ISSN:2319-2607 IISUniv.J.Sc.Tech. Vol.9(1), 29-32 (2020)
REVIEW ARTICLE

Determining the Scope of Cloud Database Platforms For Database Selection Process

Abhijeet Singh Bais, Navneet Sharma*
Department of CS & IT, IIS (deemed to be University), Jaipur

Abstract
The analysis on database system features is to establish a pattern of database systems future. These systems are enhancing in multiple dimensions, including storage distribution, query processing, faster performance and increasing type of data persistence. This study will also try to establish pattern on close integration of intelligence in modern database systems. This Analysis is an attempt to provide a point of view on latest database transformations from database systems to database Platforms.

Keywords: Advance Databases, Artificial intelligence in database, Cloud databases, Database Platforms, Database systems, Intelligent database, Modern Databases

Introduction

Actually, there are quite a hundreds of databases are available which supports various business activities. Database models are available from file-based to NoSQL (Britts, 2005). Each model fulfils some goals of business. New needs have arisen due to more data storage, better performance efficiency. Traditional relational database models cannot satisfy many of these goals of handling structured and unstructured data in an efficient and cost-effective manner. (Berg *et al.*, 2012). So, it is checking out solution in the form of database platforms.

Literature Review

According to the view point of Ellison *et al.*, 2018 selecting a cloud provider and various service option needs the estimation of cost and planning. Two stages discussed in this research paper. In the first stageworkload and structure models of the database migrated from database logs to schema, in the second stage using simulation(discrete-event) models the cost and duration estimated. They implemented software tools for both the approaches.

Siddiqui *et al.*, 2020 in their study they investigated about accurate cost models for data systems and is it possible to integrate the learned models within the query optimizer. They exploited workload pattern, Cascade framework and learned cost models within the query optimizer of SCOPE at Microsoft. Their result showed the accuracy of learned cost models.

Wu *et al.*, 2011 proposed a query optimization scheme. Hive a query optimizer is designed and query plan is proposed in this paper. To check the effectiveness of query optimizer, inhouse experiments are carried out.

Sakr, 2014 discussed about hosting the database in cloud environments, its strengths and weaknesses and suggested some opportunities.

Based on above mentioned review paper various dominant cloud databases features are determined for database selection process.

Understanding Databases on CLOUD

TRADITIONALLY software, databases and applications used to install on hardware owned by organizations. Industries are migrating on cloud-based platforms where applications and databases are installed, managed, and operated over cloud (Shehri, 2013). According to some survey almost 36% of the companies are using cloudservices.

*Corresponding Author Email: navneet.sharma@iisuniv.ac.in

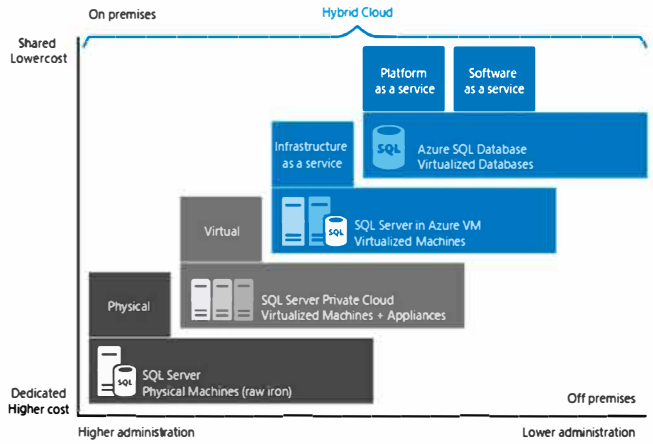

Fig. 1. Database Platform

(Source: Magenium Solutions)

Cloud enables platform as service multiple context it is considers as value-based service because it provides flexibility, higher availability, improved time to as multi dimension scaling on compute power.

Fig.1 compares on premises and off promises database usage. Database storage size can be expanded on run time. Persistence needs will determine the size of the scalable database footprints which is unlikely to non-cloud database environments where provisioning additional size to the databases had been a typical project for technicians.

Adjustable compute power is a big plus added to complex data processing systems in the cloud environments. Pay per use policies from cloud platforms has brought a value driven expenses model from organizations. Visualization tools and analytical data models makes data more discoverable but processing function often needs ability to process at scale, many times compute performance is a key for derivations, for example GPS systems in car navigation, processing at speed is equality important as accuracy of information.

