Infrastructure, Query Optimization, Data Warehousing and Data Mining in Support of Scientific Simulation

Yingping Huang

Department of Computer Science and Engineering
University of Notre Dame

Tuesday, October 29, 2002

Partially supported by NFS-ITR

Route

- Research area, results & motivation
- Background & technologies
- Modeling & simulation
- Infrastructure
- GUI & web interface
- Query optimization
- Data warehousing
- Data mining
- Summary & future work

Research Area and Results

- The domain
  - Scientific simulation
    - Natural organic matter (NOM)
    - Environmental biocomplexity
- The results: A simulation model
  - Agent-based
  - Stochastic
  - Web-based: J2EE & Oracle
  - Load-balancing and fail-over enabled
  - Data warehousing & data mining features included

Motivation

- IT: A fourth paradigm of scientific study? (J. Gray, et al, 2002; Fox, 2002)
  - Three previous approaches to scientific research:
    - Observation & theory
    - Hypothesis & experiment
    - Computational X & simulation
  - Information technologies
    - J2EE & middleware & XML
    - Databases & Data Warehouses
    - Data Mining
    - Visualization
    - Statistical analysis
- Natural organic matter (NOM)

Technology Used

- Agent-based modeling
  - SWARM: a library
- Stochastic modeling
- J2EE
  - JSP
  - Servlet
  - EJB
- Application Server
- Oracle
  - RDBMS
  - JDBC
  - PL/SQL
  - Reports Server
  - Data Warehouse
  - Data Mining

Agent-based Modeling

- Property of intelligent agents
  - Autonomous behavior
  - Individual world of view
  - Communicative & cooperative capacity
  - Intelligent behavior
  - Spatial mobility
- De-central control
  - Social insects & birds
- Emergent behavior
  - Patterns, clusters, self organization, etc

Chemical Reactions Models

- Classification criteria
  - Simulation time: discrete or continuous
    - Computers only do discrete computations
  - State-space: discrete or continuous
    - n-dimensional space containing all states of n variables
  - Evolution of system: deterministic or stochastic
    - Deterministic: State of system completely specified at all times
    - Stochastic: State of system represented by probability distributions & Evolution determined by probability events

Simulation of NOM and Microbial-Environmental Interactions

- NSF - ITR - Division of Environmental Biology
- Interdisciplinary project
  - Chemist
  - Geomicrobiologist
  - Biologist
  - Ecologist
  - Computer Scientist
- Stochastic Simulation of Environmental Transformations of Natural Organic Matter
  - In soil
  - In solution

Natural Organic Matter

- NOM is ubiquitous in terrestrial, aquatic and marine ecosystems
  - Results from breakdown of animal & plant material in the environment
- Important role in processes such as
  - compositional evolution and fertility of soil
  - mobility and transport of pollutants
  - availability of nutrients for microorganisms and plant communities
  - growth and dissolution of minerals
- Important to drinking water systems
  - Impacts drinking water treatment
  - Impacts quality of well water

Natural Organic Matter (cont)

Hardwood
Swamp

Natural Organic Matter (cont)

Open Channel

Natural Organic Matter (cont)

Cedar Swamp

Background

- Compositional evolution of NOM is an interesting problem
- Important aspect of predictive environmental modeling
- Prior modeling work is often
  - too simplistic to represent the heterogeneous structure of NOM and its complex behaviors in ecosystems (e.g., carbon cycling models)
  - too compute-intensive to be useful for large-scale environmental simulations (e.g., molecular models employing connectivity maps or electron densities)
- Hence, a Middle Computational Approach is taken …
  - Agent-based & stochastic

Previous work

- Models developed by other researchers
  - Deterministic models
    - METASIM (Park & Wright, 1973)
    - SCAMP (Saura, 1993)
  - Stochastic models
    - CKS (IBM, 1995)
    - BESS (Punch, 1997)
    - STOCHSIM (Firth & Bray, 2001)

Our Model

- Agent-based stochastic simulation
- GUI Version - Stand Alone
  - Animation of molecules
- Web-Based Version
  - OC4J/Orion Server & Oracle Reports
  - Oracle database servers
- Load-balancing & fail-over
  - Goal: efficiency, availability & reliability
- Data warehousing & Data Mining
  - Goal: data/pattern analysis

Modeling

- Object oriented: Molecules and microbes are objects
  - Molecules and microbes have attributes
    - Heterogeneous mixture: different attributes
  - Molecules have behaviors (physical & chemical processes)
    - Behaviors are stochastically determined
    - Dependent on the:
      - Attributes (intrinsic parameters)
      - Environment (extrinsic parameters)

Modeling (cont)

- Objects of interest
  - Macromolecular precursors: large molecules
    - Cellulose
    - Proteins
    - Lignin
  - Micromolecules: smaller molecules
    - Sugars
    - Amino acids
  - Microbes
    - Bacteria
    - Fungi

Modeling (cont)

- Attributes
  - Elemental composition
    - Number of C, H, O, N, S and P atoms in molecule
  - Functional group counts
    - Double-bonds
    - Ring structures
    - Phenyl groups
    - Alcohols
    - Phenols, ethers, esters, ketones, aldehydes, acids, aryl acids, amines, amides, thioethers, thiols, phosphoesters, phosphates
  - The time the molecule entered the system
  - Precursor type of molecule
    - Cellulose, protein, lignin, etc

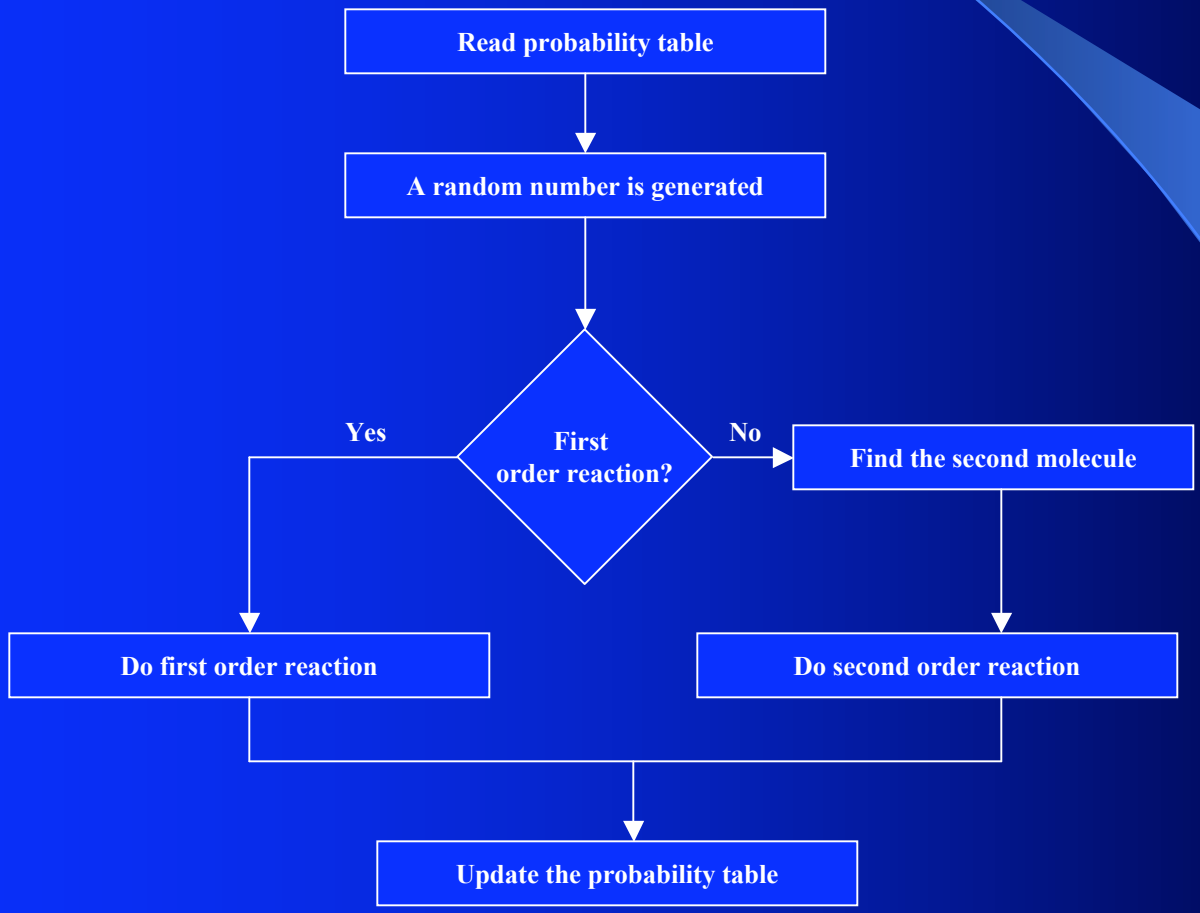
Modeling (cont)

- Behaviors (reactions and processes)
  - Physical processes
    - Adsorption (stick) to mineral surfaces
    - Aggregation/micelle formation
    - Transport downstream (surface water)
    - Transport through porous media
  - Chemical reactions
    - Abiotic bulk reactions: free molecules
    - Abiotic surface reactions: adsorbed molecules
    - Extracellular enzyme reactions on large molecules
    - Microbial uptake by small molecules

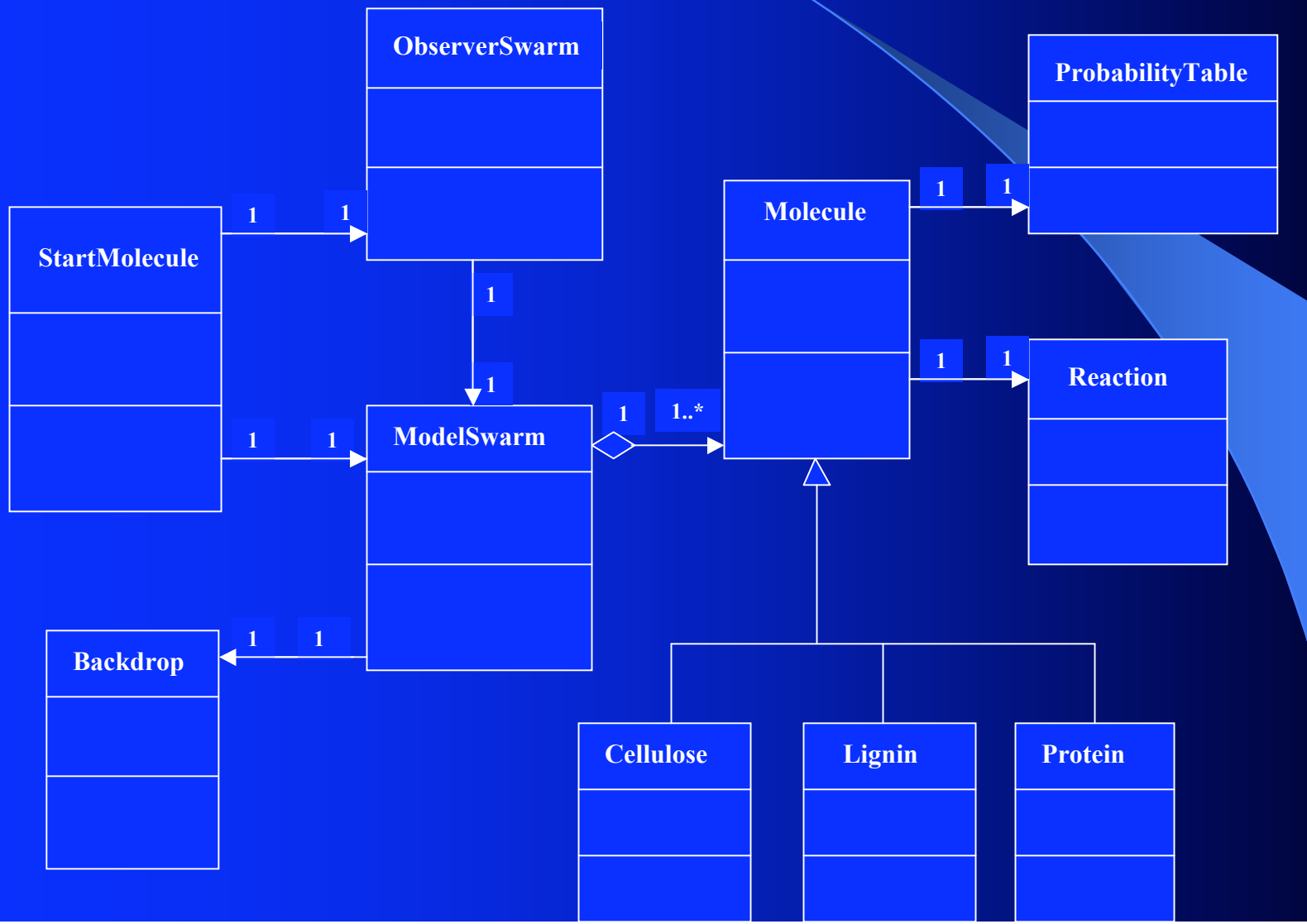
Modeling (cont)

- Environmental parameters
  - Temperature
  - pH
  - Light intensity
  - Simulation time
  - Microbial activity
  - Water flow rate/pressure gradient
  - Oxygen density

