Course Code	:	SWE309
Course Name	:	Introduction to Cloud Computing
Lecturer	:	Mahdi H. Miraz
Academic Session	:	202204
Assessment Title	:	Assignment
Submission Due Date	:	7th July 2022

Prepared by :

Student ID	Student Name
SWE1909762	Tan Yi Ying

Date Received :

Feedback from Lecturer:

Mark:

Own Work Declaration

I/We hereby understand my/our work would be checked for plagiarism or other misconduct, and the softcopy would be saved for future comparison(s).

I/We hereby confirm that all the references or sources of citations have been correctly listed or presented and I/we clearly understand the serious consequence caused by any intentional or unintentional misconduct.

This work is not made on any work of other students (past or present), and it has not been submitted to any other courses or institutions before.

Signature: 

Date: 7th July 2022

Scenario

DiDi Chuxing is a ride-hailing service in China. The company would like to expand their business in Malaysia and Singapore. The expansion plan also includes developing an enhanced website as well as mobile application (app) offering many facilities to the customers, including searching and booking online. The online services need to available 24 hours per day, 7 days per week, 365 days per year.

Question 1

Do you suggest that the company should purchase the required hardware and software to deploy the website and/or the mobile app on the internet (on-premise)? or deploy the website and/or mobile app on the internet by using the cloud computing technology (on the cloud)? (Discuss your opinion and justify your answer).

DiDi Chuxing is one of the world's largest ride-hailing companies, serving more than 493 million users across Asia-Pacific, Africa, Latin, America, Central Asia, and Russia. According to the scenario listed where DiDi Chuxing decided to expand their business in Malaysia and Singapore it is suggested to deploy the website or the mobile application by using the cloud computing technology. One of the biggest and most complex obstacles that any of the transportation business faced is lack of abilities and inefficiencies in encountering the manual ride-hailing operations. A proficient ride-hailing software ensures that it kills most of the common issues by offering the ideal solution which is sufficiently competent to give the dispatchers an intact control by allocating cabs to each user. Cloud-based ride hailing system helps in consistent working of the entire ride-hailing service. There are few advantages and features of cloud-based ride-hailing system which are as follow:

- It is highly efficient and cost effective since there is no need for physical equipment for data storage or outsourcing requirements needed, the company does not need to invest to much framework for any cloud based services. The company itself will not required any technical expertise to manage or to store their data as compared to vendor company. Hence, there is no need for expensive training or hiring experts in order to manage the data storage as well as their servers. Also, this allows the system to have lower rate of downtime, and thus it improves higher business profitability with overall efficiency.

- Cloud-based ride-hailing software provides system with a robust backup and recovery plan, since all information is not stored locally in house PC's but in the cloud file system. Cloud file system is a hierarchical storage system that provides shared access to file data. With cloud storage, the physical server locations are owned and organized by the hosting companies, who maintain, secure and run the servers to allow for the services to be provided. Besides, if the company's on-premise storage infrastructure breaks down, the cloud system makes disaster recovery easier to handle.
- Another benefit of using cloud computing technology is that the company is able to scale up and scale down their ride-hailing business or other necessities using cloud services. DiDi is able to increase or decrease the IT resources such as networking ability, processing power and also the data storage capacity without worrying about any operational disruption. According to Avram (2014), cloud computing services enable enterprises to scale their services according to client demand, since the computer resources are managed through software, thus the system can be set up easily as new requirements arise. As a matter of fact, cloud computing aims to scale resources up or down vigorously using software APIs based on client load using minimal service provider engagement.
- The cloud computing technologies provide green computing environment as compared to the others. Computing power, including server power, cooling and overhead power consumption is costly, and it can be worst when management is not efficient. As compared to that, the cloud vendors provide a better management especially on voltage conversion, spending less on cooling by locating data centers in cooler places and having better cooling facilities and less power consumption rate. According to Shayan (2014), the conventional data centers suffer from low resource utilization which approximately 15% to 20%. However, the cloud computing data which consists of multi tenants and sharing of resources for many customers increase the resources utilization up to 40% by distributing the load over many users. One virtual server is able to handle work of at least 2.5 utilized servers. Thus, higher utilization allows less power wasting which helps the environment to keep safe from pollution.

