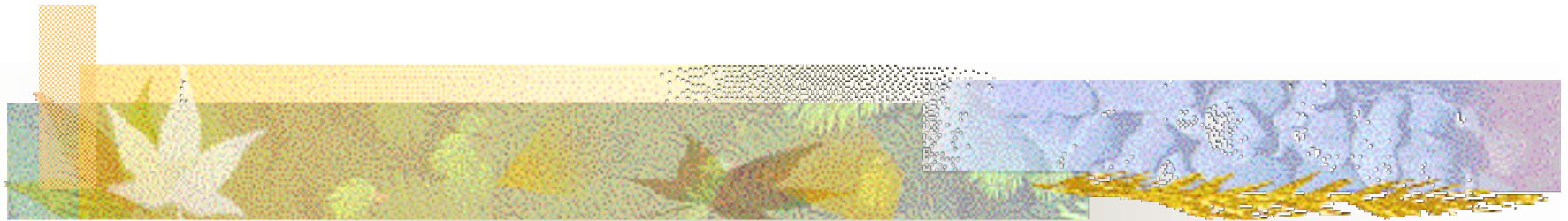


# A Web-based Collaboratory for Supporting Environmental Science Research



Xiaorong Xiang    Yingping Huang    Greg Madey

Department of Computer Science and Engineering  
University of Notre Dame

Steve Cabaniss

Department of Chemistry  
University of New Mexico

WSS'03: WI/IAT 2003 Workshop on Applications, Products of Web-based Support Systems  
October 13, 2003, Halifax

*This research was supported in part by NSF ITR Grant No. 0112820*



# Introduction

- Combination of words “collaboration” and “laboratory” first coined by William Wulf (1996):  
Richard T. Kouzes, James D. Myers, and William A. Wulf. Collaboratories: Doing science on the internet. *IEEE Computer*, 1996
  - Diesel Collaboratory: C. M. Pancerella, L. A. Rahn, and C. L. Yang: The diesel combustion collaboratory: combustion researchers collaborating over the internet. In *Proceedings of the 1999 ACM/IEEE conference on Supercomputing*
  - BioCoRE: <http://ks.uiuc.edu/Research/biocore>
  - EMSL Collaboratory: <http://www.emsl.pnl.gov:2080/docs/collab>
- An example of E-Science
  - G. Fox (2002): E-science meets computational science and information technology. *Computing & Engineering*
  - R. M. Jakobovits, J. F. Brinkley, C. Rosse, and E. Weinberger (1998): Enabling clinicians, researchers, and educators to build custom Web-based biomedical information system



# The NOM Collaboratory

- Interdisciplinary project
  - Supporting research on Natural Organic Matter (NOM)
  - Understanding NOM behavior is an important environmental research area
  - Simulations of NOM in the soil and groundwater
- Objectives
  - Information and model sharing
  - Data repository and analysis support tools
  - Electronic communication tools
- No installation or maintenance of computation resources needed by the environmental scientists



# NOM Collaboratory Components

- NOM simulators
- Search engine
- NOML upload
- Molecule editor
- Molecule validation
- Simulation finder
- Automatic restart
- Completion predictor
- Data Analysis
  - Graphical reports
  - XML reports
  - Data mining
- Communication Tools
  - Discussion board
  - Chat room
  - File sharing











NOM Collaboratory - Mozilla

File Edit View Go Bookmarks Tools Window Help

Back Forward Reload Stop [http://localhost:8888/interface\\_COLLABORATION/Collaboratory.html](http://localhost:8888/interface_COLLABORATION/Collaboratory.html) Search Print

# NOM Collaboratory

	Provide an intelligent interface to facilitate using the NOM simulator	Provide an Interface to define new molecule type	
	Provide an interface to search simulation information	Provide an administration role to validate the newly added molecule for public usage	
	Provide a threaded discussion board	Provide a XML-based Markup Language definition	
	Provide a real time chat box	Provide an Interface to upload publications	

