GIPSE :

Streamlining the Management of Simulation on the Grid

J. M. Wozniak, A. Striegel, D. Salyers, J. A. Izaguirre

http://gipse.cse.nd.edu

NSF Grant: SCI04-12633

University of Notre Dame
Computer Science and Engineering

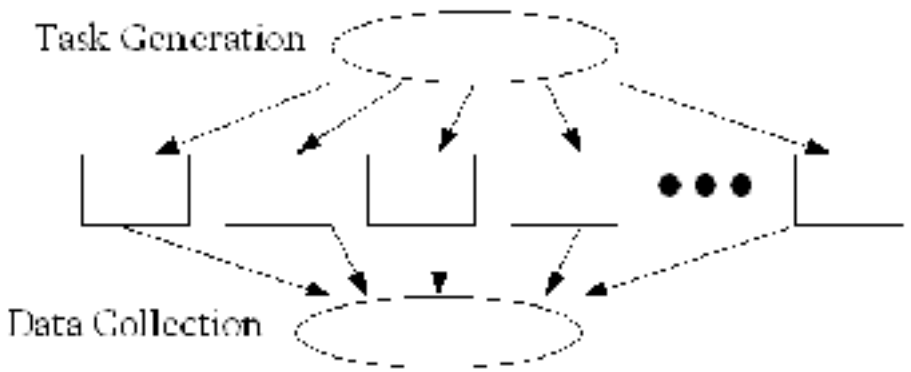
Overview

- GIPSE: Grid Interface for Parameter Sweeps and Exploration

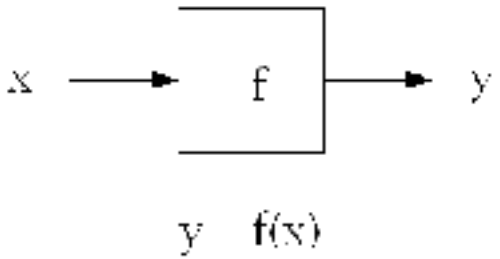
- Problem Description
  - Parameter Sweeps
  - Applications
- Existing Tools
- GIPSE's solution
  - Common Data Format
  - Intelligent Task Submission
  - Data Repository

Parameter Sweeps

- Embarassingly Parallel



- Small input/output

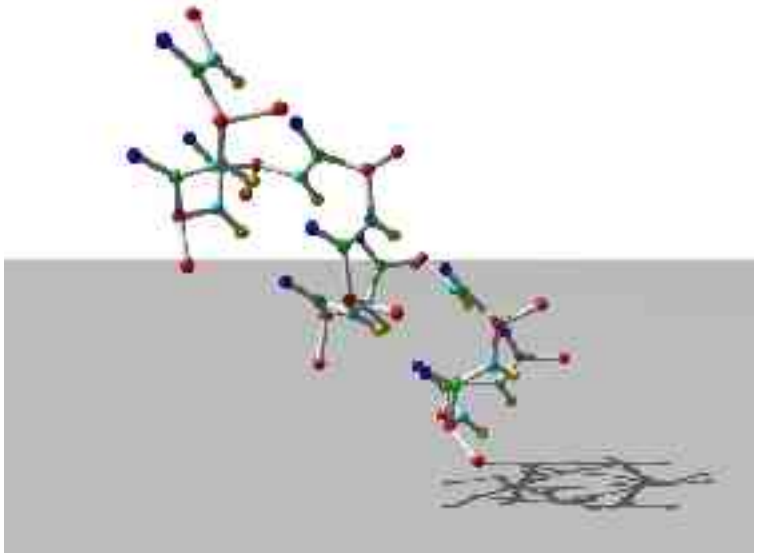


Parameter Sweep Applications

•ProtoMol

Given: a molecular dynamics simulator that accepts 10 parameters.

Plot the running time of the simulation against a tolerance input.

