EAI Endorsed Transactions
on Cloud Systems

Research Article EAI.EU

Database Migration on Premises to AWS RDS

Lakshmi Narasimhan G

M.E CSE II YEAR, Department of Computer Science and Engineering, Sri Ramanujar Engineering College, Kolapakkam, Chennai 600127, (g_l_narasimhan@hotmail.com)

Abstract

For the past four decades, the traditional relational databases have been in use in Information Technology industry. There was a phenomenal conversion in the IT industry in terms of commercial applications in the previous years. The applications that were running on a Single server in organizations IT infrastructure have been replaced or migrated with e-apps. Also, the dedicated storages are replaced with system storages. The model of pay per use, flexibility and lesser cost are the main reasons, which caused the distributed computing pick up into reality. Cloud databases, Simple DB, and Amazon RDS are getting to be more familiar to communities because they have brought up and highlighted the issues and problems of current social databases in terms of usability, flexibility, and provisioning. Basically now, the cloud databases are at present considered as a solution for programmers, designers, and architects since they need to store the information of their applications in an adaptable and exceptionally accessible from backend when required. These Database-as-a-Service (DBaaS) administrations are cloud-based information stockpiling administrations can be arranged into two principle classifications: benefits that backing conventional social databases (RDB) (e.g., Amazon RDS, Google SQL, Microsoft Azure), and key/quality pair information stockpiling administrations (e.g., Amazon Simple DB, Google Data Store), which are otherwise called NoSQL Databases [Harrison John Bhatti and Babak Bashari Rad 2017]. In this paper, we are going to analyze and perform one such On-Premi~~ses to AWS RDS To support, Cloud migration~~ which helps the users on performance, cost, and scalability

Keywords: Data migration; AWS RDS, cloud database, information accountability, database activity monitoring

Received on 08 March 2018, accepted on 02 April 2018, published on 11 April 2018

doi: 10.4108/eai.11-4-2018.154463

1. Introduction

IT (Information Technology) department is accountable and responsible for delivering the low-cost storage and network facilities and reliable computing to its users. Little and small organizations suffer from high investments in IT infrastructure facilities. Alternatively, the companies require IT solutions that could reduce their TCO in procuring and maintaining IT hardware and software to avail maximum benefits of IT services. According to a recent study [Indu Arora and Anu Gupta 2012], such organizations/institutions cloud computing (CC) becomes an appropriate and natural choice because CC delivers highly scalable infrastructure with minimal setup cost and maintenance cost comes along with the features like grid computing, server consolidation & virtualization, faster storage, etc. Using “pay-per-use" model, CC provides on-demand IT-related services anytime anywhere using Infrastructure-as-a-Service(IaaS), Platforms-as-a-Service(PaaS) and Software-as-a-Service(SaaS) over the internet.

2. Overview

A fundamental change in the concepts and experimental practices have been brought due to adoption or migration to Cloud computing in both technology and database space. Provisioning of Database-as-a-Service has been raised with demand and more usage of Cloud computing.
The changes happening in data storage requirements from different customers and business, the phenomenal growth in digital data and high bandwidth internet connectivity with cloud computing enhanced the emergence of Cloud Databases. To support the different types of users on their requirements Data management in the Cloud is offered as Database as a service (DBaaS) and Data as a service (DaaS) etc. The offerings vary from one to other based on

how data is stored and managed. Cloud storage benefits users to store objects, formats, and documents in a virtual storage. The customer data has been stored in a remote system available and connected via the Internet in the DaaS.

DBaaS provides the comprehensive database functionality which allows users to store and access their database at remote place anytime anywhere. The well-known cloud databases are Amazon RDS, Microsoft's SQL Azure Database and Google's BigTable etc, [Indu Arora and Anu Gupta 2012].