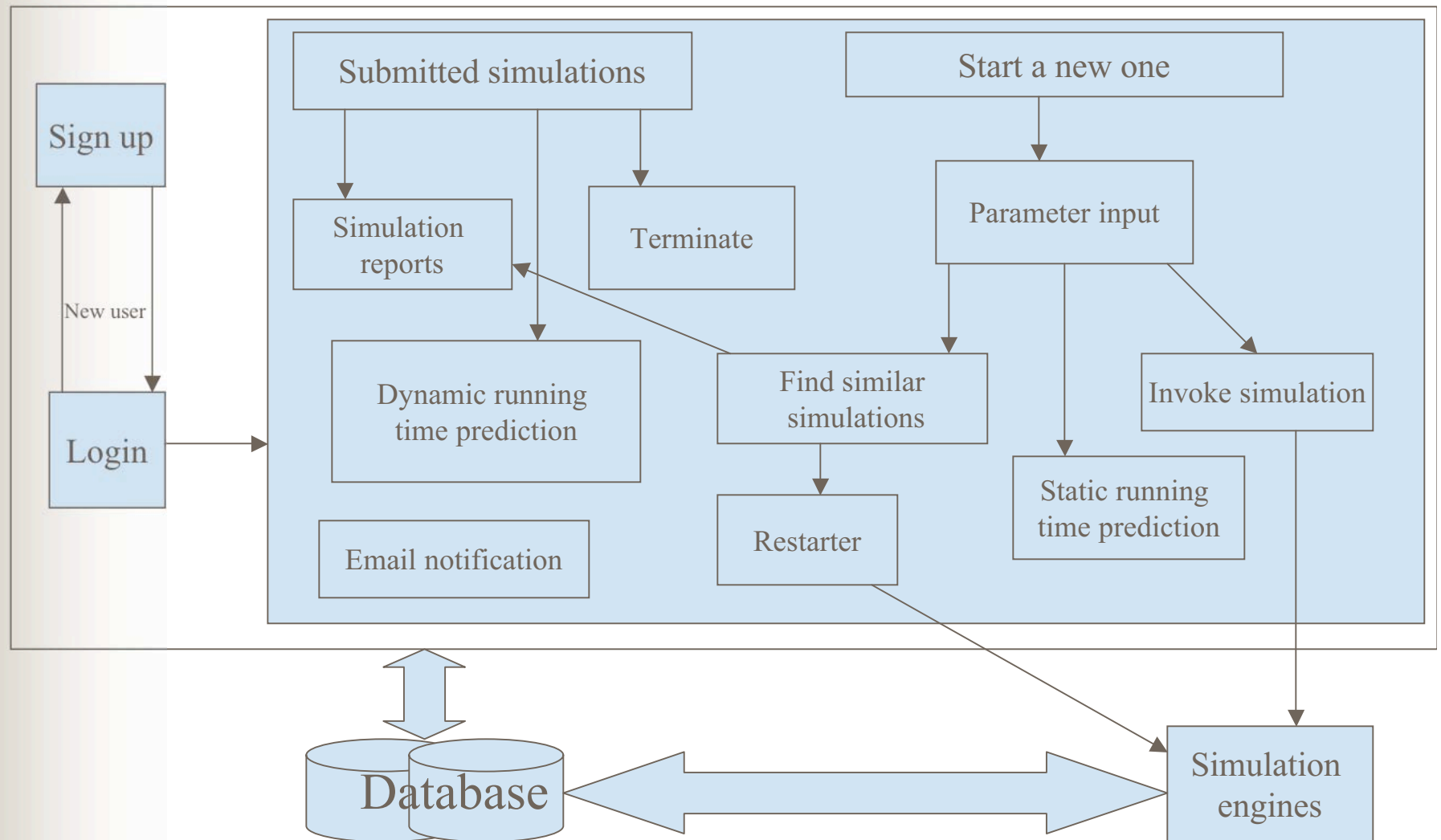
Document: Done (0.154 secs)



