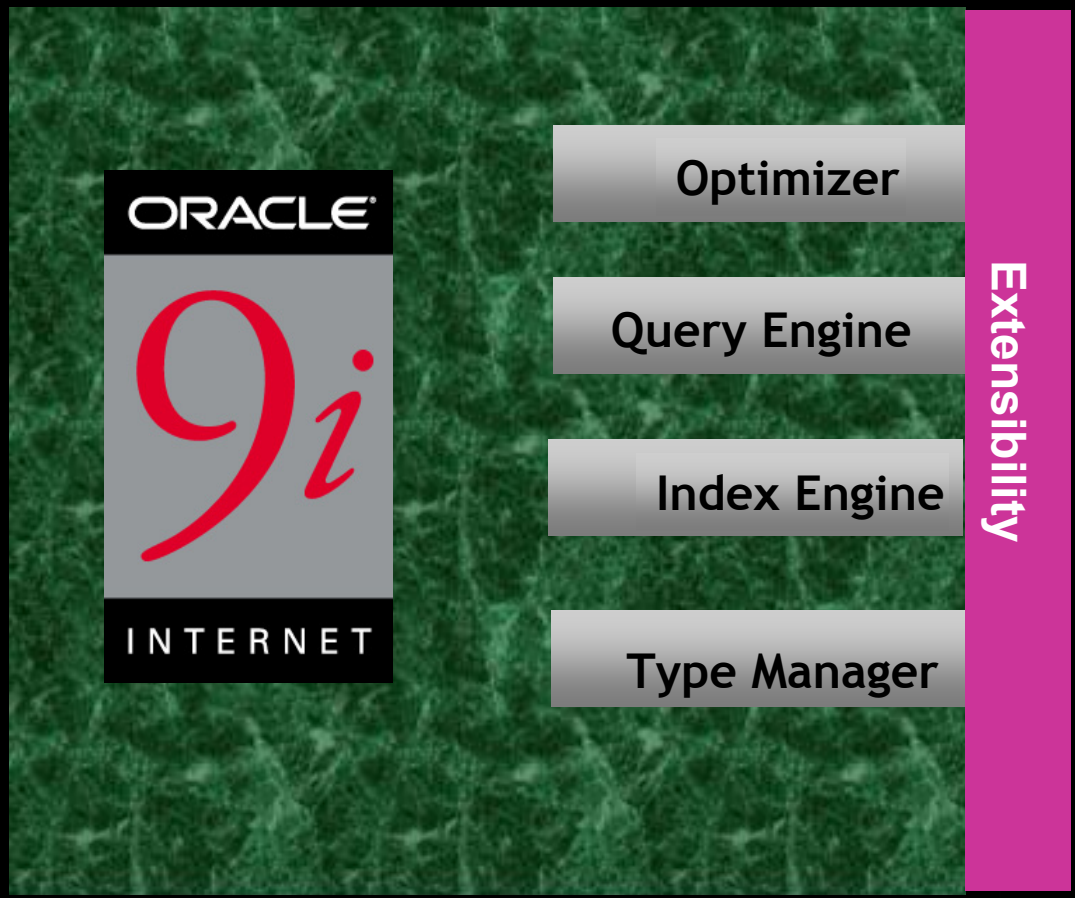
Jphone J-Navi Launch May 2000

Oracle Spatial Platform Powers:

- Worlds 1st Live Map Delivery to Phone
- Over 1M color maps delivered per day
- Vector/Raster Maps generated dynamically
- Avg. Query Processing 200ms
- Download time: Max 2 seconds
- 30,000 user sessions per hour
- 17M business listing & national map data
- Java Servlet Technology
- Prototype to Lauch: 6 Months
- Unprecedented scalability, reliability & flexibility

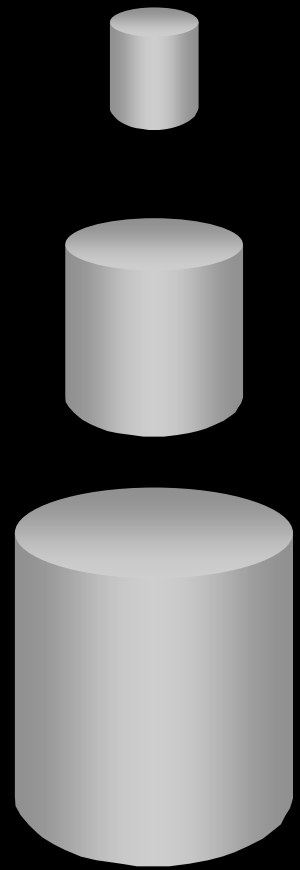
KDDI & DoCoMo: similar model

Extensible Database Framework



Dealing with large data volumes

- How *large* is *large* ?
  - 100's of thousands is normal
  - Millions is interesting
  - 10's of millions is serious
  - 100's of millions is large
- What *is* the problem with large volumes ?
  - They mean *big* structures
    - Cumbersome to manage
  - Long operations
    - Data reload, refresh
    - Index rebuilds



Partitioning: Divide and Conquer

Two reasons for partitioning

For manageability

- Break large problems into manageable pieces
- Can load / rebuild individual partitions
- Can load / rebuild multiple partitions concurrently

For performance

- Query parallelism
- Partition elimination

- Can partition tables, or indexes, or both
  - Also spatial indexes
- Transparent to applications!

