

## **PRACTICAL-13**

### **AIM:**

To perform a case study of the following platforms for solving any big data analytic problem of your choice. (1) Amazon web services, (2) Microsoft Azure, (3)Google App engine.

### **IMPLEMENTATION:**

#### **1. AWS: OLX Autos Delivers a Seamless Online Marketplace by Running Containers on AWS**

- OLX Autos' goal is to enable people to better their life through its online marketplace of services and used goods that offer superior consumer value.
- As part of the OLX Group, which runs more than 20 brands in 30 countries, OLX Autos blends the versatility of a start-up with the experience of a multinational entity.
- With a vision to shape the future of commerce, OLX Autos is excited about leveraging state-of-the-art technologies to improve consumer service.
- India is one of the world's biggest growth markets for OLX Cars.
- The organization has a highly innovative technology team running its Panamera Classified Marketplace, which is active in 12 countries, on the Amazon Web Services (AWS) cloud.
- OLX Autos is committed to investing in its people and the new technologies to unlock optimum value for its customers. "Using AWS has kept our environment running in a stable manner in the most cost optimized way." Abhishek Tomar Infrastructure Head, OLX Autos

#### **→ Modernizing Dynamic Architectural:**

- OLX Autos suffered an outage on its platform in 2018 after its internal OpenShift containerization program licenses expired.
- The organization used OpenShift version 3.5 to handle its resources and install software on its Docker container site. As part of the company's pledge to provide consumers with a seamless experience, engineers quickly found the problem and rapidly rolled out new certificates.
- However, they had to spend a substantial amount of time rebuilding the website. The team concluded that their staging environment was dysfunctional due to problems with the OpenShift Control Panel and started searching for solutions to modernize their infrastructure stack.
- Abhishek Tomar, infrastructure manager at OLX Autos, says, "Buyers and sellers were unable to check, view new advertisements, or build new listings during the shutdown. We found that even though we had just one accident a year, it would always be easier to unload the control and the guy.

### → Seamlessly Migrating 124 Microservices:

- OLX Autos launched a conversation with the AWS team a year before the migration began. "First, we reached out to AWS Business Support to provide guidance on the right migration approach," says Sharma. At the same time, OLX Autos revamped its operating system, Kubernetes, and Docker versions, and the difficulty of its OpenShift architecture, in addition to software version limitations, made migration especially difficult.
- The OLX Autos SRE team prepared a relocation approach and began shifting their staging environment to AWS in January 2020. "We received immediate assistance from AWS to address any problems in the staging area that helped us move seamlessly to the production environment," says Sharma.
- The AWS team proposed a plan for future new capabilities of Panamera using Amazon EKS in addition to other AWS offerings. OLX Autos then began constructing an Amazon EKS cluster and migrating selected workloads from OpenShift, which will be finished in only a few weeks.
- OLX Autos has now migrated all 124 of its microservices to Amazon EKS, which has improved efficiency and scalability while unlocking cost savings.

### → Achieving Efficiency:

- After relocation, OLX Autos infrastructure teams have now become well prepared to meet crucial internal product delivery deadlines.
- The organization originally used the Puppet Setup Tool to handle its OpenShift program.
- Engineers wanted to focus on their Puppet expertise and spent three to four days a month applying and tracking big improvements to the OLX Autos infrastructure.
- Through the migration to AWS, the organization removed Puppet from its design and switched container control to AWS.
- By operating on Amazon EKS, the OLX Autos website benefits from increased performance and scalability, and engineers may redirect their time to higher value-added activities.

### → Scaling in Minutes and Improving Uptime:

- When OLX Autos was running an old version of OpenShift, the company faced autoscaling difficulties, which triggered interruptions to its applications. "Now with Amazon EKS, our Amazon Elastic Compute Cloud (Amazon EC2) fleet can be auto-scaled in a few minutes if some marketing plan triggers a surge or a rapid surge in traffic.
- If online traffic declines, the fleet will be scaled down — something we haven't been able to do with OpenShift before," says Tomar.
- OLX Autos has also unloaded the vital role of handling Stable Sockets Layer (SSL)/Transport Layer Security (TLS) certificates to AWS Credential Manager. Previously, teams had to manually buy and instal new certificates each year, but with AWS, new certificates are deployed with a few quick API calls.