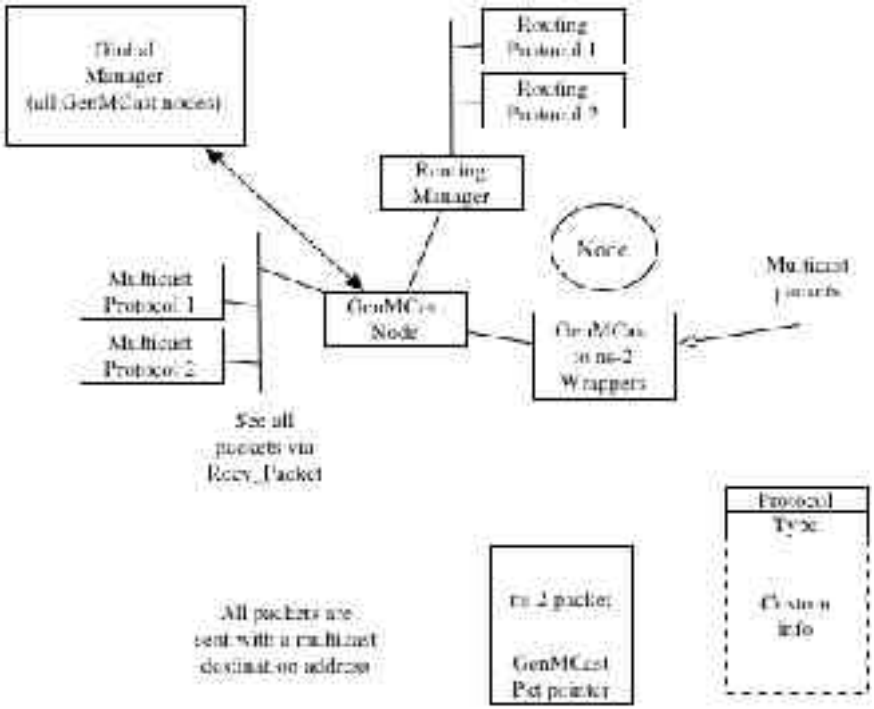


Parameter Sweep Applications

•ns-2

Given: a network simulator that accepts 30 input parameters.

Plot 3 observed throughput results.



