Hybrid Computation

Leveraging Cloud Computational Resources with On-Premesis Data

---

Tanvi Arora
Mahesh Kuklani

Examiner: Dr. Sohail Rafiqi

Southern Methodist University
Data Science
Cloud Computing

December ??th, 2018

Writing period

09. 22. 2018 – 12. ??. 2018

Project Proposal

The world we live in today is one that is increasingly data-centric. Along with our ability to generate more data must also come an increase in our capacity to make sense of that data. As organizations attempt to scale their environments, bottlenecks often arise within the underlying relational database. Some organizations use this as a reason to migrate to a cloud or hybrid environment, while others, whether from necessity or preference, remain in an on-premises environment. In this paper, we have sought to explore a new type of hybrid model that will facilitate the vertical scaling of local RDBMS systems while avoiding the more cumbersome remedies of adding hardware or offloading logic into application layers. The design underpinning this analysis was constructed to allow examination of the feasibility and performance of utilizing cloud computational resources to augment the throughput of local relational database systems while avoiding the need for additional hardware and minimizing disruption of the existing code base.
While startups and personal endeavors are typically small and agile, it is the larger enterprises that struggle against inertia and must come to grips with the long tailed transitions that would come along with cloud adoption. Through this project our team will explore reasons to migrate to cloud while also try not to get caught up in the hype. While everyone sees the cents per service unit , are there actual savings in the long run? With this research we want to explore why's and ifs to be considered before moving conventional systems to cloud.

Why this Project?

The potential benefits from that would come along with the ability to scale a local database in a way that is flexible and ultra-low cost are obvious: low barrier to entry for small organizations, mitigation of security concerns around cloud storage, inherent cost savings.

Solution Specifics

Databases provide a prime use case for on-premises private and hybrid cloud models. For our research we will use on-prem relational database and explore hybrid models that can either re-use any of the on-prem resources for storage and compute with the flexibility to be scalable to cloud on need basis or look for a potential model that may be cost effective and provide cloud advantages while moving storage and compute entirely on cloud. To test this team will explore use of APIs and serverless compute services offered by cloud providers.

Introduction

In past few years there has been a huge chatter about cloud computing. Every Organization in this era is at least reviewing or looking at resources to see if moving to cloud would same them time and efforts. Cloud as we are aware makes provisioning of new resources quickly so an organization can concentrate their efforts on tasks that create more value for them rather than concentrating their resources on procuring hardware or provisioning servers. At the same time companies do not want to invest in hardware that is hardly utilized for about 3-4 hours in a day thus making a classic case for moving to cloud. In cloud you utilize resources for the time you need and terminate the instance when your work is complete and paying only for the time a resource is up and utilized.

Background

To achieve this we are planning to use the below Amazon resources like S3, DynamoDB, EC2 instances and Lambda function.

1 S3

S3 - stands for Simple Storage Service. This service could be utilized to collect, store and analyze huge amounts of data. Data stored in S3 could be retrieved from anywhere. It provides comprehensive security and compliance capabilities. S3 is designed to deliver 99.999999999% durability.

1.1 Advantages:

1. UNMATCHED DURABILITY, AVAILABILITY, & SCALABILITY
2. MOST COMPREHENSIVE SECURITY & COMPLIANCE CAPABILITIES
3. QUERY IN PLACE
4. FLEXIBLE MANAGEMENT
5. MOST SUPPORTED BY PARTNERS, VENDORS, & AWS SERVICES
6. EASY, FLEXIBLE DATA TRANSFER

2 Dynamo DB

Amazon DynamoDB is a nonrelational database that delivers reliable performance at any scale.

2.1 Advantages:

1. PERFORMANCE AT SCALE
2. FULLY MANAGED
3. ENTERPRISE-READY

3 Elastic Compute Cloud (EC2)

Amazon Elastic Comput Cloud is a web service that provides secure, resizable compute capacity in the cloud. EC2 has changed the economics of computing by allowing companies to pay only for the capacity that is being utilized.

3.1 Advantages:

1. ELASTIC WEB-SCALE COMPUTING
2. COMPLETELY CONTROLLED
3. FLEXIBLE CLOUD HOSTING SERVICES
4. INTEGRATED
5. RELIABLE
6. SECURE
7. INEXPENSIVE
8. EASY TO START

4 Elastic Compute Cloud (EC2)

AWS Lambda can run code without provisioning or managing servers. We have to pay only for the compute time consumed.

4.1 Advantages:

1. NO SERVERS TO MANAGE
2. CONTINUOUS SCALING
3. SUBSECOND METERING

Approach

To achieve this our approach would be:

1. Create S3 bucket
2. Create a table in Dynamo DB
3. Create EC2 instance or Machine Learning to analyze a csv file and predict based on this analysis
4. Lastly if time permits upload data from S3 to Dynamo DB

1 Problem Definition

How to utilize cloud for computing when the unit of work is for a few hours instead of spending money and resources on managing on-premise resources.

2 S3

3 Dynamo DB

Dynamo DB