# Web Technologies

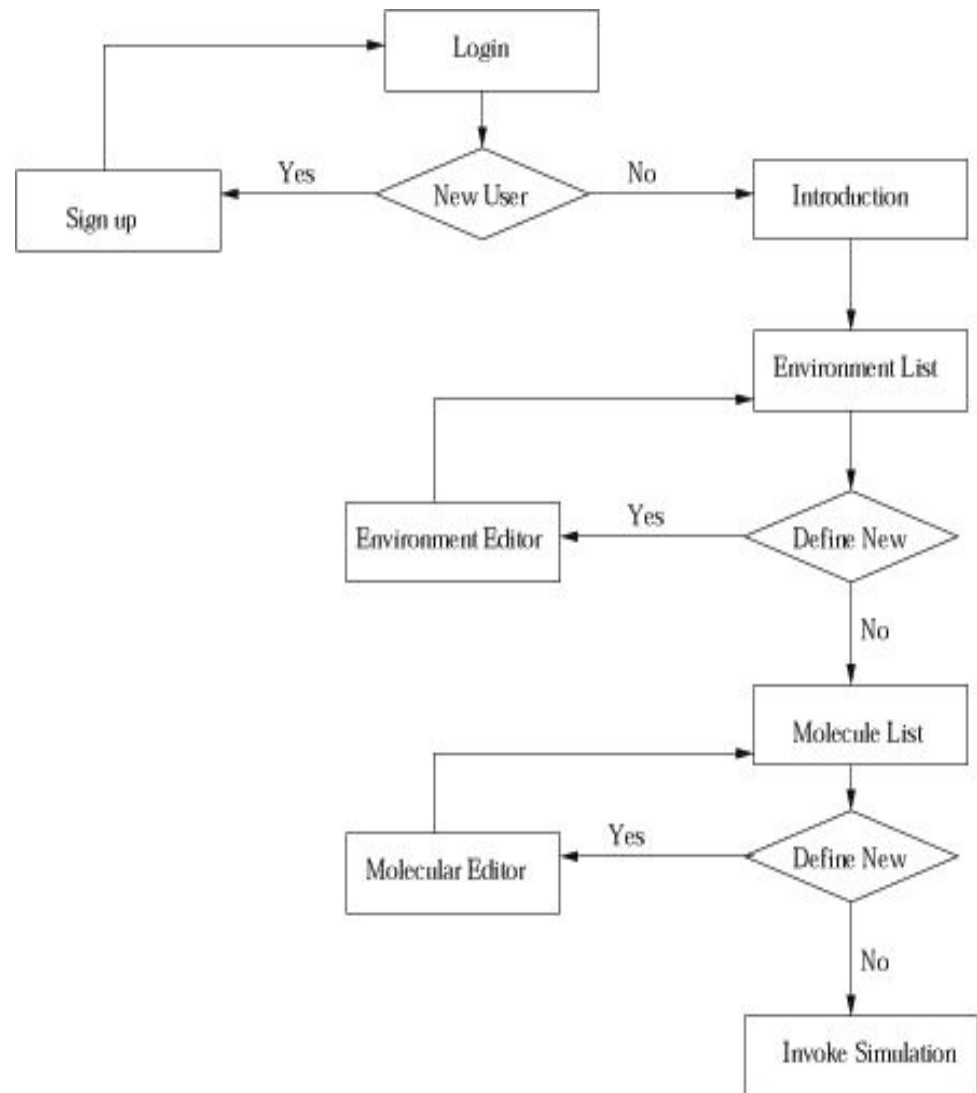
- Sun Java 2 Enterprise Edition (J2EE)
  - Java Servlets, Java Server Pages (JSP)
  - Enterprise Java Beans (EJB)
  - Java Transaction Service/API (JTS/JTA)
- RDBMS (Oracle):
  - Java DataBase Connectivity (JDBC)
  - Data analysis packages
  - Data mining technologies
- XML
  - NOML
  - XSQL

