

SOFIYA SEMENOVA

SENIOR THESIS

Contents

<i>Introduction</i>	5
<i>The Demo</i>	7
<i>Economic Methods (name?)</i>	9

Introduction

MY SENIOR THESIS focuses on leveraging machine learning methods to develop cost-effective performance management services for cloud databases. To that end, the first part of my thesis consists of developing a demo for WiSeDB, an end-to-end, machine learning based solution for workload management for cloud databases. The second part consists of developing a machine learning approach to data-driven fragmentation and fragment distribution.

Motivation

WITH THE MOVEMENT of database applications from on-premises data centers to cloud environments, database application developers are responsible for addressing the problems of resource distribution, workload distribution, and query scheduling, all while meeting specific performance criteria. To leverage the full benefits of cloud databases, an end-to-end solution that addresses all the above concerns and maintains different cost and performance constraints would, many times, create better solutions than a database developer could. (Give an example?)

Further, the cost of using IaaS cloud databases could be greatly decreased with custom data-driven fragmentation. (More information about this - tie this into WiSeDB and why it's good to develop a system that has all these features)

Previous research

PREVIOUS RESEARCH HAS addressed resource provisioning, workload distribution, and query scheduling independently, but before WiSeDB, an end-to-end solution that addresses all the problems had not yet been implemented. (Include some examples of previous research)

Several papers have been written about custom fragmentation schemes:

MARIPOSA

- Idea behind Mariposa: a centrally-planned economy wherein nodes bid for queries and a broker decides which node "wins" the bid. Nodes can also buy fragments from other nodes, sell fragments from other nodes, and delete their fragments.

- Some problems: Mariposa is the basis for the idea of "economic methods", but has far too many components to be feasible to implement. Further, a centrally planned economy with a "broker" controlling all transactions is a lot of overhead that could be simplified with a more distributed system of planning. Lastly, the Mariposa system, being an economic system, relies on monetarily quantifying transactions between nodes, the broker, and the user. However, the "money" used in Mariposa does not translate to any real money spent, which creates an unnecessary barrier for the user when attempting to put a price on queries they want to evaluate.

SWORD

- Idea behind SWORD: SWORD's method of creating fragments relies on constructing a hypergraph where rows in the database are the nodes and queries are the edges, minifying the hypergraph by combining similar rows and queries, and finding the optimal cut of the min-graph.

- Some problems: The SWORD implementation creates a lot of overhead to create the hypergraph, minify it, and find the optimal cut. We would like to decrease some of that overhead to create a more usable system. (Look into how performance is measured and given to the user. Does SWORD guarantee a Nash equilibrium?)

Summary of work done

(Should I include this section? It seems like I'm already going over a summary of what I did in the next two chapters)

The Demo

Introduction

THE FIRST HALF of my senior thesis consisted of developing a demo to allow users to visualize WiSeDB. The demo is a website written in Java (using the Spark web framework) and Vue.js. The final result allows users to play around with variables such as template types, SLA type, SLA value, and the learning method, in order to see the different outputs from WiSeDB. (Do I need to explain more in depth here?)

WiSeDB

THE DEMONSTRATION is built on WiSeDB, a machine-learning driven system for cost and performance management for cloud databases. WiSeDB is comprised of two learning-based approaches: a supervised learning approach, and a reinforcement learning approach.

THE SUPERVISED LEARNING APPROACH - Go over the construction of the decision tree and how that is used to generate a schedule. Stress that supervised learning is used for batch scheduling.

THE REINFORCEMENT LEARNING APPROACH - Stress how this approach would be used for ONLINE scheduling. Go over the tiered VM system and how decisions are made to process a query.

Demo

THE DEMO ALLOWS the user to select a subset of provided query templates, specify an SLA type and value, and choose a learning approach.

(Include screenshots of website)

IF THE USER selects supervised learning, the website generates a few SLA recommendations and asks the user for a workload specification. (Explain SLA recommendations?) Once selected, a recommended strategy is generated. The user can then view the decision tree that was used to create the strategy, view the predicted cost of WiSeDB's strategy compared with some common heuristics, and run the strategy on the cloud to create actual costs. The user can then decide to change any variables and observe any changes in the recommendation to get a sense of the intuition behind them.

IF THE USER selects reinforcement learning, the website begins generating queries (poisson process?) and uses WiSeDB to make decisions for query placement as described in the reinforcement learning section. The user can see the query queue for each VM and observe the queries being processed by each VM, the creation and deletion of VMs, and the scheduling of new queries in real-time. The user can also pause execution, and look at more in-depth stats for each VMs history.

Further, the user can see some statistics about the query execution on the left side of the screen, and observe how the query latency converges over time as the system learns.

Economic Methods (name?)

Introduction

Overview of the System

Go over:

- Value-based fragmentation
- Initially distribute fragments across nodes to create a nash equilibrium
- Splitting/joining fragments based on variance
- Distributing fragments based on cost and density estimates

Results

