WiSeDB: A Learning-based Workload Management Advisor for Cloud Databases

NTHU　DB / AI Bootcamp

2018/08/31

Wang Ning Ya

Basic information of this paper

- Issue : 2016. 6. 10
- Authors : Ryan Marcus and Olga Papaemmanouil

  of Brandeis University
- Published in : Proceedings of the VLDB Endowment

Outline

- Cloud DBMS
- About the problem
  - workload management
  - notations
  - problem definition
- Approach
  - overview
  - graph construction
  - feature extraction

Recalling the Definition of Cloud DBMS

- A ***cloud DBMS*** is a DBMS designed to run in the cloud and that could ensure ***SAE:***
  - ***high Scalability***
  - ***high Availability***
  - ***Elasticity***
- **Workload Management**

Workload Management

- What is workload management?
  - Provisioning resource
  - Assigning incoming queries to provisioned VMs
  - Query scheduling within a VM

- Specific by application workload characteristics and performance goals

Notations Definition

- Query templates T = { T1, T2, … }
- Workload Q = { q_1^x, q_2^y, … }
- VM_i = { q_1^x, q_2^y, … }

- Schedule S = { VM_i , VM_j, … }

- Fix VM start-up cost : f_s ; Running cost per unit of time : f_r
- Latency of a query q_j^x run on VM_i = $l(q_j^x, i)$.

Problem Definition

- Given performance goal R and query templates T, we try to find a S that minimize the cost:

$$cost(R,S) = \sum_{vm_j^i \in S} \left[f_s^i + \sum_{q_k^m \in vm_j^i} f_r^i \times l(q_k^m, i) \right] + p(R,S) \quad (1)$$

- Where p(R,S) is penalty function for queries didn't meet up with given deadline.
- NP-Hard problem

How?

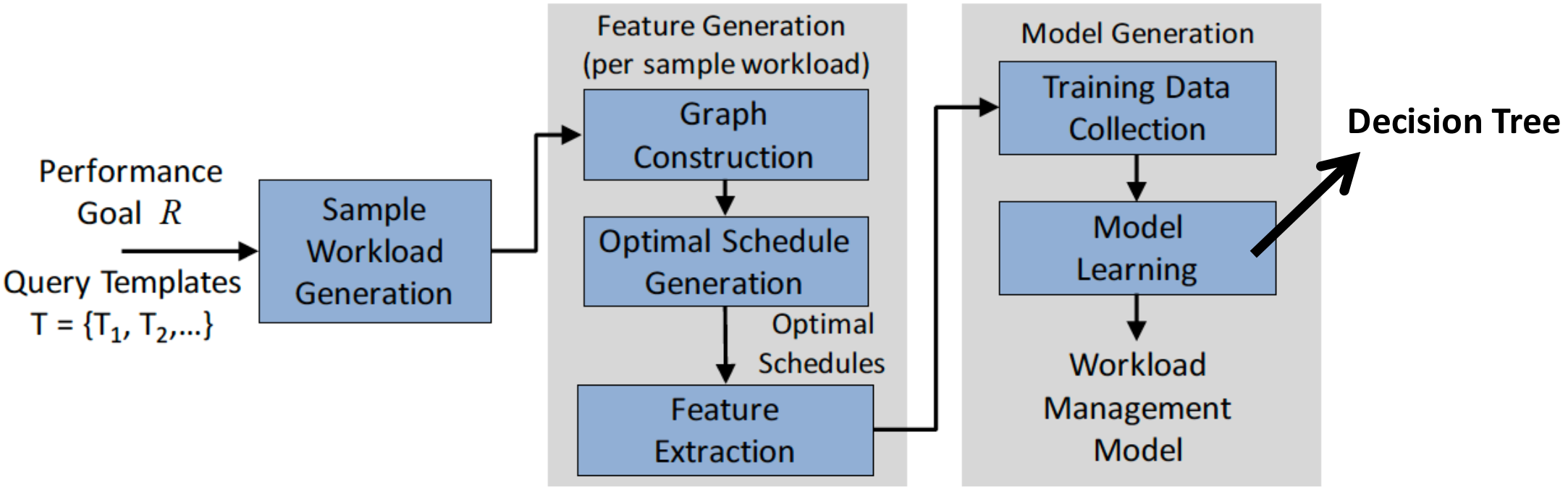
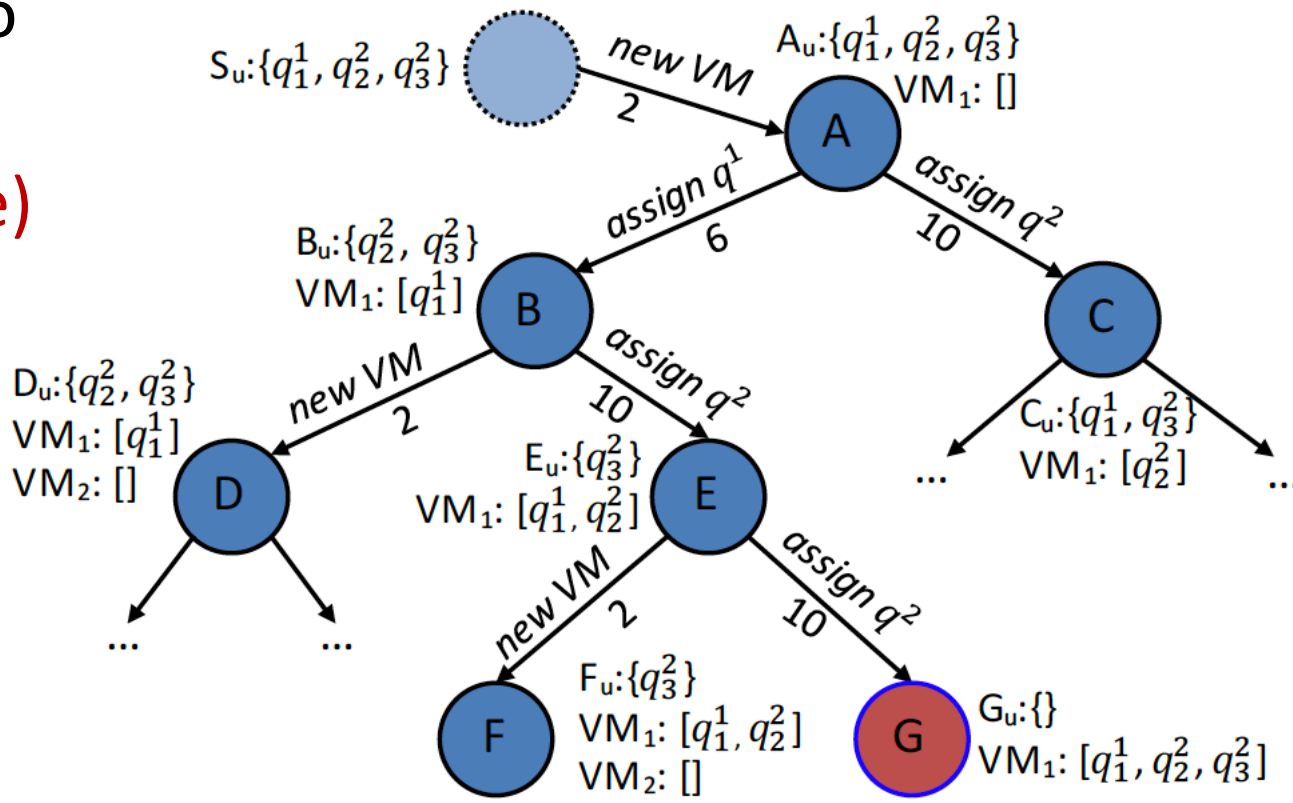