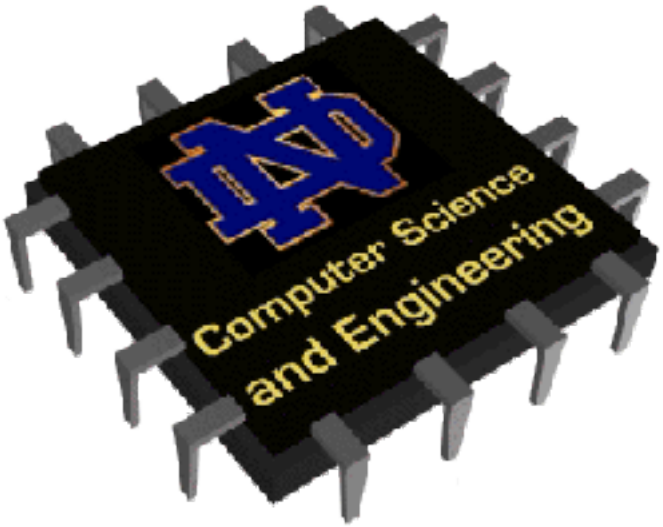
Parameter Sweep Applications

•SimpleScalar

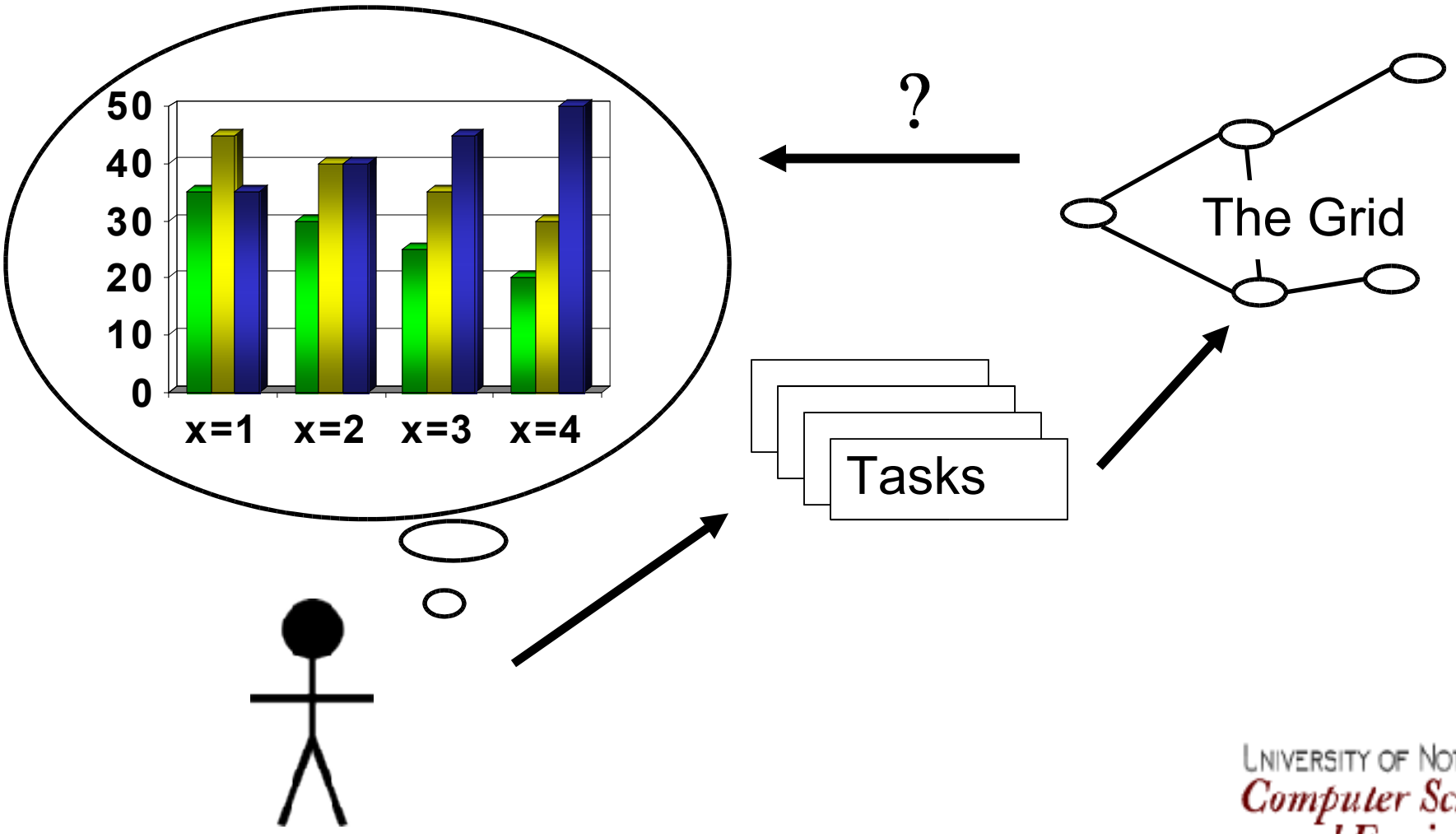
Given: a simulated CPU that accepts 8 design parameters.

Maximize a performance measurement over all parameters,

for a large set of input programs.



Task-oriented Grid Computing



Problems with Scripting

- Focus on tasks, without context
- Limited ability to isolate errors on the Grid
- Storing, indexing, re-using output data
  takes additional scripts
- Limited ability to perform advanced analysis
- Transforming a request for a data set into a script
  is not a trivial operation

Existing Tools

Condor
- Locates free compute resources
- Emphasizes fairness

Globus GRAM
- Grid Resource and Allocation Management
- Provides a virtual front-end for job submission that benefits from the Globus Toolkit

Batch systems
- PBS, LSF, etc.
- The local machine, called the fork resource

Globus GridFTP
- A superset of FTP functionality, with Globus security.
- Optimized for large data sets