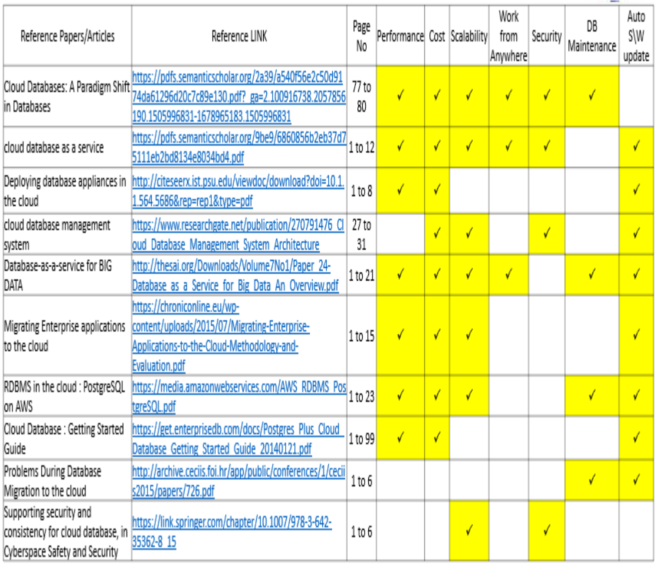
2. SYSTEM ANALYSIS

2.1. Review Stage

Some popular databases used in cloud computing have been reviewed, including advantages and disadvantages of local database systems, and the most important challenges in the development of cloud databases have been discussed. A list of references is given as below, we have reviewed the literature and analyzed the parameters which are governing the On-premises databases issues which could eventually addressed while migrating to the cloud. Find below the table which contains the different reference articles taken into consideration as listed the parent "Databases in Cloud Computing: A Literature Review" for our analysis.

The following are the most important parameters we considered:

- Performance
- Cost
- Scalability
- Work from anywhere
- Security
- DB Maintenance
- Automatic S/W Upgrade



Reference Papers/Articles	Reference LINK	Page No	Performance	Cost	Scalability	Work from Anywhere	Security	DB Maintenance	Auto S\W update
Cloud Databases: A Paradigm Shift in Databases	https://pdfs.semanticscholar.org/2a39/a540f56e2c50d9174da61296d20c7c89e130.pdf?_ga=2.100916738.2057856190.1505996831-1678965183.1505996831	77 to 80	✓	✓	✓	✓	✓	✓	
cloud database as a service	https://pdfs.semanticscholar.org/9be9/6860856b2eb37d75111eb2bd8134e8034bd4.pdf	1 to 12	✓	✓	✓	✓	✓		✓
Deploying database appliances in the cloud	http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.564.5686&rep=rep1&type=pdf	1 to 8	✓	✓					✓
cloud database management system	https://www.researchgate.net/publication/270791476_Cloud_Database_Management_System_Architecture	27 to 31		✓	✓		✓		✓
Database-as-a-service for BIG DATA	http://thesai.org/Downloads/Volume7No1/Paper_24-Database_as_a_Service_for_Big_Data_An_Overview.pdf	1 to 21	✓	✓	✓	✓		✓	✓
Migrating Enterprise applications to the cloud	https://chroniconline.eu/wp-content/uploads/2015/07/Migrating-Enterprise-Applications-to-the-Cloud-Methodology-and-Evaluation.pdf	1 to 15	✓	✓	✓				✓
RDBMS in the cloud : PostgreSQL on AWS	https://media.amazonwebservices.com/AWS_RDBMS_PostgreSQL.pdf	1 to 23	✓	✓	✓			✓	✓
Cloud Database : Getting Started Guide	https://get.enterprisedb.com/docs/Postgres_Plus_Cloud_Database_Getting_Started_Guide_20140121.pdf	1 to 99	✓	✓					✓
Problems During Database Migration to the cloud	http://archive.ceciis.foi.hr/app/public/conferences/1/ceciis2015/papers/726.pdf	1 to 6						✓	✓
Supporting security and consistency for cloud database, in Cyberspace Safety and Security	https://link.springer.com/chapter/10.1007/978-3-642-35362-8_15	1 to 6			✓		✓		

Figure 1. Snapshot of system analysis

2.2. Real Issues

The following are the summary of real issues identified during our literature survey analysis as listed in the table above.

