**Seer: Profile-Driven Prefetching and Caching**
for Interactive Visualization of Large Multidimensional Data

Justin DeBrabant, Chenggang Wu, Ugur Cetintemel, and Stan Zdonik

### Introduction
- Big data visualization popular across many domains
- Challenge: High data-fetch latency
- Seer: Predictive database middleware layer
  - Used for large multidimensional datasets
  - Hides data-fetching latency/improve interactivity

### System Architecture and Features

**Visualization Application**

- Query
- Data tile

**Buffer Manager**

**Hierarchical Prediction Model**

- Split query into two parts: template and parameter
- Two-stage prediction:
  - Variable order Markov model predicts query template
  - Sequential association rule mining predicts query parameters
- Train models with query logs
- Probabilistic prediction tree:
  - Root: current tile
  - Predictions at arbitrary depth

**Prefetching and Caching**

- Tunable prefetching confidence threshold to meet user-specified accuracy requirements
  - Supports both aggressive prefetching and conservative prefetching
- Seer cache separate from system’s cache
- Size of cache block equals to size of data tiles
- Predictive LRU eviction policy increases cache hit rate

### Hierarchical Prediction Model

- Probabilistic Prediction Tree

### Experimental Analysis

- Incorporate Seer into imMens visualization platform
- User study:
  - five visualization tasks
  - Brightkite dataset
- Metrics: prediction precision, cache hit rate, and latency

### Experimental Results

- Prediction model provides up to 5× better precision
- Predictive LRU cache eviction policy gives 3× cache hit rate boost
- Improved prediction accuracy and cache hit rate translate to 2× lower average latency across all workloads