Existing Middleware Tools

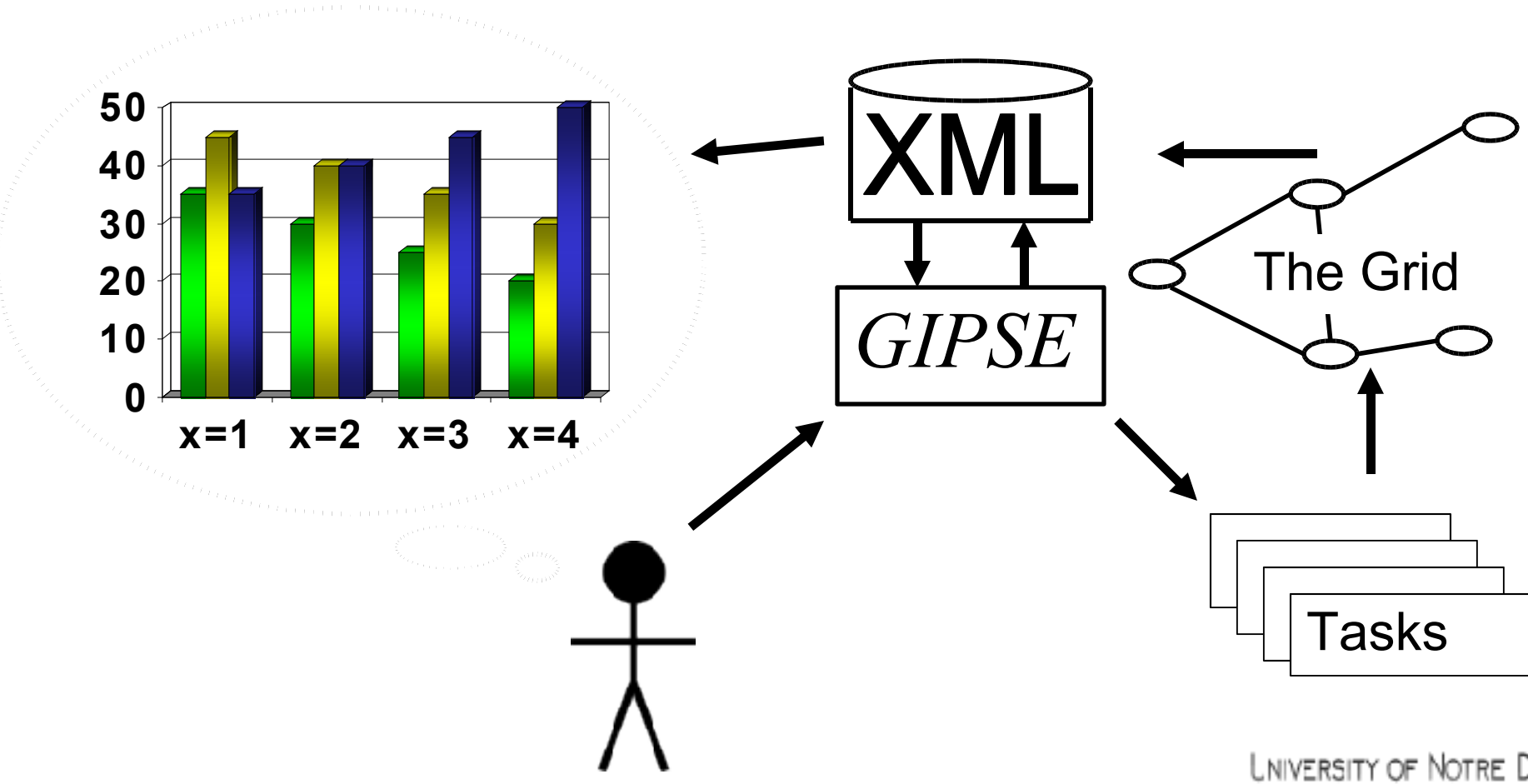
APST
- Allows tasks to be distributed over various resources
- Coordinates data transfer

Nimrod/O
- Distributes tasks over compute nodes
- Helps greatly with the parameter optimization problem