On-Premises Database Issues and Constraints:

- Currently, the Databases are hosted on virtual servers in On-Premises which incurs a lot of running cost on Infrastructure like licensing, EOL, SW/ HW upgrade etc.
- Frequent performance issues encountered due to high CPU, memory, I\O latency, Blocker process, Long Running Queries & Deadlocks
- The issue on storage capacity due to data growth and a lot of time spent on capacity planning.
- When the Hardware reaches End of Life, the huge capital cost is required to buy a new server.
- The effort required to manage the hardware, operating systems are high as these are hosted on Premises.

2.3. The Problem Statement

From the summary of On-Premises Database Issues and Constraints, now we have framed the simplified Problem Statement as below.

On-Premises Databases suffers from

- Very high running cost
- Frequent performance issues
- Storage capacity issues and
- Database management issues

3. SYSTEM DESIGN

Based on the above System Analysis (Fig. 1) and the Problem Statement we concluded that most of the Performance, Cost, Scalability and Security issues & constraints can be addressed by migrating the On-premises DBs to Cloud. And we preferred to AWS RDS as AWS is vendor neutral and agnostic

3.1. Proposed Algorithm/Technique – AWS Cloud RDS

To resolve the above-said issues, the proposal is to move the databases to Amazon Cloud RDS (Relations Database service) which is a Managed service offered by AWS.

- A Cloud Database Management Systems (CDBMS) is hosted in cloud environment which we can access it over the internet from anywhere.
- On-Premise database systems are installed on a server and data is accessed over a local area network(LAN)
- AWS RDS is Easy to set up, operate and scalable in the cloud.
- It provides cost-effective and resizable capacity while automating the administration tasks like, scheduling backups, apply patching, software & hardware provisioning etc.
- Applications will run on fast performance, high availability, security, and compatibility.
- No Capital cost is needed as AWS RDS uses pay as you go model. There are many reports in AWS to keep a check on running costs and reduce the running costs.

3.2. Amazon RDS

Amazon RDS (Fig. 2) allows the user to create and operate a relational database in the cloud which is a cost-effective and easy to manage solutions compared to On-Prem databases and also helps us to reduce the manual intervention on the database tasks. It offers you different types of RDBMS like Oracle, Microsoft SQL Server, PostgreSQL, MySQL, and MariaDB.

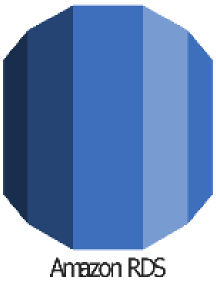
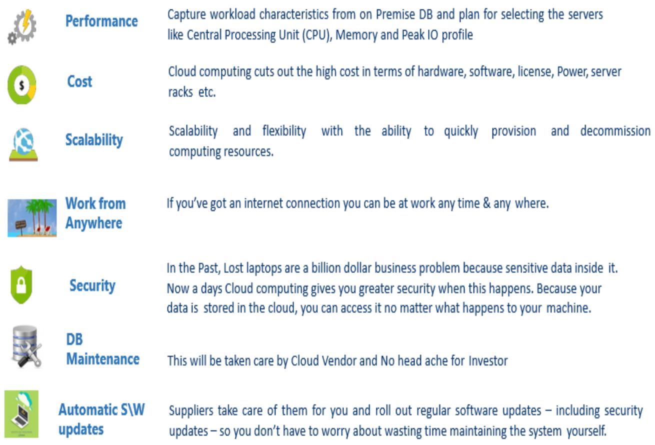