Question 2

Demonstrate at least three scenarios where application of Hadoop can help DiDi Chuxing improve their services and/or business activities. You answer should outline which aspects of Hadoop to use and how. It may additionally include Hadoop application related design and planning.

Hadoop is an open-source software framework that is used in storing data and running applications on clusters of commodity hardware. Hadoop is increasingly being made use of for application that are essential to business operations as it provides large storage for various kind of data, massive processing power and the ability to handle virtually limitless concurrent tasks or jobs. In order to understand how Hadoop can be used to help DiDi Chuxing to improve their services or business, we should first understand the system requirements for the ride-hailing system. In this system, it is broken down into several subsystem which are as follow:

Functional Requirements

1. Booking System
    - Track the location of the register driver and the passenger.
    - Keep track of the passenger search and request.
    - Keep track of passenger payments.
    - Upon a passenger request for a ride, a matching process to locate those nearest driver and issue notification to these drivers, system will select a fast and save traffic route to both the driver and the passenger.
    - Upon a driver accept the passenger request, information of the driver will be available in times at the passenger's device and this passenger request is block from other driver.
    - Upon a driver accepted the passenger request, the passenger information and a ride request should be available at the driver's device.
    - Upon the start of the journey, the system provides traffic signal or the road condition to both the driver and the passenger in order to optimize cost, time and safety will lead to improve riders and passengers experience. An update to both the driver and passenger with the estimate time need to reach destination.
2. Driver Traffic Reporting System
    - Site to track drivers traffic report such as accidents, police traps, blocked roads, weather conditions and much more.
3. Traffic Signal Collection System

- A system to collects driver movement information such average speed, check for errors, improve road layout and learn road and turn direction while drivers are on the road with the system running.

4. Traffic Analysis System
    - The combination of the information collected from the Driver Traffic Reporting and Traffic Signal Collection, this system is responsible to analyse the activity and status of the road.

Non-functional Requirements

- The processing speed in the Ride-Hailing system should runs within milliseconds in order to cater for real-time processing.
- The system needs to deal with high volume of structure, unstructured and semi-structure data.
- It is expected that the velocity is high as it needs to track of the real-time information of the driver and passenger.
- Tools available are useful and requires minimum learning curve.
- The Ride-Hailing should be able to run on multi platforms.
- Scalability for large data set as well as exponential future growth.
- The system should run under 24x7 daily with zero failure.
- The system also emphasis on high throughput of data access.

The 3 major aspect in Hadoop which applied in the development of DiDi Chuxing Ride-Hailing System are as follow:

1. Hadoop Distributed File System

    The data velocity for the Ride-Hailing System for DiDi Chuxing are huge and would expend and exceed the storage limit in very short period. The Hadoop Distributed File System (HDFS) is used to store the structured (business data), unstructured data (such as traffic images, video, sound recording) as well as the semi structure (XML file) or so call Big Data (GeeksforGeeks, 2022). The HDFS enable streaming data accessing pattern as well as low hardware cost. Data files are divided into smaller data block and stored in different data notes. Smaller data block requires shorter accessing and processing time. Upon receiving a request to access or process the data, each block of data belongs to the data file can be access or process simultaneously or parallel in different data notes. HDFS also ensure that high data availability and high fault tolerance since a piece of the data block is stored in multiple data nodes.

2. Yarn in Hadoop

   There are 3 major components in Yarn which are (Pedamkar, 2021):

   - Resources manager that runs on a master server in a cluster in which it performs several responsibilities such as resources management, scheduling management, applications management & etc. During the management of resources, it will communicate with the node manager to obtain and keep the usae of all the resources such as RAM, CPU, and network available within the cluster.
   - Node manager which is running on each data server. It function is to manage the hardware status and usage. Node Manager tracks the usage and status of the CPU, memory, and network on the local data server. It will then inform Resource Manager of the hardware status.
   - Application Master in which a background process a.k.a. daemon will be initiated by the Resources Manager upon starting a jod in the data server. The Application Master issue requests to the NameNode in the master server to get locate the data storage. It then communicates with the scheduler in the Resource Manager the containers.