# Web-based Interface

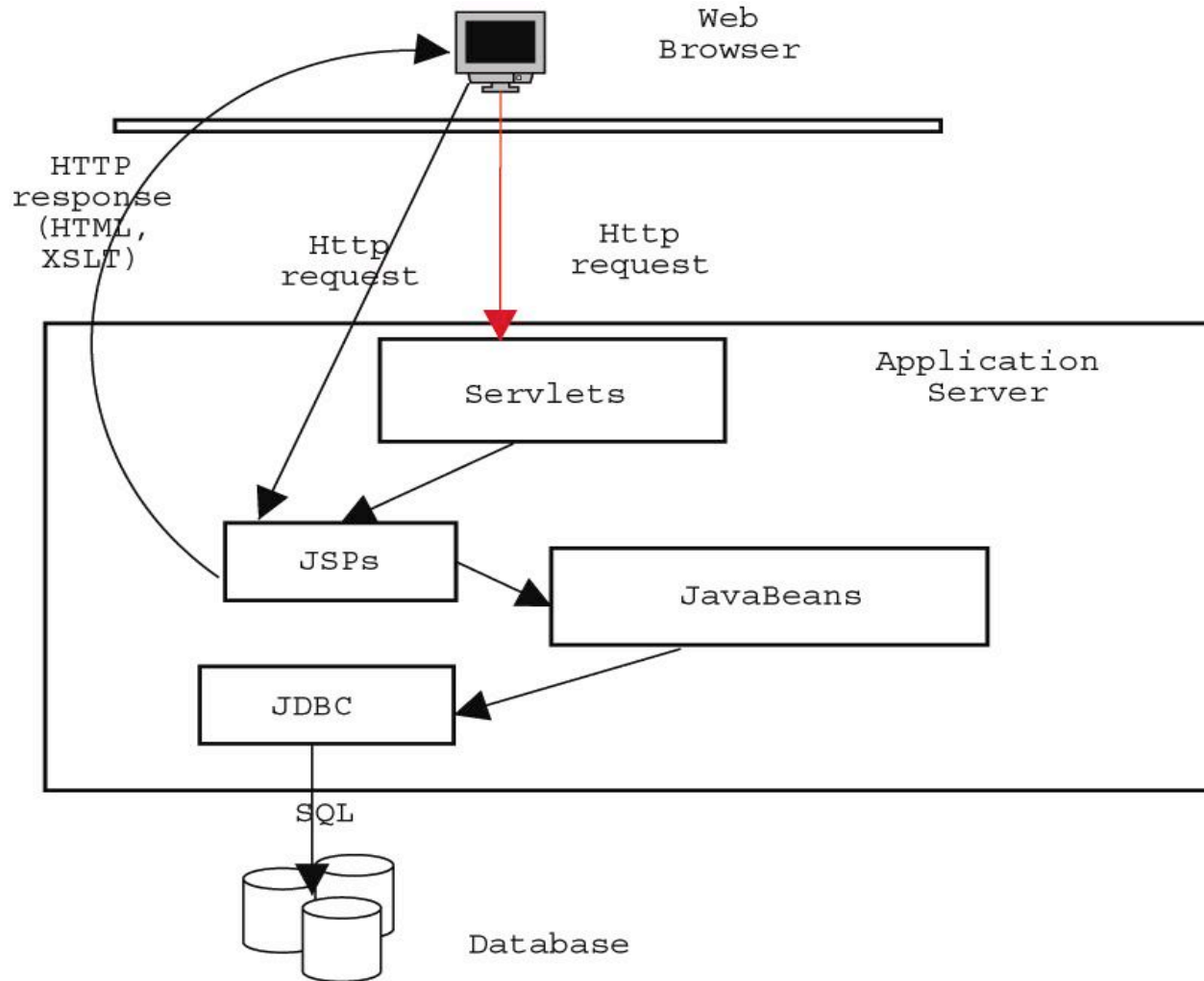


# Web-based Interface Logic

- Input the simulation parameters
- Invoke the simulation
- Stop the simulation
- View the real-time simulation results



# Web Interface Implementation



# Example of Interface



Welcome to NOM Research Group!

Leilani Arthurs

(\* Required fields)

<b>Molecule Name*</b> 3ficFA	<input type="checkbox"/> Make it available to public				
<b>Atoms of the molecule</b> Each field must be a non-negative integer. Default value is 0.	<b>(Atom) C</b> 66	<b>(Atom) H</b> 64	<b>(Atom) N</b> 0		
	<b>(Atom) O</b> 38	<b>(Atom) S</b> 0	<b>(Atom) P</b> 0		
<b>Functional groups of the molecule</b> Each field must be a non-negative integer. Default value is 0.	<b>DoubleBond</b> 0	<b>Rings</b> 0	<b>Phenyl</b> 0	<b>Alcohols</b> 0	<b>Phenols</b> 0
	<b>Ethers</b> 0	<b>Esters</b> 0	<b>Ketones</b> 0	<b>Aldehydes</b> 0	<b>Acids</b> 0
	<b>Arylacids</b> 0	<b>Amines</b> 0	<b>RingN</b> 0	<b>Amides</b> 0	<b>Thioethers</b> 0
	<b>Thiols</b> 0	<b>Phosphoesters</b> 0	<b>HPhosphoesters</b> 0	<b>Phosphates</b> 0	

Create New Molecule Type



# Interface Features

- Email notification
- Running time prediction
  - Static
    - number of molecules
    - number of time steps
  - Dynamic
    - current time step
    - current wall clock time