Figure 2. Amazon RDS

- **Highly Scalable**: Scaling up the database's computer and storage resources with a few mouse clicks or through API call and doesn’t require downtime. Read Replicas are used to offload the read traffic from the primary database instance.
- **Fast**: Faster access to relational databases enabled using AWS RDS Command-Line Interface, or simple API calls in minutes. No need for infrastructure provisioning, and no need for installing and maintaining database software.
- **Available and Durable**: Amazon RDS synchronously replicates the data in a different Availability Zone to a standby instance. Amazon RDS has features which include automated backups, a database snapshot, and auto host replacement.
- **Easy to Administer**: Amazon RDS can be managed easily from the AWS Management Console or the RDS CLI or through simple API calls to access the capabilities of a production-ready relational database in minutes. No need to provision any infrastructure and database software.
- **Inexpensive**: On-Demand pricing using Reserved Instance pricing helps users with no up-front or long-term cost commitments.
- **Secure**: Amazon RDS makes it easy to control network access to your database. Amazon RDS database instances secured using Amazon VPC, which enables you to connect to your existing IT infrastructure using encrypted IPsec VPN. Many Amazon RDS engine types offer “in transit” and “at rest” encryption capabilities

3.3. Advantages of Algorithm/Technique – AWS Cloud RDS



Performance	Capture workload characteristics from on Premise DB and plan for selecting the servers like Central Processing Unit (CPU), Memory and Peak IO profile
Cost	Cloud computing cuts out the high cost in terms of hardware, software, license, Power, server racks etc.
Scalability	Scalability and flexibility with the ability to quickly provision and decommission computing resources.
Work from Anywhere	If you’ve got an internet connection you can be at work any time & any where.
Security	In the Past, Lost laptops are a billion dollar business problem because sensitive data inside it. Now a days Cloud computing gives you greater security when this happens. Because your data is stored in the cloud, you can access it no matter what happens to your machine.
DB Maintenance	This will be taken care by Cloud Vendor and No head ache for Investor
Automatic S\W updates	Suppliers take care of them for you and roll out regular software updates – including security updates – so you don’t have to worry about wasting time maintaining the system yourself.

Figure 3. Advantages of cloud RDS algorithm/technique

3.4. Design Diagram

Find the design diagrams below (Fig. 4., Fig. 5, Fig. 6.) on preferred AWS RDS to take care of the Performance, Cost, Scalability.