Chimera
- Virtualizes output data
- Allows for complex workflows and data dependencies

GridAnt
- Allows for advanced Grid programming
- Helps greatly with the parameter optimization problem

Shift focus to Visualization



Goals

- Application independent tool
- Make the Grid easier, better
- Allow for some advanced features
- Provide and exploit an abstract storage model
- Emphasize overarching result, generally a visualization

Approach

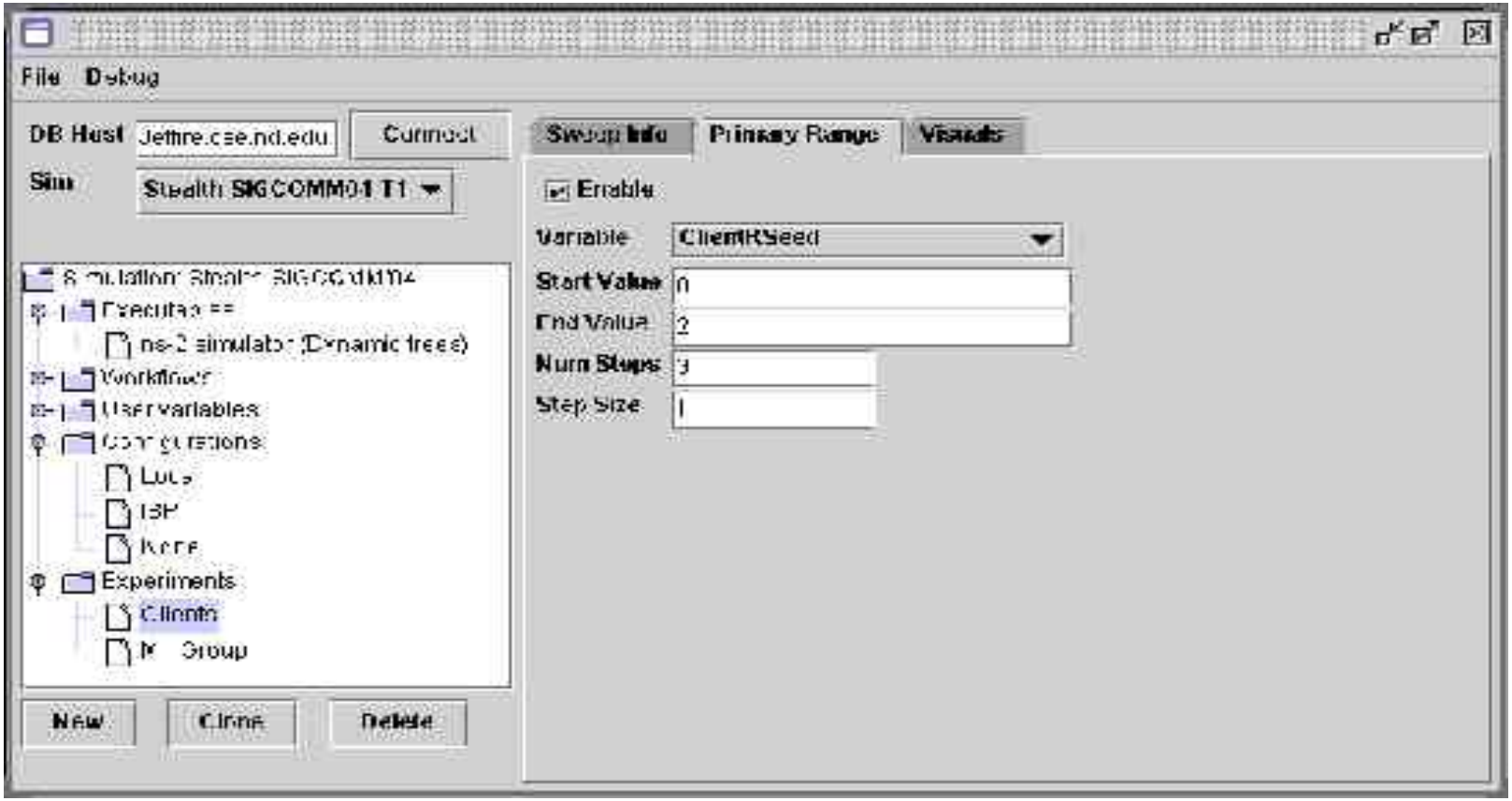
- Common Data Format
  - User libraries or data conversion
  - Uniform namespace across tasks
  - Used for simulation output as well as system information
- Intelligent Task Submission
  - Use functions of output data to generate new input parameters
  - Allows for automatic optimization, exploration
- GIPSE Data Repository
  - SQL-based system
  - Provides data and metadata management capabilities

