E-cloud Database Management

Yiqiao Yin
W.Y.N. Associates, LLC

Abstract

This assignment serves as an executive report for e-cloud business to advocate for organization cloud database management. The report summarizes the main effect of using e-cloud based database management system and analyze the benefits as well as the challenges of the potential transition. The information of this report is also summarized into a presentation slides in a separate document.

1 Data Safety and Machine Learning Practice

Cloud database management platform are becoming a significant component of our lives for a wide variety of business practices (Grozev and Buyya, 2014). E-clouds enable end-users to manage on-demand computing instances and jobs Mendez et al. (2013). It is also an emerging trend to develop efficient healthcare protocols and applications (Sodhro et al., 2019). In addition, wearable technologies have been making more and more application-based modules feasible for us. This paper evaluated recent low-power wireless and wearable devices and have shown great potential for mobile healthcare and e-cloud opportunities (Sonawane and Sutar, 2017). The e-cloud convenient infrastructure provides cost efficient access and management of database from anywhere in the world. From Google Drive of an individual to AWS to an organization, the cloud database management has been quietly transition our day to day database management platform.

This direction comes with the following benefits: (1) flexibility, (2) cost-effective, (3) reliability, (4) scalability and high performance, and (5) security.

1.1 Flexibility

The cloud service and database management platform enables users to select and choose any database management language to operate. On a cloud database management platform, a virtual machine is created of which each user can create their own environment. Upon the creation of an desired environment, each virtual machine can also be designed using multiple different versions of repositories of code. This brings forth the Quality Assessment and Quality Control (QAQC) which is a crucial component of nowaday data science practice.

1.2 Cost-effective

The concept of cost in data science practice for today's e-cloud or online healthcare solution is based on two basic concepts: money and time. The first perspective is rather

easy to understand. Instead of owning a room full of hard drives of which no one uses or everyone uses them together, the service is contracted to a third party cloud base management system of which the payment is billable by time used of running a machine. Argument can be made for existing healthcare companies that are sufficiently large. In that case, a privately owned cloud database management system can be extremely cost-effective. However, for smaller scale healthcare companies, it is much more efficient to outsource this service.

The second perspective is time. There are many different ways of running an algorithm and it is not always the case that it is faster to run programs on parallel. Sometimes it is a judgement of the experiment design and how to define the usage on each core. This is an area where time space analysis in algorithmic design is essential to deploying such a project onto a cloud database management platform. For example, if it is desired to search for each item from an array of items, a running time of $\mathcal{O}(n)$ is required while n is the length of the array. If it is desired to search and compare every two items of an array, then a running time of $\mathcal{O}(n^2)$ is possible for some programming solutions. When n is sufficiently large, it is challenging to finish the job on one core. This is an area where programmers debate whether to put the job on one core or multiple cores. If the job is running on multiple cores, then space analysis also needs to be discussed because the returning output from these jobs need to be processed efficiently as well.

1.3 Reliability and Consistency

In the discussion of QAQC, two other concepts need to be addressed are reliability and consistency. Reliability is referring to the maintenance of quality over time for a machine. Consistency is referring to the level of quality across different machines.

Reliability is the most essential concept when it comes to software programming, specifically at the deployment of a machine. Without the aid of cloud database management system, the deployment of a machine is almost always locally and rely on an installation software that is designed by tailoring the requirements of the program to a specific sets of environment. This is extremely difficult to scale up and even more difficult to maintain over time. With the aid of the cloud technology and centralized virtual machine, it provides programmers more conviction that machines can duplicate results with a fixed quality throughout time.

Consistency is the next important concept in QAQC and also essential at deployment of a machine onto an e-cloud platform. The main attribute when diagnosing the consistency is the effectiveness of a certain program throughout different machines. This can also bring value when a program is deployed on different environment to serve customers' needs when desired to ensure version control and consistent performance.

