

# Principal Associate Directorate for Science, Technology, and Engineering (PADSTE)

## Theory, Simulation & Computation (ADTSC) Theoretical (T) Division

### A partial listing of ISTI Interactions

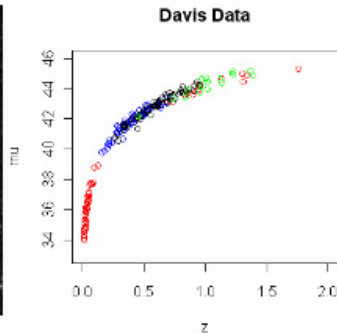
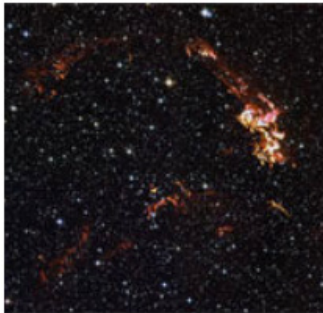
#### IS&T DATA EXPLORATORY - DATA INTENSIVE SUPERCOMPUTING AT LANL

The IS&T Data Exploratory is a unique data intensive supercomputing facility designed to explore the challenges of working with large datasets. Sensors, Internet packet filters, telescopes and satellites all generate massive amounts of data. To solve problems in areas like energy security, bio-security, and cosmology we need effective ways to analyze this data. The Data Exploratory will provide a trial facility in an open environment where researchers can experiment with data intensive methods.

One of the challenges of large datasets is representing the data in a manageable format that scientists can learn from. The IS&T Data Exploratory will include a visualization environment, the DISC Visualization Collaboratory, to provide an alternative method for looking at results from data intensive applications. The goal is to find new ways of using visualization for information applications, to provide an abstraction of the data rather than a simulation of it.

#### ISTI COLLABORATIVE RESEARCH PROGRAM

##### CRP Project (LANL/UCSC): Cosmic Calibration - Statistical Modeling of Dark Energy



We propose to work on the problem of calibrating the parameters of computer code used for simulation of physical phenomena. We explore statistical methods based on a Bayesian approach implemented with Sampling Importance Resampling (SIR).

**LANL (T) Collaborators/mentors:** Salman Habib, T-2

##### Related Papers/presentations

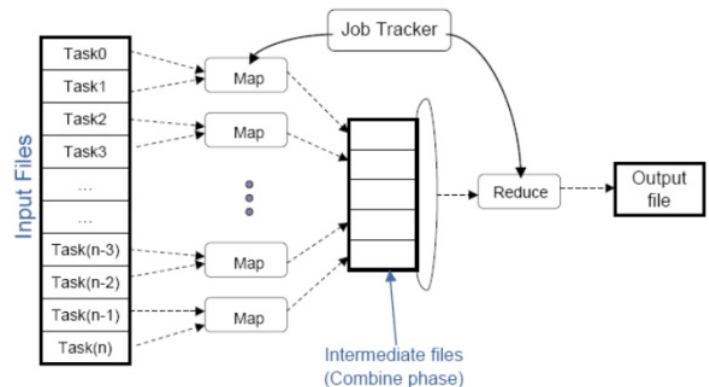
- ◇ Paper: Simulations and Cosmological Inference: A Statistical Model for Power Spectra means and Covariances, January 2010
- ◇ Paper: The Case for Deep, Wide-Field Cosmology
- ◇ Presentation: Cosmic Calibration - Statistical Modeling for Dark Energy (for UCSC Engr Days), May 2008
- ◇ Presentation: Discriminating Theoretical Models of Dark Energy, August 2009
- ◇ Presentation: ISSDM Day: Comic Calibration - Statistical Modeling for Dark Energy, October 2009
- ◇ Presentation: Cosmic Calibration - Statistical Modeling for Dark Energy, April 2010

##### Related Proposals

- ◇ Proposal: NSF: III Medium Collaborative Research Statistical and Feature Comparisons for Large Scale Simulations, February 2009
- ◇ Proposal: NSF Cosmic Calibration: Meeting the Precision Cosmology Challenge, May 2009
- ◇ Proposal: LDRD ER: Venturing Beyond the Cosmological Constant, April 2009
- ◇ Proposal: LDRD DR: Cosmological Signatures of Physics beyond the Standard Model: Petascale Cosmology Meets the Great Surveys, March 2009
- ◇ Proposal: NSF Understanding Uncertainties in Petascale Applications, October 2008
- ◇ Proposal: NSF Probabilistic features in petascale data with uncertainty, April 2009

