

Introduction

The effective and efficient management of data and their transformation into information and knowledge is considered a key requirement in modern industries and businesses. Data mining is the technology addressing this information need. Modern scientific and industrial complex-problem solving environments are characterized by:

- ◆ Increasing amounts of distributed digital data and,
- ◆ Rising demands for co-ordinated resource sharing across geographically widely dispersed sites.

Grid computing promises to address the requirements of future distributed data-mining applications.

Grid services & Test bed

Capabilities of the test bed are the following:

- ◆ **Ability to grid-enable a variety of existing programs.** Over 24 programs are currently stored in executable repositories in the test bed. These programs may be combined into complex-workflows.
- ◆ **Meta-scheduling.** Dynamic and automatic allocation of optimal computational servers in the grid environment, is achieved through the use of the Resource Broker, the Information Integrator service and the MDS4 service.
- ◆ **Program and data movement** across different administrative domains is achieved through the use of the GridFTP service.

Partners, Software

The DataMiningGrid software is freely available.

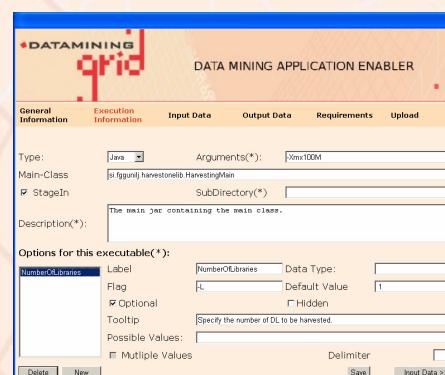


References

Triana, Globus Toolkit, GridBus, DataMiningGrid tools and services. See paper for a complete list of references.

The use of the Application Description Schema

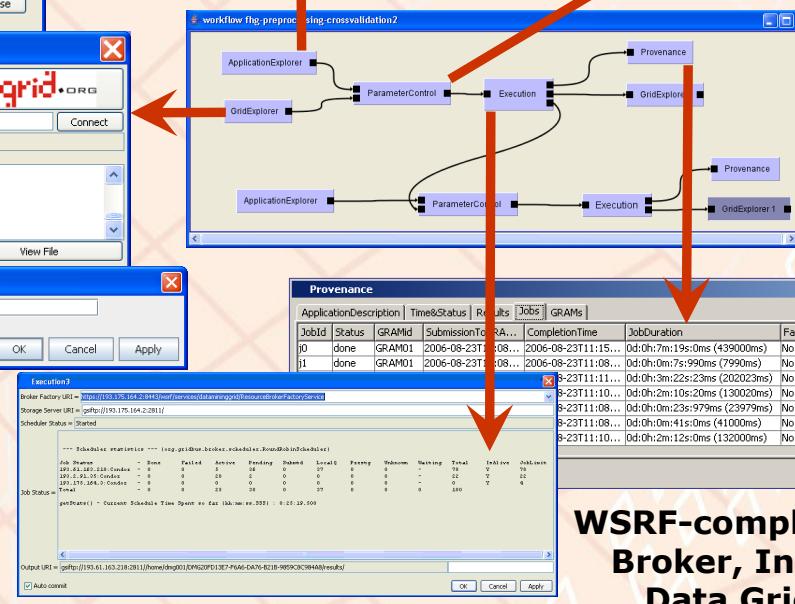
Grid-enabling existing applications



Data grid resources files, databases, directories

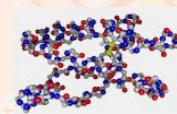


Resource brokering & execution monitoring



A variety of applications

Protein unfolding massive data



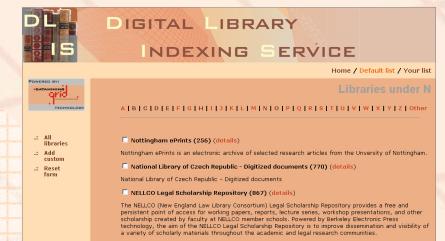
Nutritional habits distributed medical databases



Gene reengineering evolutionary algorithms



Personalized federated digital library text-mining



Grid monitoring log mining



Modeling lake eco-systems equation discovery & civil engineering applications



Customer relationships management text-mining & ontology learning



The DataMiningGrid technology facilitates the development of novel business intelligence applications.

Conclusion

The Application Description Schema provides for conceptual simplicity, interoperability, flexibility and extensibility of the DataMiningGrid system.

Consequently, a variety of grid resources have been grid-enabled and combined into complex grid applications. Two user interfaces are possible: (1) a workflow editor for more demanding users and (2) specialized grid portals.

The software represents a serious attempt at developing a conceptual and technical framework for data mining in grid computing environments.