1.4 Scalability and High Performance

By moving the machine learning system from local server to cloud, there is a higher magnitude of potential for scalability and better performance. A neural network developed 20 years ago may have thousands of neurons, however, a neural network developed today may have billions of neurons. To train machines and tune the performance at this scale, it might be challenging for local machines to serve the purpose. Even if they do, it is not scalable at a global level especially in healthcare because healthcare does not have national borders. In this case, the software scalability is not only required to be efficient, but also needs to be applicable across different sites when there is a rising

need of demand. In other words, when a certain program or healthcare product is mass produced, there should not be a decline of the performance.

1.5 Security

The last important concept to discuss is the data security. There can be potential debate to argue that there is a lack of data security when moving everything up to a cloud server. However, the debate could easily be turned around, because all data security issues come down to the trade-off between safety and effort of authentication, which is the same for local and cloud server. Only because the platform is online does not mean that the trade-off somehow disappeared. There can be insecure online cloud platform if there is little effort applied to build its authentication process. Conversely speaking, a secured online database management is entirely possible if sufficient resource is dedicated to constructing such cloud platform.

Another aspect worth noting is the legislation and regulatory compliance. In some occasion when it comes to data sharing across different teams and members of the same company, strict protocols need to be applied and be enforced.

2 Challenges

Despite the apparent benefits gained from the deployment of cloud database management. There are some potential challenges as well. This section we discuss these challenges in the following.

2.1 Lack of Access

The online cloud-based database management can have power outage problems. Major vendors could be left without the access of their database access for as long as the main factory is fixed and back online. The cloud database management also requires constant connection of internet. For customers or patients in developing world where the internet of things (IoT) are not as widely spread as those in the developed world, this could pose great challenge because cloud access is simply unavailable for these audiences.

In addition, it is worth worth to mention that database security is another issue when it comes to national boundaries without borders, because the protocols need to follow different regulations before a data can simply be “shared”. Beyond this point, the cloud database management becomes a jurisdiction discussion of which it can involve national security.

2.2 Component Interaction

In the design of cloud database management platform, the clients interact with a programming Integrated Development Environment or IDE where a particular environment s created to tailor to the purpose of the job. Then entry points where clients can use Application Programming Interface or API to communicate with server to extract certain functions or programs from the cloud. At the step of data center where authentication occurs, an admission ticket is issued when the data requested pass the protocols and is indeed accepted. If the data request is approved, the access of a piece of data is processed and the data is extracted and ready for retrieval. Depending on different database management system, sometimes database can be insured using insurance provider.

2.3 Scalability

Though scalability is raised as an important benefit, it can also be a potential cause of a major challenge by itself. Despite its benefits, scalability could be at risk when virtual machines are utilized in a completely different way. The program is designed using the environment in a virtual machine where this environment dictates the feasibility of the software program. The success of the deployment is out of the equation and feasibility faces difficulty when the virtual machine on one cloud database management is different than another one.

This leads to another issue at large. It is commonly known as "time-out" problem. No cloud server will provide on-going and continuous cloud access forever. This is because it breaches the protocol of data security. Based on level of management, pay grade, and depth of the project, each program will run have its own designated time-out period. The time-period is different can provide great challenge for feasibility and scalability.

3 Proposal

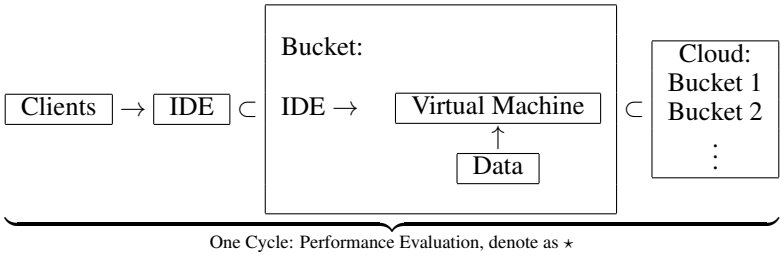
3.1 Protocols

Based on the discussion of potential benefits and challenges from previous section, we recommend cloud computing and deploying database management platform on cloud services. In general, it is shown that model-driven and data-driven development can be helpful in this level of context Ardagna et al. (2012). The process of instantiating environments in online virtual machine is proven to be more efficient at developing machine learning models as well as other related machine learning based platforms. Major technology companies such as Microsoft, Google, Amazon, HP, AT&T, and IBM contribute to the e-cloud platforms and they serve at an enterprise level for small and large corporation clients.