- ◇ Proposal: LDRD ER: Venturing Beyond the Cosmological Constant, April 2009
- ◇ Proposal: NSF: Understanding Uncertainties in Petascale Applications, October 2008
- ◇ Proposal: DOE ASCR: Probabilistic Features in Petascale Data with Uncertainty, April 2009

##### CRP Project (LANL/CMU): Data Intensive Super Computing for Science



This project explores how well suited data intensive computing programming/run time paradigms like map reduce and other graphs apply to scientific applications.

**LANL (T) Collaborators/mentors:** Salman Habib, T-2

##### Related Papers/presentations

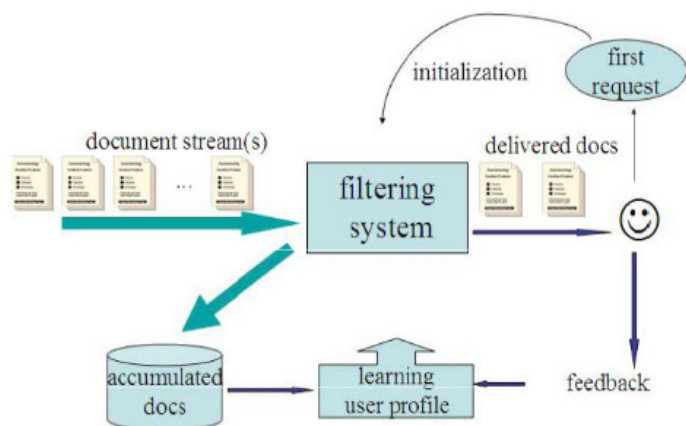
- ◇ Paper: Introducing Map-Reduce to High End Computing PDSW08, November 2008
- ◇ Paper: DiskReduce: RAID for Data-Intensive Scalable Computing, November 2009
- ◇ Paper: In Search of an API for Scalable File Systems: Under the Table or Above It?, June 2009

##### CRP Project (LANL/MIT): Energy Infrastructure: Smart Grids

- ◇ Grid Design - develop efficient and robust integration of geographically distributed renewable generation by solving cutting-edge optimization problems that incorporate difficult constraints imposed by the transmission of electrical power.
- ◇ Grid Control - develop distributed control techniques and algorithms based on queuing theory to enable high penetration of small-scale distributed generation and plug-in hybrid vehicles.
- ◇ Grid Stability - develop a toolbox capable of detecting instabilities and failures for preventing costly outages.

**LANL (T) Collaborators/mentors:** Michael Chertkov, T-4

**CRP Project (LANL/UCSC): Filtering Internet Information for Use in Bio-threat Scenarios**



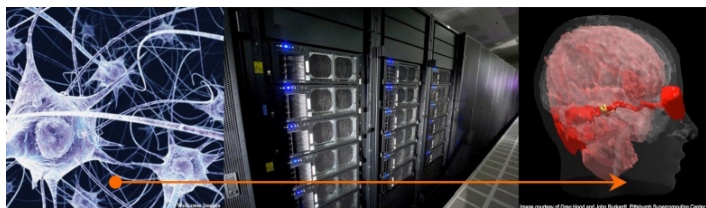
We propose to apply information technology developed for personalized recommendation systems and collaborative filtering (e.g. amazon.com's "Customers who bought this also bought" or Netflix's "You might also like") to ill-defined scientific issues where choosing the most relevant information from a large amount of slightly-relevant information becomes literally of vital importance. The scenario is an outbreak of an infectious pathogen, where the nature of the pathogen is known through previous analysis. The question then becomes: what to do next? Quarantine, triage, treat, limit further exposure, locate drug and vaccine supplies, determine their likely efficacy and the most efficient way to use them as well as the time they will take to arrive, find subject experts, identify possible origins, and countermeasures, etc. Much information that can be highly relevant in this scenario exists on the internet, but locating it and assessing its reliability and relevance is extremely difficult. We believe that personalized recommendation systems can play a role in focusing the searches and finding the most important and urgently needed information.