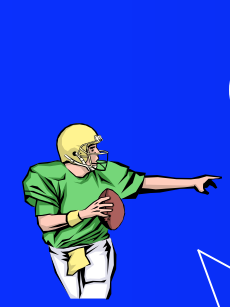
A Molecule at a Time Step



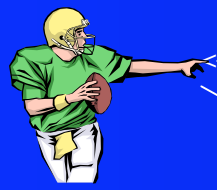
UML Class Diagram



UML Use Case Diagram



ObserverSwarm



ModelSwarm

Update Probe Display

Update World Display

Execute GUI Update Schedule

Execute Simulation Schedule

Move to New location

Update Molecule

Update ProbabilityTable

Update World

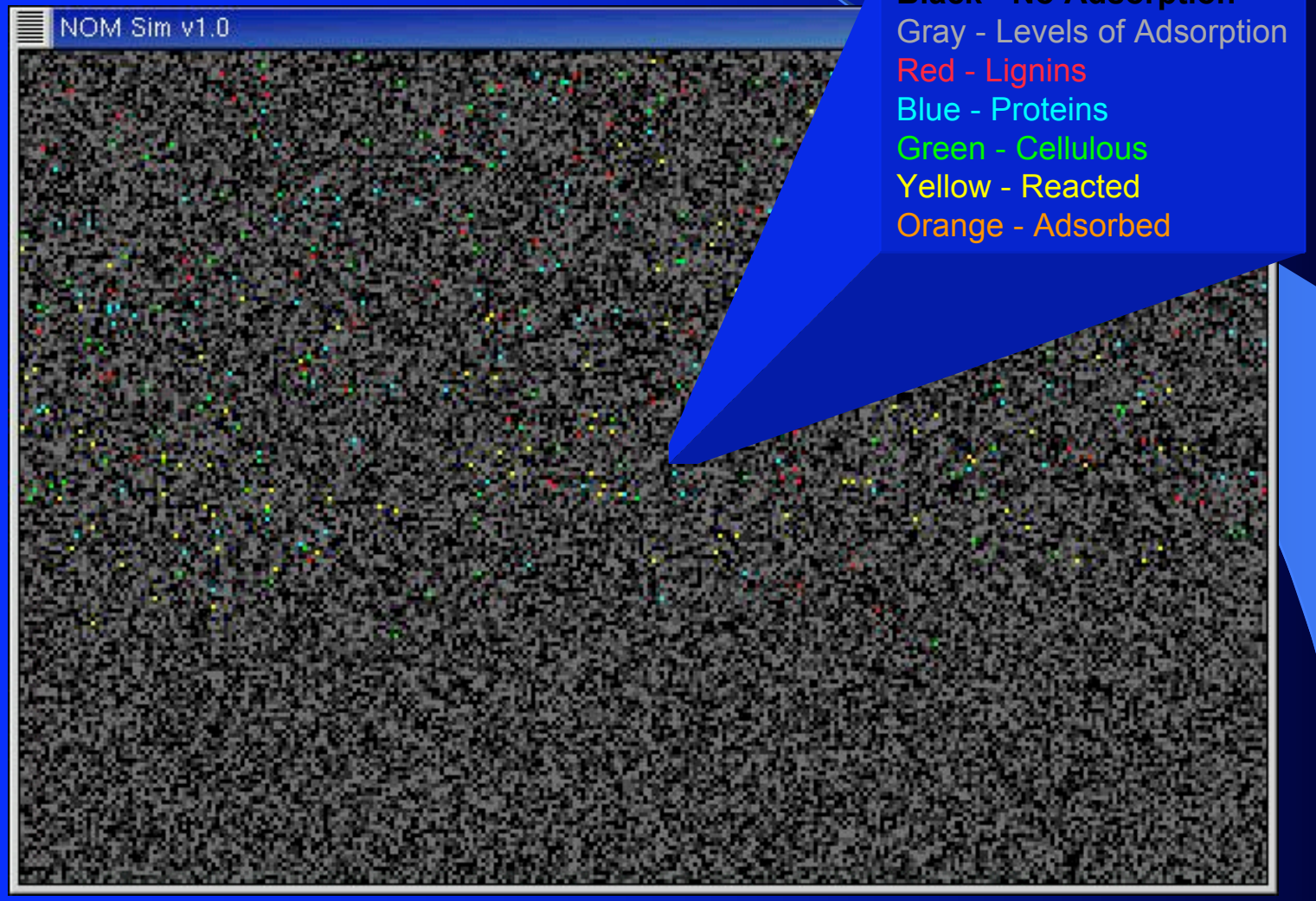
Write Database

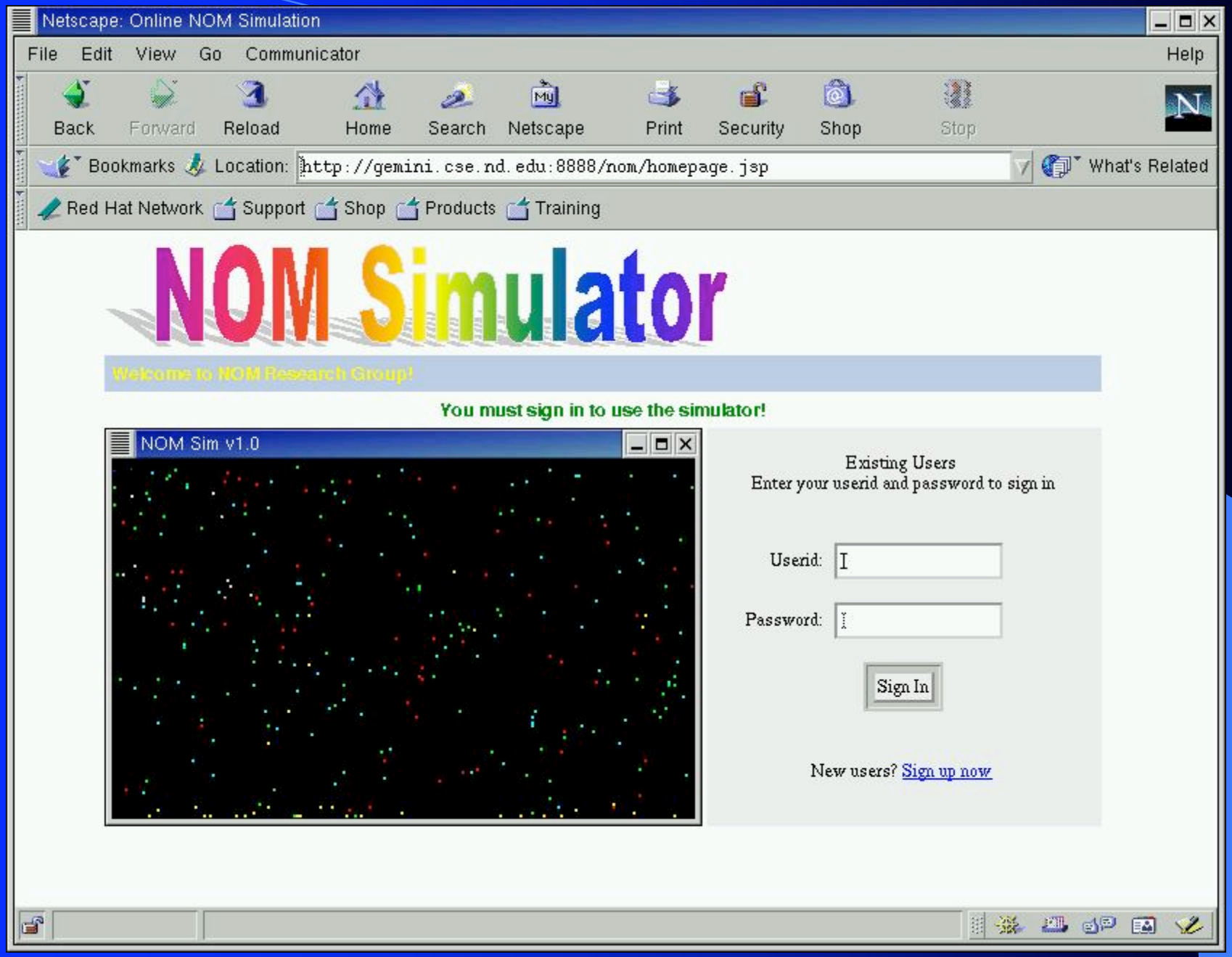
<<executes>>

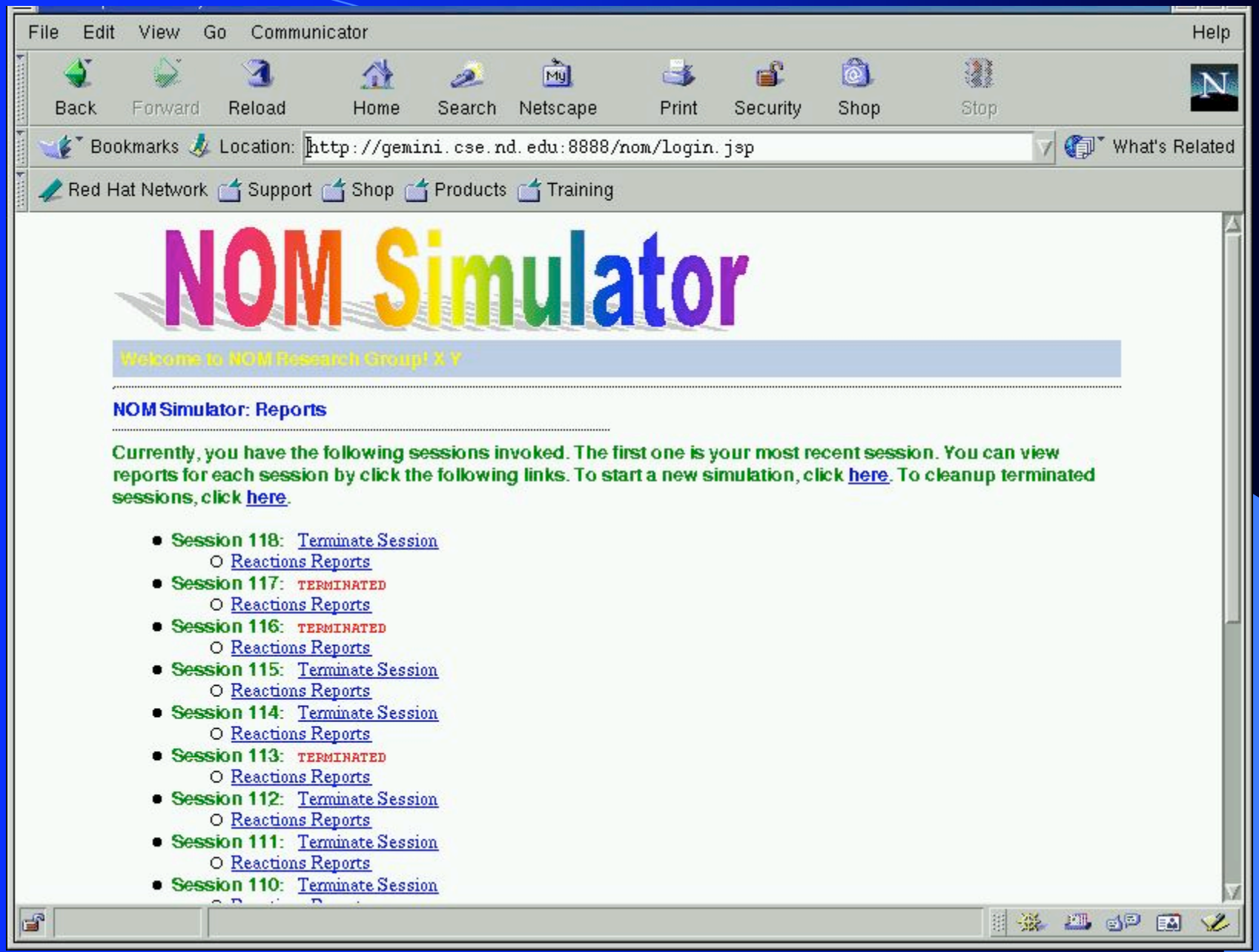
<<uses>>

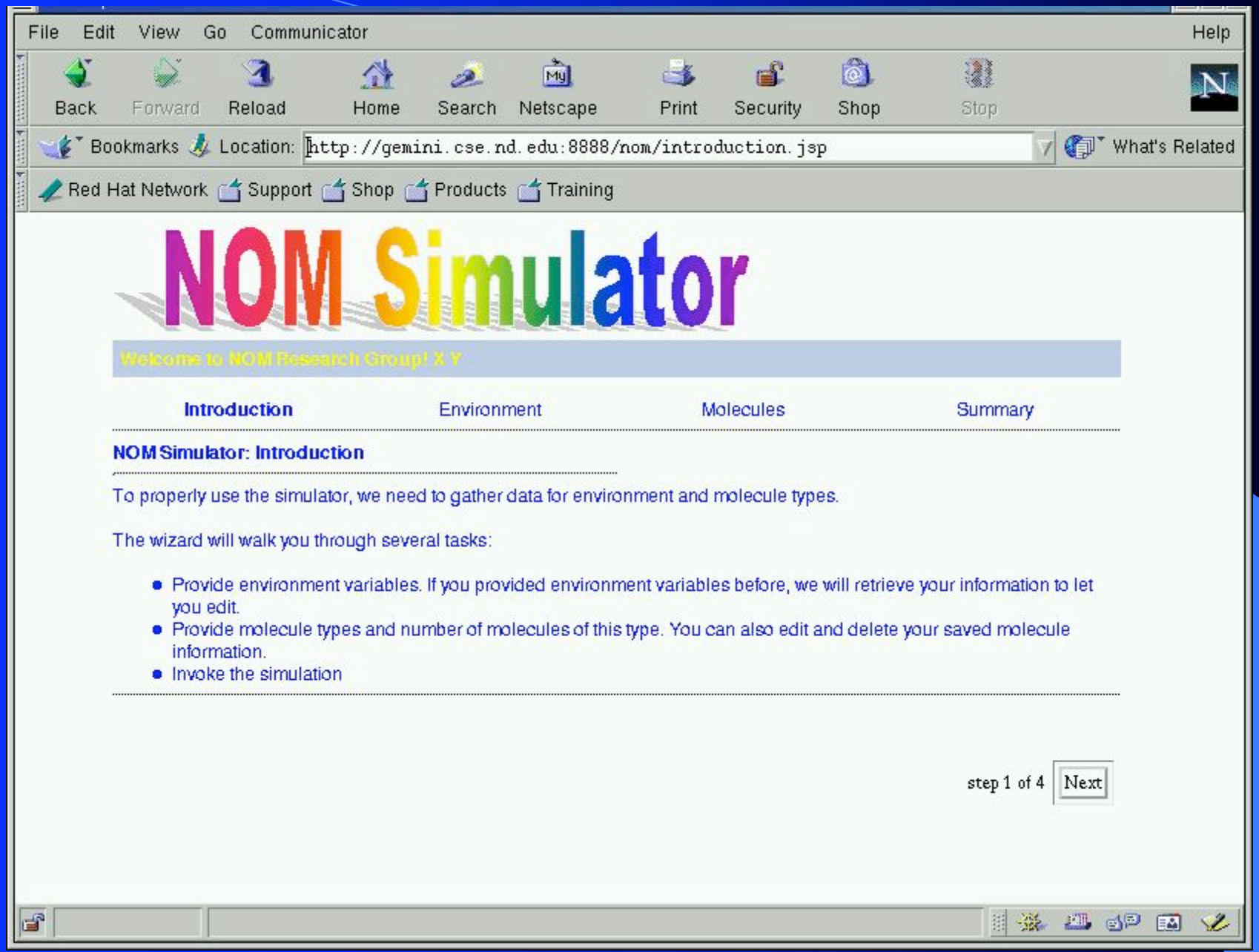
<<extends>>

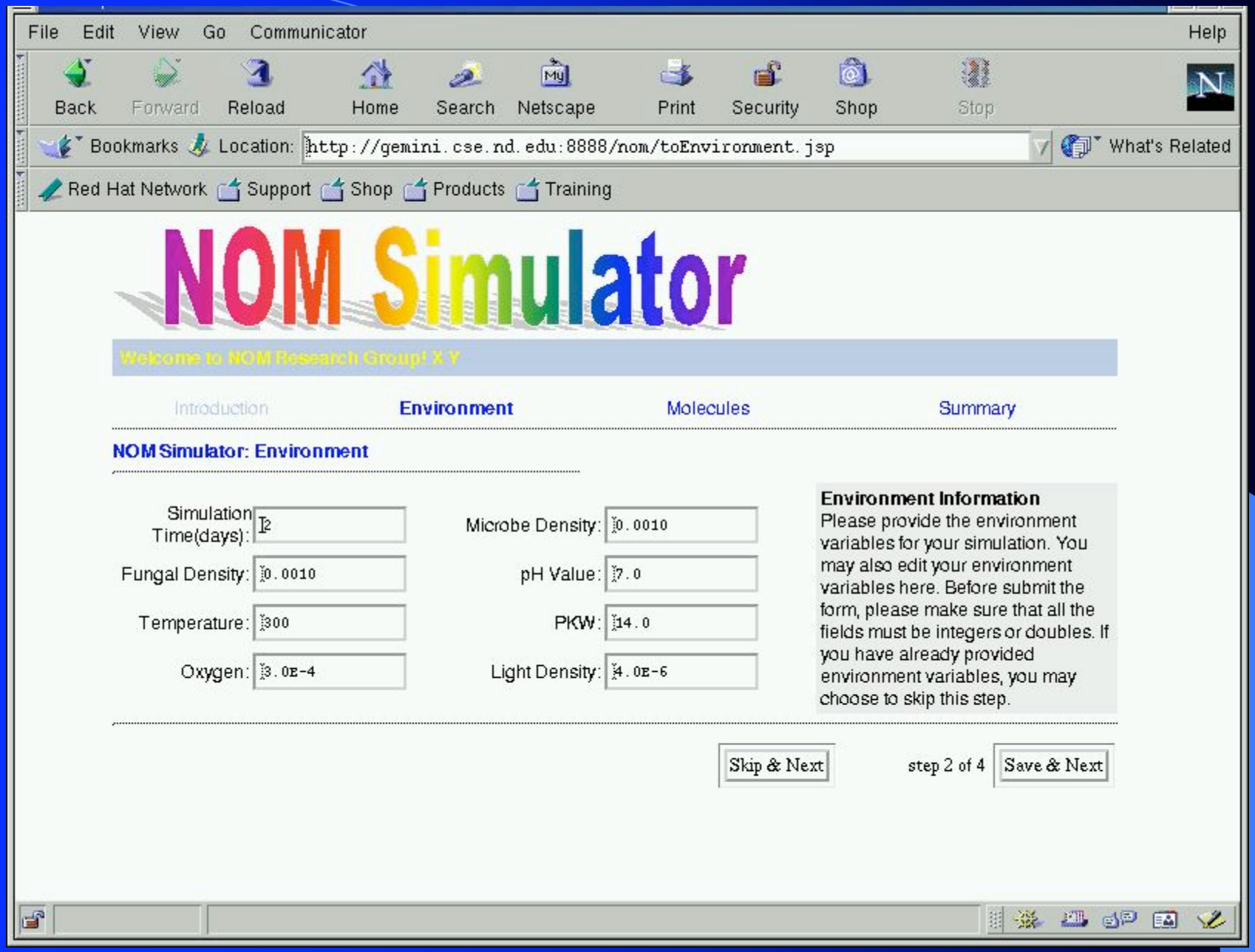
GUI Animation

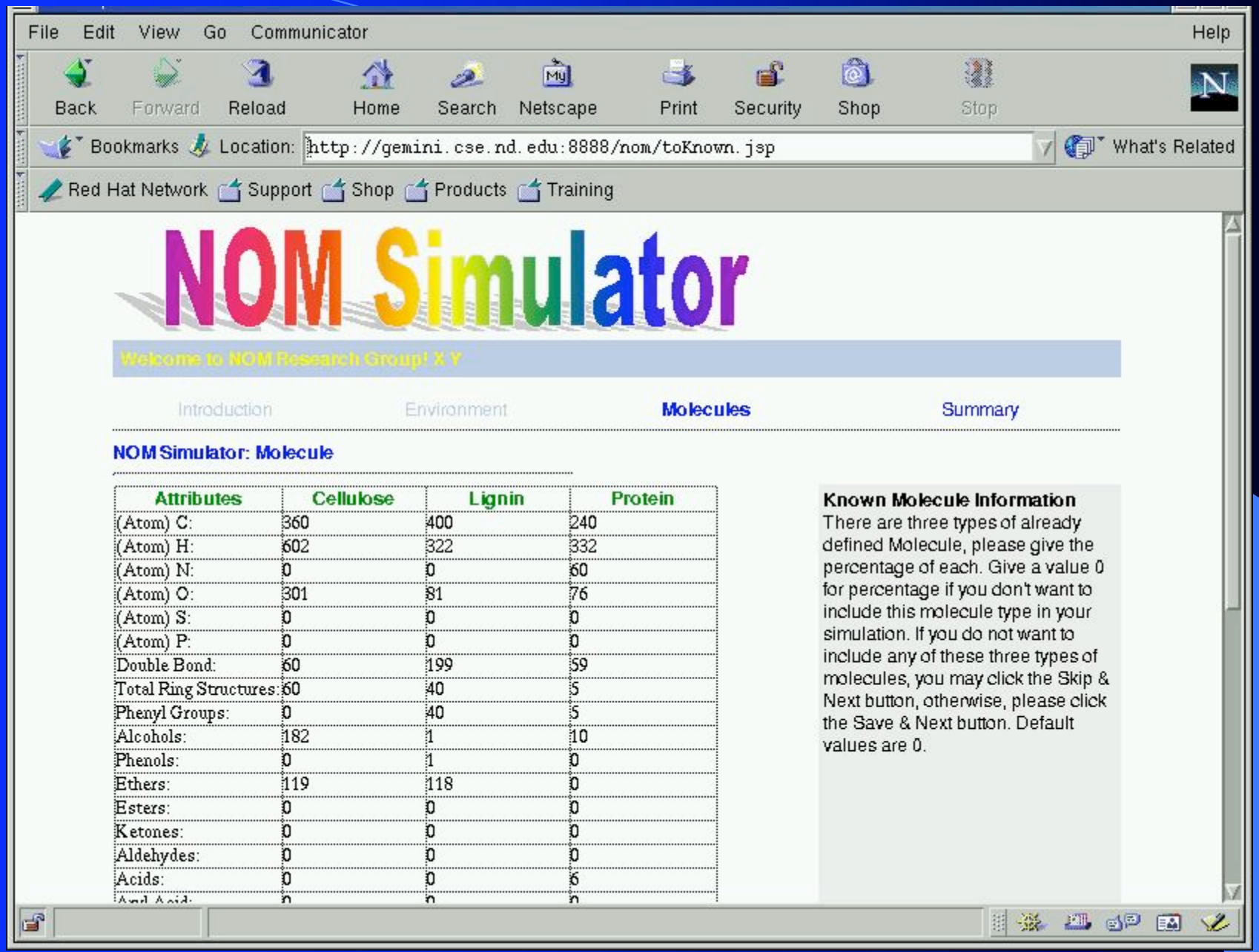



Netscape: Online NOM Simulation
File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Shop Stop
Bookmarks Location: http://gemini.cse.nd.edu:8888/nom/homepage.jsp What's Related
Red Hat Network Support Shop Products Training
NOM Simulator
Welcome to NOM Research Group!
You must sign in to use the simulator!
NOM Sim v1.0
Existing Users
