Towards Workload-Aware Self-Management: Predicting Significant Workload Shifts

Marc Holze, Ali Haschimi, Norbert Ritter
Distributed Systems and Information Systems
Department of Informatics
University of Hamburg

5th International Workshop on Self Managing Database Systems
Long Beach, March 1, 2010
Cyclic DBS Workload

- DBS configuration depends on workload
- Prior Work: Detect significant workload changes to trigger tuning by DBA or self-management functions
- DBS workload often shows cyclic characteristics
- Goal: Identification of periodic workload shifts
Stage 1: Workload Monitoring

- Workload monitoring levels
  - OS workload metrics (CPU usage, I/O, memory)
  - DBS-internal metrics (page requests, sorting, tablescans, index reads, ...)
  - **Statement-level monitoring**

- Statement-level monitoring
  - Internal characteristics (source: execution plans)
    - identify semantically equivalent statements
    - identify changes in statement processing over time
  - **External characteristics** (source: SQL statement texts)
    - standardized
    - represents *usage* of DBS
Stage 2: Workload Classification

- Reduction of workload diversity
- Self-Management
  - Classification techniques inappropriate
  - Clustering (k-means)
  - Class defined by nearest medoid
- Feature selection for SQL statements
  - subject to experimental evaluation
  - e.g. tables accessed, grouping columns
- Distance function
  - nominal, set-valued features
  - Generalized Minkowski metric
- Workload Evolution
  - Add or extend clusters
  - Prevent stealing of feature vectors
Stage 3: Workload Models & Shift Detection

- n-gram models
  - approximation of Markov models
  - statistically dependent events

- Assessment of observed workload: perplexity

\[
PP(X) = \left( \prod P(X_t | X_{t-n+1}, \ldots X_{t-1}) \right)^{-1/T}
\]
Workload Shift Prediction

- **Requirements**
  - Different periodicity types
  - Self-Management
  - Robustness

- **Prerequisite: Identification of recurring workload models**
  - exploit existing similarity measure (perplexity)
  - store workload used for model creation
  - bi-directional comparison

---

**Model Pool**

- Current Model
- Model Pool

**Pattern Workload**

**Interval Workload**

**Regular Workload**

- Monday 06:00
- Monday 18:00
- Tuesday 06:00
- Tuesday 18:00
Periodicity Detection

- **Workload history**
  - nominal-scaled
  - time-series analysis techniques not applicable

- **Model histories**
  - Split workload history by models
  - Transform each model history to interval-scaled activation graph

- **Periodicity Analysis**
  - Solution 1: Convert to frequency domain and employ existing periodicity measures
  - Solution 2: "Manual" activation interval analysis
Prediction and Pattern Validation

- **Periodicity Information**
  - periodic *patterns*
  - average activation intervals

- **Prediction of DBS workload shift**
  - when periodic pattern of *any* model indicates activation

- **Pattern validation**
  - patterns may evolve or become invalid
  - adapt patterns to deviations within allowed fluctuation range
  - skip premature arrivals; count late arrivals
Summary & Outlook

Summary

- Identification of periodic patterns in DBS workload
- Usage
  - knowledge for DBA or self-management functions
  - reactive application of DBS configurations
  - proactive application of DBS configurations?

Outlook

- Rule-based predictions
- Link workload models with DBS configurations
- Trigger self-management functions
- Examine overlapping predictions