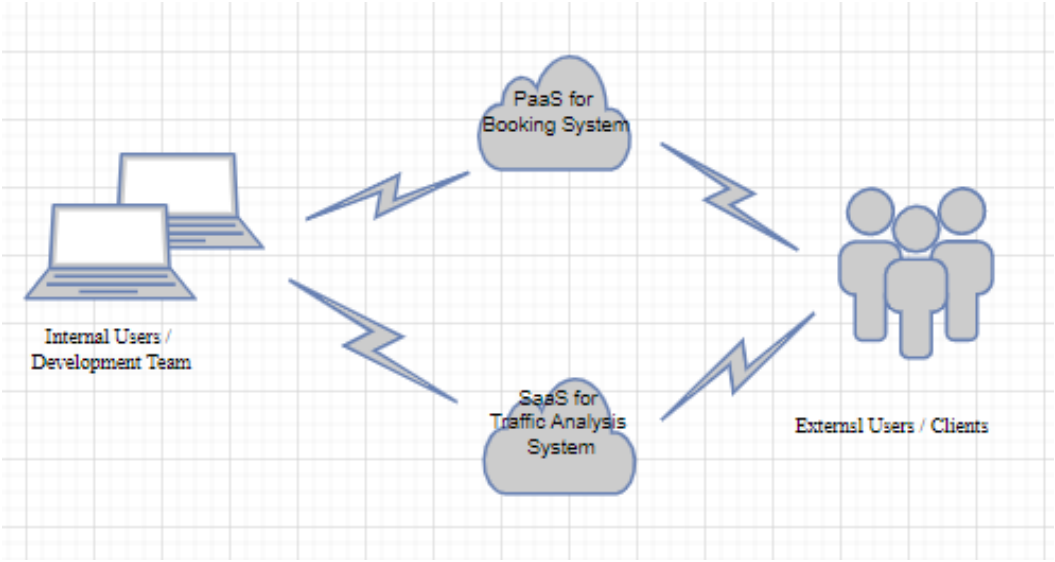
   Yarn in Hadoop for job scheduling and resource management will suit to the requirement for Ride-Hailing System to handle large volume of real time request from rider as well as from the passenger. With Yarn, it allows various processing methods and runs interactive queries that require continuous flow of real-time information. This is the Hadoop components that manage, schedule and distribute jobs to available resources to ensure that the real-time requests are processed within the shortest time and produce accurate result. Job are segmented into smaller function that parallel on different processing unit or nodes.

3. HBase of Hadoop (GeeksforGeeks, 2021)

   HBase is an open source handling Big Data storage. It is a distributed database running on top of the HDFS. HBase bundle with the HDFS provide a better combination for the Ride-Hailing System as they are from the same source Hadoop.

During the initial phase, at least 2 computing clusters is assigned for each state in West Malaysia and Singapore. A cluster to host the Booking System and runs it's background processes. A Private Cloud (PaaS) in needed to store the Structured data input from the Booking System as most of these data are confidential data. Another cluster to host the Driver Traffic Reporting System and Traffic Signal Collection System. This cluster nodes also run the background processes for Traffic Analysis System. A Hybrid Cloud (SaaS) is uses to store

the unstructured and semi structure data. Both cluster nodes will be equipped which Hadoop HDFS, Yarn and HBase.



Question 3

To solve the above problem, choose an appropriate service provided by any of cloud providers (i.e., GCP, AWS, Microsoft Azure, etc.) to host the DiDi Chuxing website. You should specifically clarify why, amongst other services, you have chosen this particular cloud service and/or provider.

The three most popular cloud vendors – Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) have appeased the concern of security and data sovereignty by adopting strict company's regulations continuously when deploying cloud services.

Amazon Web Services, the current maker leader of the cloud computing platforms, offers a wide range of services to everyone including individual developers, large enterprises, and even governments. AWS was able to stand out from all other cloud computing technologies as it can be distinguished from other vendors based on several key elements (Varia and Mathew, 2013). AWS is flexible as it enables organizations to use operating systems, databases, programming models and architectures in which the internal organization are familiar with it. The flexibility helps organizations to mix and match architectures in order to serve their business needs and also there is no need for additional expensive training or hiring experts in order to manage cloud system. Organizations can quickly add and subtract AWS resources to their applications in order to meet customer demand and manage costs. Besides, AWS provides services in accordance with best security practices which also includes appropriate features in order to provide end-to-end security as well as privacy. However, the only problem is the price to afford AWS will not be cheap (Carey, 2020).