Enter your userid and password to sign in
Userid:
Password:
Sign In
New users? Sign up now

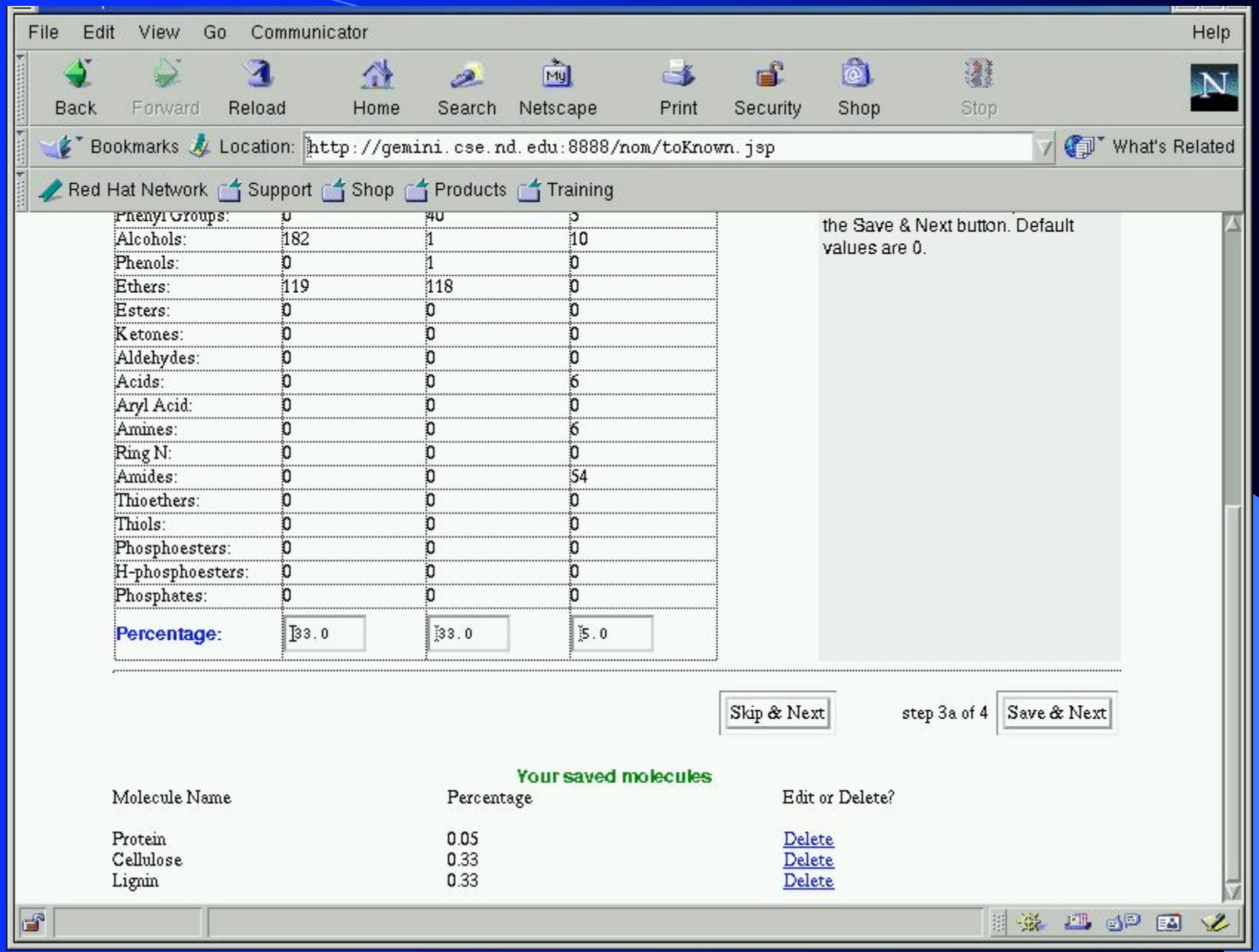

File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Shop Stop
Bookmarks Location: http://gemini.cse.nd.edu:8888/nom/login.jsp What's Related
Red Hat Network Support Shop Products Training
NOM Simulator
Welcome to NOM Research Group! X Y
NOM Simulator: Reports
Currently, you have the following sessions invoked. The first one is your most recent session. You can view reports for each session by click the following links. To start a new simulation, click here. To cleanup terminated sessions, click here.
Session 118: Terminate Session
Reactions Reports
Session 117: TERMINATED
Reactions Reports
Session 116: TERMINATED
Reactions Reports
Session 115: Terminate Session
Reactions Reports
Session 114: Terminate Session
Reactions Reports
Session 113: TERMINATED
Reactions Reports
Session 112: Terminate Session
Reactions Reports
Session 111: Terminate Session
Reactions Reports
Session 110: Terminate Session