Figure 3: Generation of the decision model

Graph Construction

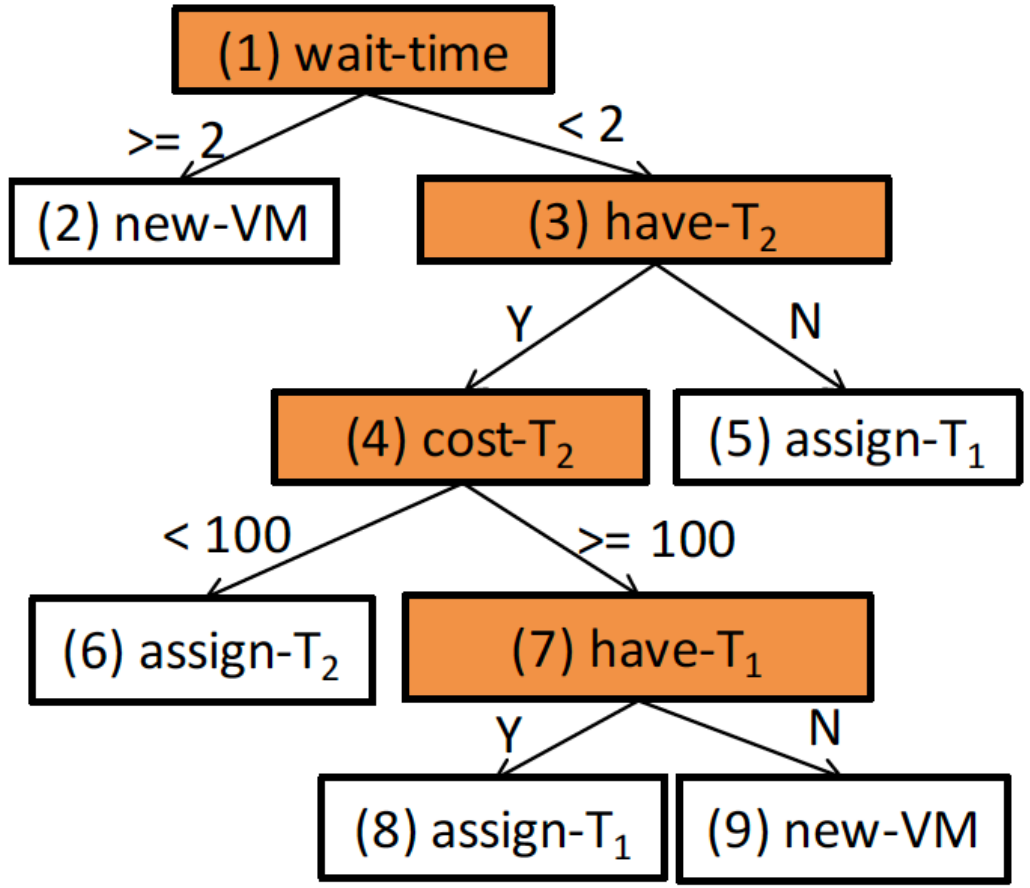
- Every node of the graph represents a scheduling step.
- To simplify question, we have two choices in every step:
- Renting a new VM (start-up edge)
- Assign a query to the newest VM (replacement edge)



Feature Extraction

- Edge selection for a vertex u only depends on the unassigned queries u_u and the scheduled u_s so far.

- Features :
  - wait – time
  - proportion – of – x
  - cost – of – x
  - have - x



Wrap up

- input :
  - query templates T
  - Performance goal R
- output : a schedule S that could minimize cost function

$$cost(R,S) = \sum_{vm_j^i \in S} \left[f_s^i + \sum_{q_k^m \in vm_j^i} f_r^i \times l(q_k^m, i) \right] + p(R,S) \quad (1)$$