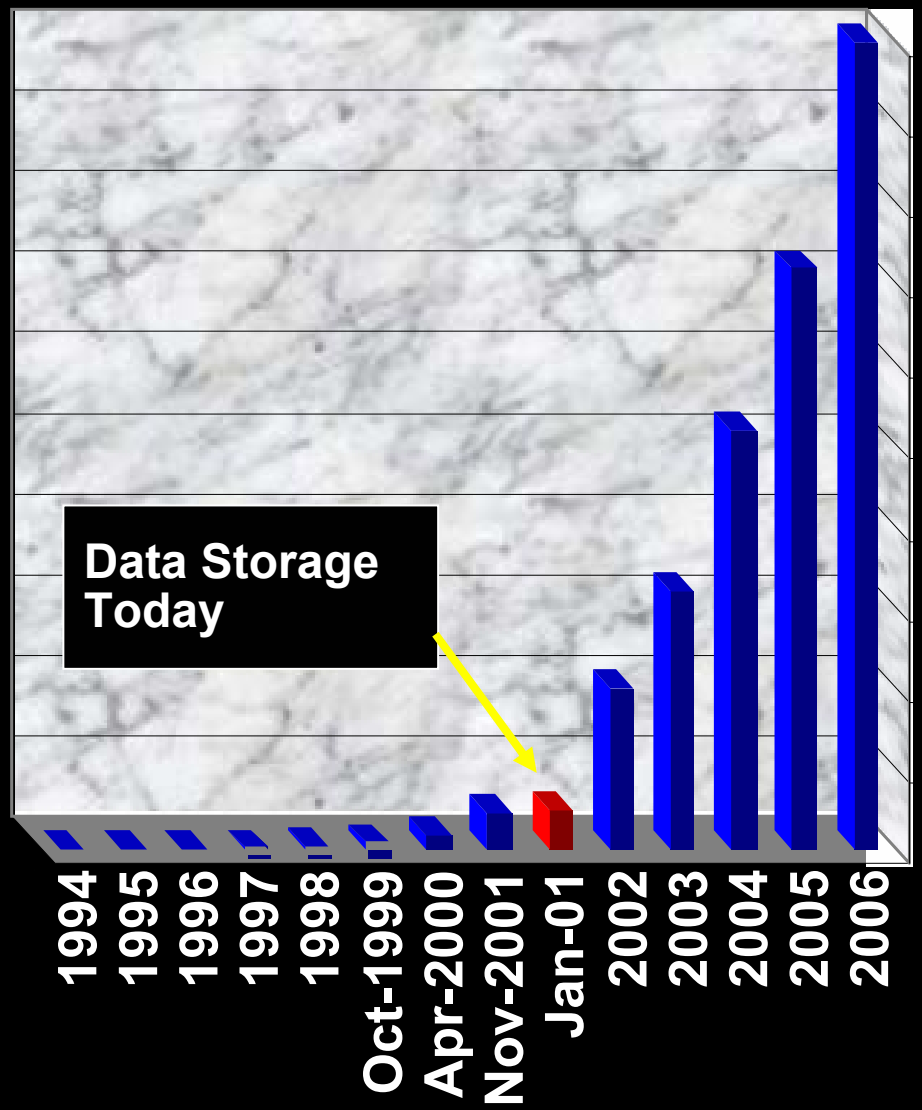
Oracle9i Spatial Features

- Spatial Reference System
- Spatial Operators
- Versioning/Long Transactions
- Linear Referencing
- Quadtree/R-tree index
- Parallel Index create
- Geodetic Support
- Spatial Aggregates
- Topology *
- Raster/Grid Management *
- Spatial Data Mining *

