CSCI B669 Scientific Data Management and Preservation, Spring 2011

Instructor: Beth Plale, SoIC IUB http://www.cs.indiana.edu/~plale
Times/Venue: Wed 5:30 - 8:00 p.m.  Informatics East 122

Readings:

The Fourth Paradigm (http://research.microsoft.com/en-us/collaboration/fourthparadigm/) as well as journal and conference papers.

Course Description:

As supercomputers and modern scientific instruments allow scientists to generate data at vast rates on everything from the human genome to the origin of distant planets and the changing climate of our own planet, we find ourselves awash in digital data - a problem often referred to as the data deluge. Data produced by these technologies are precious and irreplaceable, holding the potential for greater scientific knowledge and understanding in perpetuity. In this seminar course we will explore multiple dimensions of scientific data management including at-scale and issues of preservation of.

Topics Covered:

- provenance and metadata collection
- linked data and semantics of; ontologies for scientific workflow systems
- large-scale file systems
- data storage systems at scale
- performance monitoring and the implications of alternate compute architectures such as GPUs on data centric-computing.
- archiving scientific data; OAIS model for principles of data preservation
- data-centric computing in domains within earth and environment, and health and well-being.

The course utilizes lectures, presentations, and discussions. If student interest and background merits, students will get some hands-on experience with research tools, web services, and cloud computing around a class project. See pti.iu.edu/d2i for kinds of research tools to be explored.

Prerequisite:

Moderate level of mastery with programming in traditional programming language such as Java or C++.