- "It's a relief that AWS is now going to take care of that," says Tomar. "We don't have to spend a single minute in testing and upgrading certificates that may have an effect on crucial deadlines for our company."

→ Unlocking savings with a reserved instance:

- OLX Autos had reserved a number of Amazon EC2 reserved instances for form C5 instances but was unable to use them for the older version of OpenShift. "As a part of this conversion and version update, we have transferred all of our C4 instances to C5 and will leverage the price of reserved instances.

- That eventually saved us 33% of the cost of computation, "says Tomar.

- OLX Autos have benefited from the AWS Container Network Interface (CNI) plug-in, which was blocked by its previous OpenShift setup.

- Not only has the CNI plug-in increased device latency, but OLX Autos now hopes to save at least 10% of its average monthly AWS bill for more efficient applications. Tomar says, "Using AWS has kept our climate secure in the most cost-optimized way."

Google App Engine: Travlytix:

Turns to Google Cloud to run a customer data platform and personalization engine

- With big data products running on Google Cloud Platform, Travlytix created an intelligent customer data platform for airlines in six months with a small engineering team.

- Travelers generate data across a range of touchpoints, presenting opportunities for businesses to understand more about how they behave. Malaysia-headquartered Travlytix decided to create a platform to consolidate real-time data from these touchpoints into a single location. The business also had to ensure the data was secured from leakage, disruption, or theft. "Google Cloud Platform services like Cloud Pub/Sub, Cloud Dataflow, and Big Query allow us to minimize the effort and resources needed to build data pipelines for airline customers and focus instead on the quality, volume, and velocity of data." —Srinivas Sri Perumbuduru, Head of Product, Travlytix

- Airlines, online travel agents, and travel eCommerce companies face challenges in capturing, processing and utilizing data. The business turned to the cloud to deliver its platform. "The cloud gave us an opportunity to work with large volumes of data while becoming more agile," says Travlytix.

- When COVID-19 forced lockdowns all over the world, the re-engineered Dr. Foster landscape eased the sudden transition for its employees.

- The company continues to meet via Microsoft Teams, using the whiteboarding functionality to collaborate. Dr. Foster is beginning to process 60,000 test results from parent company Telstra Health in Australia each morning.

- The solution allows the company to react as the market requires, says Bayliff. "Because of the SQL Server 2019 Big Data Clusters solution and the architecture we've put into place, we can react to the market," he says.

→ Making data transparent:

- "Over the history of the industry, people have used data to make more educated choices," adds Srinivas Sri Perumbuduru, Product Manager at Travlytix.
- "Our goal in Travelytix is to make the data clearer."
- Travlytix needed a highly available, resilient, and secure cloud platform to deliver what Sri Perumbuduru calls "one of the most workable customer data platforms in the travel industry." The corporation then switched to the Google Cloud Network.

→ Google Cloud Platform supports architecture principles:

- Travlytix's review showed that Google Cloud Platform will support its founding principles of keeping its architecture basic, microservice-based, and highly secure.
- "The documentation and expertise of the Google Cloud Platform and simple contact with Google's customer engineers have made this possible," states Sri Perumbuduru.
- With Google Cloud Technology, the company was able to develop consumer data technology in just six months.
- Travlytix uses Google Cloud Platform services to collect, process, and store traveler data.
- The platform captures data in real-time, processes it through data pipelines, and stores it in a data warehouse.
- TravlyTix also relies on Compute Engine to provide compute resources; Google Kubernetes Engine to manage and orchestrate Docker containers; App Engine to develop and host its applications, and Cloud Memorystore to provide caches to enable sub-millisecond data access.
- "Google Cloud managed services give us more time to strategize our data pipeline architecture and allow us to be less concerned about maintaining the infrastructure," says Kevin Chin, Data Engineer, Travleytix.

→ Low-latency services:

- Travlytix also utilizes several Google Cloud Network zones to offer ongoing, low latency support to clients around the globe.
- In addition, applications such as the Cloud Key Management Service, which helps companies to control cryptographic keys and secure confidential data on the Google Cloud Network, facilitate compliance with global security regulations.
- These support Travlytix's own data encryption during collection and processing to protect it from intrusion

## CONCLUSION:

In this practical, we perform a case study on big data analytic problems on Amazon web services, Microsoft Azure, Google App Engine