* Planned Release 10i

Life Sciences
Data Management Trends

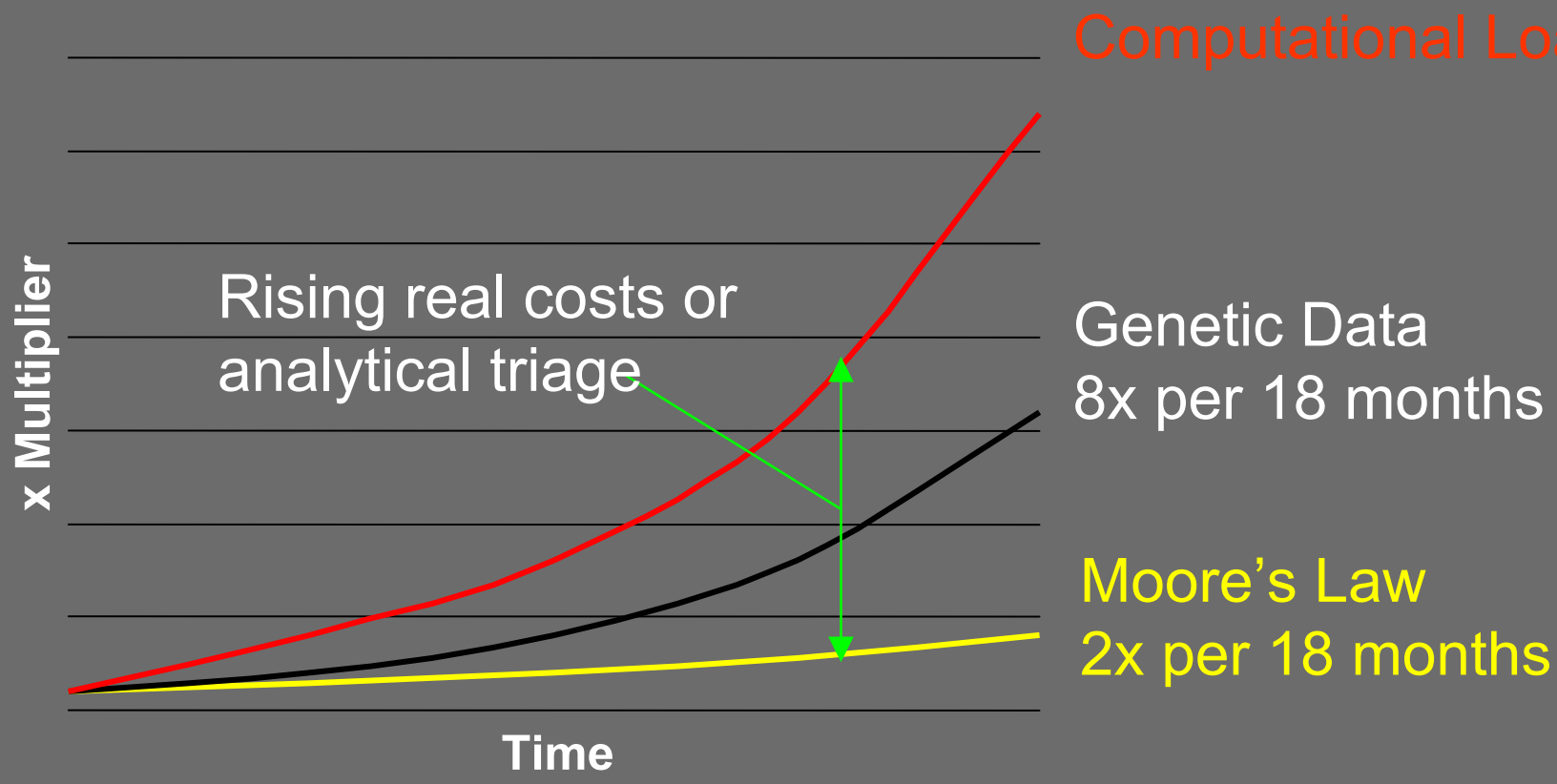
Expanding Data Storage Needs



"To meet the scientific goals we believe we need to add around 80 - 100TB of storage each year for the next 5 years"

P. Butcher,
The Sanger Centre

Increasing Computational Load

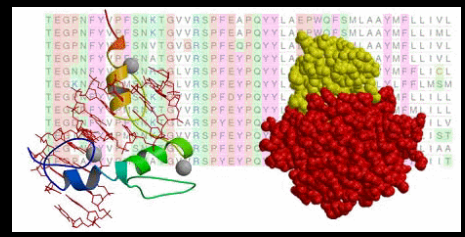


ORACLE

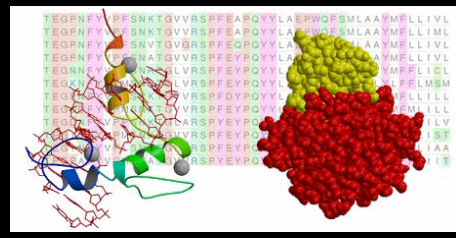
Source: Sun Microsystems Life Sciences marketing collateral

What does DBMS technology bring?