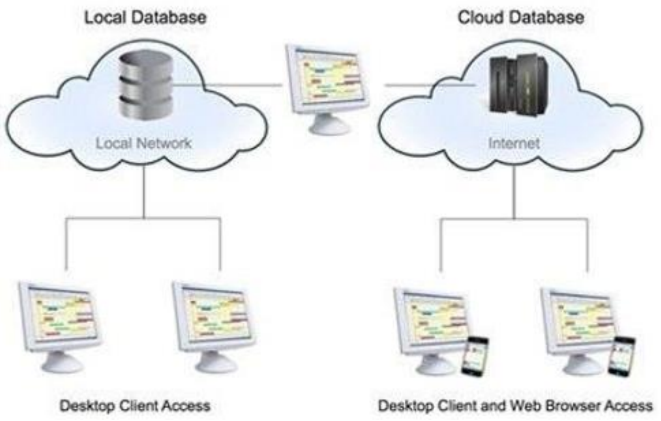


Figure 4. Cloud Database Migration Design Block Diagram

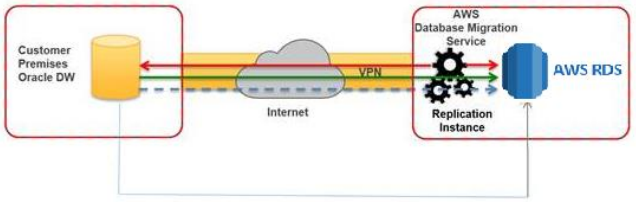


Figure 5. Cloud Database Migration Design Diagram – Schematic

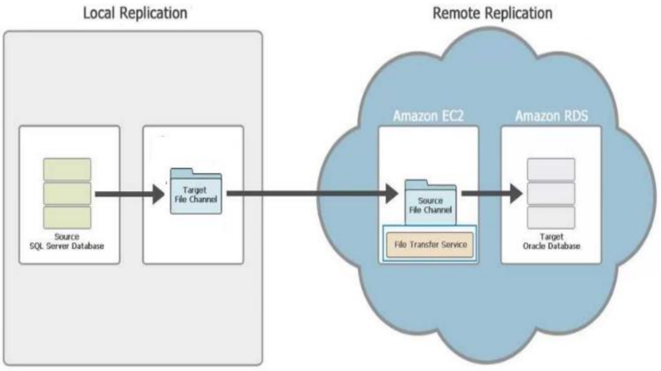


Figure 6. Cloud Database Replication Diagram

3.5. Disaster Recovery

Almost every organization in the world has realized the tolerance for downtime should be minimized and decreased day by day. They prepare the Business Continuity and Disaster recovery plan with the objectives of improving high availability.
The purpose of Disaster Recovery(DR) services is to prepare the organizations in the event of extended service outages caused by factors beyond our control (e.g., natural disasters, man-made events), and to restore services to the widest extent possible in a minimum time frame. A DR plan emphasizes the structured approach for responding to eventualities that threaten the organizations' IT infrastructure comprising of hardware and software, networks and people.

The important measurements in disaster recovery services for ensuring business continuity are
Recovery Point Objective and Recovery Time Objective. (Fig. 7)

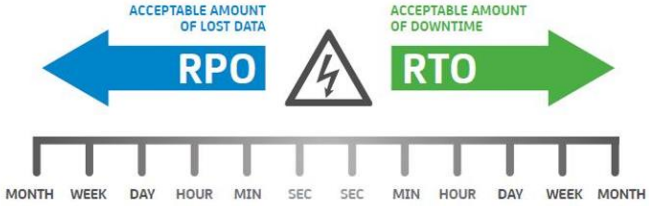


Figure 7. Disaster Recovery Services

RPO (Recovery Point Objective) is the age of lost data that an organization must recover from backup storage to perform its normal operations to resume when a disaster strikes. RPO is about "From When". Thus, the RPO estimate sand arrives at the minimum frequency of backups like in hours, daily, weekly, monthly and real time. For instance, if a company has eight hours RPO, the system must back up every eight hours.

RTO (Recovery Time Objective) is about how much time following a disaster, an organization tolerates or needed to resume normal operations. So, RTO is the maximum amount of downtime an organization can afford to have. If an organization has an RTO of 30 minutes, the systems should be up and running in 30 minutes and should not prolong beyond that. The right determination of RPO and RTO help administrators choose the appropriate optimal disaster recovery strategies, technologies, and procedures.

Cloud Disaster Recovery as a service.
Disaster recovery as a service (DRaaS) is a cloud-based DR that has got the popularity in the current period as it provides the following benefits.

- Lower cost by running on a shared infrastructure
- Faster and easy deployment so that the ability to test plans regularly and testes can be carried out by spinning up temporary instances simply

- Scalability that supports an increase in the bandwidth requirements etc.

4. CLOUD RDS IMPLEMENTATION

This migration from On-Premises to AWS RDS Cloud will be carried out in multiple phases and modules. There are Modules from Module 0 to Module 5. Find below the Modules with their proposed set of activities.

- Requirement Gathering (HARDWARE & SOFTWARE Requirement) (Fig. 8.)
- Configuration & Setup
- RDBMS Implementation:
- MIGRATING DATA to Cloud

4.1. Requirement Gathering

Requirements gathering is an essential part of any implementations and migration. Understanding fully what a solution implementation will deliver is critical to its success. Requirements gathering sounds like common sense, but surprisingly, it's an area that is given far too little attention.

To be successful in requirements gathering and to have increased likelihood of success, we follow few of these rules:

- Type of Application (Banking, public sector etc.)
- Application criticality
- Application Transactions flow rate by Quarterly(approx.)
- Disaster Recovery setup

Cloud selection (AWS, MS Azure, Google)

- Financial health.
- Organization, governance, planning and risk management
- Trust.
- Business knowledge and technical know-how.
- Compliance audit.

Administration support:

- Service Level Agreements (SLAs).
- Performance reporting, Billing, and accounting.
- Resource monitoring and configuration management.

