Hybrid-Cloud Leveraging Cloud Computation Resources With On-premises

Mahesh Kuklani Southern Methodist University, Dallas, TX,USA e-mail: mkuklani@smu.edu
Tanvi Arora Southern Methodist University, Dallas, TX,USA e-mail: tanvia@smu.edu
Submitted To : Dr Sohail Rafiqi, Ph.D. in Computer Science and Engineering

***Abstract*—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. This paper explores the transition to cloud. The design underpinning this analysis allows 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 start-ups 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.**

I. Introduction

THIS paper goes through the why's and if's to be considered before moving conventional systems to cloud. Research uses 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.

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 model makes provisioning of new resources quick 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. Companies that do not want to invest in hardware that is hardly utilized for about 3-4 hours in a day make a classic case for moving to cloud. Cloud follows pay-per use model.Companies pay for the time resources are utilized. The potential benefits that come along with the ability to scale resources 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.

Businesses are often confused by the thought of moving to cloud. Do their concerns outweigh the advantages of moving to cloud [1]. The present paper focuses on hybrid cloud computing architecture and how it helps solve some of these concerns.

Our experimental evaluations are conducted on the three top cloud providers - Amazon AWS , Google Cloud and Microsoft Azure.To conduct this experiment with ease we have given the role of on-premise environment to one of the cloud providers and tested connectivity between resources available on different cloud .

A. Advantages of Cloud

[2]

1) Flexibility : Cloud solutions are ideal for businesses with the requirement of fluctuating bandwidth of resources. It is easy to scale up or scale down your resources on cloud. This level of agility gives businesses using cloud computing a real advantage over competitors.
2) Disaster recovery : This comes associated with cost. Larger companies with wide-scale IT investment have the cash and expertise to set up a robust disaster recovery. Cloud-based backup and recovery solutions save time, avoid large up-front investments allowing even small companies to invest in setting up a robust disaster recovery.
3) Capital-expenditure Free : Cloud computing cuts out the high cost of hardware. It is a subscription based model.Another point to note is the servers are off-premise, out of sight . Suppliers take care of them for you and roll out regular software updates - including security updates.

B. Business Concerns

1) Security : Cloud environments experience – at a high level – the same threats as traditional data center environments, the threat picture is the same. Both run softwares, softwares have vulnerabilities, and there is someone out there waiting to exploit these vulnerabilities. However security on cloud is a shared responsibility model of security. While cloud provider takes care of the security of the cloud , some aspects of security remain sole responsibility of the consumer. Effective cloud security depends on knowing and meeting these consumer responsibilities. [3]
2) Data Privacy : Cloud computing involves the dispersal of data across servers located anywhere in the world. Like globalization of networks. By crossing borders,

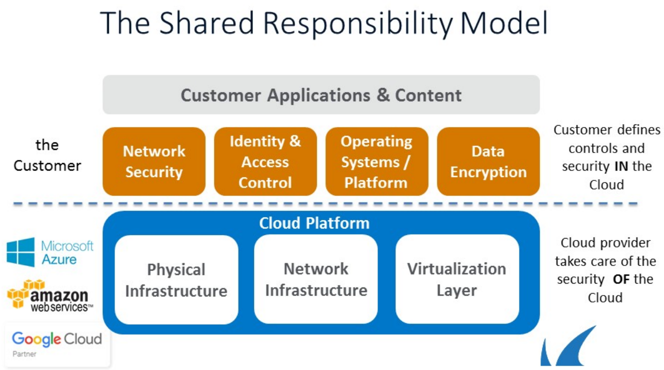

Fig. 1: Shared responsibility model

involves considering countries with restrictive privacy and protection laws . This is somewhat covered as part of customer responsibility towards cloud security. For this corporations need to understand what kind of data will they load into cloud and who will have access to this data. [4] Other aspect of data privacy is handling sensitive data. Yes one of the problems is that this technology is light years ahead of the law and there are questions that need to be answered. Who owns the data, consumer or the hosting cloud provider? Can a cloud deny the consumer access to their own data or can it share this data with marketing firms Obviously , the safest approach is data privacy is more a consumer responsibility. Keep data under proper control and apply data encryption methods. Regarding the question about laws, each company wants to protect their reputation. To get more clients and maintain them, cloud providers would uphold your data privacy. There are getting more managed and include all the necessary provisions that one should take while setting up their data on cloud. [5]

3) ROI : ROI or Return of Investment is widely a measure of financial success and can be a measured in a variety of ways. If you move to public cloud, you generally decrease investment but increase cost. With private cloud, it is vice-verse. But what matters is value to business, customer value, seller value, market brand value, corporate value. In case of cloud services, these relate to productivity, speed, size and quality. [6]
4) Integration Issues : Most enterprises would apply an incremental model of implementation. It is less risky than big-bang. This requires integration of services.The risk of not being able to integrate is critical. If you cannot build a system, you cannot use it. This also adds to the cost of including glue-software to connect various interfaces .It could involve rewrite of code or existing process models. Not to forget significant skills are required to assemble and customize multiple cloud services , requiring the applications to be loosely coupled, programmed to perform in an integration layer instead of underlying infrastructure.