The Google Cloud Platform (GCP) is the cloud offering by none other than Google, it is the smallest of the Big 3 cloud providers. The GCP provides a robust set of cloud services to power and support any kinds of application with large amount of workloads and organization are able to decide the zone to initiate the resources; however, it can be only open by an invitation (Mirghani and Hajjdiab, 2017). In addition, GCP storage offers high availability for any data that is stored in Google Cloud Storage can be replicated in other remote data centers. As compared to AWS, GCP includes smaller number of services, but it has huge competitive advantages which is Big Querry under Big Data as it helps in reducing time to analyze enormous chunks of data and delivering dataset insights simultaneously in an easy way.

Microsoft Azure, the second largest cloud platform with C-level executives that provides a wide array of services tailored particularly for Microsoft-centric enterprises allowing many

organizations to easily shift to a cloud-based or hybrid cloud environment. Azure offers the combination of IaaS and PaaS services which are attractive to the organization. They may outsource their infrastructure to Azure and just pay for the resources they use with IaaS, they are able to build their own apps and solutions without having the need to purchase and manager the underlying infrastructure. Besides, Azure offers a wider selection of hybrid connections, such as virtual private network (VPNs), content delivery networks (CDNs) ExpressRoute connections and caches (10 Reasons Why to Choose Azure for Your Enterprise, 2022)

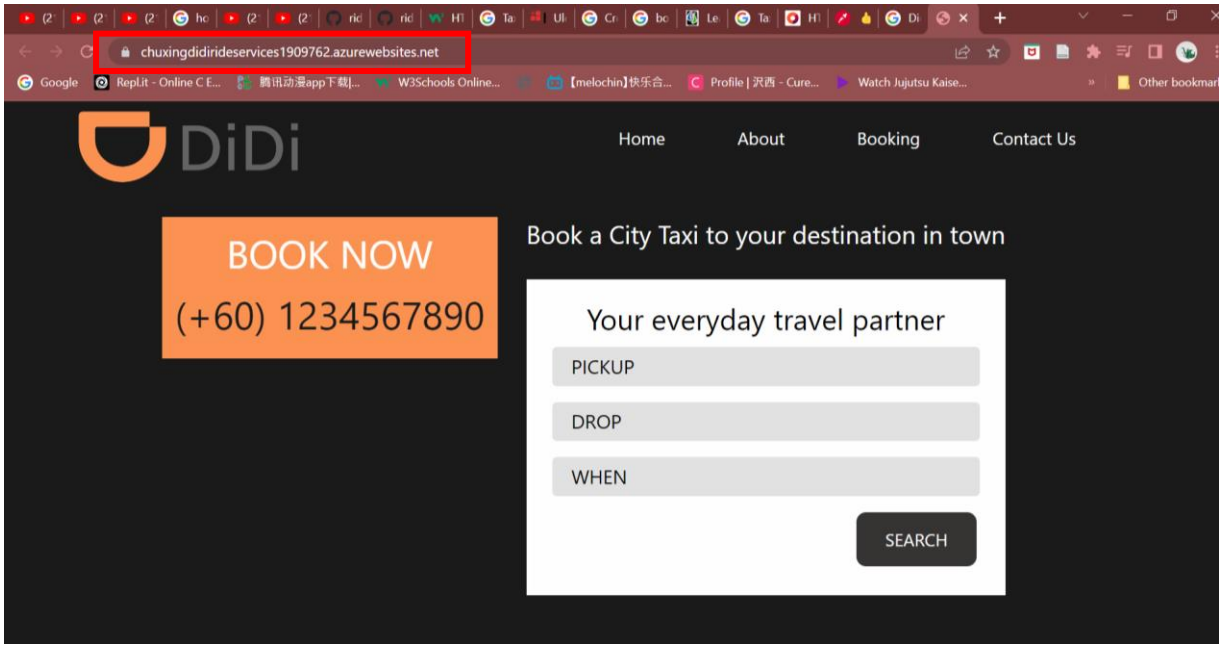
Before clarifying the most suitable cloud platform to host DiDi Chuxing website, we have to clearly identify the important features for a ride-hailing system to be implemented. So, how exactly should a ride-hailing system work? The general application execution can appear in 5 straightforward steps which are request, matching, ride, payment, and rating. Besides, there are several key elements a ride-hailing system should have, one of the most important feature is the technology of GPS tracking which is used to identify the current location, to find and track for cars. Also, it should provide an estimation of journey tracking. Besides, having payment gateway allow user to pay for the ride from the app is necessary. The integration of various payment services such as debit, credit card, online bank transfer makes the payment process easier for both parties. Furthermore, dispatch and tracking panel should be included in the system so that it enable the organization owner to track drivers with their status, track orders created from the ride-hailing application or website. To select a appropriate cloud service for the DiDi Chuxing company, we must consider few factors:

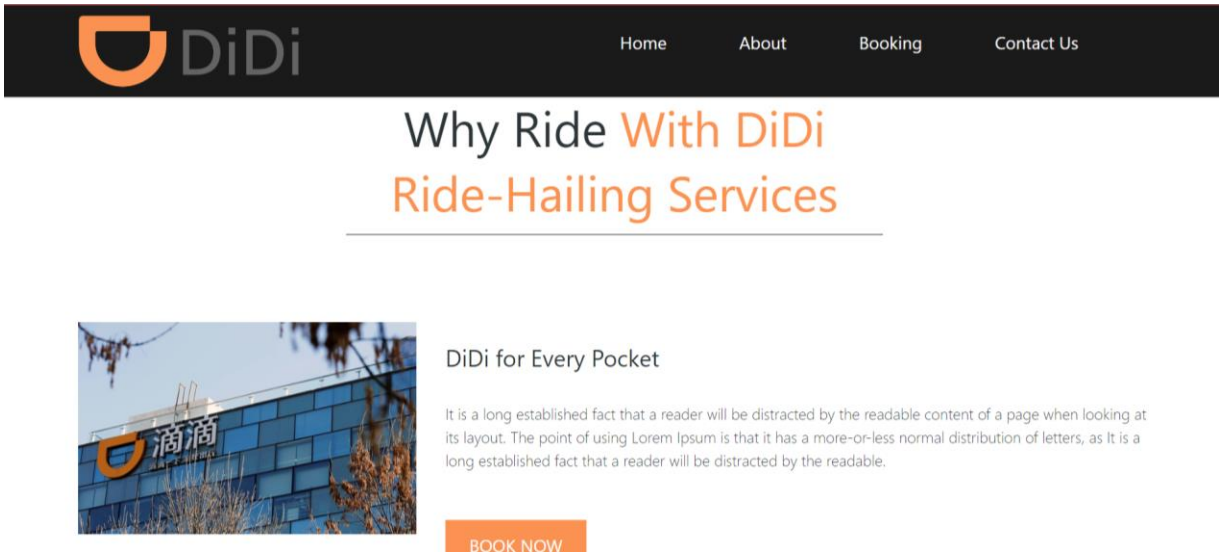
- When choosing a cloud provider, we first have to consider it's supported region and availability as it directly impacts the performance of the cloud such as latency and compliance requirements, especially dealing with data (Wickramasinghe, 2019). The big "three" cloud platforms provide specialized cloud solutions for the government. In addition, AWS and Azure provides specialized services that supply to the Chinese market with data located in China. This shows that Azure and AWS is suitable for DiDi Chuxing organization to manage since the headquarters is located at China and also able to provide services in Malaysia as well as in Singapore, it is more convenient to use AWS and Azure as compared to Google Cloud Services.
- Another factor to consider is the price to afford the appropriate cloud services, Azure may be cheaper than AWS in the long run as it is a zero maintenance solution. The organization just need to deploy the ride-hailing application and Microsoft will handle the software, patches as well as backups. The cost maintenance is needed when using AWS as compared