Security practices

- Security infrastructure and Security policies.
- Identity management
- Data backup and retention.
- Physical security
- Based on the parameters/criteria, we have arrived at the various Hardware and Software requirements

SOFTWARE/HARDWARE Requirements:

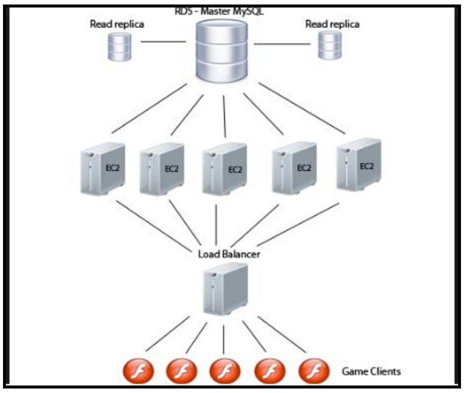
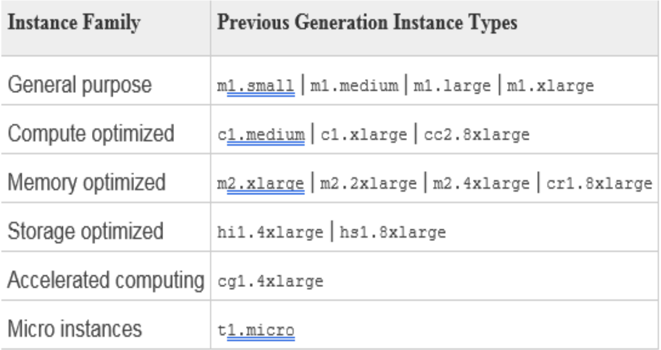
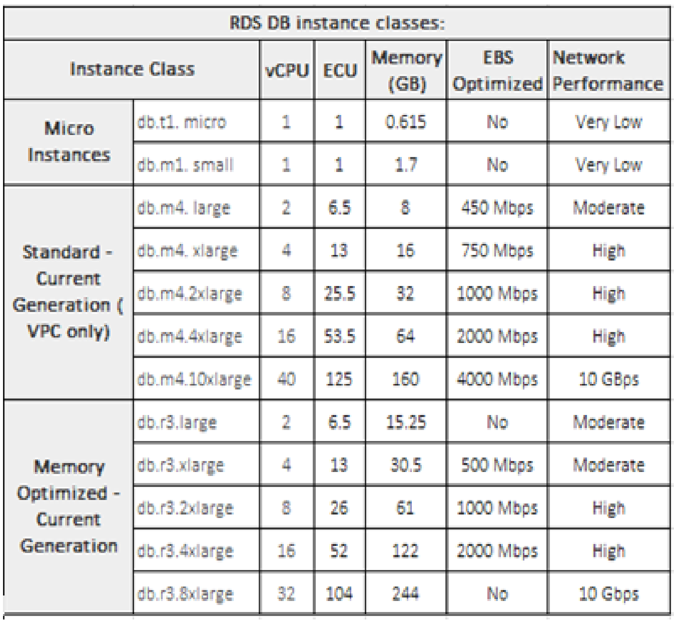


Figure 8. SYSTEM Requirements



Instance Family	Previous Generation Instance Types
General purpose	m1.small | m1.medium | m1.large | m1.xlarge
Compute optimized	c1.medium | c1.xlarge | cc2.8xlarge
Memory optimized	m2.xlarge | m2.2xlarge | m2.4xlarge | cr1.8xlarge
Storage optimized	hi1.4xlarge | hs1.8xlarge
Accelerated computing	cg1.4xlarge
Micro instances	t1.micro

