Provenance for Interactive Visualizations

Fotis Psallidas
fotis@cs.columbia.edu

Eugene Wu
ewu@cs.columbia.edu
Provenance Primer

Fine-Grained Provenance

(Connections between input and output tuples)
Provenance Primer

Fine-Grained Provenance
(Connections between input and output tuples)

\[ O = \gamma_{\text{state, avg(delay)}}(\text{Airports} \Join \text{Flights}) \]

Airports

<table>
<thead>
<tr>
<th>name</th>
<th>state</th>
</tr>
</thead>
<tbody>
<tr>
<td>LGA</td>
<td>NY</td>
</tr>
<tr>
<td>JFK</td>
<td>NY</td>
</tr>
<tr>
<td>IAH</td>
<td>TX</td>
</tr>
</tbody>
</table>

Flights

<table>
<thead>
<tr>
<th>from</th>
<th>delay</th>
<th>state</th>
</tr>
</thead>
<tbody>
<tr>
<td>LGA</td>
<td>30</td>
<td>NY</td>
</tr>
<tr>
<td>LGA</td>
<td>40</td>
<td>NY</td>
</tr>
<tr>
<td>JFK</td>
<td>50</td>
<td>NY</td>
</tr>
<tr>
<td>IAH</td>
<td>60</td>
<td>TX</td>
</tr>
</tbody>
</table>

\[ \gamma \rightarrow \begin{align*}
O_1 &= \text{NY} \\
O_2 &= \text{TX}
\end{align*} \]

Columbia Engineering
The Fu Foundation School of Engineering and Applied Science
Provenance Primer

Fine-Grained Provenance
(Connections between input and output tuples)

Airports

<table>
<thead>
<tr>
<th>name</th>
<th>state</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1</td>
<td>LGA</td>
</tr>
<tr>
<td>a2</td>
<td>JFK</td>
</tr>
<tr>
<td>a3</td>
<td>IAH</td>
</tr>
</tbody>
</table>

Flights

<table>
<thead>
<tr>
<th>from</th>
<th>delay</th>
<th>state</th>
</tr>
</thead>
<tbody>
<tr>
<td>f1</td>
<td>LGA</td>
<td>30</td>
</tr>
<tr>
<td>f2</td>
<td>LGA</td>
<td>40</td>
</tr>
<tr>
<td>f3</td>
<td>JFK</td>
<td>50</td>
</tr>
<tr>
<td>f4</td>
<td>IAH</td>
<td>60</td>
</tr>
</tbody>
</table>

\{f_4\} = \text{backward\_trace}(\{o_2\}, \text{Flights})
Provenance Primer

Fine-Grained Provenance
(Connections between input and output tuples)

### Airports

<table>
<thead>
<tr>
<th></th>
<th>name</th>
<th>state</th>
</tr>
</thead>
<tbody>
<tr>
<td>a₁</td>
<td>LGA</td>
<td>NY</td>
</tr>
<tr>
<td>a₂</td>
<td>JFK</td>
<td>NY</td>
</tr>
<tr>
<td>a₃</td>
<td>IAH</td>
<td>TX</td>
</tr>
</tbody>
</table>

### Flights

<table>
<thead>
<tr>
<th></th>
<th>from</th>
<th>delay</th>
<th>state</th>
</tr>
</thead>
<tbody>
<tr>
<td>f₁</td>
<td>LGA</td>
<td>30</td>
<td>NY</td>
</tr>
<tr>
<td>f₂</td>
<td>LGA</td>
<td>40</td>
<td>NY</td>
</tr>
<tr>
<td>f₃</td>
<td>JFK</td>
<td>50</td>
<td>NY</td>
</tr>
<tr>
<td>f₄</td>
<td>IAH</td>
<td>60</td>
<td>TX</td>
</tr>
</tbody>
</table>

### States

<table>
<thead>
<tr>
<th></th>
<th>state</th>
<th>avg(delay)</th>
</tr>
</thead>
<tbody>
<tr>
<td>o₁</td>
<td>NY</td>
<td>40</td>
</tr>
<tr>
<td>o₂</td>
<td>TX</td>
<td>60</td>
</tr>
</tbody>
</table>

\[
\{o_2\} = \text{forward_trace}([\{f_4\}, O])
\]
Provenance Primer

Fine-Grained Provenance
(Connections between input and output tuples)
Provenance Primer

Fine-Grained Provenance
(Connections between input and output tuples)

• Navigation of the input-output connections
  • \{records\} = backward_trace(...)  
  • \{records\} = forward_trace(...) 

