SDMIV2: Workshop Introduction

Bob Mann

Institute for Astronomy
University of Edinburgh

Outline

- Welcome and Introduction
- SC4DEVO "international sister project"
- First SDMIV workshop
- Workshop goals and programme
- Advert: journal special issue

Outline

- **Welcome and Introduction**
- SC4DEVO "international sister project"
- First SDMIV workshop
- Workshop goals and programme
- Advert: journal special issue

Welcome to eSI

- **e-Science Institute**
  - This building
  - The workshop-organising arm of NeSC
- **eSI Programme**
  - 2001-2005: community-building
    - Organise meetings in response to requests
  - 2005-: longer-term research support
    - Most meetings are part of "themes"

eSI Themes

- Themes
  - Coherent programme over 6-12 months
  - Theme Leader seconded part-time to eSI
- Theme-1:
  - Information services for smart decision making
  - Theme Leader: Jennifer Schopf
- Theme-2:
  - **Exploiting diverse sources of scientific data**
  - Theme Leader: Jessie Kennedy
- More details at www.nesc.ac.uk/esi

Outline

- Welcome and Introduction
- **SC4DEVO "international sister project"**
- First SDMIV workshop
- Workshop goals and programme
- Advert: journal special issue

Service Composition For Data Exploration in the Virtual Observatory (SC4DEVO)

- e-Science "international sister project"
  - One of first four funded
- Data exploration in astronomy…and beyond
  - Data mining and visualization
- Organise four workshops in 2004/2005
  - Two astronomy-specific
    - Caltech (July 2004) and Sydney (July 2005)
  - Two general
    - Edinburgh (Dec 2004 and Dec 2005)

Outline

- Welcome and Introduction
- SC4DEVO "international sister project"
- **First SDMIV workshop**
- Workshop goals and programme
- Advert: journal special issue

"Scientific Data Mining, Integration & Visualization"

- NeSC, October 2002
- www.nesc.ac.uk/talks/sdmiv/report.pdf
- Participants from wide range of domains
  - astronomy, atmospheric science, bioinformatics, chemistry, digital libraries, engineering, environmental science, experimental physics, marine sciences, oceanography, and statistics…plus CS researchers and software engineers

- But a common set of problems

Problems from SDMIV

- Lots of DM packages, lots of data formats
- How to mine distributed data sources?
- How to mine large data volumes?
  - and high-dimensional datasets
- How to do data exploration?
  - Coupling data mining and visualization
- How to work iteratively & interactively?
  - Tracking provenance, building workflows...

Outline

- Welcome and Introduction
- SC4DEVO "international sister project"
- First SDMIV workshop
- **Workshop goals and programme**
- Advert: journal special issue

Workshop Goals

- **What's happened since SDMIV1?**
  - Are there solutions to its problems?
  - Are they generic or domain-specific?
- Learn about work in other domains
- What do we do next?
  - Are meetings like this useful?
    - How often? Covering what?

Workshop Programme

- Introduction
- Data Integration
- SDMIV Technologies

Wednesday

- Data Mining
- Visualization
- Panel-led Discussion

Thursday

Outline

- Welcome and Introduction
- SC4DEVO "international sister project"
- First SDMIV workshop
- Workshop goals and programme
- **Advert: journal special issue**

Concurrency and Computation: Practice and Experience

- Special Issue: Computational analysis and exploration of distributed data

- Deadline: 15 January 2006

- More information
  - Call for Papers: research.bioinformatics.ulster.ac.uk/article.php3?id_article=81
  - Editors:
    - Werner Dubitzky (*w.dubitzky@ulster.ac.uk*)
    - Bob Mann (*rgm@roe.ac.uk*)
    - Omer Rana (*o.f.rana@cs.cardiff.ac.uk*)

Session 1: Data Integration

- 10.00 Tasos Kementsietsidis (Edinburgh)
- 10.30 Rasmus Fogh (Cambridge)
- 11.00 *Coffee*
- 11.30 Diego Prina Ricotti (Roma)
- 12.00 Neil Chue Hong (Edinburgh)
- 12.30 Anatoly Sorokin (Edinburgh)
- 13.00 *Lunch*