

How to Apply Modeling to Compare Options and Select the Appropriate Cloud Platform

Boris Zibitsker, PhD[†]
BEZNext
bzibitsker@beznext.com

Alex Lupersolsky, PhD
BEZNext
alupersolsky@beznext

ABSTRACT

Organizations want to take advantage of the flexibility and scalability of Cloud platforms. By migrating to the Cloud, they hope to develop and implement new applications faster with lower cost. Amazon AWS, Microsoft Azure, Google, IBM, Oracle and others Cloud providers support different DBMS like Snowflake, Redshift, Teradata Vantage, and others. These platforms have different architectures, mechanisms of allocation and management of resources, and levels of sophistication of DBMS optimizers which affect performance, scalability and cost. As a result, the response time, CPU Service Time and the number of I/Os for the same query, accessing the similar table in the Cloud could be significantly different than On Prem.

In order to select the appropriate Cloud platform as a first step we perform a Workload Characterization for On Prem Data Warehouse. Each Data Warehouse workload represents a specific line of business and includes activity of many users generating concurrently simple and complex queries accessing data from different tables. Each workload has different demands for resources and different Response Time and Throughput Service Level Goals. In this presentation we will review results of the workload characterization for an On Prem Data Warehouse environment.

During the second step we collected measurement data for standard TPC-DS benchmark tests performed in AWS Vantage, Redshift and Snowflake Cloud platform for different sizes of the data sets and different number of concurrent users.

During the third step we used the results of the workload characterization and measurement data collected during the benchmark to modify BEZNext On Prem Closed Queueing model to model individual Clouds.

And finally, during the fourth step we used our Model to take into consideration differences in concurrency, priorities and resource allocation to different workloads. BEZNext optimization algorithms incorporating Graduate search mechanism are used to find the AWS instance type and minimum number of instances which will be required to meet SLGs for each of the workloads. Publicly available information about the cost of the different AWS instances is used to predict the cost of supporting workloads in the Cloud month by month during next 12 months.

CCS CONCEPTS

Data Analytics, Cloud, Performance

KEYWORDS

Cloud Platform, Service Level Goals, Workload Characterization, Workload Forecasting, Seasonality Determination, Benchmarking, Modeling, Optimization.

ACM Reference Format:

Boris Zibitsker, Alex Lupersolsky. 2020. How to Apply Modeling to Compare Options and Select the Appropriate Cloud Platform. *In Eleventh ACM/SPEC International Conference on Performance Engineering Companion (ICPE '20 Companion)*, April 20–24, 2020, Edmonton, Canada. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3375555.3384938>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

ICPE '20 Companion, April 20–24, 2020, Edmonton, AB, Canada

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-7109-4/20/04...\$15.00

<https://doi.org/10.1145/3375555.3384938>