The proposed diagram of a general idea of deploying cloud based management platform is as shown in Figure 1. The clients (programmers and other users who interact with the program) write code in an IDE that is launched on a virtual machine. The IDE executes code and programs in the virtual machine with a designated software environment. The software, as it is running, may pull data from the bucket as well. The virtual machine and the data lives inside a bucket. The bucket is a folder-like location or directory that lives on a cloud server. The cloud server provides storage and computation capability for the end-users which are clients. The clients code and pulls data (assuming it is approved and does not violating protocols of data security) executes ideas to solve certain business problems. The problems can sometimes take a few turns of which a draft program is deployed as a test case. The idea can be considered as some type of information flow (represented by arrows, i.e. $\rightarrow$ in the diagram). The information flow goes from the left to the right. On the left, the information is in a natural form in the clients' mind. The information becomes solid and transparent when it is processed as computer program. The program accomplish a test job and becomes a test case. The test case lives on the cloud server for potential deployment. The cycle occurs when there is more than once the diagram executes from the left to the right.

The protocols designed in the diagram presented in Figure 1 then follows certain performance evaluation metrics. The entire process of the diagram is called a cycle, i.e. denoted as $\star$. The cycle is then evaluated with some benchmark performance. The threshold used for performance valuation is denoted as t. The threshold is then compared

Figure 1: **Diagram of Cloud-based Data Platform**. This diagram illustrates a general idea of a cloud-based data management platform.



to sets of error protocols, which is denoted with ϵ. A global algorithm can be developed to walk through the cycle and it is presented in Algorithm 1.

Algorithm 1 Data Science Cycle

Require: $\star$
Ensure: $\star$ is completed
  $\star$ is valuated according to some threshold t
  ϵ is a set of protocols for error
  if $t \subset \epsilon$ **then**
    restart $\star$
  else if $t \not\subset \epsilon$ **then**
    deployment
  end if

3.2 Machine Learning Containment

In addition to the algorithmic design in previous subsection, the e-cloud database management system is proposed to be accompanied with module containment with each software production. A module containment is a management console that can launch user interface or UI. It can simply be considered as a folder live in a bucket described in Figure 1 and a machine learning job can be launched inside. There is a gateway checkup for kernel initialization. This is important because common machine learning tools such as tensorflow and sci-kit learn all depend on this setup. A studio-like interface can then be presented if the end-user is programmer or employees with the technical background to initiate the IDE for desired programming languages such as Python.

The containment could also be equipped with multiple cores to enable parallel computing. Common machine learning modules can run on Central Processing Unit (CPU) and Graphical Processing Unit (GPU). GPUs are usually faster at handling tensor objects especially if the job is a deep learning based module. In some occasion, Tensor Processing Unit (TPU) can also be installed where multiple GPUs are loaded to process the job simultaneously.

References

Ardagna, D., Di Nitto, E., Mohagheghi, P., Mosser, S., Ballagny, C., D'Andria, F., Casale, G., Matthews, P., Nechifor, C.-S., Petcu, D., et al. (2012). Modaclouds: A model-driven approach for the design and execution of applications on multiple clouds.

In *2012 4th International Workshop on Modeling in Software Engineering (MISE)*, pages 50–56. IEEE.

Grozev, N. and Buyya, R. (2014). Multi-cloud provisioning and load distribution for three-tier applications. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 9(3):1–21.

Mendez, D., Villamiazr, M., and Castro, H. (2013). e-clouds: Scientific computing as a service. In *2013 Seventh International Conference on Complex, Intelligent, and Software Intensive Systems*, pages 481–486. IEEE.

Sodhro, A. H., Luo, Z., Sangaiah, A. K., and Baik, S. W. (2019). Mobile edge computing based qos optimization in medical healthcare applications. *International Journal of Information Management*, 45:308–318.

Sonawane, P. D. and Sutar, R. G. (2017). A schematic review on body area networks for e-health systems. In *2017 International conference on intelligent computing and control (I2C2)*, pages 1–5. IEEE.