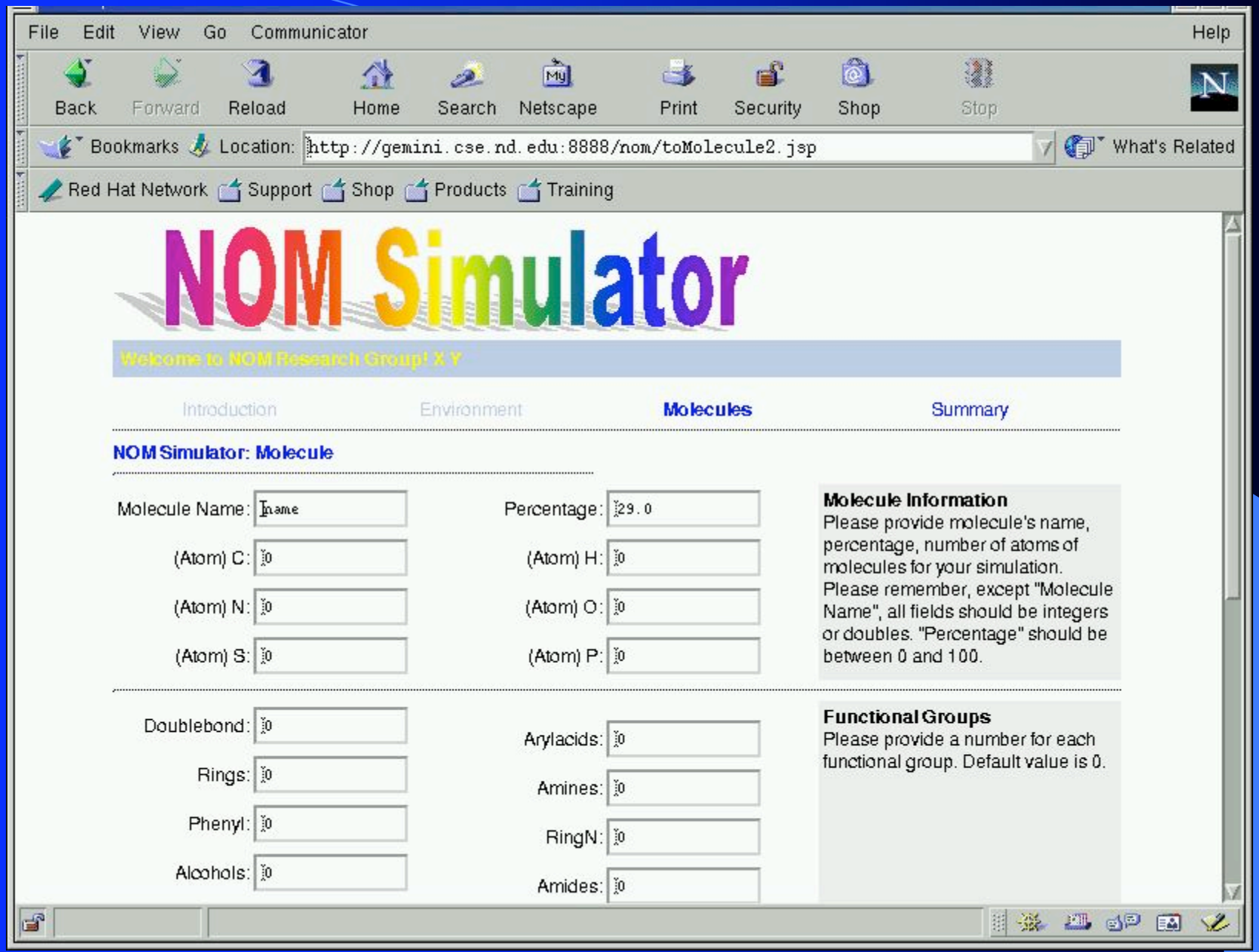

File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Shop Stop
Bookmarks Location: http://gemini.cse.nd.edu:8888/nom/introduction.jsp What's Related
Red Hat Network Support Shop Products Training
NOM Simulator
Welcome to NOM Research Group! X Y
Introduction Environment Molecules Summary
NOM Simulator: Introduction
To properly use the simulator, we need to gather data for environment and molecule types.
The wizard will walk you through several tasks:
Provide environment variables. If you provided environment variables before, we will retrieve your information to let you edit.
Provide molecule types and number of molecules of this type. You can also edit and delete your saved molecule information.
Invoke the simulation
step 1 of 4 Next

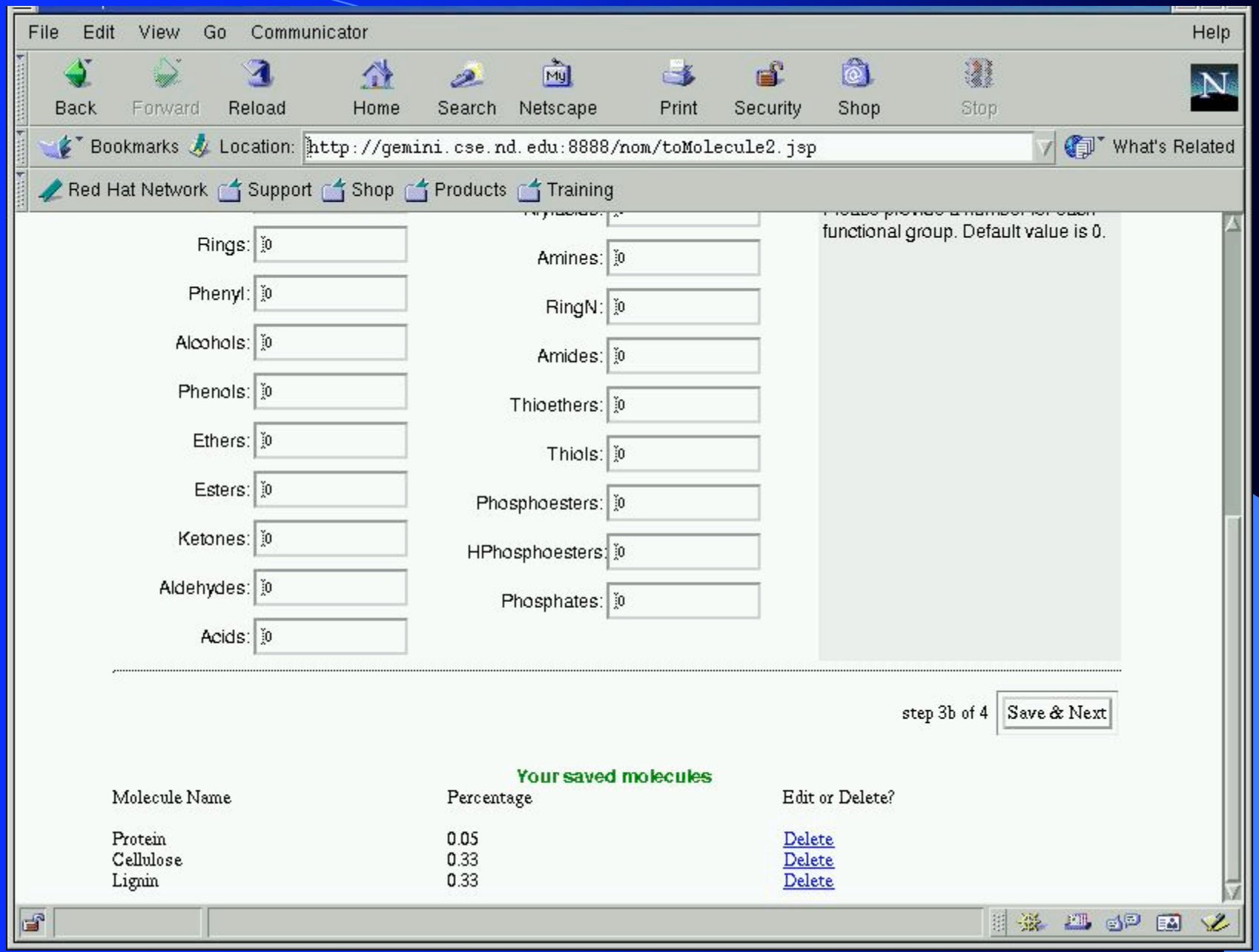

File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Shop Stop
Bookmarks Location: http://gemini.cse.nd.edu:8888/nom/toEnvironment.jsp What's Related
Red Hat Network Support Shop Products Training
NOM Simulator
Welcome to NOM Research Group! X Y
Introduction Environment Molecules Summary
NOM Simulator: Environment
Simulation Time(days): 2
Microbe Density: 0.0010
Fungal Density: 0.0010
pH Value: 7.0
Temperature: 300
PKW: 14.0
Oxygen: 3.0E-4
Light Density: 4.0E-6
Environment Information
Please provide the environment variables for your simulation. You may also edit your environment variables here. Before submit the form, please make sure that all the fields must be integers or doubles. If you have already provided environment variables, you may choose to skip this step.
Skip & Next
step 2 of 4
Save & Next