# Interface Features (cont)

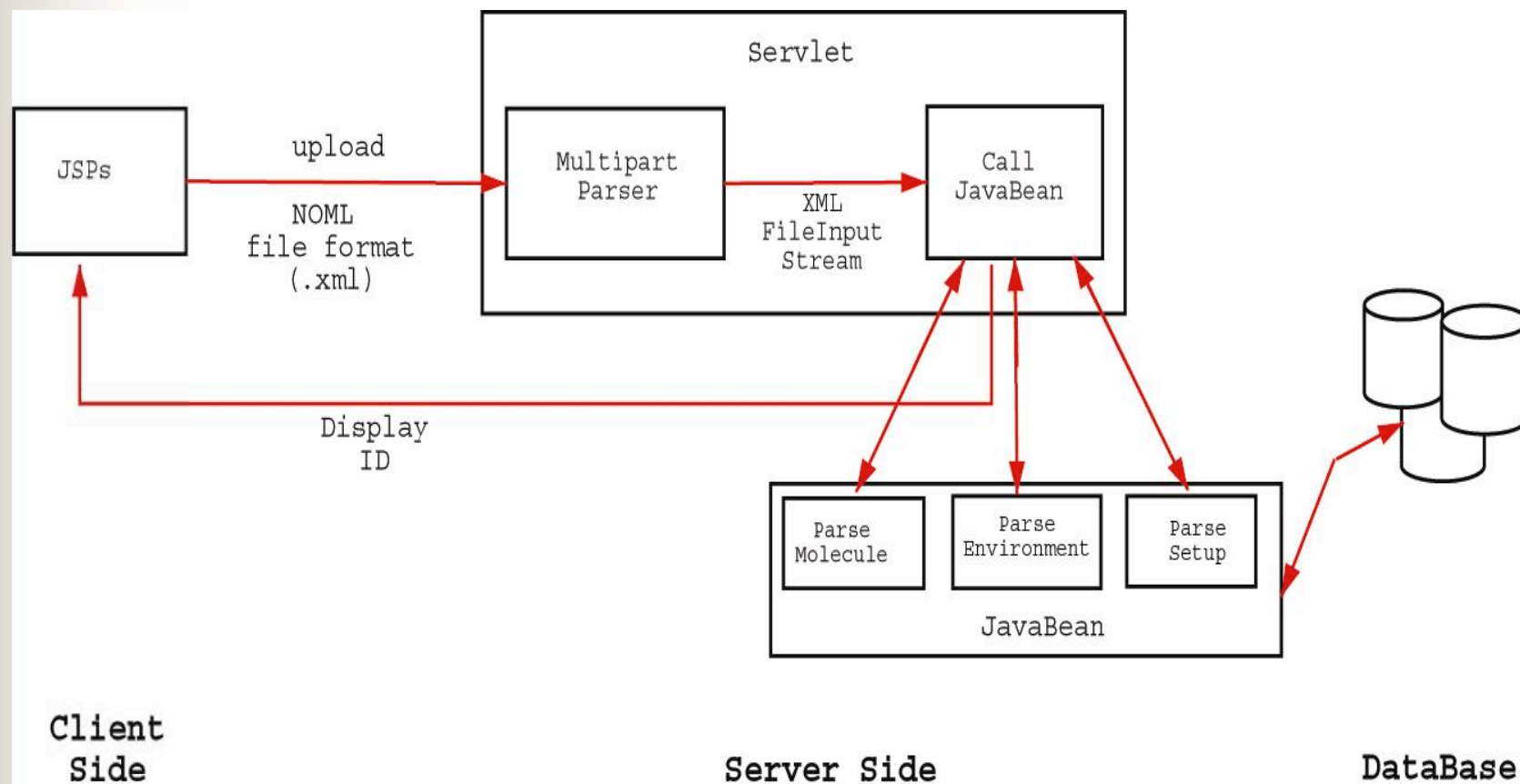
- Find similar simulations
  - Environment parameters
  - Molecule types and distributions
  - Retrieve the data sets from database
  - Points on a high dimension space
  - Euclidean distance
  - Ordered list
  - Review the simulation results or restart
- Automatic restarter
  - Save the state of each objects in the system to database every check point
  - Load the state to the core simulation engine