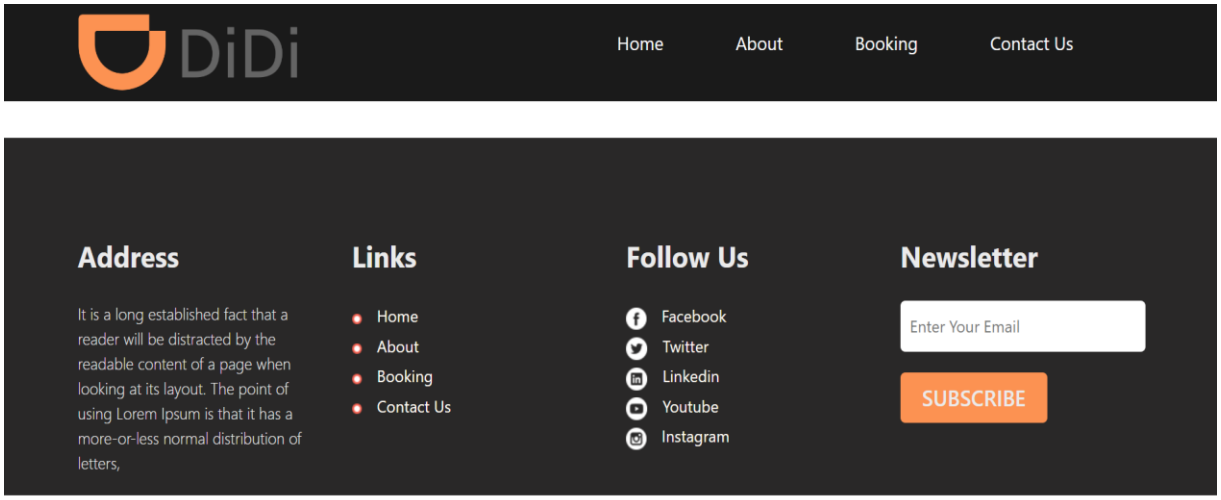
to Azure. In order to reduce the cost of maintenance, Azure is more suitable for Didi since it requires lower maintenance costs.

Question 4

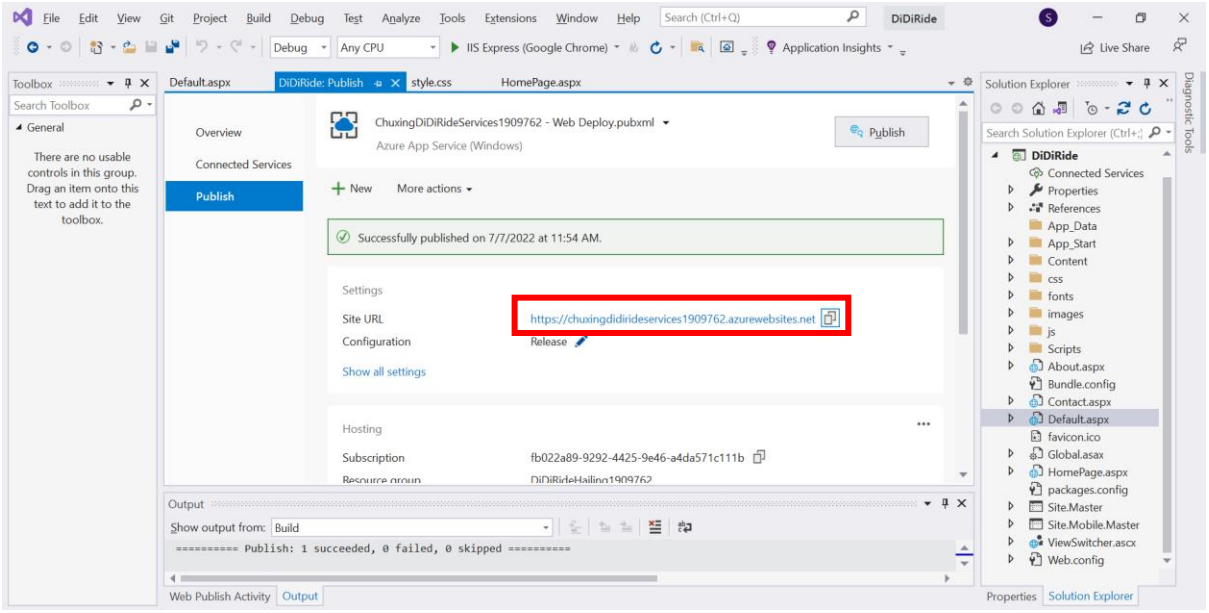
On the chosen platform (of task #3), create a cloud customer account if you do not already have one. A free/trial account shall suffice. Host your developed solution on the cloud platform and submit your work as outlined in the 'Important Notes' below.