Cost of data persistence looks cheaper in the cloud environment, which is one of the major factors of higher cloud platforms adoption across industry lines. Maintenance costs have been shifted to shared costing model in a way because major database maintenance activities have been carried out by cloud providers. Activities including backup, recovery, disaster recovery, monitoring, logs etc. are generally carried out by cloud platforms. Which is not only a big cost saver, but it is saving lot of maintenance efforts and troubleshooting on day-to-day basis. More important is the system downtime for maintenance or upgrades, backward compatibility etc. are major issues organizations had been facing frequently before could base databases were introduced.

Every database system has been influenced by surrounding systems. Extraction Load and Transformations (ELT) are typical operations on data warehouse systems. Transactional (OLTP) systems and analytical (OLAP) systems have many online or batch transformations. Cloud eco systems have multiple state of art systems which is catering all database operations including ELT, OLTP and OLAP.

Database transactions have been classified as batch processing, real time and streaming solutions. Each of the transaction types have multiple cloud base tools, products and platforms to simplify these transactions using modern algorithms and intelligence. Interfacing with modern data processing systems on cloud platform has indulge the need to move traditional databases to cloud and enhance.

Relational Databases system (RDBMS), which have been serving in majority, cater to schema base data storage. RDBMS structures limit the data storage, schema is often predefined and majority of times context about data get missed while capturing data (Shende and Chapke, 2015). In real world data is contextual to the event or action. Each event of actions has variations of schema and because fix schema of RDBMS limit capturing all aspects of data and store. According to research, 90% of worlds data going to be digitalized with the two years.

Modern database systems provide ability to store unstructured, structured, and semi structured data storage. Which need flexible and scalable storage and allow distributed complex processing for effective business utilization of data storage. Combination of many aspects of database storage, operations, extraction, performance, maintenance, scalability, flexibility and pay per use cost structure has modernized database over cloud.

Today , database and DBMS are an integral part of any kind of work. It may be business related or managing accounts (Berg *et al.*, 2012).

Some Dominant CLOUD Databases

SNOWFLAKE

It is a Cloud Managed Database, which does not need any installation neither on virtual machines nor on physical infrastructure. Snowflake has fully managed Software as service. enterprise does not need to perform maintenance activities. Replications failovers and disaster recovery have been provided as Software service. Snowflake is cloud only database, which has database persistence, ELT, OLTP and OLAP supporting tools, intelligent analytical tools. Snowflake also provides partitioned data persistence as well as advanced database features like time travel. Snowflake support multiple

cloud-based platforms. Some examples are Azure, Google cloud and Amazon web services. The cloud database is easy to use and manage . Mostly it should reduce the costs as well (Curino *et al.*, 2011).

Connectivity to multiple state of art visualization tools, analytical models, in built data pipelines, data import & export features as well as data sharing ability across multiple customers across secure connections. For example: snowflake provides tableau connector which provides unstructured as well as structure data to the visualization tools. Snowflake operates scalable compute power and distributed computing. Snowflake query performance is recognized as one of the best-in-class complex query performance.

Azure Databases (CosmosDB and SQL Datawarehouse)

This is one of most adopted database services on Microsoft Azure platform. It is a Fully managed platform. Cosmosdb provides automated failover and business continuity provisions with time to live feature. CosmosDB has multitenant container. It can be accessed using SQL, Gremlin, Mongo dB, Casandra and Table API. SQL works as relational database acid transactions. These transactions are like traditional RDBMS systems except autoscaling, implicit indexing and auto failover (Jain and Alam, 2017). Gremlin is used when graph or network related use cases need to access. it supports vertex and relations; it also supports graph query language finding anomalies and semantic co relation base applications.

Mongodb tenant facilitates document based No-Sql implementations. Mongo is Json base document database to facilitate high available and high consistence data extractions, mainly suitable for unstructured or semi structured data processing. Casandra tenant has been utilized for analytical use cases, where key value pair-based data processing can be performed. CosmosDB is a Microsoft Azure native tool, which has integral connectivity with rich analytics library i.e., Microsoft Synapse. Architecture of synapse also includes data bricks (Distributed Parallel Data Processing) and State of Art data visualization tool i.e., Power BI.

BIGQUERY

Google Clouds native database system is known as BigQuery.. Like other cloud databases, big query is also a fully managed platform as service. Where failovers, higher availability, business continuity, higher scalability and other features are integral to the cloud platform. BigQuery supports in built machine learning libraries to perform real time analytics. it is designed to support geospatial queries. Natural Language Processing services are part of analytical system. BigQuery provides data transformation applications to extract load and transform the data. There are prebuilt products to transform data from Teradata or Amazon Red shift. This enables BigQuery platform more connected to other cloud platforms.