II. Experiments

Cloud computing is a change and if used in the right way, can be a great accelerator of collaborated architecture. Cloud is an Internet phenomenon and trying to use cloud the traditional enterprise way will not achieve real returns from the cloud. In short Enterprises need to embrace new architecture. One approach is using an incremental approach. This is less risky than big-bang and also gives transition time to employees. This approach uses a hybrid cloud where existing on-prem resources are connected to new resources on cloud. New external cloud services can be incorporated in the in-house solutions, leading way to gradually get rid of on-prem resources by their end of life.

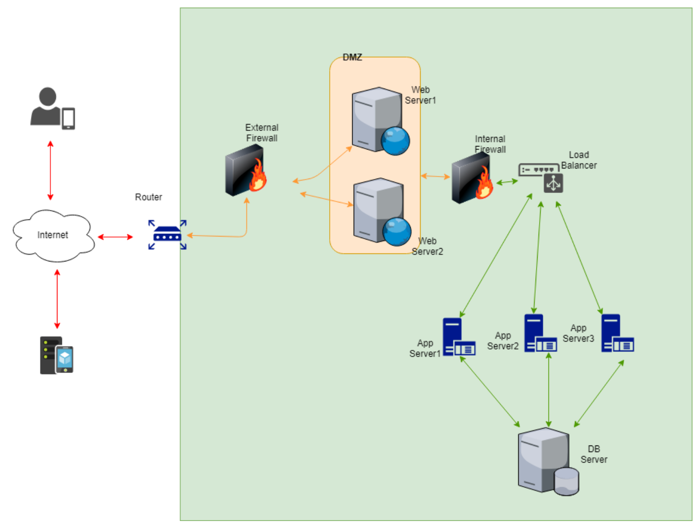

Fig. 2: Sample on-premise architecture

Consider a general on-premise architecture as shown in figure 2. The different hybrid models among others that can be built are :

1) Moving/extending storage and compute to cloud
2) Moving backup/archive to cloud
3) Moving business continuity or Disaster Recovery plan to cloud

Depending on the flexibility of the existing on-premise architecture, storage and compute can be moved to cloud entirely , so that its basic advantages of scalability and reliability can be leveraged they can be used as an extension. In order to reap the benefits of cloud, moving components to cloud would be advisable rather than extending them. Moving on to the experiments of using the hybrid model, database from one of the cloud providers has been utilized to replicate the on-premise resource due to the limitations of creating a VPN or direct-connect as these seem to be costly options for an individual.For practical purposes, any corporation or industry will be already hosting a VPN or direct-connect to reduce latency.

1) Azure Data Factory and SQL Database : As part of this experiment compute is moved to Azure Cloud by using Azure Data Factory. Azure Data Factory is a fully managed service provided by Microsoft that composes data storage, processing and movement services

into streamlined, scalable, and reliable data production pipelines. To utilize compute in Azure Cloud a table sales_records , is created in Azure SQL Database Cloud and an SSIS service in AzureDataFactory. Once SSIS service is deployed to Cloud it is scheduled to run daily which computes sales data for every quarter and sales for every product based on region.

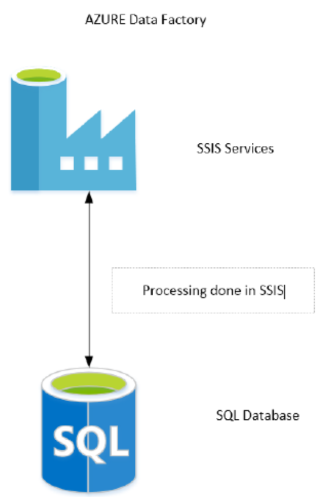

Fig. 3: SSIS Azure Data Factory

a) Create SQL Database.
   i) Login to Azure portal and click create SQL Database
   ii) Create Cloud Computing database with a blank database
   iii) Once DB is created, get the connection string for this database.
b) Create Azure Data Factory.
   i) Login to Azure portal and click Data + Analytics and click Data Factory
   ii) Create unique name for SSIS Data Factory
   iii) Click Author and Monitor.
   iv) Click Configure SSIS Integration Runtime tile.
   v) On the SQL Settings page enter the configuration of the above SQL Database.
   vi) Select Catalog Database Server Endpoint to host SSISDB.