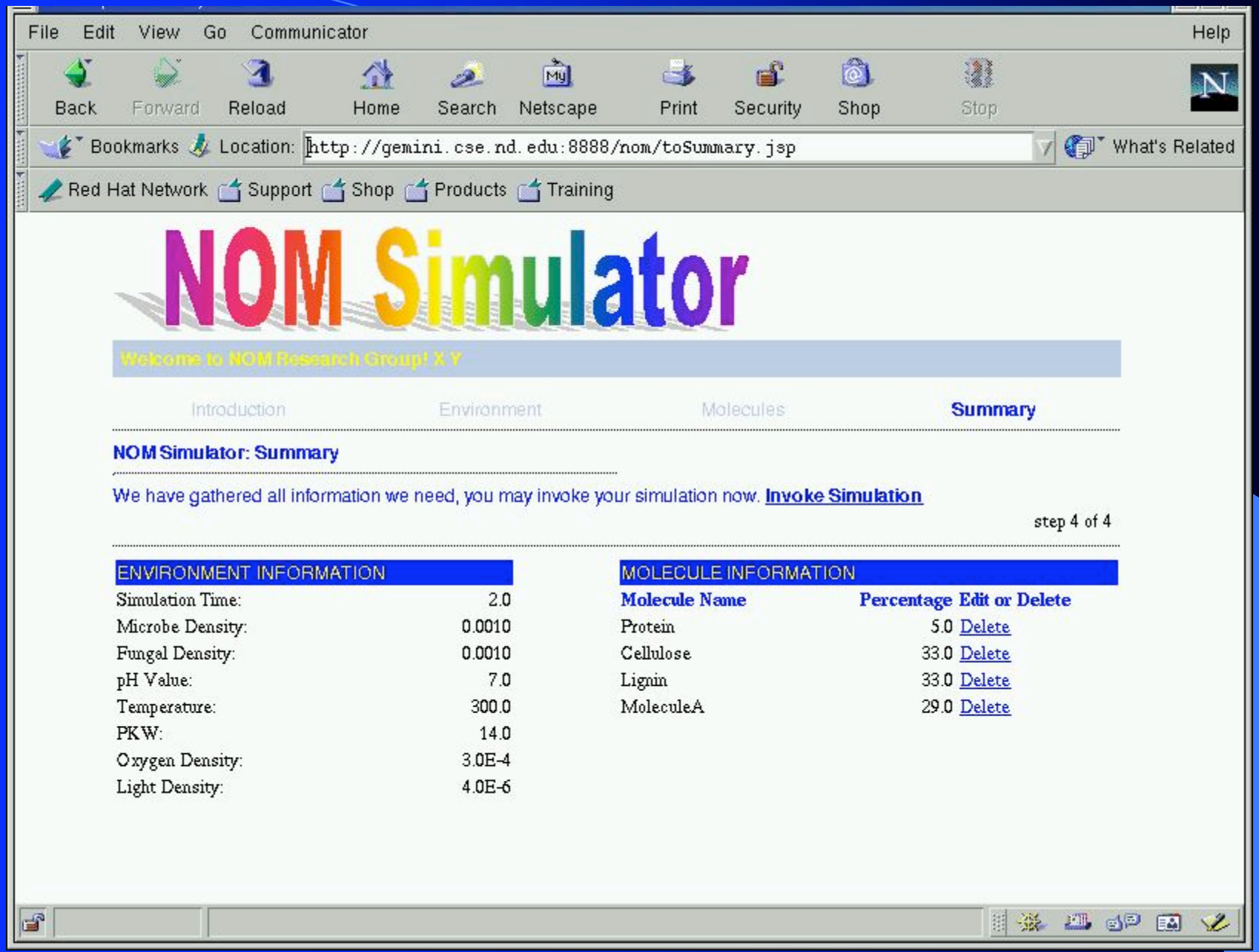


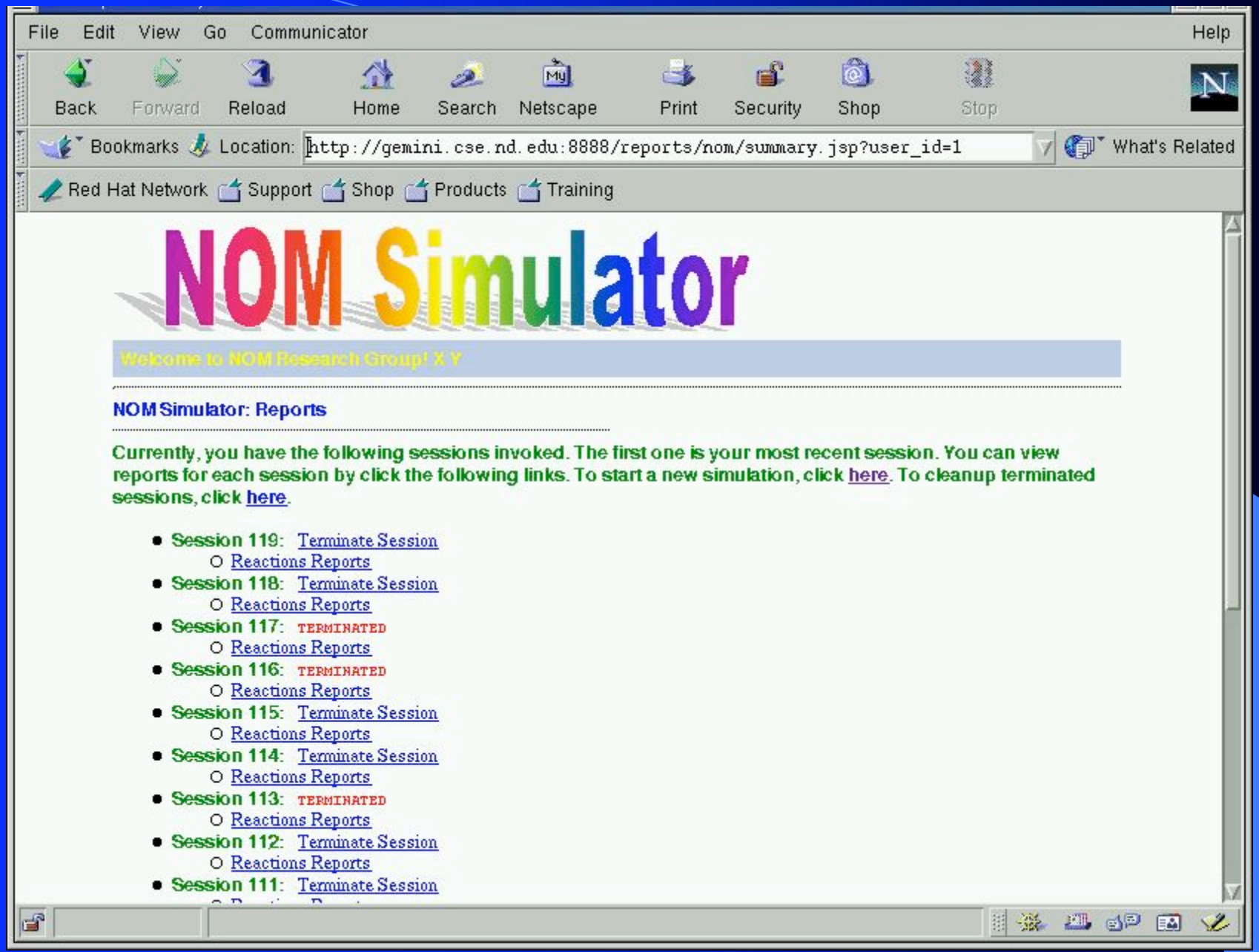
Attributes	Cellulose	Lignin	Protein
(Atom) C:	360	400	240
(Atom) H:	602	322	332
(Atom) N:	0	0	60
(Atom) O:	301	81	76
(Atom) S:	0	0	0
(Atom) P:	0	0	0
Double Bond:	60	199	59
Total Ring Structures:	60	40	5
Phenyl Groups:	0	40	5
Alcohols:	182	1	10
Phenols:	0	1	0
Ethers:	119	118	0
Esters:	0	0	0
Ketones:	0	0	0
Aldehydes:	0	0	0
Acids:	0	0	6
[illegible] Acid:	0	0	0

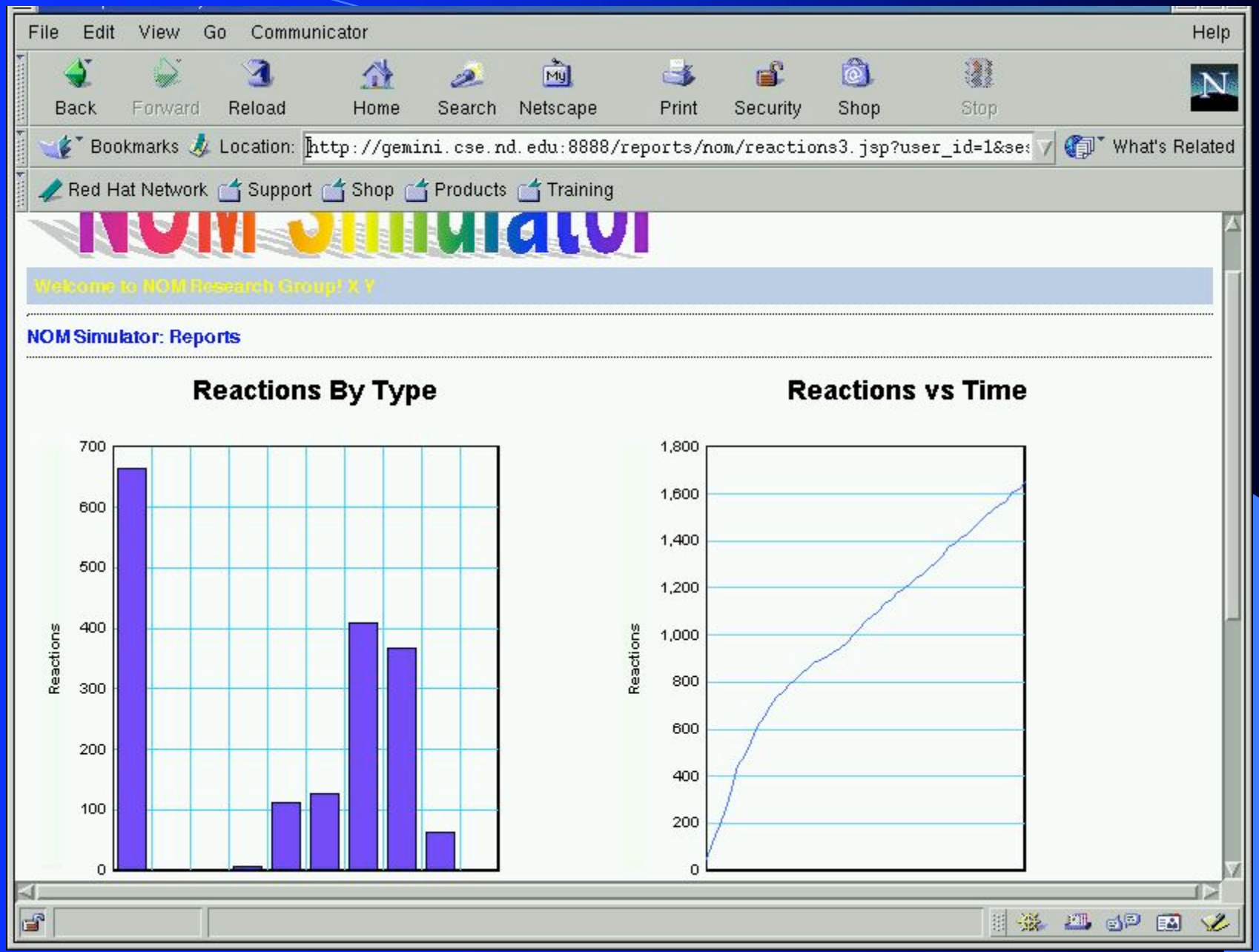

File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Shop Stop
Bookmarks Location: http://gemini.cse.nd.edu:8888/nom/toKnown.jsp What's Related
Red Hat Network Support Shop Products Training
Phenyl Groups: 0 40 5
Alcohols: 182 1 10
Phenols: 0 1 0
Ethers: 119 118 0
Esters: 0 0 0
Ketones: 0 0 0
Aldehydes: 0 0 0
Acids: 0 0 6
Aryl Acid: 0 0 0
Amines: 0 0 6
Ring N: 0 0 0
Amides: 0 0 54
Thioethers: 0 0 0
Thiols: 0 0 0
Phosphoesters: 0 0 0
H-phosphoesters: 0 0 0
Phosphates: 0 0 0
Percentage: 33.0 33.0 5.0
the Save & Next button. Default values are 0.
Skip & Next
step 3a of 4
Save & Next
Your saved molecules
Molecule Name
Percentage
Edit or Delete?
Protein 0.05 Delete
Cellulose 0.33 Delete
Lignin 0.33 Delete


File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Shop Stop
Bookmarks Location: http://gemini.cse.nd.edu:8888/nom/toMolecule2.jsp What's Related
Red Hat Network Support Shop Products Training
NOM Simulator
Welcome to NOM Research Group! X Y
Introduction Environment Molecules Summary
NOM Simulator: Molecule
Molecule Name: name
Percentage: 29.0
(Atom) C: 0
(Atom) H: 0
(Atom) N: 0
(Atom) O: 0
(Atom) S: 0
(Atom) P: 0
Molecule Information
Please provide molecule's name, percentage, number of atoms of molecules for your simulation. Please remember, except "Molecule Name", all fields should be integers or doubles. "Percentage" should be between 0 and 100.
Doublebond: 0
Rings: 0
Phenyl: 0
Alcohols: 0
Arylacids: 0
Amines: 0
RingN: 0
Amides: 0
Functional Groups
Please provide a number for each functional group. Default value is 0.