Figure 9. Instance Types



RDS DB instance classes:						
Instance Class		vCPU	ECU	Memory (GB)	EBS Optimized	Network Performance
Micro Instances	db.t1. micro	1	1	0.615	No	Very Low
	db.m1. small	1	1	1.7	No	Very Low
Standard - Current Generation (VPC only)	db.m4. large	2	6.5	8	450 Mbps	Moderate
	db.m4. xlarge	4	13	16	750 Mbps	High
	db.m4.2xlarge	8	25.5	32	1000 Mbps	High
	db.m4.4xlarge	16	53.5	64	2000 Mbps	High
	db.m4.10xlarge	40	125	160	4000 Mbps	10 GBps
Memory Optimized - Current Generation	db.r3.large	2	6.5	15.25	No	Moderate
	db.r3.xlarge	4	13	30.5	500 Mbps	Moderate
	db.r3.2xlarge	8	26	61	1000 Mbps	High
	db.r3.4xlarge	16	52	122	2000 Mbps	High
	db.r3.8xlarge	32	104	244	No	10 Gbps

Figure 10. RDS DB Instance Classes

HARDWARE Requirement:

- Network bandwidth
- Number of CPU's
- Size of Memory
- Storage Type & Capacity

- General IOPS
- Provisioned IOPS & Magnetic

SOFTWARE Requirements:

- Types of OS Version (HP UNIX, CentOS, Ubuntu, RHEL Sun Solaris WINDOWS)
- Types of DBMS (MS SQL, ORACLE or MySQL etc)

4.2. CLOUD Configuration & Setup

- Create Active Directory
- Create VPC
- Create IAM role
- Create S3 bucket storage

4.3. RDBMS Implementation

This section will demonstrate how Amazon RDS is used to establish a connection to the DB instances (Fig. 9, Fig. 10). and same also applies to other RDBMS RDS instances too.

- Instance creation
- CSSMS to connect RDS Instance

4.3. Migrating Data to Cloud

We have four methods to migrate the data from on-premises to RDS or EC2 Instances.

- Using NATIVE BACKUP and RESTORE
- Using Microsoft Azure tool
- Using AWS Schema Conversion Tool
- Using NATIVE Import/Export Wizard

Native Backup and Restore Method. There were no options to use native backup and restore methods in RDS before AWS introduced this option in July 2016. This was a major issue faced by users while migrating the data from their On-Prem to AWS cloud (Fig. 11.).

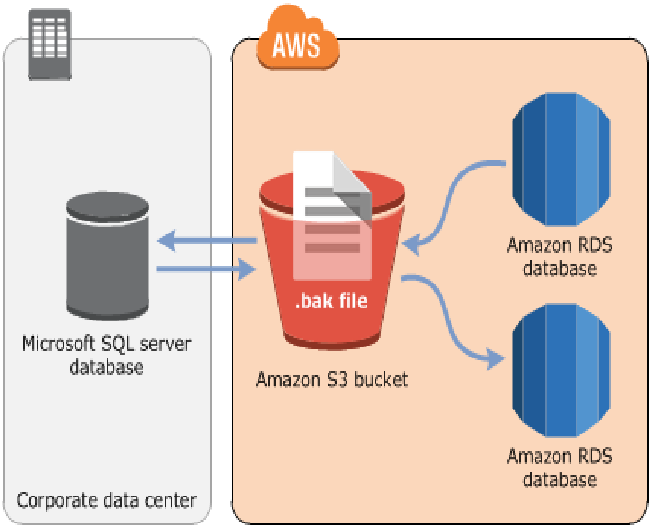


Figure 11. Backup & Restore Method

Backup databases
The following three commands should be executed to back up the databases.

```
(i)   exec msdb.dbo.rds_backup_database;
(ii)  @source_db_name='database_to_be_backed_up',;
(iii) @s3_arn_to_backup_to='arn:aws:s3:::S3_bucketfold
      er/backup_name.bak',@overwrite_S3_backup_file=;
```

Using Microsoft Azure tool
To minimize the time and reduce errors, SQL AZURE migration tool is used to migrate the data from On-Prem to AWS cloud. Please follow the below steps to migrate the data from SQL Server data to AWS RDS SQL DB.

