



AWS Architecture Design for Database Specialist

– Assignment #1 & #2

BRIEF

ABC company is a ride hailing company, they have large volume of subscribe users using their mobile app to get transportation services from local drivers. The mobile app for passengers and drivers will upload activities data to server for data analyst. ABC company has an operation team to manage the performance of each fleet.

Imagine that you are a Solution Architect from AWS, had a meeting with ABC company and you were required to design a proof-of-concept operation analysis platform. This platform is designed to illustrate the benefits to use cloud based database components.

The customer asked you to use the New York City Taxi and Limousine Commission (TLC) Trip Record Data from AWS public datasets as source data (<https://registry.opendata.aws/nyc-tlc-trip-records-pds/>). You are required to produce the following use cases from the sample data:

1. Business operation statics. You need to run query to illustrate the performance for each provider (vendorid). You need to summary the analysis of each vendor by date, week and month for the income/cost
2. Perform operational analysis for top performers
3. Establish a real-time analytic mechanism to generate alerts. The alerts in each location should be a continuous increase or decrease within 15 minutes. Since our data is in files, you need to create some mechanism to do the show case
4. Visualization of key operation metrics. You need to provide week to week, month to month, quarter to quarter and location to location trends analysis based on your design

Note: the data format in the open data is not consistent from Y2015 to Y2018. You need to do some pre-treatment to polishing the data. You could just use Y2017 and/or Y2018 data for yellow or green cabs. Detail description of meanings could be found on NYC transportation agency website.

This is a customer demo. So, besides the application/data stores, you need to provide a presentation to walkthrough your solutions with proper design principals, codes, architecture and key assumptions. You need to submit the artifacts together with the hosted environment on AWS. Please focus to use AWS hosted database platforms for the assignment.

The architecture design would like to include:

- ✚ Key components on cloud to support the deployment of customer access
- ✚ Use AWS manage services as wisdom choice
- ✚ AWS services to support elastic of 50 million users and 10 thousand drivers
- ✚ Security, reliability and performance services to support the deployment

OBJECTIVE

Recommend a manageable, secure, scalable, high performance, efficient, elastic, highly available, fault tolerant and recoverable architecture that allows ABC to grow with AWS.

Build a demo system to prove the concept you design.

DELIVERABLES

(1) A **well-written** document to illustrate your great works. You could put diagrams in appendix, which will not occupy the total length.

- Proposed **architecture diagram** (you could use [AWS Products Icon](#)), give a detailed description for your architecture diagram and explained why you choose this solution.
- Clearly state all **assumptions** made during the design and explicitly state the referenced **Amazon Web Services**.
- If you would like open source components together with AWS services, please describe the reason and identify where the reviewers could find the reference on internet in appendix.

(2) A **workable Demo deployed on AWS** demonstrating data ETL and data visualization.

- A deployable map-reduce package, which performs ETL action on source data, with detail description to deploy it.
- Source code of the deployable map-reduce package.
- Visualization result of the formatted data: select one or more visualization formats which can demonstrate the business value of the data.
- Source code of visualization component if any program language is used, or detail description of the steps if any UI tools is used.

Once complete, send the document and demo to manyuson@amazon.com.