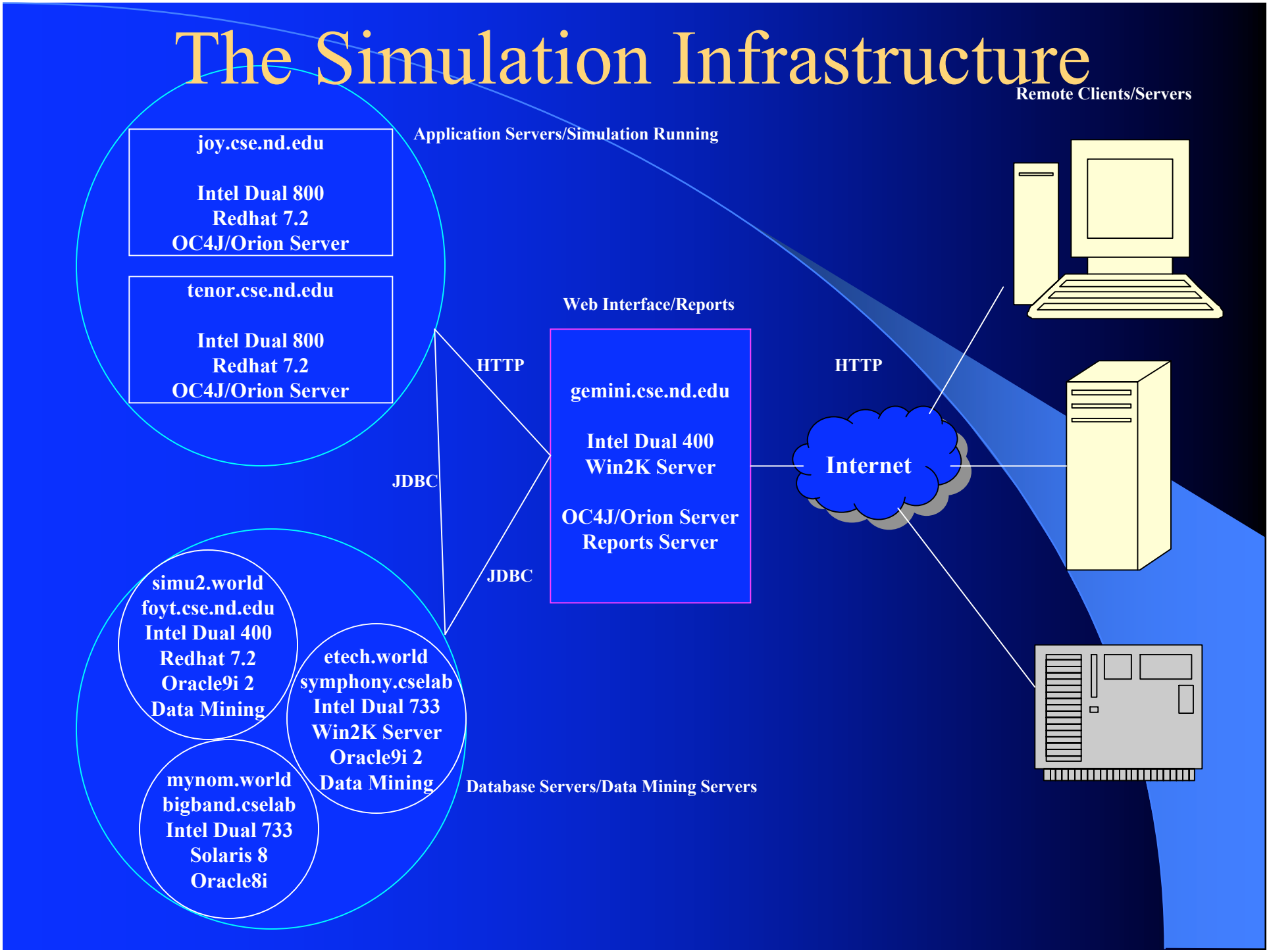

File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Shop Stop
Bookmarks Location: http://gemini.cse.nd.edu:8888/nom/toMolecule2.jsp What's Related
Red Hat Network Support Shop Products Training
Rings: 0
Phenyl: 0
Alcohols: 0
Phenols: 0
Ethers: 0
Esters: 0
Ketones: 0
Aldehydes: 0
Acids: 0
Amines: 0
RingN: 0
Amides: 0
Thioethers: 0
Thiols: 0
Phosphoesters: 0
HPhosphoesters: 0
Phosphates: 0
functional group. Default value is 0.
step 3b of 4
Save & Next
Your saved molecules
Molecule Name
Percentage
Edit or Delete?
Protein 0.05 Delete
Cellulose 0.33 Delete
Lignin 0.33 Delete


File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Shop Stop
Bookmarks Location: http://gemini.cse.nd.edu:8888/nom/toSummary.jsp What's Related
Red Hat Network Support Shop Products Training
NOM Simulator
Welcome to NOM Research Group! X Y
Introduction Environment Molecules Summary
NOM Simulator: Summary
We have gathered all information we need, you may invoke your simulation now. Invoke Simulation
step 4 of 4
ENVIRONMENT INFORMATION
Simulation Time: 2.0
Microbe Density: 0.0010
Fungal Density: 0.0010
pH Value: 7.0
Temperature: 300.0
PKW: 14.0
Oxygen Density: 3.0E-4
Light Density: 4.0E-6
MOLECULE INFORMATION
Molecule Name Percentage Edit or Delete
Protein 5.0 Delete
Cellulose 33.0 Delete
Lignin 33.0 Delete
MoleculeA 29.0 Delete


File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Shop Stop
Bookmarks Location: http://gemini.cse.nd.edu:8888/reports/nom/summary.jsp?user_id=1 What's Related
Red Hat Network Support Shop Products Training
NOM Simulator
Welcome to NOM Research Group! X Y
NOM Simulator: Reports
Currently, you have the following sessions invoked. The first one is your most recent session. You can view reports for each session by click the following links. To start a new simulation, click here. To cleanup terminated sessions, click here.
Session 119: Terminate Session
Reactions Reports
Session 118: Terminate Session
Reactions Reports
Session 117: TERMINATED
Reactions Reports
Session 116: TERMINATED
Reactions Reports
Session 115: Terminate Session
Reactions Reports
Session 114: Terminate Session
Reactions Reports
Session 113: TERMINATED
Reactions Reports
Session 112: Terminate Session
Reactions Reports
Session 111: Terminate Session


File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Shop Stop
Bookmarks Location: http://gemini.cse.nd.edu:8888/reports/nom/reactions3.jsp?user_id=1&ses What's Related
Red Hat Network Support Shop Products Training
Welcome to NOM Research Group! X Y
NOM Simulator: Reports
Reactions By Type
Reactions
700
600
500
400
300
200
100
0
Reactions vs Time
Reactions
1,800
1,600
1,400
1,200
1,000
800
600
400
200
0

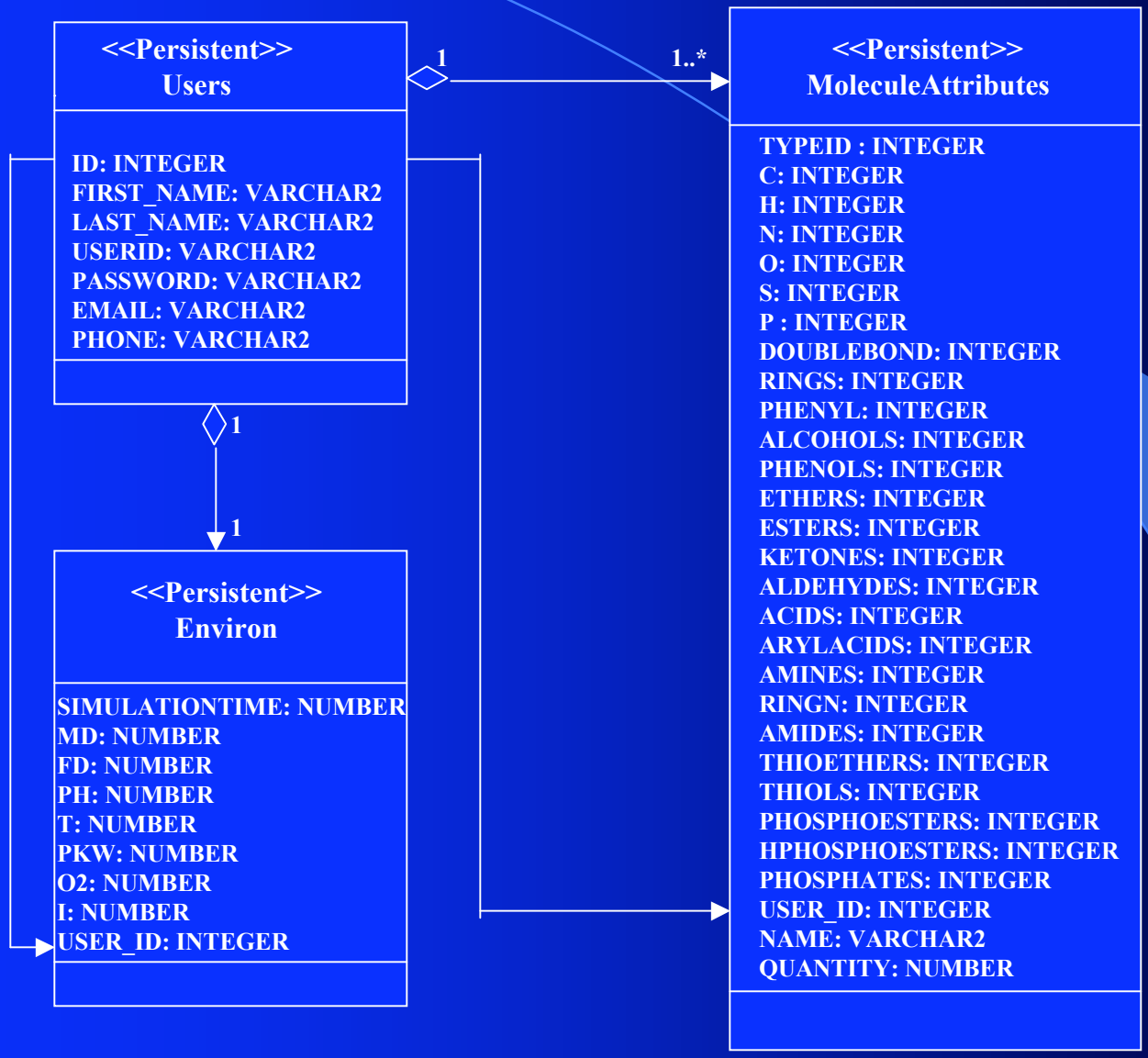
The Simulation Infrastructure



NOM 1.0

- Loosely coupled distributed systems
  - 2 Application servers (Orion Servers)
  - 3 Database servers (Oracle)
  - Reports server (OC4J Server/Reports Server)
- Load balancing (round robin based on computational needs)
  - application servers & database servers
- Fail over
  - application servers & database servers
  - Multi-master replication of important tables
- Why fail-over (Assume down probability p for each machine)
  - No fail-over
    - Simulation system down probability: $1-(1-p)^2 = 2p-p^2$
  - With fail-over
    - Simulation system down probability: $1-(1-p^2)(1-p^3) = p^2 + p^3 – p^5$
  - Improvement:
    - $2/p = 200$ if $p=0.01$ (the smaller p, the larger improvement)

Simulation Configuration Data Model



Simulation Data

- Molecule_ID
  - All molecule entered the system or produced by chemical reactions have a molecule_id
- Session_ID
  - Each simulation session has a unique ID
- TimeStamp
  - Each time step of the system is associated with molecules
- xPos & yPos

Simulation Data (Cont)

- Parent1 & Parent2
  - If first order reaction, parent2 is NULL
- Reaction probabilities
  - After a chemical reaction, probability tables are updated
- Molecule structures
  - After a chemical reaction, molecule structures are updated

Query Optimization

- Insertion performance
  - Disable indexes
  - Disable constraints
- Query performance
  - Indexes
  - Aggregation tables
- Space utilization
  - PCTFREE & PCTUSED & INITRANS & MAXTRANS
  - Drop indexes

Query/Report Examples

- Example 1:
  - Show the number of chemical reactions for each of the ten reaction types so far in the simulation using bar charts
- Example 2:
  - Create a line graph which shows the trend of the total number of chemical reactions vs time steps.

Example 1

```
SQL> select nom.reactiontype "Reaction Type",
  2              reactiontype.rname "Reaction Name",
  3              count(nom.moleculeid) "Reactions"
  4       from nom, reactiontype
  5      where nom.reactiontype=reactiontype.rtype
  6        and sessionid=:session_id and user_id=:user_id
  7      group by nom.reactiontype, reactiontype.rname
  8       order by nom.reactiontype;


Elapsed: 00:00:10.03
```

Example 2

```
SQL> select t1.timestamp “Time Step”,
2                 sum(t2.total) “Reactions”
3           from (select timestamp,
4                                count(moleculeid) total
5                      from nom
6                     where sessionid=:session_id
7                       and user_id=:user_id
8                      group by timestamp ) t1,
9                    (select timestamp,
10                                count(moleculeid) total
11                     from nom
12                    where sessionid=:session_id
13                      and user_id=:user_id
14                     group by timestamp ) t2
15          where t2.timestamp <= t1.timestamp
16    group by t1.timestamp;

Elapsed: 01:20:10.23
```

Aggregation Tables

- Example 1
  - REACTIONS_BY_TYPE
    - Session_ID & Reaction Type & Reactions
  - Updated at the end of every time step
- Example 2
  - REACTIONS_BY_TIME
    - Session_ID & Time Step & Total Reactions
  - A new row inserted at the end of every time step

Insertion and Query Performance Comparison

Scenario (>16million)	Insertion (sec/row)	Query Time (example 2)
No indexes No aggregation	0.0106	>1 hour
With indexes	0.0122	>0.5 hour
With aggregations	0.0107	5 seconds