1. Access and storage of vast quantities of life science data from a variety of sources
2. High throughput loading, indexing, processing and update of information
3. Data integration from a variety of sources
4. Scalability and reliability problems
5. Find patterns & insights through queries, analyses and data mining
6. Collaboration & security challenges



1. Vast quantities of data, types & sources

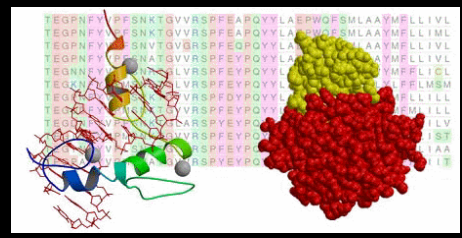


Benefits

- Access and integration from variety of sources/types of data
- Efficient handling of new data types
- Ability to search data using SQL and/or XML
- Ability to manage external files within database

Gateways, XDB & XML, iFS, Extensible indexing, Spatial

2. High Throughput Process

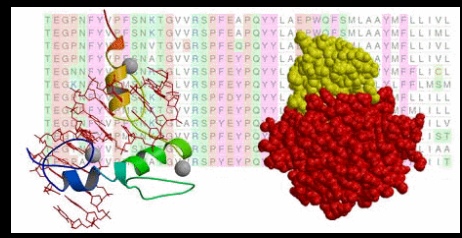


Benefits

- Scalability across multiple CPUs and cluster nodes
- Fast uploads of new life sciences data
- Build life science applications
- Ability to speed up compute intensive operations
- Linear scaling with cheap (Lintel) hardware

RAC, Partitioning, Advanced Queuing, Workflow,
Table functions, UpSert, Linux

3. Scalability & Reliability

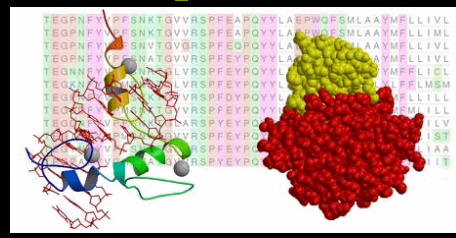


Benefits

- Increasing fault tolerance from system failures
- Protecting data from site failure and storage failure
- Identifying and quickly resolving human errors
- Eliminating the need for planned downtime

Oracle9i RAC, Data Guard

4. Hidden Patterns & Relationships

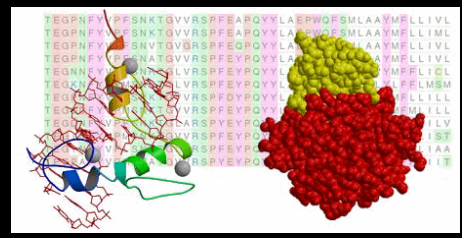


Benefits

- Find patterns and clusters e.g. base pairs associated with healthy and diseased states
- Classify and predict diseases likely to respond to certain treatments
- Classify documents relevant to area of interest

Oracle9i Data Mining, Oracle Discoverer & Oracle Text, Spatial

5. Collaboration & Security



Benefits

- Build departmental portals for common activities and favorite genes and proteins
- Integrate and automate common tasks and functions
- Revision control
- Row level access control that enables multiple users to share the same database, yet only access the row(s) of data that pertain to each individual user

Oracle Portal, Thesaurus, VPD, JDeveloper, Workflow

Some Additional Proteomics Challenges:

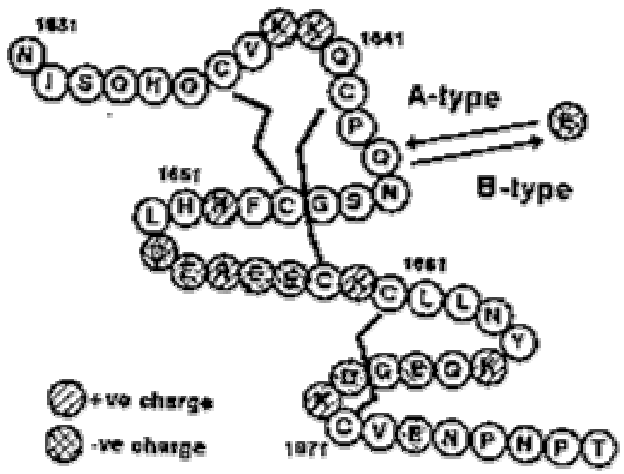
- High-throughput crystallography generating large volumes of complex protein structure data
- Small molecule (structure) databases growing to tens of millions of compounds
- 3D and pharmacophore analysis require efficient storage, indices and operators of structure data
- Integrated visualization & computation tools with DBMS