Common Data Format

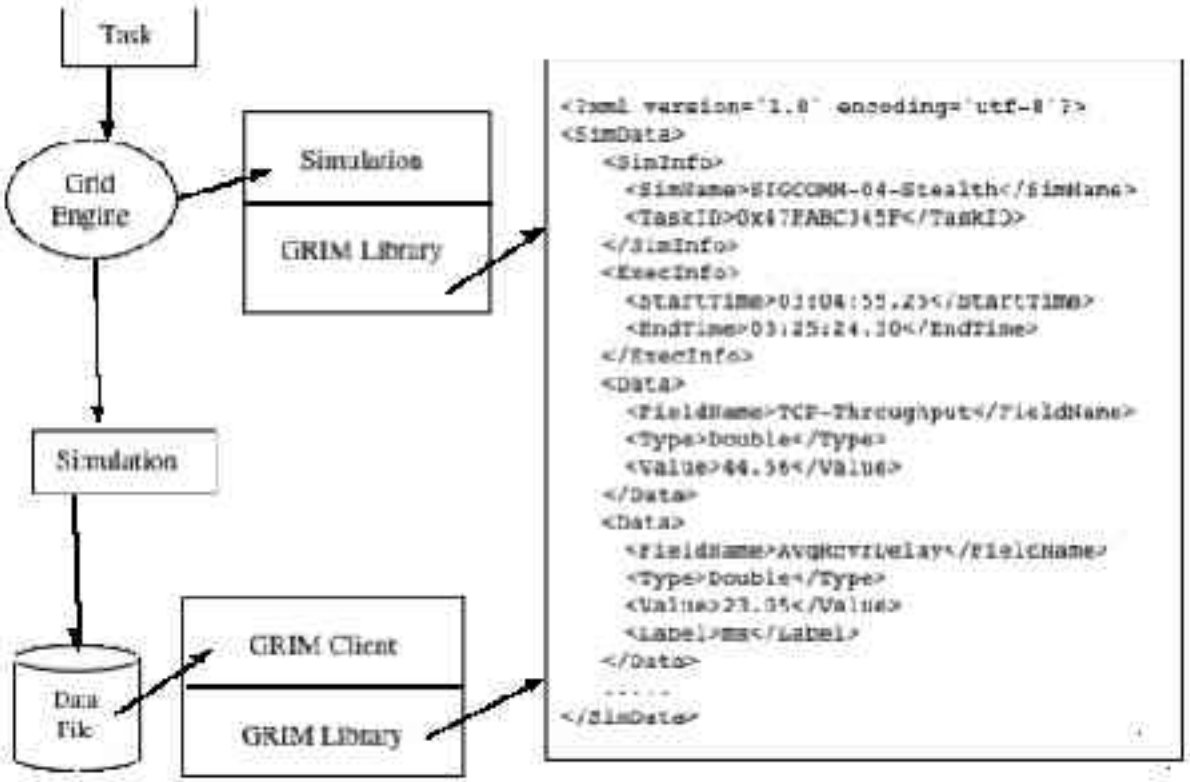
Network simulation example:

```
<gipseData>
    <dataTask name="TestRun">
      <property name = "gipse:SimName" value = "SmallTest-3" />
      <property name = "gipse:TaskID" value = "1" />

      <dataList name="DSMCast-Router">
        <data name = "gipse:time" value = "0.10" />
        <dataPoint name = "Count_Packets" value = "35" />

       ...

      </dataList>
    </dataTask>
</gipseData>
```