2) S3 –> $Lambda$ –> $DynamoDB$: As part of this experiment a csv file is uploaded to S3. AWS Lambda has a trigger whenever a new item is added to S3. Lambda function kicks in, it does its processing and uploads the data in csv file in S3 to DynamoDB. DynamoDB can process this data for end user or any analytic requirement. AWS Lambda is a server less architecture which utilizes resources required for processing the file in S3 to DynamoDB. Once the processing is complete the client is not charged unless the Lambda function is triggered again. To achieve this we followed the following steps: [7]
   a) Create Policy in IAM.

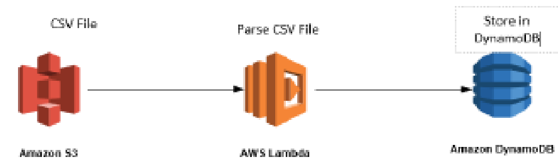

Fig. 4: AWS Lambda S3 to DynamoDB

   i) Select S3, then select all actions and all resources
   ii) Add additional permissions for DynamoDB, select all actions and all resources
   iii) Add additional permissions for Cloudwatch, select all actions and all resources.
b) Create Role
   i) Attach the above policy to this role
c) Create Lambda Function
   i) Create function, author from scratch
   ii) Give function name
   iii) Select runtime as Python 3.0
   iv) Choose existing role
   v) Select role created above
   vi) Add trigger for S3, select bucket, event type (object created), prefix and filter if any.

3) S3 –> $Lambda$ –> $AzureSQLDatabase$: As part of this experiment connectivity is tested between Amazon and Azure by invoking a lambda function to unload data from Azure database. AWS Lambda has a trigger whenever a new item is added to S3. Lambda function kicks in, it does its processing and performs query on Azure database. This is another solution of using serverless compute in a hybrid model. Once the processing is complete the client is not charged unless the Lambda function is triggered again. To achieve this we followed the following steps: [7]

Fig. 5: Lambda S3 to Azure

a) Create Lambda Function
   i) Create function, author from scratch
   ii) Give function name
   iii) Select runtime as Python 3.0
   iv) Choose existing role
   v) Select role created above

vi) Add trigger for S3, select bucket, event type (object created), prefix and filter if any.

4) AWS/Google Cloud –> $AzureSQLDatabase$: As part of this experiment, connectivity is tested between Amazon EC2 instance used for compute with SQL Server on Azure.

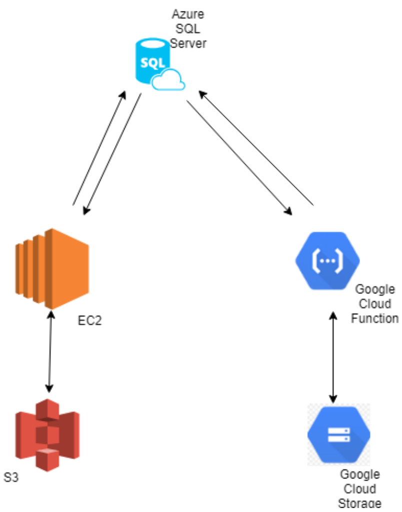

Fig. 6: AWS to Azure and Google to Azure

a) Connect to Azure SQL Server
   i) Create SQL Database. Steps remain same as discussed during Experiment (1)
b) Connecting from AWS : Includes setting up EC2 instance with ability to connect to S3 and Azure.
   i) Create user/IAM Roles to allow EC2 to connect to S3.Required role is "AmazonS3FullAccess" to the user.
   ii) Create EC2 instance with default configurations
   iii) login to EC2 instance with ec2-user and configure IAM role on EC2 to connect to S3 using aws configure. Use the access key ID and secret Access Key generated while creating IAM Roles. This is a one-time set up on EC2, even after a stop and restart the role is valid and available.
   iv) connection to s3 from EC2 can be now tested using aws s3 ls commands
   v) Install sqlcmd on AWS EC2 and update firewall settings on Azure Sql server db to enable connections from EC2 instance
   vi) EC2 is ready to connect to SQl Server on Azure
c) Connecting from Google : Includes setting up python to connect to Azure from cloud shell. Google provides a shell environment for managing resources hosted on Google Cloud Platform. This allows us to test the connectivity without creating a Compute Engine.
   i) Setting up Google cloud shell environment, requires installing python with some of its additional packages like pip,pyodbc and ODBC drivers including unixODBC and MSsql odbc
   ii) Update firewall settings on Azure SQL Server DB to enable connections from Google shell.
   iii) connect using python pyodbc.connect

III. Conclusions and Future Work

In this paper we have tried different connectivity options to connect 1 cloud environment to a database in another cloud. We considered simple approaches , single connect options with comparatively less amount of data. Same options can be used in a scaled resource on cloud without much ado.

