Comparison of NoSQL on the Cloud Software, Systems and Platforms

J. McFadden
Univ. of Washington: Tacoma
Tacoma, WA
mcfaddja@uw.edu

Y. Tamta
Univ. of Washington: Tacoma
Tacoma, WA
yashaswitamta@gmail.com

J. N. Gandhi
Univ. of Washington: Tacoma
Tacoma, WA
jugalg@uw.edu

April 11, 2017

Project coordinator indicated by ***

Abstract

The software/systems chosen for comparison in this project are two different NoSQL database system. These systems will be deployed/run/operated in several different ways. These include *SaaS*[1] implementations, *containerized* implementations, and *native installations*. The goal of the project is to understand the performance characteristics of each deployment method ***and*** to quantify the costs of each deployment method. These costs will be calculated based on the hourly cost to operate, the initial time & costs required for setup, and the maintenance requirement of a deployment. Additionally, performance of the systems and deployments will be measured using the time required to carry out various database operations, under a set of several different conditions, as well as the CPU, memory, and network loads imposed by the various deployments under the same set of conditions.

1 Systems and Platforms

We will be using two NoSQL database software packages. The first software package is **DynamoDB** from Amazon Web Services (*AWS*), while the second software package will be **Cassandra**, an open-source NoSQL database software package. These software packages will be deployed using several different systems and platforms, as

[1] **SaaS :** Software as a service.

described below.

1.1 Systems

This project will use four different systems for running these software packages. These systems range from hosted *SaaS* through various degrees of virtualization and then, finally, to non-virtualized machines. These systems are

- **AWS *SaaS* system(s)**
- **Virtualization using Docker containers**
- **Virtualization on AWS's EC2 VMs**
- **Dedicated, non-virtualized machines**

1.2 Platforms

We have chosen three different platforms on which to deploy the software packages, with the platforms spanning the range of cloud service paradigms from *SaaS* to *PaaS*[2] to *IaaS*[3]. These platforms are

- **AWS's DynamoDB Service** (*SaaS*)
- **Containerized implementations** (using *Docker*) running on
  - ***AWS's Container Service*** (*PaaS*)
  - ***AWS EC2 Machines running the docker run-time in Linux*** (hybrid *Pass/IaaS*)
- **AWS EC2 Machines running native installations of the software in Linux** (*IaaS*)

2 Deployment

[2] **PaaS :** Platform as a service.
[3] **IaaS :** Infrastructure as a service.