Parameter Sweep Example



Data Conversion



Data Control

GIPSE provides data organization and traceability

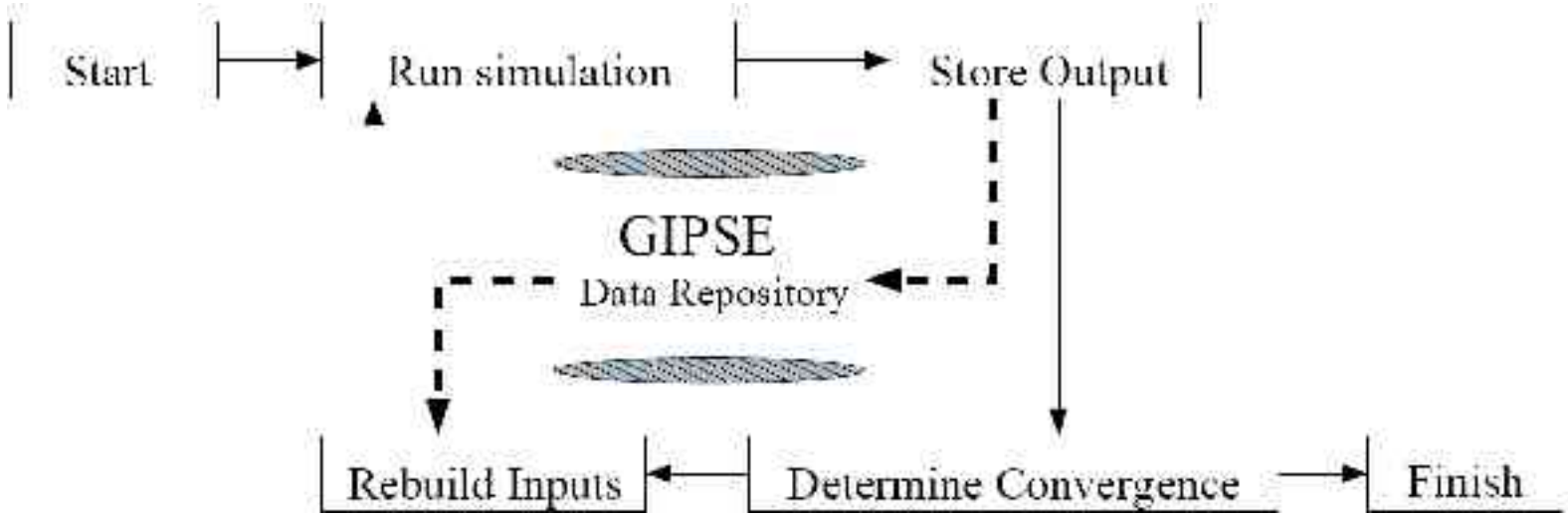
- Data organized in a global namespace
- Allows for the management of incomplete datasets
- Automatic "logging" of system information
- Traceability records how data was produced
  - May help to isolate errors in a dataset
  - Benefits data reuse