**LANL (T) Collaborators/mentors:** Carla Kuiken, T-6

Related Papers/presentations

- ◇ Presentation: ISSDM Day: Adaptive Information Filtering and its Application in Medical Domain, October 2009
- ◇ Presentation: SIGIR' 10: Interactive Retrieval Based on Faceted Feedback, July 2010
- ◇ Presentation: 2010 ISSDM Day: Adaptive Information Filtering, October 2010

**CRP Project (LANL/MIT): Synthetic Cognition through Petascale Models of the Primate Visual Cortex**

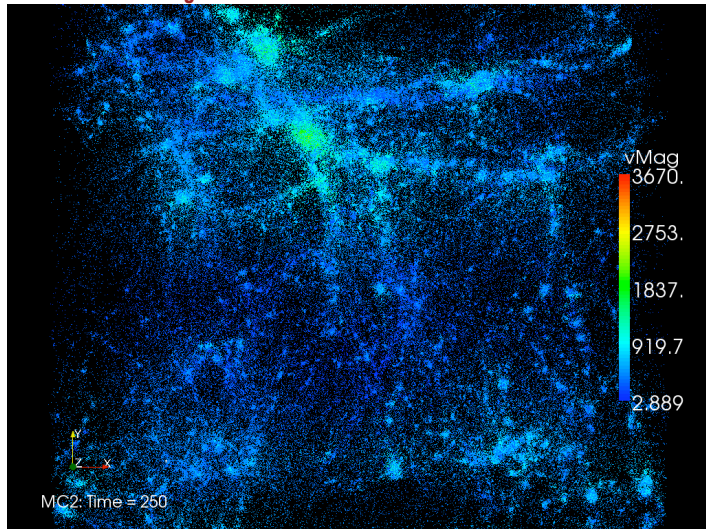


We propose to apply information technology developed for personalized recommendation systems and collaborative filtering (e.g. amazon.com's "Customers who bought this also bought" or Netflix's "You might also like") to ill-defined scientific issues where choosing the most relevant information from a large amount of slightly-relevant information becomes literally of vital importance. The scenario is an outbreak of an infectious pathogen, where the nature of the pathogen is known through previous analysis. The question then becomes: what to do next?

Quarantine, triage, treat, limit further exposure, locate drug and vaccine supplies, determine their likely efficacy and the most efficient way to use them as well as the time they will take to arrive, find subject experts, identify possible origins, and countermeasures, etc. Much information that can be highly relevant in this scenario exists on the internet, but locating it and assessing its reliability and relevance is extremely difficult. We believe that personalized recommendation systems can play a role in focusing the searches and finding the most important and urgently needed information.

**LANL (T) Collaborators/mentors:** Luis Bettencourt, T-5

**CRP Project (LANL/UCSC): Understanding Multi-scale Multi-streaming Events in Cosmological Simulations**



The science goal of this project is to seek a better understanding of the multi-streaming phenomenon in the universe. The specific aim is to develop tools that will analyze, extract and visualize multi-streaming events from cosmological simulations. Both technical and scientific challenges must be addressed for a successful outcome of the proposed work, including the ability to handle very large data sets and to perform comparative studies on different formulations of the rather imprecise definition of what constitutes multi-streaming. Multi-streaming is said to happen when at some (Eulerian) location, particles arrive from different (Lagrangian) locations and at different velocities. Places where multi-streaming happen are singularities in coordinate space, and are also referred to as caustics. This behavior is analogous to how light rays are reflected and/or refracted in a swimming pool to produce a brighter intensity where the light rays are convergent. In cosmological structure formation, cold flows in dark matter merge and interact in multi-streaming regions. Hence, the identification and characterization of these regions are of great importance in extending our understanding of cosmological evolution.

**LANL (T) Collaborators/mentors:** Salman Habib, T-2

Related Papers/presentations

- ◇ Presentation: 2010 ISSDM Day: How do multi streaming regions form and evolve?, October 2010

**CRP Project (LANL/UCSC): Web Based Pathogen Information System**

We are planning to develop a web-based system to help 'decision makers' quickly identify and process relevant web-based information in case of a disease outbreak. We will work on identifying the pathogen based on sequence information. We will also develop an adaptive information filtering to find, filter and condense the information available on the web.