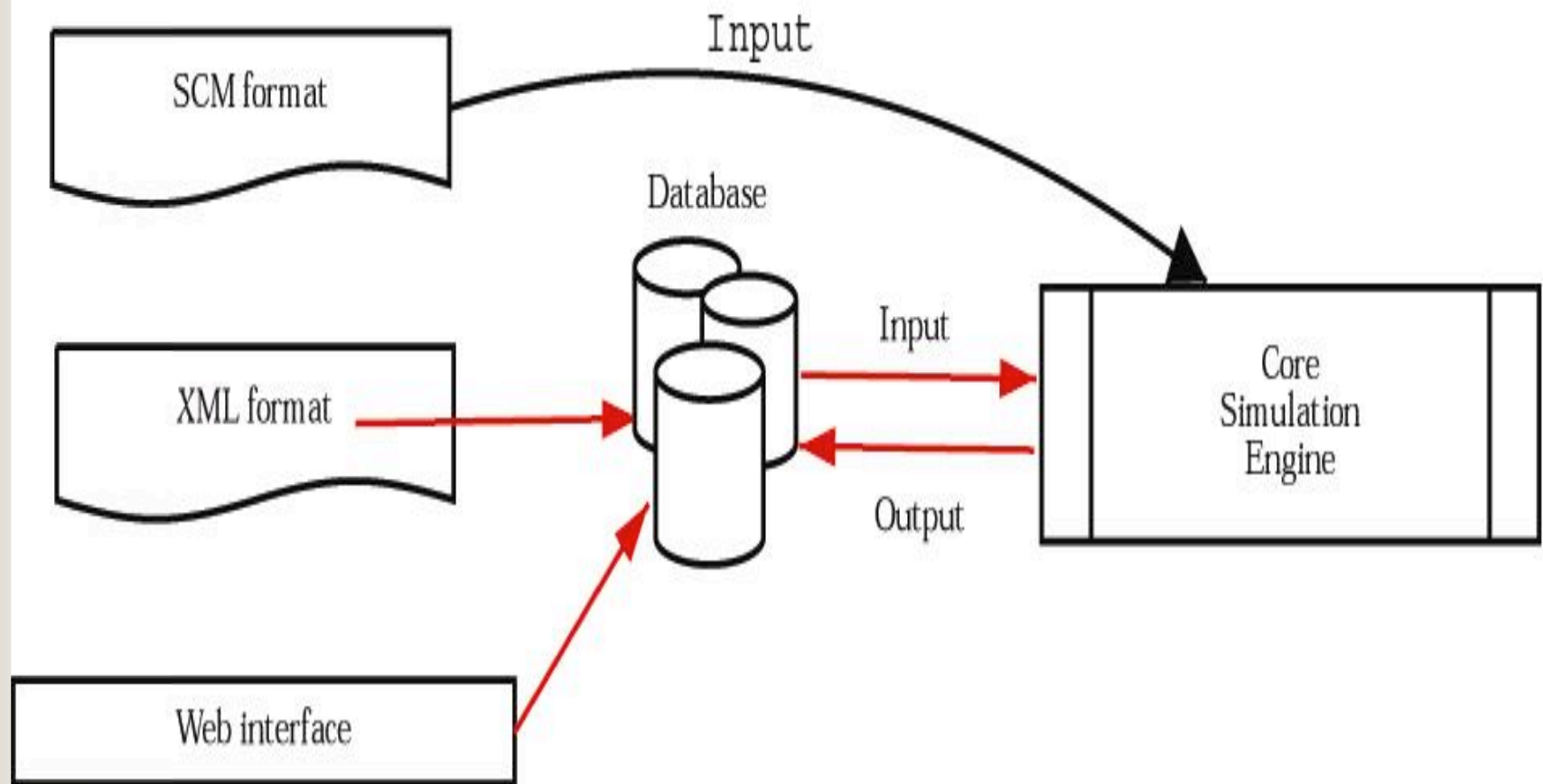
# XML-based NOM Markup Language

- NOML:
  - Standard data format
  - Environment.dtd, Molecules.dtd, Setup.dtd
  - Environment.xml, Molecules.xml, Setup.xml
- Facilitates communication
  - User ===== User
  - Application ===== Application
  - User ===== Application
- Extensions planned