• Provenance consuming queries
  • SQL(backward_trace(...))  
  • SQL(forward_trace(...))
Goal of this talk
How to use fine-grained provenance to express core interactive application functionality

Why though?
- **(Expressivity)** Logic over provenance is expressed declaratively
- **(Performance)** Provenance management systems are becoming *fast*

See [Smoke, VLDB18] or pass by our demo on Wednesday/Thursday
Connections

Core interaction logic with provenance
• Selections
• Logic over selections
• Multi-view linking
**Goal:** Get subset of inputs that correspond to selected visual outputs

**Example:** Find the airports that operate at the selected states
**Interactive Selections**

**Goal:** Get subset of inputs that correspond to selected visual outputs

**Example:** Find the airports that operate at the selected states

\[
\text{airports} = \text{backward_trace}(..., \text{airports})
\]
Logic over Selections

Goal: Express application logic over the selected inputs

Example: Find the #airports that operate at the selected states

SQL(backward_trace(airports))

54 airports operate in this area
**Multi-View linking**

**Goal:** Look at the relationships between different views

**Example:** Show the distribution of #flights per carrier only for selected states

```
backward_trace(airports, flights)
```

```
selective_refresh(V2, flights)
```
Provenance for Interactive Visualizations

Interactive Selections
- Item selection
- **Group selection**
- Range selection
- Generalized selections

Logic over Selections
- **Tooltips**
- Details-On-Demand
- Semantic Zooming

Multi-View Linking
- Linked Brushing
- **Crossfilter**

```
backward_trace(...) SQL(backward_trace(...)) selective_refresh(backward_trace(...))
```
What next?

Traditionally provenance systems have been at the core of several applications

- Data Integration
- Debugging
- Network Diagnostics
- Resource Scheduling
- Auditing

- Interactive Visualizations
- Interactive Data Profiling
- Multi-Application Linking
- Interactive Query Specification
- What-if Provisioning

- Query Explanations
- Why-not Analytics
- Iterative Analytics
- Viz Workflow Debugging
- Interactive Data Cleaning

- Interaction Debugging
- ML Interpretability
- Visualization Deconstruction and Restyling
- Interaction By Example
- Application Design Search

- Collaborative Communication
- Action Recovery
- Sense-Making
- Meta-Analysis
- Replication and Reproducibility
Traditionally provenance systems have been at the core of several applications

Data Integration  Debugging  Network Diagnostics  Resource Scheduling  Auditing  ...

(Fast) Provenance management systems can make a difference on several other domains

Interactive Visualizations  Interactive Data Profiling  Multi-Application Linking  Interactive Query Specification  What-if Provisioning

Query Explanations  Why-not Analytics  Interactive Analytics  Viz Workflow Debugging  Interactive Data Cleaning

Interaction Debugging  ML Interpretability  Visualization Deconstruction and Restyling  Interaction By Example  Application Design Search

Collaborative Communication  Action Recovery  Sense-Making  Meta-Analysis  Replication and Reproducibility
Multi-Application Linking

• Many applications are built over the same database (esp. in enterprises)
• Extend multi-view linking to multi-application linking
• Powerful for: connecting data across apps, reuse app logic
Multi-Application Linking

- Many applications are built over the same database (esp. in enterprises)
- Extend multi-view linking to multi-application linking
- Powerful for: connecting data across apps, reuse app logic

Search App

Data Store

Vis App

KAYAK
HOUSTON

2:50-5:60 DL33
IAH, TX

3:50-6:60 UA22
IAH, TX

Average Departure Delay
60 minutes

airports
airlines
flights
Takeaway

Interaction Logic as Provenance

⇓

Declarative wins & Holistic optimization
Thank You
//Q