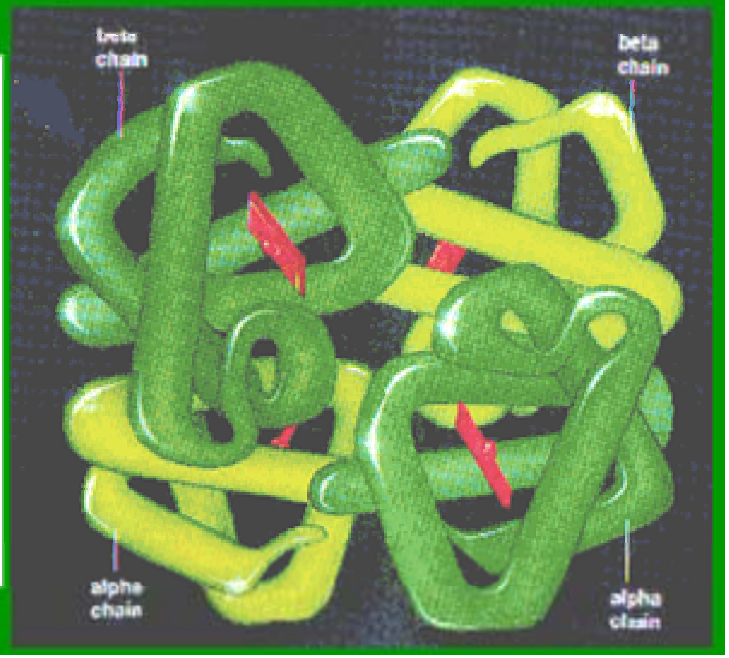
How do spatial databases help?

- Object-relational model and extensibility enable 2D data types and indices
- Powerful and growing operator set for sophisticated location/structure queries
- Validation by Geographic Information Systems (GIS) and CAD Community
- Common query language – SQL- that all data banks and tool vendors leverage
- Security, reliability, scalability and flexibility
- Faster, bigger, better, cheaper