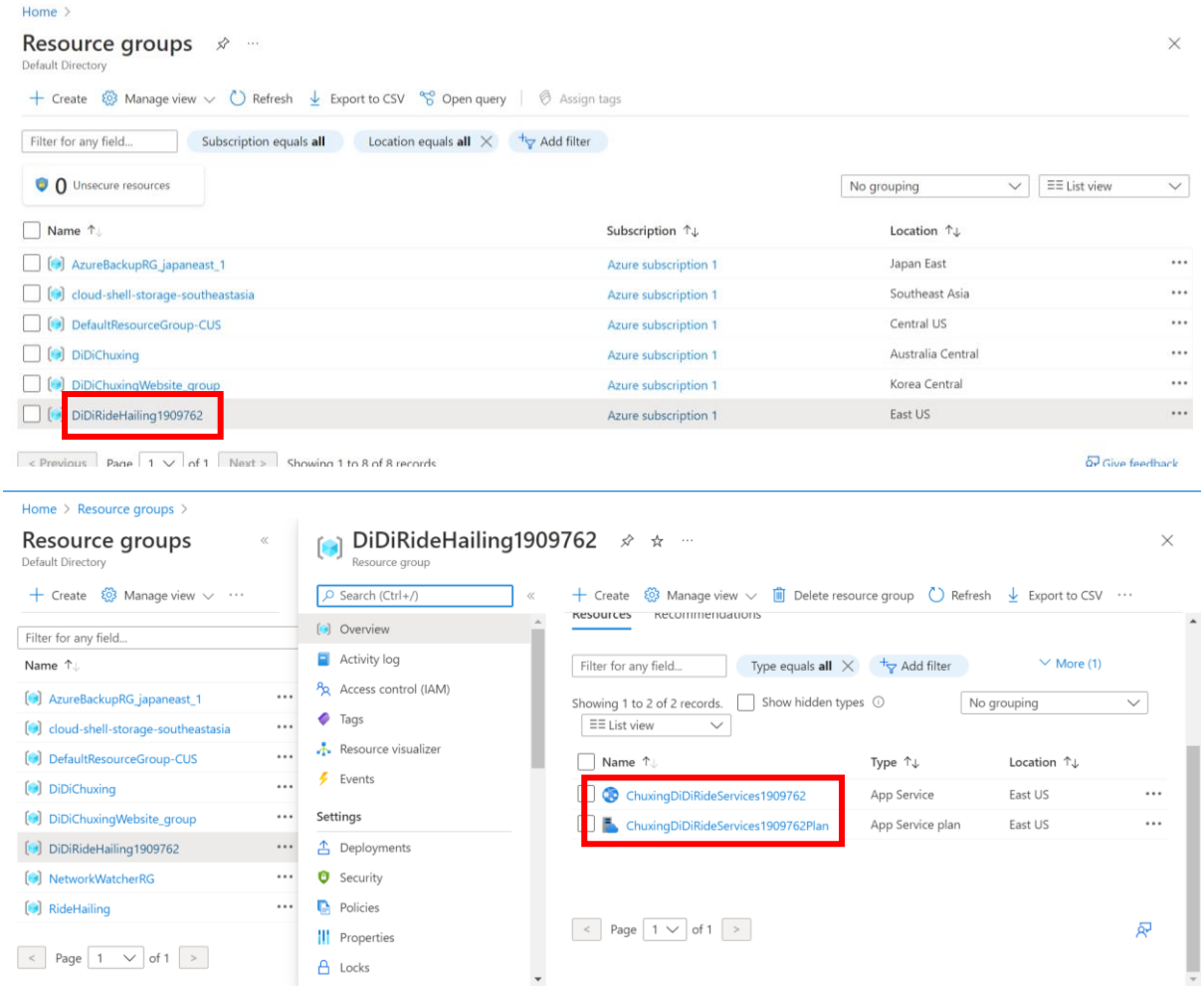




This DiDi Chuxing Ride-Hailing System is deployed on the Microsoft Azure cloud platform using ASP.NET web application. To do so, a ASP.NET Web Application (.NET Framework) is created on the Visual Studio. Once we have completed the code, the web application project will be published to Azure App Service (Windows).



The images above show that the website is deployed successfully in the cloud using the given URL (https://chuxingdidirideservices1909762.azurewebsites.net).



As you can see, the resources group we created is available on the Microsoft Azure Cloud Platform. Both App Services and App Services Plan are all presented in the "DiDiRideHailing1909762".