**LANL (T) Collaborators/mentors:** Carla Kuiken, T-6

**RESEARCH PROPOSALS**

- ◇ NSF HECURA Informatics for Astronomy, 2009 (Salman Habib, T-2)
- ◇ Google Research Informatics for Astronomy, February 2009 (Salman Habib, T-2)
- ◇ NSF Cosmic Calibration: Meeting the Precision Cosmology Challenge, May 2009 (Salman Habib, T-2; Zarija Lukic, T-2)
- ◇ LDRD ER: Venturing Beyond the Cosmological Constant, April 2009 (Salman Habib, T-2; Suman Bhattacharya, T-2)
- ◇ LDRD DR: Cosmological Signatures of Physics beyond the Standard Model: Petascale Cosmology Meets the Great Surveys, March 2009 (Salman Habib, T-2; Suman Bhattacharya, T-2)

- ◇ NSF Understanding Uncertainties in Petascale Applications, October 2008 (Salman Habib, T-2)
- ◇ LDRD: Data to Knowledge at Scale: The Power of Data-Intensive Super-Computing (LDRD-DR Preproposal #20100029DR), January 2009 (Salman Habib, T-2)
- ◇ LDRD: Data to Knowledge at Scale: The Power of Data-Intensive Super-Computing (LDRD-DR Full Proposal #20100029), March 2009 (Salman Habib, T-2)
- ◇ LDRD-DR Optimization and Control Theory for Smart Grids, June 2009 (Jason Johnson, T-4; Nikolai Sinitsyn, T-4; Eli Ben-Naim, T-4; Rajan Gupta, T-2; Timothy Wallstrom, T-4)

## STAFF DEVELOPMENT AND RETENTION AND ORGANIZATIONAL OUTREACH ACTIVITIES

### **Cluster and Network Summer Institute Guest Lectures:**

Salman Habib (T-2)

- ◇ SI 2008 Guest Lecture: Surveying the Universe

- ◇ SI 2009 Guest Lecture: Surveying the Universe

Timothy German (T-12)

- ◇ SI 2008 Guest Lecture: Computing on Roadrunner Materials Science Application

- ◇ SI 2009 Guest Lecture: Computing on Roadrunner Materials Science Application

### **Mentor Opportunity**

- ◇ Mentor of Student for dissertation: Information-driven Cooperative Sampling Strategies for Spatial Estimation by Robotic Sensor Networks (Salman Habib, T-2)

### **Seminars/presentations/lectures**

Salman Habib, T-2

- ◇ Paper: Introducing Map-Reduce to High End Computing PDSW08, November 2008

- ◇ Paper: LSST Science Book, January 2010

- ◇ Paper: Simulations and Cosmological Inference: A Statistical Model for Power Spectra means and Covariances, January 2010

- ◇ Paper: The Case for Deep, Wide-Field Cosmology, no month

- ◇ Paper: Exploring Dark Energy with Next-Generation Photometric Redshift Surveys, January 2010

- ◇ Paper: Nonparametric Reconstruction of the Dark Energy Equation of State, November 2010

- ◇ Paper: Nonparametric Dark Energy Reconstruction from Supernova Data, December 2010

- ◇ Paper: Gaussian process modeling of derivative curves, November 2010

- ◇ Presentation: Supercomputing the Universe for the ISTI Summer School, July 2008

- ◇ Presentation: Introducing Map-Reduce to High End Computing PDSW 08, November 2008

- ◇ Presentation: Early Experiences with Disc-style Computation PDL08, October 2008

- ◇ Presentation: Early Experiences with Disc-style Computation PDSW08, October 2008

- ◇ Presentation: ISSDM Day: Cosmic Calibration - Statistical Modeling for Dark Energy, October 2009

- ◇ Presentation: Cosmic Calibration - Statistical Modeling for Dark Energy, April 2010

- ◇ Presentation: 2010 ISSDM Day: Statistical Modeling of Dark Energy and the Cosmological Constants, October 2010

- ◇ Carla Kuiken, T-10

- ◇ Presentation: ISSDM Day: Adaptive Information Filtering and its Application in Medical Domain, October 2009

- ◇ Presentation: 2010 ISSDM Day: Adaptive Information Filtering, October 2010

### **UCSC Graduate Classes are offered through ISTI (for credit toward a graduate degree in computer science or for professional development):**

- ◇ Pragneshkuman Patel, T-10: Attended course: Computer Networks, fall 2008