# NOML Uploader



# Data Input Options

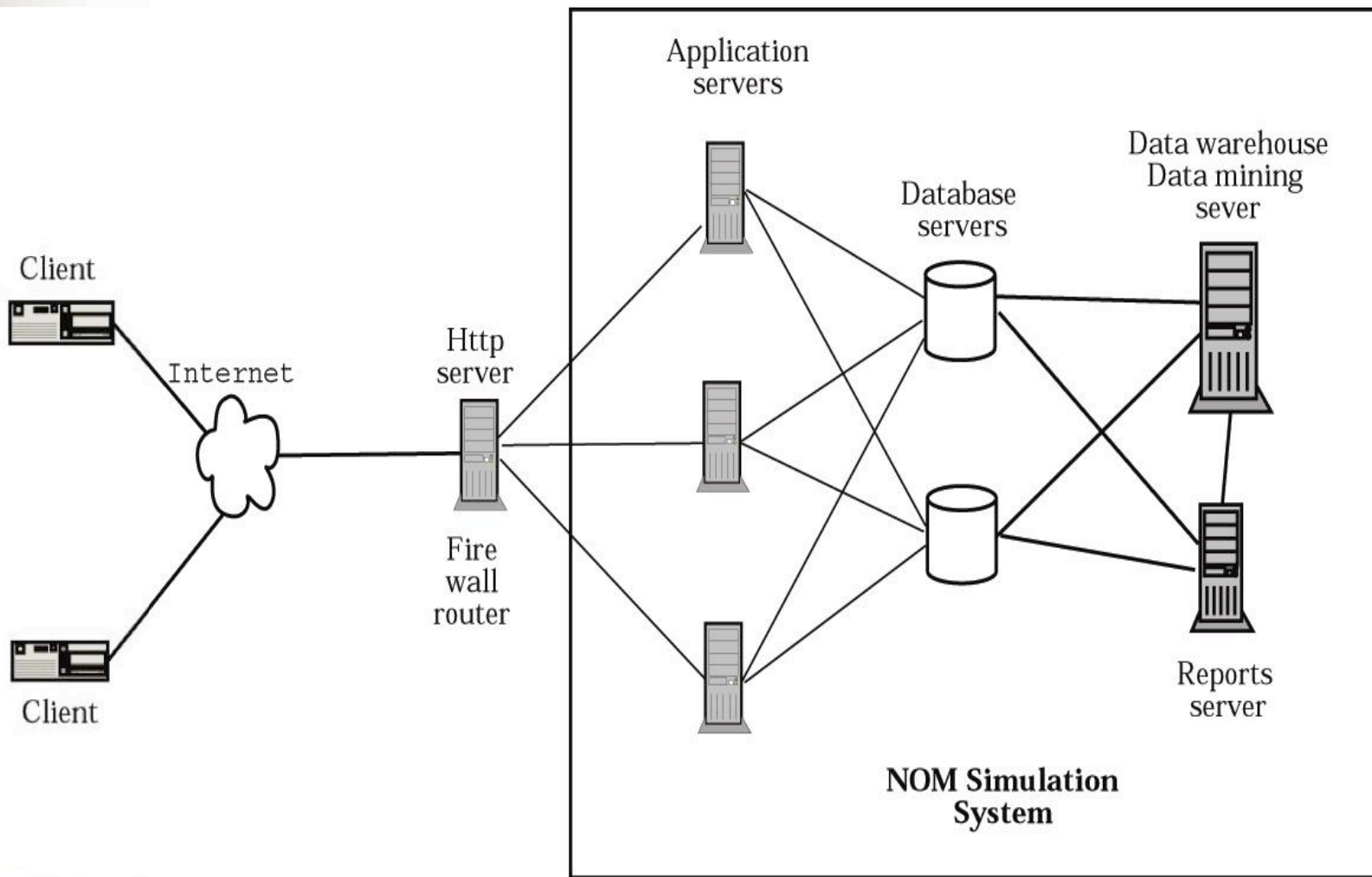




# Other Tools

- Molecule editor
  - Define new molecule type
- Molecule validator
  - Authorized persons (Chemists) to validate data
  - Share the molecule type
- Search engine
  - Ad-hoc query
  - View results of the completed simulations
  - Restart some simulations

# Architecture



# NOM Simulation Engine

- Design

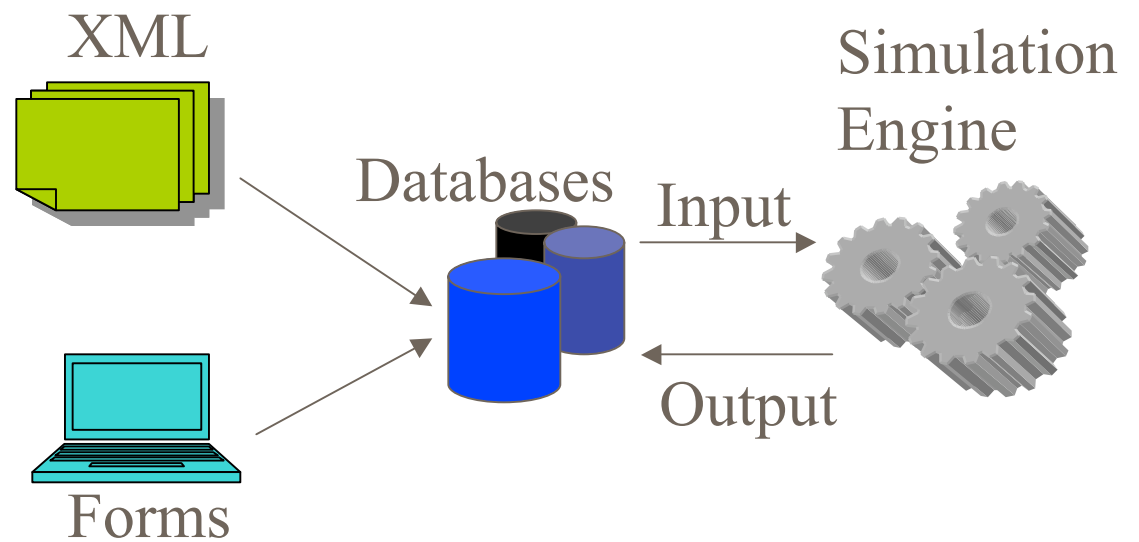
- Agent-based
- Stochastic
- Discrete event

- Packages

- Swarm
- RePast

- Languages

- Java
- SQL & PL/SQL





# NOM Simulation Engine

- Read simulation parameter from the database (JDBC)
  - Environmental parameters (pH, temperature, light intensity, and so on)
  - Molecule types and distributions
- User defined time has been separated to a large number of equal size time steps
- Write relevant data into the database every time step (JDBC)
  - Trace the dynamic properties of individuals and the system over time

# Data Analysis

- Analysis
  - SQL
  - Statistics
  - Data mining
- Presentation
  - Oracle Reports
  - XML/XSLT
  - XSQL

Please provide the following information.