References

10 reasons why to choose Azure for your Enterprise. (2022). Saviant. https://www.saviantconsulting.com/blog/10-reasons-why-choose-microsoft-azure.aspx

Avram, M. (2014). Advantages and Challenges of Adopting Cloud Computing from an Enterprise Perspective. *Procedia Technology*, *12*, 529–534. https://doi.org/10.1016/j.protcy.2013.12.525

AWS vs Azure vs Google Cloud—Detailed Cloud Comparison. (2019, June 25). *Intellipaat Blog*. https://intellipaat.com/blog/aws-vs-azure-vs-google-cloud/

Carey, S. (2020, January 23). *AWS vs Azure vs Google Cloud: What's the best cloud platform for enterprise?* Computerworld. https://www.computerworld.com/article/3429365/aws-vs-azure-vs-google-whats-the-best-cloud-platform-for-enterprise.html

GeeksforGeeks. (2021, July 27). *Apache HBase*. https://www.geeksforgeeks.org/apache-hbase/

GeeksforGeeks. (2022, June 17). *Introduction to Hadoop Distributed File System(HDFS)*. https://www.geeksforgeeks.org/introduction-to-hadoop-distributed-file-systemhdfs/

Hajjat, M., Sun, X., Sung, Y.-W. E., Maltz, D., Rao, S., Sripanidkulchai, K., & Tawarmalani, M. (2010). Cloudward bound: Planning for beneficial migration of enterprise applications to the cloud. *ACM SIGCOMM Computer Communication Review*, *40*(4), 243–254. https://doi.org/10.1145/1851275.1851212

Kotas, C., Naughton, T., & Imam, N. (2018). A comparison of Amazon Web Services and Microsoft Azure cloud platforms for high performance computing. *2018 IEEE International Conference on Consumer Electronics (ICCE)*, 1–4. https://doi.org/10.1109/ICCE.2018.8326349

Mirghani, S., & Hajjdiab, H. (2017). Comparison between Amazon S3 and Google Cloud Drive. *Proceedings of the 2017 2nd International Conference on Communication and Information Systems*, 250–255. https://doi.org/10.1145/3158233.3159371

Pedamkar, P. (2021, March 3). *What is Yarn in Hadoop?* EDUCBA. https://www.educba.com/what-is-yarn-in-hadoop/

Shayan, J. (2014, January 21). *Identifying Benefits and risks associated with utilizing cloud computing*. arXiv.Org. https://arxiv.org/abs/1401.5155

Varia, J., & Mathew, S. (2013). *Overview of Amazon Web Services*. 22.

Wickramasinghe, S. (2019). *AWS vs Azure vs GCP: Comparing The Big 3 Cloud Platforms*. BMC Blogs. https://www.bmc.com/blogs/aws-vs-azure-vs-google-cloud-platforms/

新闻详情-滴滴官网. (2018, March 16). DiDi. https://www.didiglobal.com/news/newsDetail?id=323&type=news

APPENDIX 1

MARKING RUBRICS

Task	Criteria	(Below Expectation)	(Beginning)	(Developing)	(Exemplary)	Marks
1	Migration to Cloud	**0-5**	**6-10**	**11-15**	**16-20**	
		Unable to provide relevant analysis	Provide relevant analysis	Provide mostly relevant and correct analysis	Provide comprehensive and correct analysis	
2	Hadoop	**0-6**	**7-14**	**15-22**	**23-30**	
		Unable to provide relevant analysis	Provide relevant analysis	Provide mostly relevant and correct analysis	Provide comprehensive and correct analysis	
3	Selection of cloud service provider	**0-5**	**6-10**	**11-15**	**16-20**	
		Unable to provide relevant analysis	Provide relevant analysis	Provide mostly relevant and correct analysis	Provide comprehensive and correct analysis	
4	Cloud-based web hosting	**0-6**	**7-14**	**15-22**	**23-30**	
		No service has chosen to host the website. There is no explanation provided. The screenshot showed an unsuccessful deployment of the website on the chosen service.	A service has been chosen to host the website with no clear explanation. The screenshot showed an unsuccessful deployment of the website on the chosen service.	An appropriate service has been chosen to host the website with some explanation. The screenshot showed a successful deployment of the website on the chosen service.	An appropriate service has been chosen to host the website with a very clear explanation. The screenshot showed a successful deployment of the website on the chosen service.	
					Total	

Note to students: Please print out and attach this appendix together with the submission of coursework.