System-Level Provenance

Dr. David Koop
<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Presenter</th>
</tr>
</thead>
<tbody>
<tr>
<td>9/16</td>
<td>Provenance in Scientific Workflows</td>
<td>Shrinivass Balasubramanian</td>
</tr>
<tr>
<td>9/18</td>
<td>Workflow Evolution Provenance</td>
<td>Manideep Gunnam</td>
</tr>
<tr>
<td>9/23</td>
<td>System-Level Provenance</td>
<td>Vaishnavi Guduguntla</td>
</tr>
<tr>
<td>9/25</td>
<td>Database Provenance</td>
<td>Zach Sylvia</td>
</tr>
<tr>
<td>9/30</td>
<td>Provenance Storage</td>
<td>Sachin Patel</td>
</tr>
<tr>
<td>10/7</td>
<td>Querying Provenance</td>
<td>Avin Myneni</td>
</tr>
<tr>
<td>10/9</td>
<td>Map-Reduce Provenance</td>
<td>Kevin Lydon</td>
</tr>
<tr>
<td>10/9</td>
<td>Provenance Analytics</td>
<td>Sivaraj Srinivas Busayavalasa</td>
</tr>
<tr>
<td>10/14</td>
<td>Provenance Mining</td>
<td>Rutvi Dave</td>
</tr>
<tr>
<td>10/16</td>
<td>Secure Provenance</td>
<td>Bharathi Kotharu</td>
</tr>
<tr>
<td>10/21</td>
<td>Provenance &amp; Semantics</td>
<td>Yashwanth Nandanam</td>
</tr>
<tr>
<td>10/23</td>
<td>Provenance Standards</td>
<td>Srinidhi Pesara</td>
</tr>
<tr>
<td>10/28</td>
<td>Visualization &amp; Provenance</td>
<td>Surya Kiran Juthuka</td>
</tr>
<tr>
<td>10/30</td>
<td>Visualization &amp; Provenance</td>
<td>Prudhvi Pathri</td>
</tr>
<tr>
<td>11/4</td>
<td>Reproducibility</td>
<td>Gursharanpreet Singh</td>
</tr>
<tr>
<td>11/6</td>
<td>Reproducibility</td>
<td>Drushti Gawade</td>
</tr>
<tr>
<td>11/8</td>
<td>Reproducibility</td>
<td>Laxman Pothamshetti</td>
</tr>
<tr>
<td>11/18</td>
<td>Graph Databases</td>
<td>Bhargav Thatikonda</td>
</tr>
<tr>
<td>11/18</td>
<td>Graph Databases</td>
<td>Kishore Chikkam</td>
</tr>
<tr>
<td>11/20</td>
<td>Graph Indexing</td>
<td>Dhvani Patel</td>
</tr>
<tr>
<td>11/25</td>
<td>Scientific Databases</td>
<td>Jeffrey Rezendes</td>
</tr>
</tbody>
</table>
Reading Presentations

• Most received their first or second choice
• If you are interested in switching, you and person you are switching with must email me ahead of time
• Remember:
  - Understand the material, look up background material if needed and include it in the presentation for the benefit of all
  - Use the figures and tables the authors include if they are informative
  - Try to highlight issues or topics that we can discuss
  - No reading response on the day you present
Course Project Proposal

• Due October 2 at 12pm, also a few slides for October 2 class

• Requirements:
  - Standard metadata: Title, Authors, Date
  - Introduction with description of the project goals
  - Motivation: why did you choose this project and how does it tie into the topics of the class
  - Background: what do we need to understand as background
  - Design: what are the components and for your project
  - Implementation Details: (language, platform(s), system requirements)
  - Project plan: what are the different components and what is your timeline for implementing the different pieces?
Review & Roadmap

• Scientific Workflows and Provenance
  - Systems: Kepler, Karma, VisTrails
  - Retrospective Provenance:
    • Coarse: based on the definition of actors/services/modules
    • Captures details as actors execute (the engine itself, each service, or hybrid)
  - Prospective Provenance:
    • Workflow specification and instance
    • Workflow evolution (provenance of the workflow design)
• This Week: Other Types of Provenance:
  - System-Level Provenance
  - Database Provenance

} Fine-Grained
Provenance-Aware Storage Systems


Presented by: Vaishnavi Guduguntla
PASS Notes

• “A provenance-aware storage system (PASS) is a storage system that automatically collects, stores, manages, and provides search for provenance.” [K.-K. Muniswamy-Reddy et al.]

• “...ability to uncover system mysteries”

• Potential Issues: overhead, cycles, clutter

• Suggestion (from the paper): combine system-provenance with other types of provenance

• Level of provenance: Filesystem
  - Is this best?
  - What about other levels of the system?
ES3

- Another system-level provenance solution
- Software environment for data-intensive Earth science
- Multiple approaches: strace, instrument libraries, annotated code
- “Plugins” all report provenance
- Potential for detecting dependency differences from a coarse-grained specification
ES3 Example

10. slicer
   ▼
   Atlas X Slice

11. slicer
   ▼
   Atlas Y Slice

12. slicer
   ▼
   Atlas Z Slice

13. convert
   ▼
   Atlas X Graphic

14. convert
   ▼
   Atlas Y Graphic

15. convert
   ▼
   Atlas Z Graphic

Fig. 2. convert operation in challenge workflow

Fig. 3. ES3 provenance for convert operation

[ES3, Frew and Slaughter]
ES3: A Demonstration of Transparent Provenance for Scientific Computation

J. Frew and P. Slaughter
Reminders

• Course Project Proposals
  - If you would like me to look over a draft of proposal, please email it to me

• Thursday: Database Provenance
  - Reading Response (respond to 1st Paper, skim 2nd Paper)