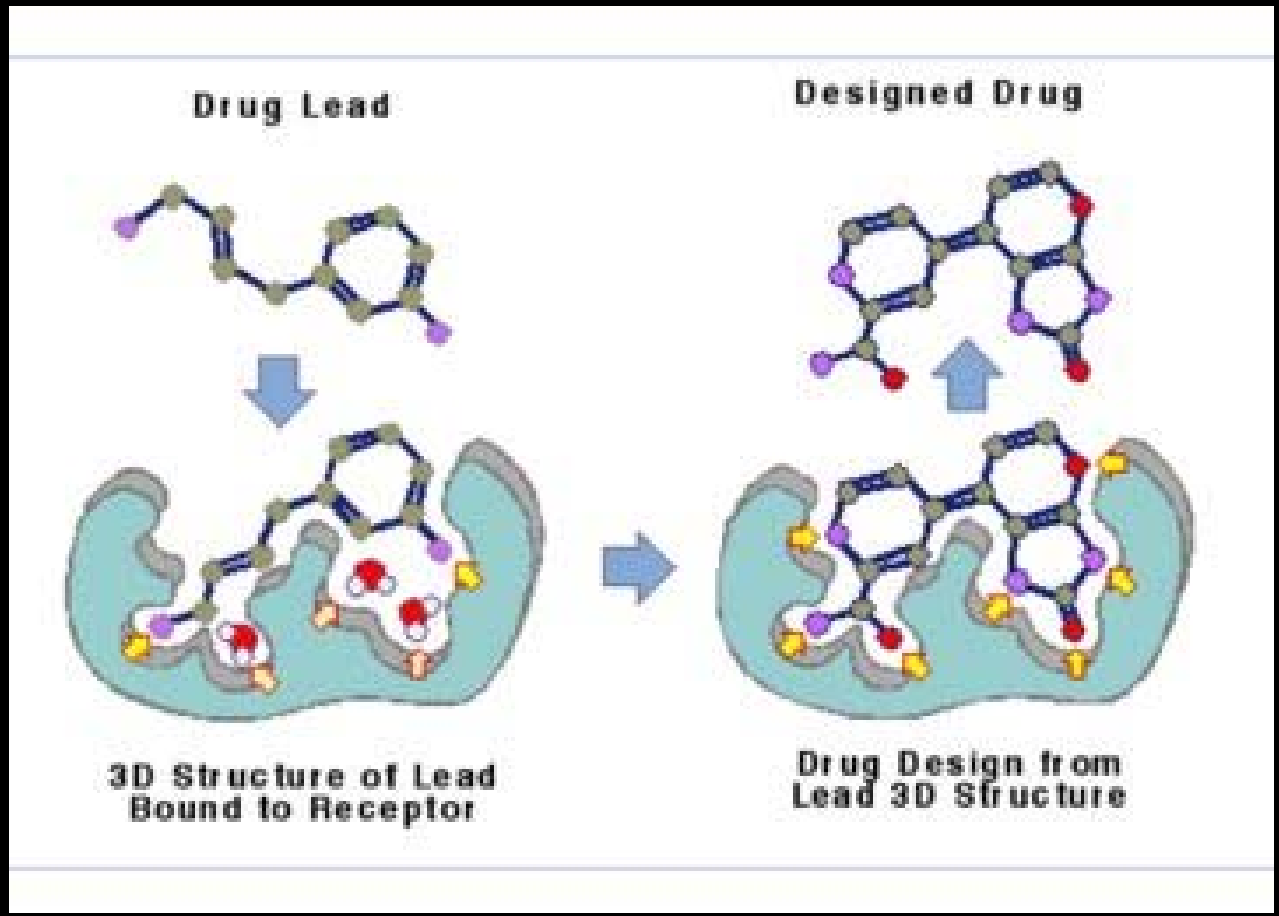
Protein Structure

Amino acids form linear polypeptide chains that fold into 3-dimensional structures.



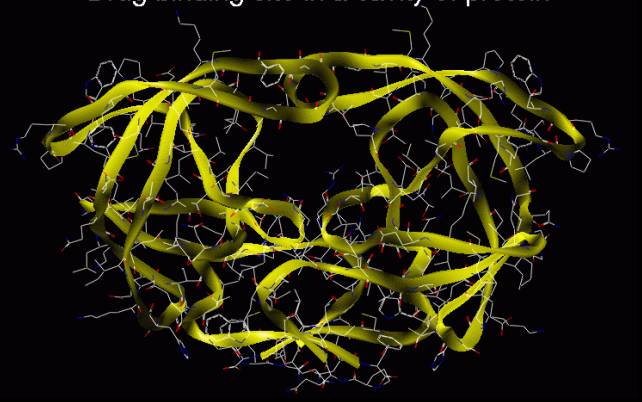


Structural Bioinformatics and Rational Drug Design

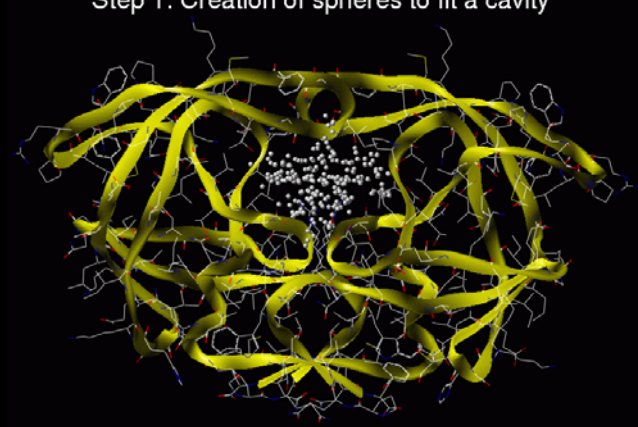


Virtual High-throughput Screening Ligand-Protein Docking Simulation

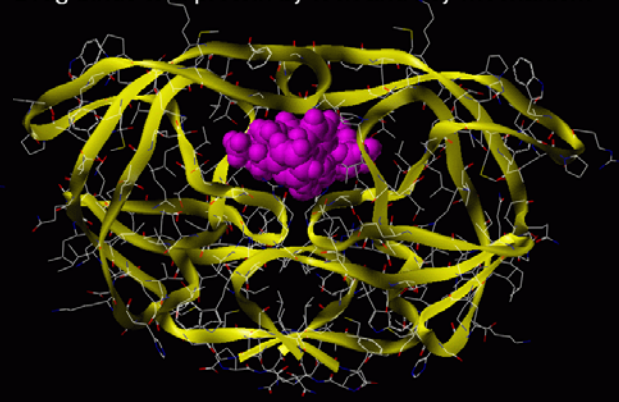
Drug binding site in a cavity of protein



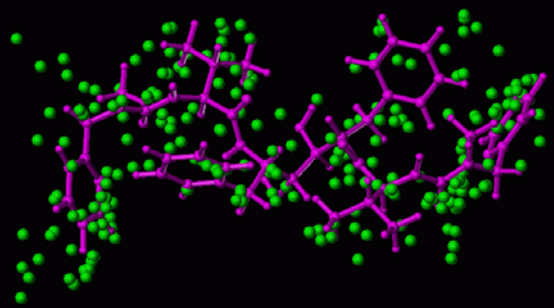
Ligand–protein docking:
Step 1: Creation of spheres to fit a cavity



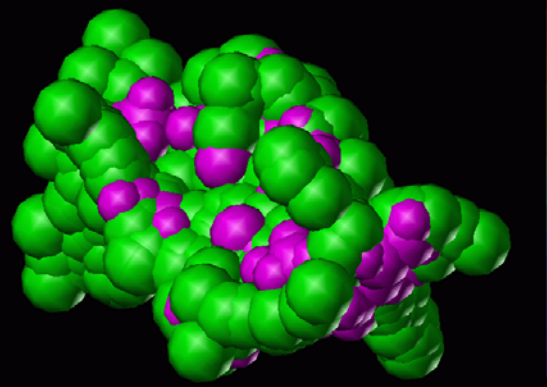
Drug binds to a protein by lock and key mechanism