While the three cloud environments considered have almost similar type of features or resources available, in order to reap the benefits of cloud, a use-cased based architecture should be considered. Working on cloud does require added skills as segregation of duties is lessened. Needless to say, cloud does not eliminate the need of your platform team. Security of cloud is still a shared responsibility.

In future work we will experiment on creating VPC tunnel between the different cloud environments. This will be close to real application of a hybrid cloud. Hybrid cloud allows collaboration of multiple cloud services as per requirement . Another feature that we will like to study is the use of multi-cloud structure for backup and recovery.

Appendix A
Cloud Resources

A. 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. Some of the other features/advantages are [8]

1) Unmatched Durability, Availability & Scability
2) Most comprehensive security & compliance capabilities
3) Query in place
4) Flexible management
5) Most supported by partners, vendors & AWS serices
6) Easy, Flexible data transfer

B. Google Cloud Storage

Google Cloud Storage is a RESTful online file storage web service for storing and accessing data on Google Cloud Platform infrastructure. It is comparable to Amazon S3 . Contrary to Google Drive, this is more suitable for enterprises.Some of the other features/advantages are [9]

1) designed for 99.999999999% durability. It stores data redundantly with automatic checksums to ensure data integrity.
2) highly scalable . Objects larger than 5MB should be uploaded with multipart or resumable uploading.Google Storage provides a resumable data transfer feature that allows user to resume upload operations after a communication failure has interrupted the flow of data.
3) Google Storage is interoperable with other cloud storage tools and libraries

4) Upload operations to Google Storage are atomic , providing strong read-after-write consistency for all upload operations

C. Dynamo DB

Amazon DynamoDB is a non-relational database that delivers reliable performance at any scale. Some of the other features/advantages are : [10]

1) Performance at scale
2) Fully managed
3) Enterprise-ready

D. Elastic Compute Cloud (EC2)

Amazon Elastic Comput Cloud is a web service that provides secure, re sizable 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.Some of the other features/advantages are [11]

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

E. AWS Lambda

AWS Lambda can run code without provisioning or managing servers. We have to pay only for the compute time consumed.Some of the other features/advantages are [12]

1) No servers to manage
2) Continuous scaling
3) Subsecond metering

F. Azure SQL Database

This is a relational database-as-a-service (DBaaS) with the latest version of Microsoft SQL Server Database Engine. It is high performance, reliable and secure database on which data-driven applications and websites can be built in the programming language of choice without needing to manage infrastructure. Some of the other features/advantages are [13]

1) Fully managed
2) Advanced security
3) Built-in intelligence

Acknowledgment

The authors would like to thank Dr Sohail Rafiqi for giving us the opportunity to work on this paper and showing us a liberal approach towards cloud . There is no right answer and there is no right model.

References

[1] [Online]. Available: https://www.marutitech.com/5-reasons-why-cloud-can-transform-your-business/
[2] [Online]. Available: https://www.salesforce.com/uk/blog/2015/11/why-move-to-the-cloud-10-benefits-of-cloud-computing.html
[3] [Online]. Available: https://insights.sei.cmu.edu/sei_blog/2018/03/12-risks-threats-vulnerabilities-in-moving-to-the-cloud.html
[4] [Online]. Available: https://legal.thomsonreuters.com/en/insights/miscs/understanding-data-privacy-and-cloud-computing
[5] [Online]. Available: https://www.privacyrights.org/blog/privacy-implications-cloud-computing
[6] [Online]. Available: http://www.opengroup.org/cloud/cloud_for_business/p6.htm
[7] [Online]. Available: https://www.youtube.com/watch?v=5nuUmlezrHs
[8] [Online]. Available: https://aws.amazon.com/s3/
[9] [Online]. Available: https://cloud.google.com/shell/docs/
[10] [Online]. Available: https://aws.amazon.com/dynamodb/
[11] [Online]. Available: https://aws.amazon.com/ec2/
[12] [Online]. Available: https://aws.amazon.com/lambda/
[13] [Online]. Available: https://docs.microsoft.com/en-us/azure/sql-database/

Mahesh Kuklani Mahesh Kuklani received his bachelor's degree in Mechanical Engineering from University of Pune, India. He is currently working towards his Master's degree in Data Science at Southern Methodist University, Dallas , TX. He works as a Database Administrator at ECFMG Philadelphia PA and has over 20 years of experience working in this field.

Tanvi Arora Tanvi Arora received the bachelor's degree in Computer Science Engineering from University of Rajasthan, India . She is currently working towards her Master's degree in Data Science at Southern Methodist University, Dallas , Texas.She works full time as a Senior Software Developer in USA and has 11 years of work experience in this field .