Data Warehousing

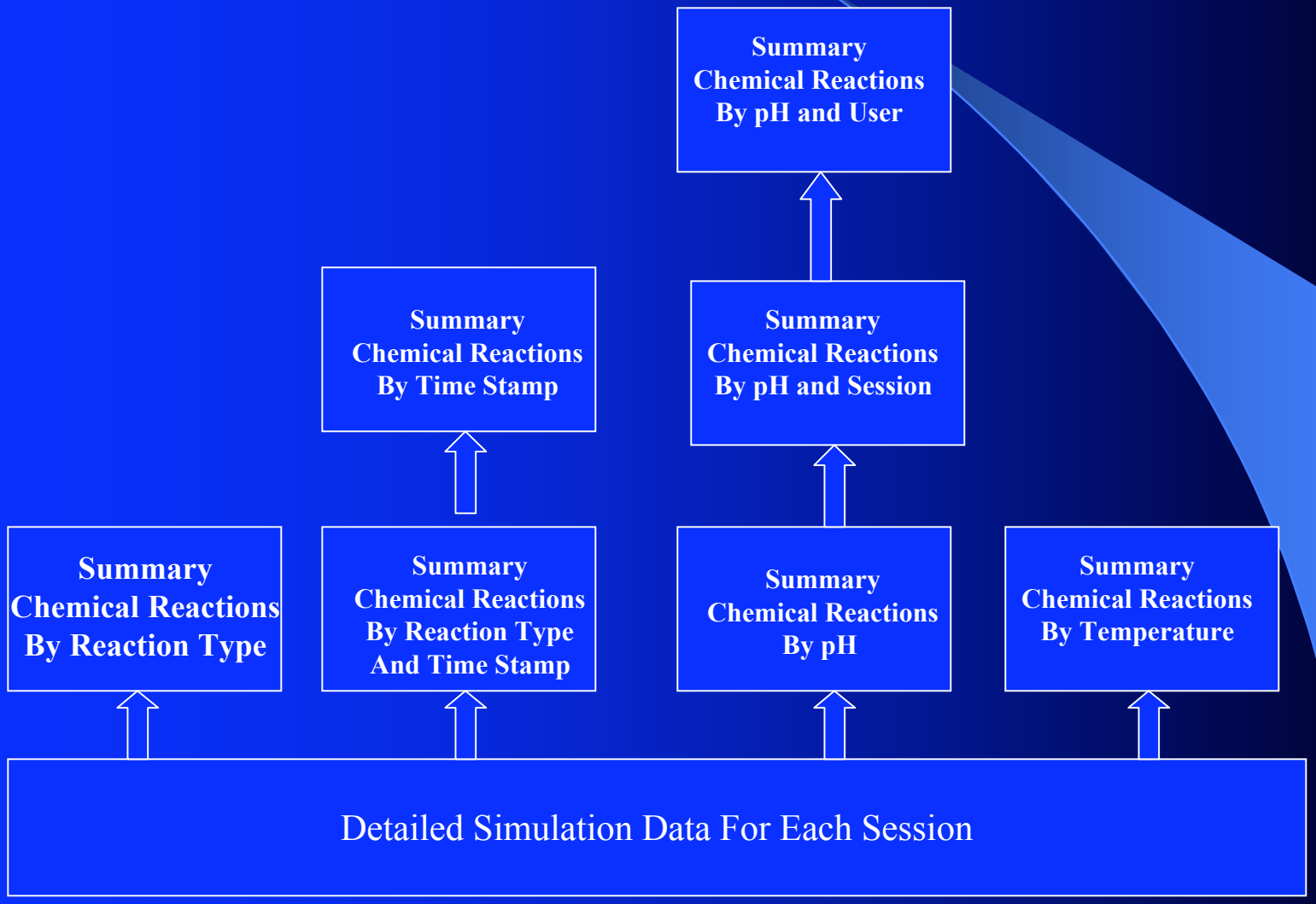
- A data warehouse is a database with the following properties:
  - Subject oriented
    - Define data warehouse by subject matter
  - Integrated
    - Consistent format, data integrity
  - Non-volatile
    - Rarely update
  - Time-variant
    - Data collected over time, temporal attributes

Inmon, 1996

Logical Design of The Data Warehouse

- Conceptual & abstract
  - Define the metadata
  - Entity-relationship modeling
  - Using Oracle Designer to generate DDL
- Two design approaches
  - Detail and Summary Schema
  - Star Schema

Detail and Summary Schema



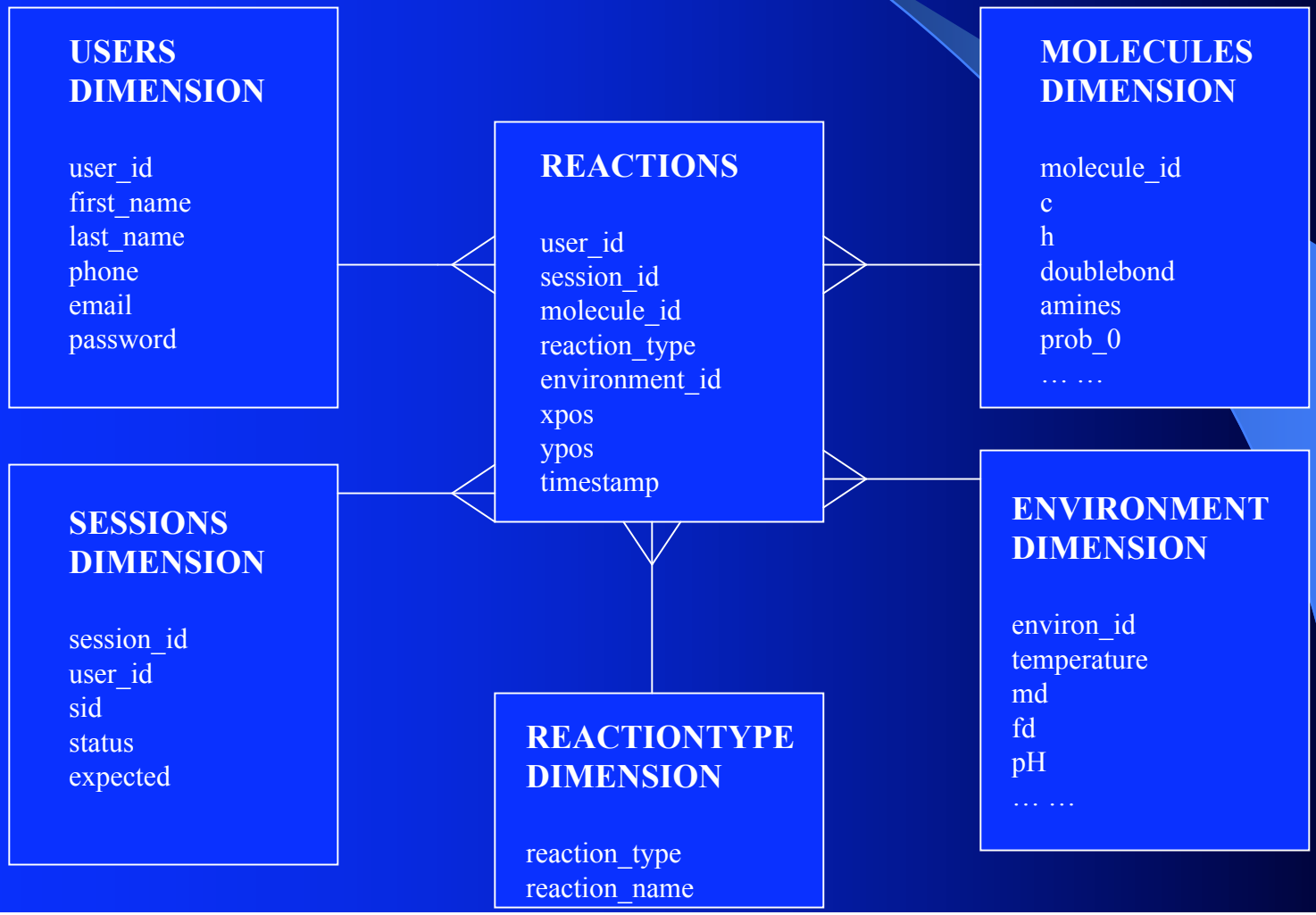
Advantages and Disadvantages of Detail and Summary Schema

- Advantages
  - Easy to navigate
    - Incorporate data from other related tables to avoid join operation from the summary
    - For example, The REACTIONS_BY_TYPE avoids join of NOM and REACTIONTYPE.
- Disadvantages
  - What summarizations are anticipated?

Star Schema

- Derived from multidimensional database design (Kimball, 1996)
- Facts tables
  - Central large tables
- Dimension tables
  - Descriptive attributes about a dimension in facts tables
- Fact table has a foreign key relationship to each dimension table
- More flexible than Detail and Summary Schema
  - Summary and GROUP BY in Detail and Summary Schema

A Star Schema



Build the Data Warehouse

- Oracle database as the data warehouse
- Tablespaces design
  - I/O contention reduction
    - Files associated with each tablespace are striped across multiple disks
- Predict space requirement
  - Load sample data
  - ANALYZE command
  - STATSPACK
- Space availability insurance
  - AUTOEXTEND
- Partitioned tables and indexes

Populate the Data Warehouse

- Tools
  - SQL*Loader
  - Export/Import
  - SQL*Plus copy command
  - CREATE TABLE … AS SELECT command
  - JDBC
- Data preprocessing
  - Data Cleansing
  - Resolve name and format inconsistencies
- Summary & aggregation

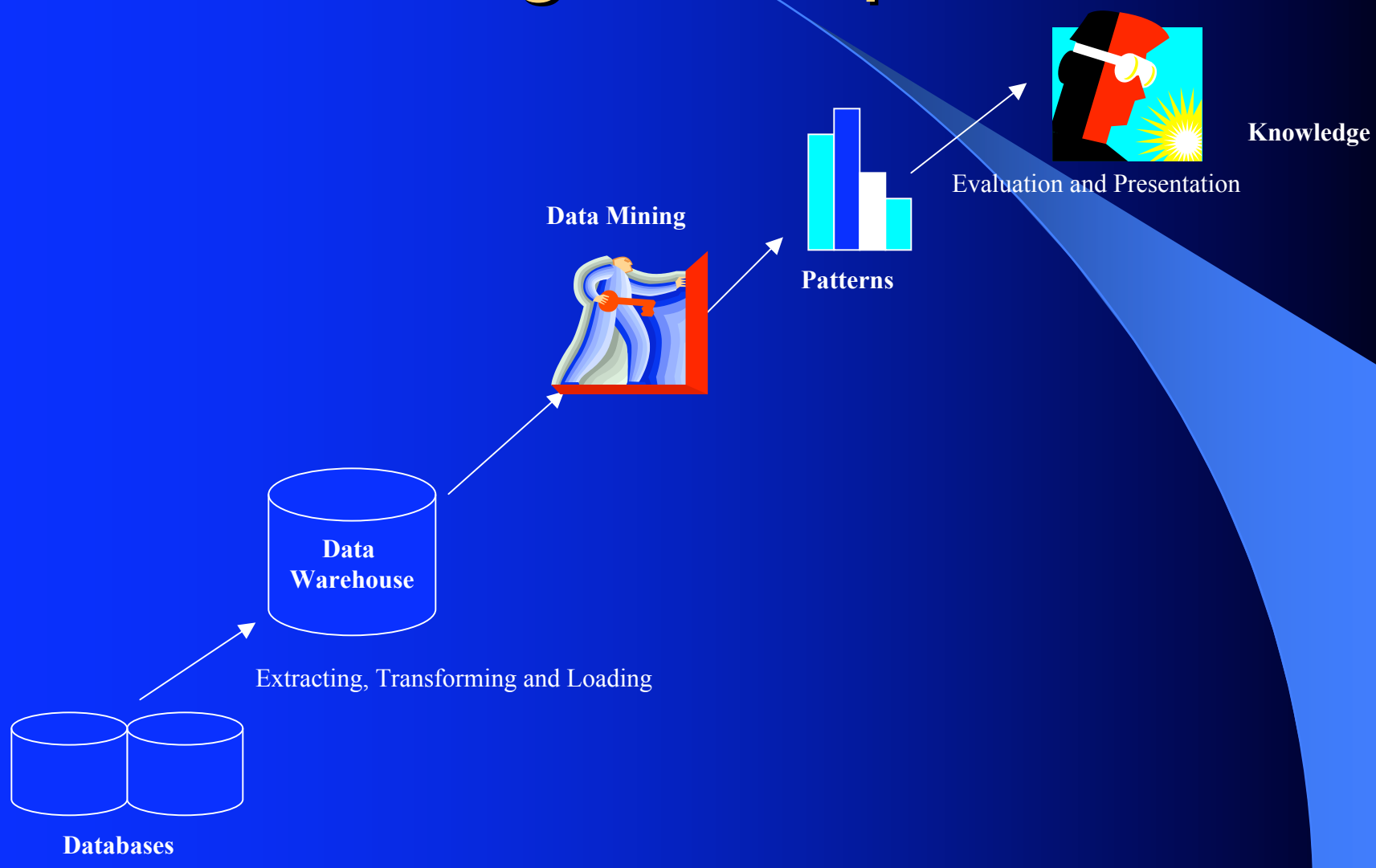
Query Optimization for the Data Warehouse

- Optimization techniques involved
  - Ordered hint: SELECT /*+ordered*/ …
    - reducing parsing time
    - For example, join of 9 tables has 8!=40320 join combinations; parsing takes more than 30 minutes
  - Star hint: SELECT /*+star*/ …
    - Hash join
    - Bitmap indexes
    - Result in reducing I/O
  - Partitioning
    - Join devided into small joins

Data Mining

- Data mining refers to extracting or mining knowledge from a large amount of data
- Other terms
  - Knowledge discovery in database (KDD)
  - Data/pattern analysis
  - Information retrieval
  - Machine learning

Data Mining as Step of KDD



Oracle Data Mining

- ODM has two components
  - Data Mining API
    - Provides an early look at concepts and approaches begin proposed for the emerging standard Java Data Mining (JDM)
    - Based on data mining standards
      - Object Management Group's Common Warehouse Metadata (CWM)
      - Data Mining Group's Predictive Model Markup Language (PMML)
      - International Standards Organization's SQL/MM for Data Mining
  - Data Mining Server
    - Server-side in-database component that performs data mining

