**SQL Query Visualization Using Data Provenance**

Robert Ward, Charlie Summers, Haneen Mohammed, Professor Eugene Wu

### Datasets and SQL Query Input

SQL is a language that uses **Queries** to interact with and manipulate databases.

- **Declarative**
- **High-level**

**Query Execution**

A query engine converts it into a **Physical Plan** which is executed on the data.

- **Restructured**
- **Engine-specific**
- **Lower-level**

**Data Provenance**

Data provenance is metadata describing the origin of data values and how it was processed throughout the execution.

Once provenance is captured, **How should we convey it?**

### Related Implementations

**LANTERN** [1] – Physical Plan Execution

- Natural language descriptions
- Physical operator explanations
- Can arrange to full sequence steps

**SAVI** [2] – Query Syntax Breakdown

- Abstracted to query-level operations
- Generates intermediate datasets
- Animates transformations

**Perfopticon** [3] – Distributed Database Performance Analysis

- Physical plan overview
- Performance details by operator

### Considerations

- Be able to follow along the entire query execution
- Show how values are derived and processed
- Allow to focus on a singular operation
- Visuals should resemble actual operator behavior

### Implementation

**Query Plan Tree Diagram**

Query Plan displayed as an interactive diagram, acts as point of reference for entire query.

**Operator Transformations**

Visualizations for each physical operator, with unique annotations for each operator type.

**Intermediate Datasets**

All visualizations display inputs (output of the previous operators) and how it is transformed by this operator to the output for the next operator.

**Operator Steps**

Can view operators in isolation, or all at once arranged in post-order to show entire sequence.

### References