- Download SQLAzureMW Tool
- Start Migration
- Start the migration by download SQLAzureMW Tool
- Source database tasks and Destination database tasks
- Data Verification

Schema Conversion Tool.
The schema conversion tool is very useful for heterogeneous database migrations. Following are the steps for Schema Conversion tool migration:

- Connecting to the Source Database
- Connecting to the Destination Database
- Running the Database Migration
- Assessment Report
- Converting the Source Schema and apply
- Data Verification

Using Import and Exports.
Following are the steps for Native Import/Export process:

- Connecting to the Source and Destination Database
- Data Verification

5. IMPLEMENTATION/MIGRATION BENEFITS

While migrating from On-Prem servers to AWS Premise, we achieved the following parameters (Fig. 12.) in AWS Premise

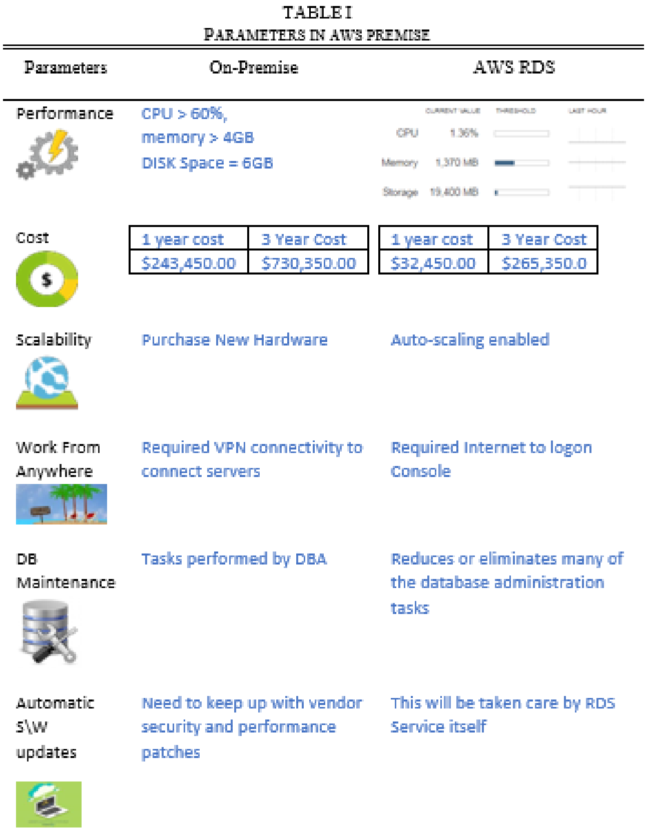


TABLE I
PARAMETERS IN AWS PREMISE

Parameters	On-Premise	AWS RDS
Performance	CPU > 60%, memory > 4GB DISK Space = 6GB	CURRENT VALUE THRESHOLD LAST HOUR CPU 1.36% Memory 1,370 MB Storage 19,400 MB
Cost	1 year cost: $243,450.00 3 Year Cost: $730,350.00	1 year cost: $32,450.00 3 Year Cost: $265,350.0
Scalability	Purchase New Hardware	Auto-scaling enabled
Work From Anywhere	Required VPN connectivity to connect servers	Required Internet to logon Console
DB Maintenance	Tasks performed by DBA	Reduces or eliminates many of the database administration tasks
Automatic S\W updates	Need to keep up with vendor security and performance patches	This will be taken care by RDS Service itself

Figure 12. AWS Premise Parameters

5. CONCLUSION

As we know every organizations, companies, customers and end users cannot spend money to construct data centers to manage their respective databases. Also, we know very high and large no. of online and real-time transaction generate a huge volume of data. For which cloud databases are the solution for managing the data flows in/out and stored. Thus, the emerging new cloud databases are creating a new era in the database world.

Acknowledgment

I am indebted to my Head of the Department of Computer Science and Engineering and Guide cum Supervisor Dr. S. Anithaa, M.E., Ph.D for her excellent support and extended guidance which made me in completing this on time.

References

[1] Harrison John Bhatti and Babak Bashari Rad. Databases in Cloud Computing: A Literature Review. In *I.J. Information Technology and Computer Science*. 2017. 4: 9.17 p. 1.

[2] Arora I. and A. Gupta, Cloud databases: a paradigm shift in databases. International J. of Computer Science Issues, 2012. 9(4): p. 1.