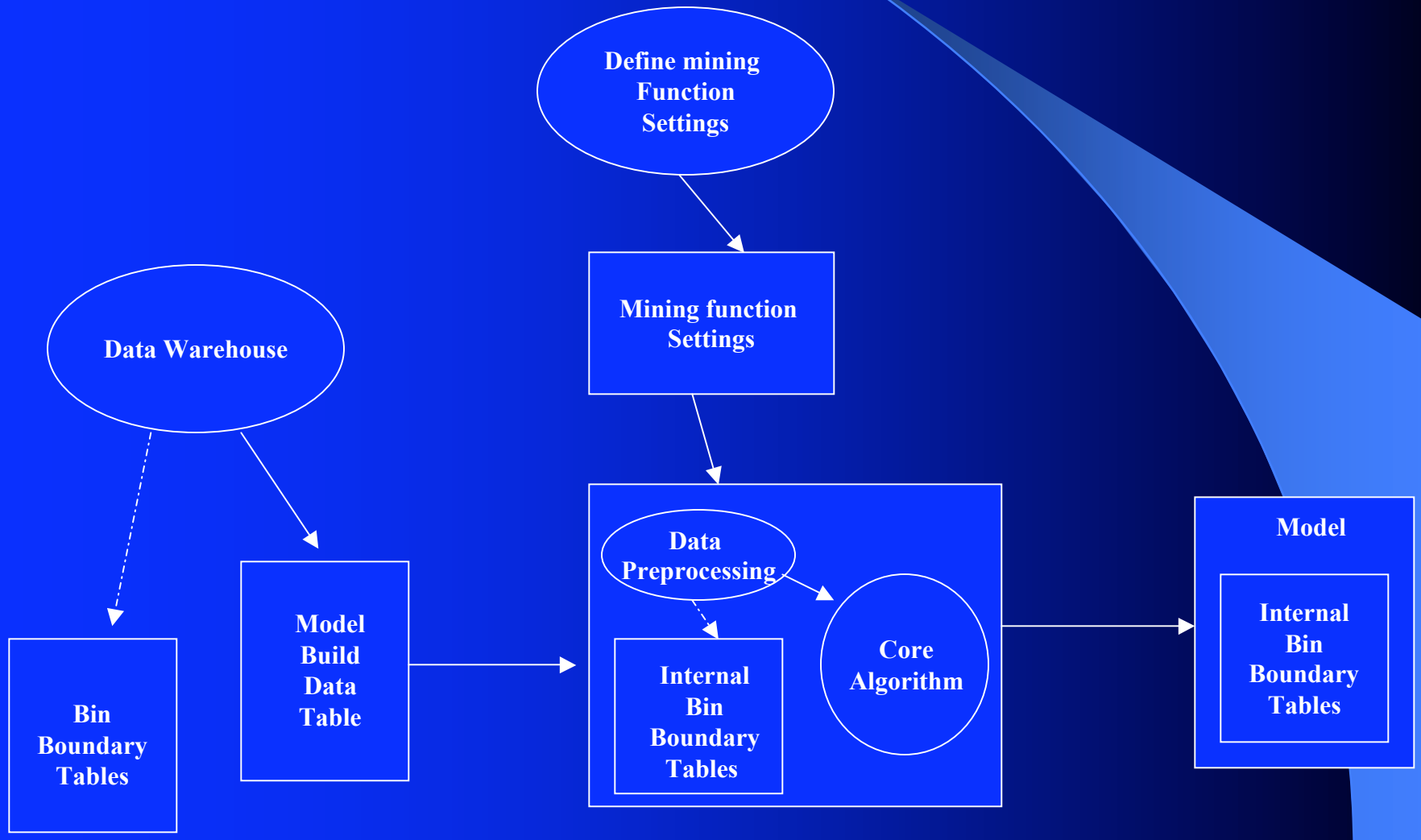
Data Mining Functions

- Classification (supervised)
  - Naïve Bayes algorithm
  - Decision tree algorithm: CART & C5.0
- Clustering (unsupervised)
  - Low inter-cluster similarity
  - High intra-cluster similarity
- Association Rules (unsupervised)
- Attribute Importance (supervised)

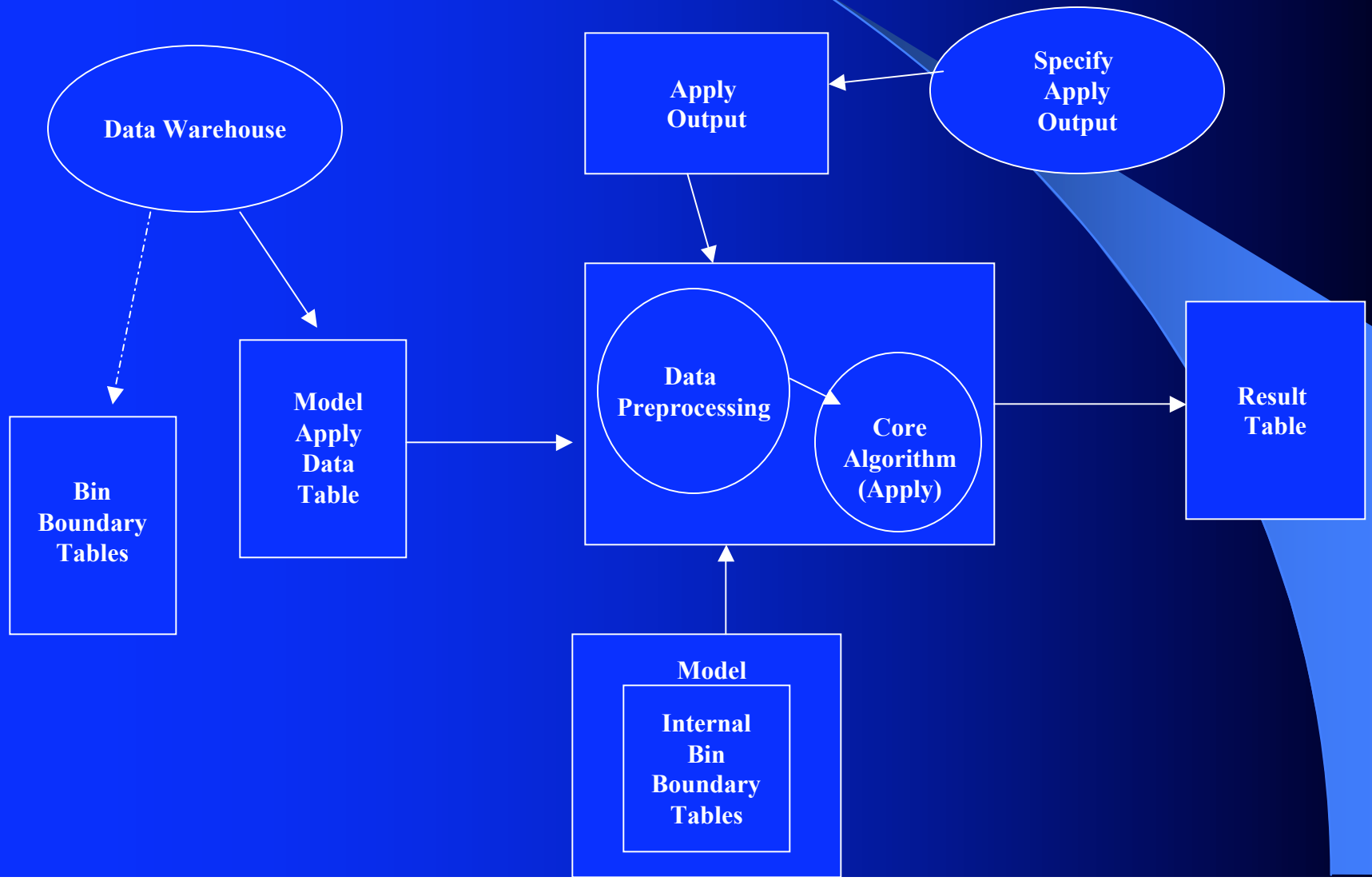
Data Mining Steps

- Build model
  - Build model using training set
- Test model
  - Data has same format as model-build data
- Compute lift (if applicable)
  - Usually for classification
  - To test whether the model is useful
- Apply model
  - Data has same format as model-build data

Model-Build Process



Model-Apply Process



Clustering Algorithms

- Partitioning
  - K-means: each cluster represented by the mean value of the objects in the cluster
  - K-medoids: each cluster represented by one of the objects located near the center of the cluster
- Hierarchical
  - Agglomerative: bottom-up
  - Divisive: top-down

Clustering Algorithms (cont)

- Density-based
  - Continuing growing cluster as long as the density in the neighborhood exceeds a threshold
- Grid-based
  - Quantize the object space into finitely many cells that form a grid structure
  - Fast processing time
- Model-based
  - Statistic approach
  - Neural network approach

Oracle Clustering Algorithms

- Enhanced k-means algorithm
  - Hierarchical k-means algorithm
  - Top-down approach
  - The cluster with largest distortion (sum of distances to the cluster center) is split until desired number of clusters reached
- O-Cluster algorithm
  - Grid-based
  - Hierarchical
  - A unit (cell) is dense if the density exceeds SENSITIVITY

Build Clustering model for Data Warehouse

- Clustering model build steps
  - Data is standardized
  - Connect to the data mining server
  - Create a PhysicalDataSpecification object for model build data
  - Create a MiningFunctionSettings object which specifies the algorithm settings
  - Build the model

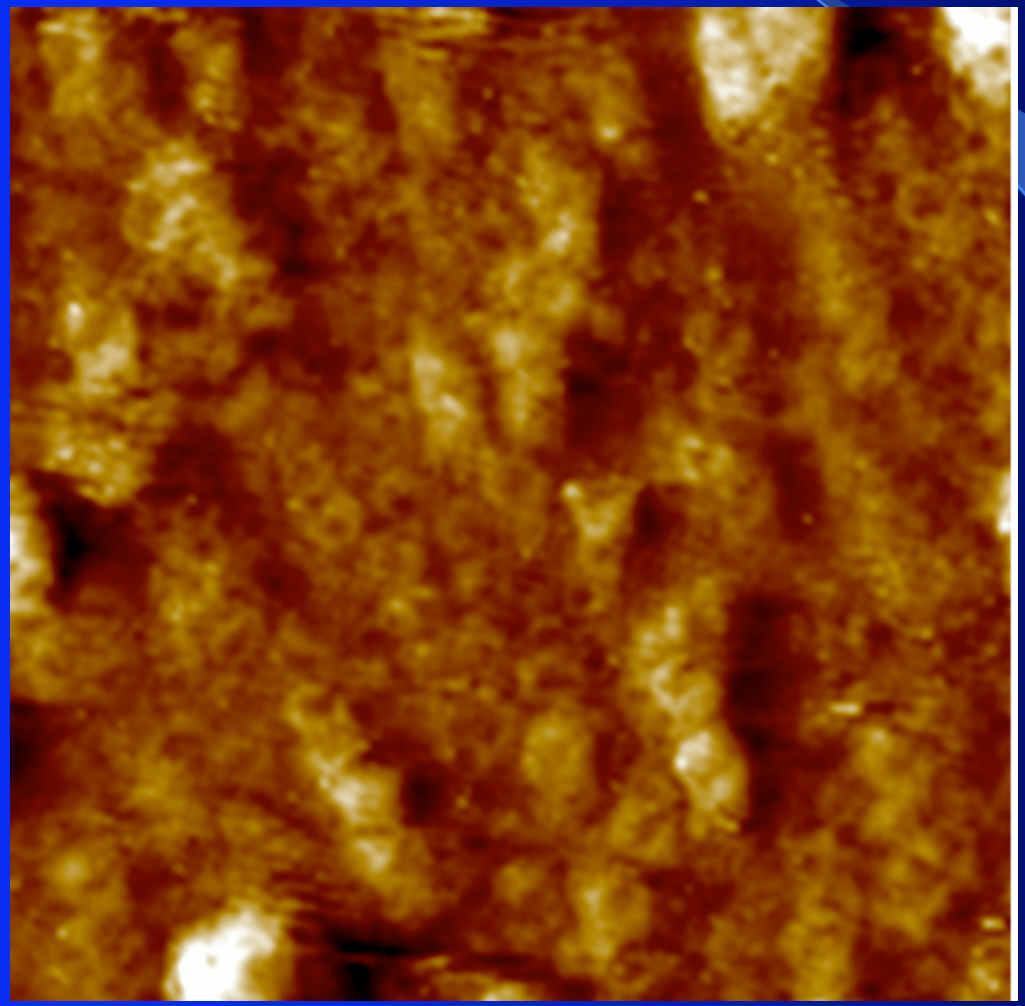
Apply Clustering Model to Data Warehouse

- Programming steps
  - Model apply data is standardized
  - Connect to the data mining server
  - Create a PhysicalDataSpecification object for input data which is the data to be scored
  - Create a LocationAccessData object for output data, which is the table to store the scoring results
  - Create a MiningApplyOutput object for output data, capturing the format of output
  - Score the data

Apply Clustering

- Model-build data format
  - A table POINTS with attributes x & y
    - Points are chosen from the data warehouse
    - Standardized: x & y are in [0,1)
    - 16 million records
- Clusters explanation
  - Dense areas in soil or solution
  - Emerging behavior of random molecules

Aggregation & Micelle Formation



NOM Rings

Maurice, 1999

Comparison of Enhanced k-means and O-Cluster

algorithm	build time (16M rows)	Cluster shape	Clusters
Enhanced k-means	34 min	Spherical	8 (specified)
O-Cluster	14 min	Rectangular	15 (auto)

Summary

- Contributions are
  - New model which treats NOM as a heterogeneous mixture
  - Simulation system with advanced web & database tools
  - System aspects of implementation of load-balancing and fail-over
  - Basic data mining features

Future Work

- Simulation system
  - More features
  - Reliability
  - Efficiency
  - Intelligent simulation configuration wizards
- Simulation data analysis
  - More data mining algorithms
  - Ad hoc queries
- Collaboration tools
  - Oracle Collaboration Suite