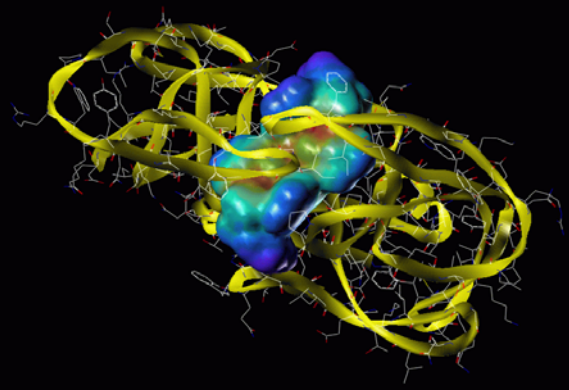
Ligand–protein docking:
Step 2: Place a ligand to match the positions of spheres



Ligand–protein docking:
Step 2: Place a ligand to match the positions of spheres



Ligand-protein docking:
Step 3: Check chemical complementarity.



ORACLE

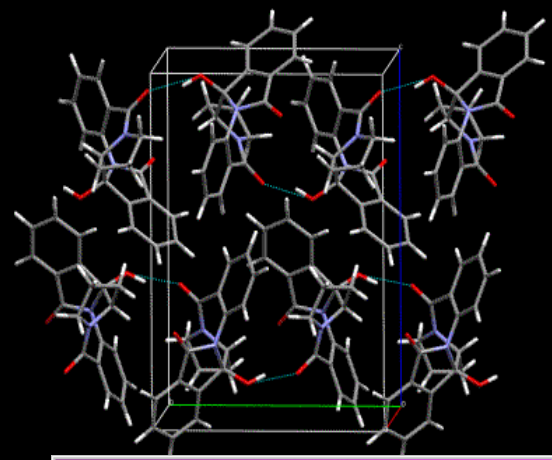
Planned Oracle BioSpatial Types and Functions

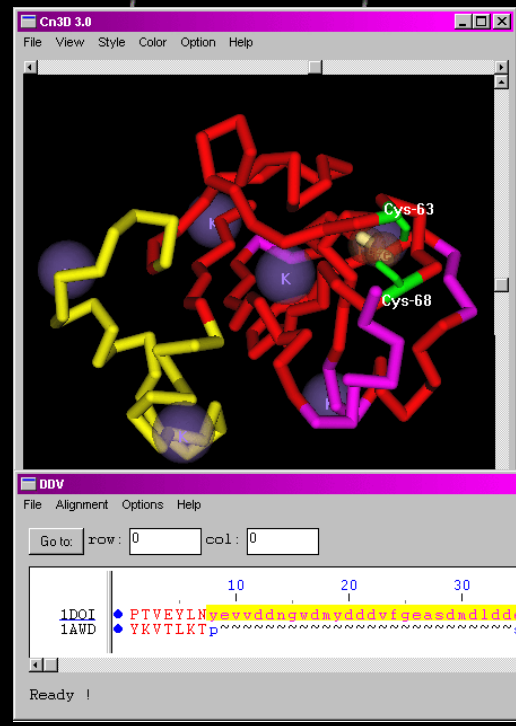
Managing Protein Structures in DBMS

- **Extend Oracle DBMS with custom 3D structure features**
- **Provide BioSpatial types and an object-relational schema for large & small molecule data in Oracle**
  - Compliant with mmCIF; SQL interface
- **Provide a low-level interfaces consistent with OMG standard (RCSB)**
- **Integration with leading visualization and analytical tools (commercial, shareware)**

Rich BioSpatial Operato

- Support the SQL query and computation requirements from needed by biotechs and pharmas and independent software vendors
- Implement indices and operators in the server to meet requirements
- Begin with simple operators and those that serve as foundations for extension
- Integration with 3rd party visualization tools



