A Demonstration of the BigDAWG Polystore System

**Motivation**

- Hybrid applications are becoming common
- Use many back-ends, lots of data models, single dataset
- Example: MIMIC II, intensive care unit dataset
- Workflow contains:
  - Streaming for heartbeat monitoring
  - Complex analytics for data mining
  - Browsing schemas and patients individually
- Data Exploration or “tell me something interesting”
- Need a unified interface for diverse users and their workloads

**The BigDAWG Polystore**

- A polystore is a federated db system for multiple, disparate data models
- Polystore requirements:
  - Location transparency
  - Semantic completeness
  - N:N relationship between user semantics and back-ends
- BigDAWG achieves data independence with islands of information. Each contains a
  - Data model, query language, shims to dbs
- Users pose queries by invoking islands with scope and casting between disparate semantics. Example: RELATIONAL

```sql
SELECT * FROM R, CAST(A, relation) WHERE A.v = R.v;
```

**Technologies for Visualization of Big Medical Text Data**

Lelani Leite, Lauren Edmonds, Vigna Godbole, Brandon Harms, Nadine Makarenko, Dylan Hitchcock, Jeremy Kaper, Anil Moran

**ABSTRACT**

The SPT BigData Pipeline, aiming to process biomedical data, includes a modular platform to transform raw data into structured, human-readable data. This project uses a modular design to visualize data from multiple sources, including heterogeneous data types.

**System Prototype**

The pipeline can be configured to support processing with a range of query analyzers, including relational DBMSs and graph databases. The system can be extended to support streaming data from multiple sources.

**Future Work**

- Explore new data sources
- Implement new query analyzers
- Improve performance

**SeeDB: Efficient Data-Driven Visualization Recommendations to Support Visual Analytics**

Manasvi Varrak, Sajjad Rahaman, Samuel Maddess, Aditya Parameeswaran, Neoklos Polyzois

**Motivation**

- Data visualization first step in analysis
- High-dimensional data 100s of visualizations Manual specification
- Automation for rapid analysis
- Automatically identify and recommend “interesting” visualizations

**Approach**

- Interestingness metrics defined via metrics: Visualization Image saliency in showdevs
- Techniques for interactive visualization
- Detection of interesting insights
- Automatic visualization recommendations

**System Architecture**

- Visualizations on a “what’s new” list
- Require interactive latency
- Automatic visualization recommendations

**Evaluation**

- User Study: Comparison of manual vs. automated suggestions
- SeeDB ranking along with number of bookmarks [F1 vs. 18.398; p = 0.001]