REDSHIFT

Amazon Redshift is widely used could data warehouse. This is the data warehouse designed to migrate on premise data warehouse systems over cloud like Oracle data warehouse. Typical data warehouse systems are used to create faster and flexible data analysis. Amazon redshift has provided cloud based better option to migrate on-premises warehouses. It includes business analytics, operational analytics, and predictive analytics. Redshift compute cluster has provided scalable, flexible cloud platform. With strong integration with Amazon Sagemaker, Amazon EMR, Amazon Athena and 3rd party services. Amazon Redshift has ability to store query response on Amazon S3. This cloud warehouse system has been established 10 of thousands analytical clusters on AWS cloud platform.

MARKLOGIC

Marklogic Datahub is multi cloud data hub which can simplify data curation, transformation, integration for unified data hubs. Marklogic is designed to store document, which can cater structured, unstructured and semi structural documents. Marklogic have three components storage, webserver and analytical services. There are multiple set of tools are available to ingest the data, migrate, import and export of the data. It includes dashboards, machine learning libraries as well as analytical data models.

Advantages of Database Platform

Cost Saving: Organization can invest in the resources they truly need, without worrying about the maintenance of database.

Rapid Provisioning: Comparatively it takes very less time to process and in cost effective manner.

Outsourcing: Various operations like Backups, Optimization are handled by outsourcing.

High Security: Security breaches avoided by by-default security mechanism of Database platform.

Tracking: Easily can track usage time, space, resource consumption.

Manpower: Freeing up staff is the biggest advantage of this. Manpower can focus on their development.

Server Space: Lot of server space frees up.

Scalability: on-demand scalability is possible.

Conclusion

Databases on cloud platform are generally fully managed and performing operations more like a platform rather

than traditional software applications. Disaster recovery, Intelligent data insight systems, easy extract load and transformation products are integrated in cloud database platforms. Microsoft Synapse architecture is built on SQL DW as well as Power BI, Marklogic machine leaning libraries integration with persistence layers and AWS providing Redshift with EMR as Data platform or Snowflake Data sharing among customers establish a strong pattern indicating that Database systems are evolving towards Database platforms. Data Technologies are migrating towards modern Era of Database platforms. Database platforms are tightly coupled integrations of Data Analytics, Governance, Persistence, Data Data Applications and Intelligent Data processing products hosted on cloud which are ready to consume. Cloud database selection will primarily dependent on how intelligent data integration, management or data consumption features are integrated with database platform. This review paper introduced the basic knowledge and explained the important features of various cloud databases. their pros and cons of database as a service have been explored and allow users to help in database selection process.

References

Berg, K. L., Seymour, T., Goel, R. (2012) History of Databases. *International Journal of Management & Information Systems* **17**(1):29-36.

Britts, W. (2005) Select Database (SeDB) A Database Selection Process Model. Dissertation, Uppsala Universitet.

Curino, C., Jones, E.P.C., Madden, S., Balakrishnan, H. (2011) Workload-aware database monitoring and consolidation. In Proceedings of the 2011 *International Conference on Management of data http://doi.acm.org/10.1145/1989323.1989357*.

Ellison, M., Calinescu, R., Paige, R.F. (2018) Evaluating cloud database migration options using workload models. *J Cloud Comp* **7**. https://doi.org/10.1186/s13677-018 0108-5.

Jain, S., Alam, A. A. (2017) Comparative Study of Traditional Database and Cloud Computing Database. *Int J Adv Res Comput Sci* **8** (2): 80-87.

Sakr S. (2014) Cloud-hosted databases: technologies, challenges and opportunities. *Cluster Comput* **17**: 487502. https://doi.org/10.1007/s10586-013-0290-7.

Shehri, W.A. (2013) Cloud Database- Database as Service. *Int J Database Manag Syst* **5** (2): 1-12.

Shende, S. B., Chapke, P. P. (2015) Cloud Database Management System (CDBMS). *COMPUSOFT Int J Adv Comput Technol* **4** (1). https://ijact.in/index.php/ijact/article/view/61.

Siddiqui, T., Jindal, A., Qiao, S., Patel, H., Le, W. (2020) Cost Models for Big Data Query Processing: Learning, Retrofitting, and Our Findings. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data* 99-113 *ttps://doi.org/10.1145/3318464.3380584*.

Wu, S., Li, F., Mehrotra, S., Ooi, B.C. (2011) Query optimization for massively parallel data processing. In Proceedings of the 2nd Symposium on Cloud Computing (SOCC 11) Article 12: 1-13 https://doi.org/10.1145/2038916.2038928.