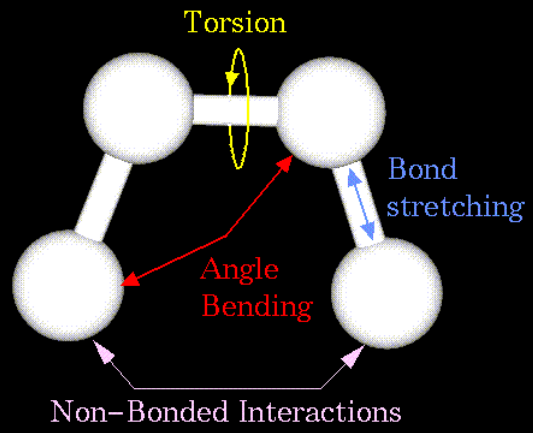


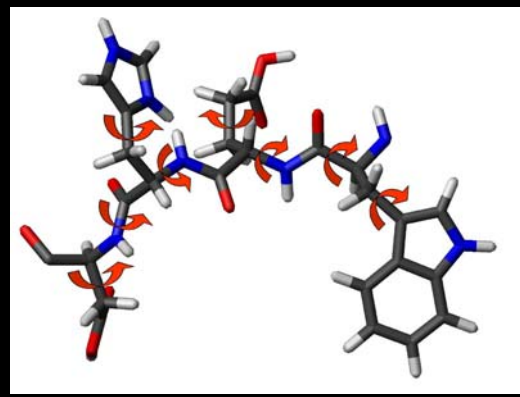
ORACLE

Foundation Operators

Sample BioSpatial Operators:

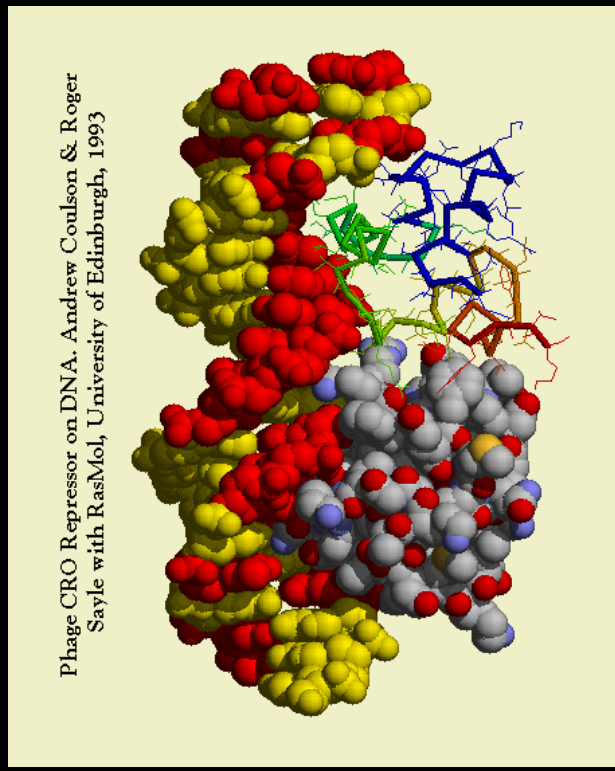
- **Nearest atom(s) to a specified position or residue in a structure**
  - Embedded atomic position index
- **Retrieve polypeptide skeleton list**
- **On-the-fly bond and bond-order computation**





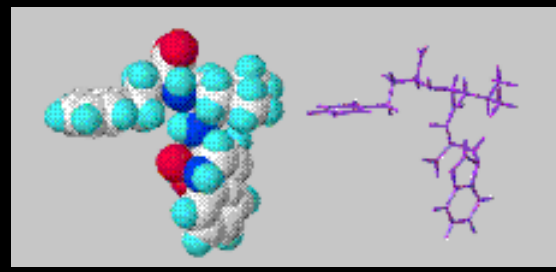
Advanced Operators

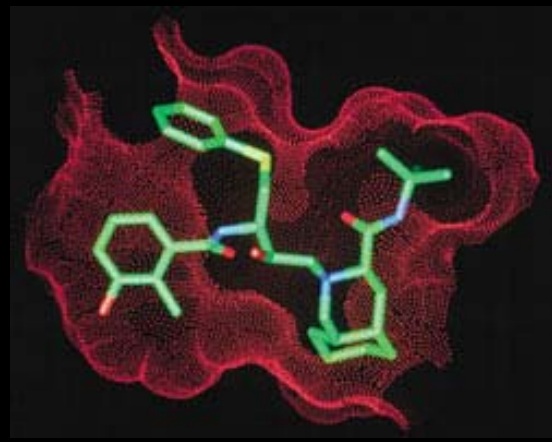
- **Protein active site identification**
- **Protein surface representation**
  - van der Waals; solvation.
- **Surface classification, abstraction**
  - Charges; hydrophobicity; H-bond donors/acceptors
  - Extraction of pharmacophore keys



Integrate with Existing Tools

- Current visualization tools based on PDB format parsers
  - Integrate with popular public domain tools and make available
- Deposition tools
  - Support transition with PDB-to-CIF conversion tool
- Protein 3rd party docking and homology applications





Oracle Life Sciences Product Directions

- Better support for life sciences data types
- Improved support for life science specific analytics
- Improved support for data import and incremental update
- Enhanced XML (XDB) & Java support in the Database and Application Server (IAS)
- Enhanced support for distributed data
- Partner with ISVs and researchers to deliver “solution”
- Customer Advisory Board participation

http:// technet.oracle.com/products/spatial
http://technet.oracle.com/products/iaswe