Intelligent Task Submission

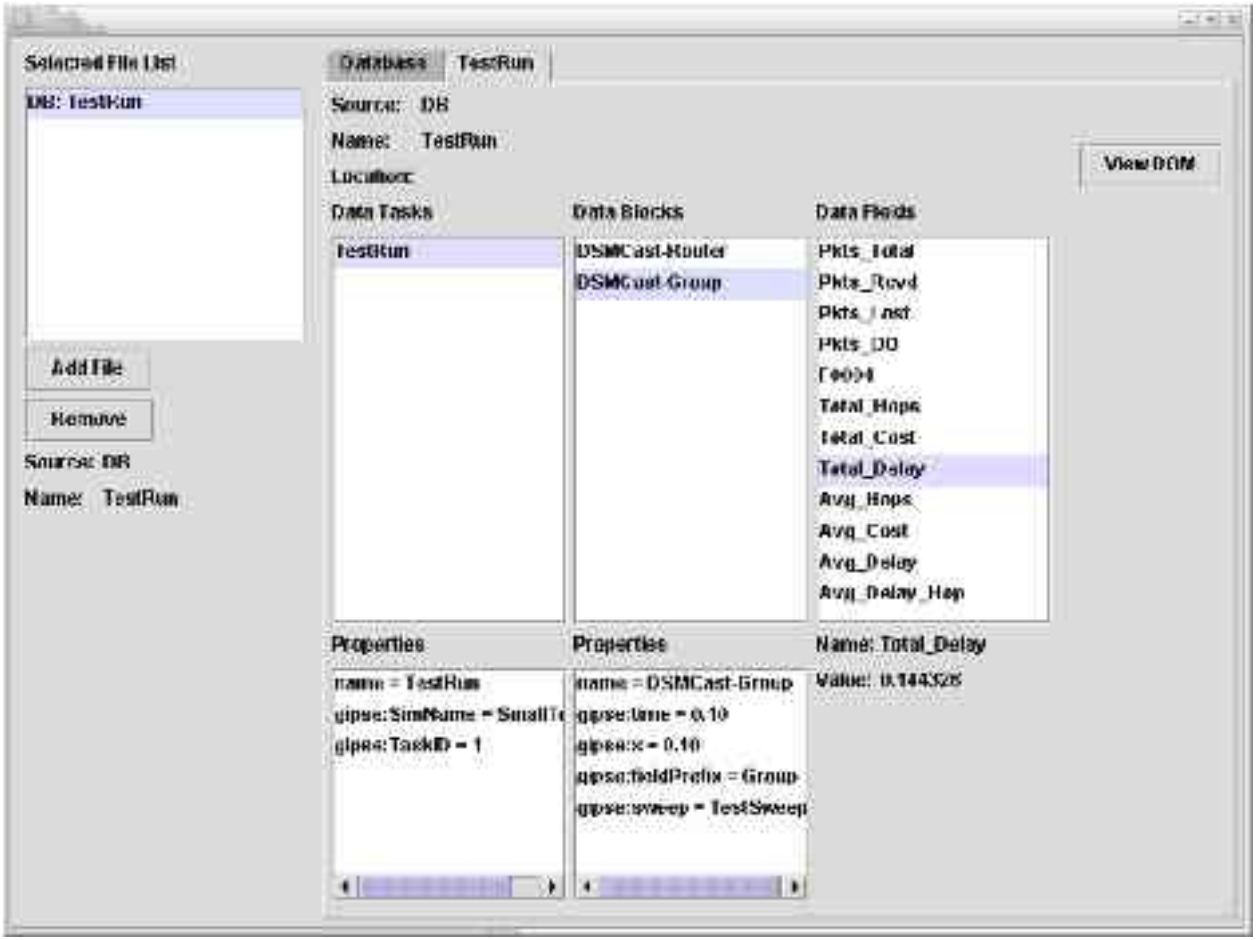
Knowledge of output data allows the system to provide data-driven services, such as:

- Parameter optimization
- Deadline-driven scheduling
- Data set maximization with constraints

GIPSE Data Repository



GIPSE Data Repository



Summary

GIPSE = Grid Interface for Parameter-Driven Simulation and Exploration

Goals

- Free users from extra scripting
- Interact with existing grid tools
- Research-centric, not task-centric
- Manage data flow with XML
- Provide scheduling tools

Prototype Status

- Basics are together
- Downloadable by mid-summer
- Undergoing GUI development (Swing)

`http://gipse.cse.nd.edu`

Future Work

- Develop a portal interface for GIPSE in the OGCE framework
- Application: network simulation
- Finalize interaction with the GIPSE database

`http://gipse.cse.nd.edu`

Discussion

- Questions about GIPSE?
- What are your needs from Grid middleware?
- Suggestions?

http://gipse.cse.nd.edu

jwozniak@nd.edu

University of Notre Dame
Computer Science and Engineering