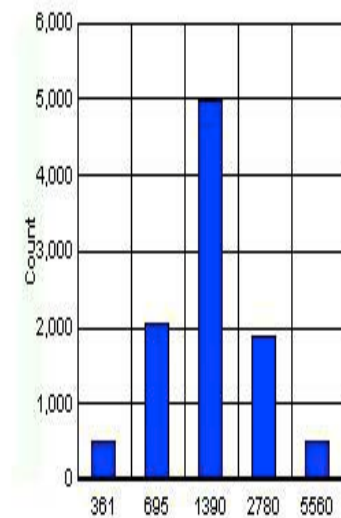
Current Time Step is: 1969 [Refresh](#)

Time Step: 1969  
Status: adsorbed  
Start Y: 0  
End Y: 299

Generate Reports

Generate XML file

**Weight Distribution**



# Report Example

Please provide the following information.

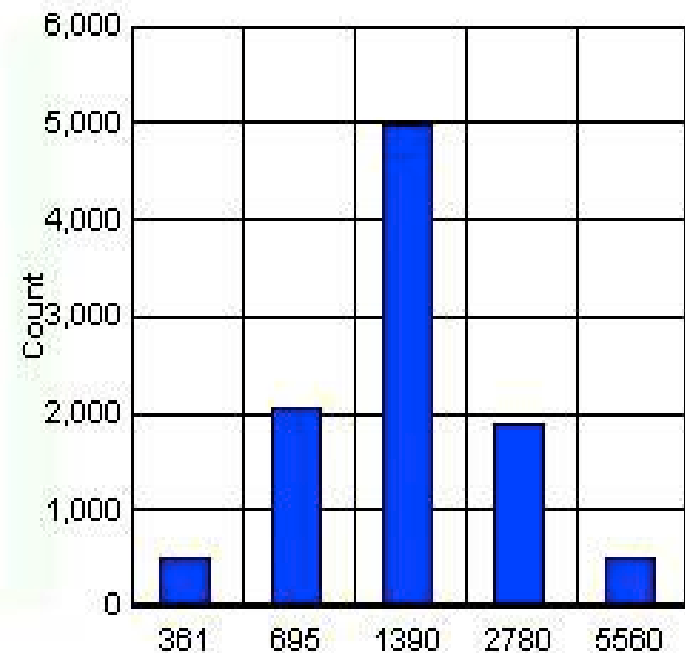
Current Time Step is: 1969 [Refresh](#)

Time Step:   
Status:   
Start Y:   
End Y:

Generate Reports

Generate XML file

## Weight Distribution





# Conclusion

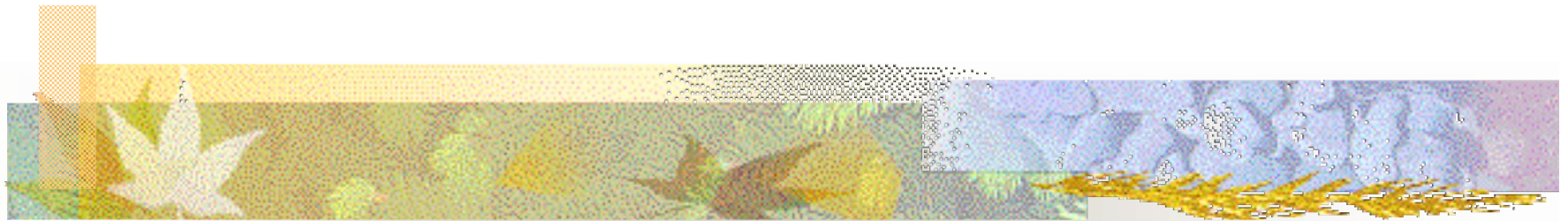
- Web-based Collaboratory
- A Web-based simulation architecture (multiple simulation servers, database servers, and data mining technologies)
- A Web-based configuration interface
- NOML data upload



# Future Work

- Collaboratory
  - More communication tools
    - Audio
    - Video
- More simulation models for NOM study
- NOML extensions
- JDBC performance
  - AutoCommit issues
  - Batch inserts
  - Use of sqlloader
- User testing

**Thank You !**



**Questions?**