

MSiA 423: Cloud Engineering For Data Science

Assignment -1

Individual Assignment (100 points)

Instructions

- Include your full name in the title slide of the presentation
- The focus of this deck should primarily be around technology architecture and cost estimates
- Make your own assumptions about overall storage requirements
- Make your own assumptions about analytical needs and the involved computation
- Make your own assumptions around SLAs, server uptime, etc. based on industry/market research
- You may assume the use of additional technologies beyond the ones discussed in the classroom
- You may research reference architectures, but you need to create your own diagram

Problem

Hospital costs are rising partially because of [high readmission rates](#) within 30 days of patient release. Readmission rates have long been a trusted measure of effective and responsible care and have become a primary assessment driver in the healthcare industry. How can data and analytics be leveraged to reduce readmission rates and improve overall health outcomes for patients.

Solution

Your goal as a data science consultant is to design an analytics platform that will meet the analytics needs of a large health care provider with 500,000 customers (patients) around the world. The solution would identify at-risk patients based on past history, chart information, and patient trends. The health care provider should be able to use this data to identify at-risk patients and provide the necessary care to improve patient care by reducing readmission rates.

Provide your solution as a powerpoint slide deck (no more than 10 slides) that address the following :

1. What type of data would you want to store about each patient and how often is it refreshed
2. What type of data sources would feed into the analytics platform
3. What type of analytics would you like to perform on the patient data and how often. Provide a list of data science solutions that would drive decision making.
4. Using AWS as your primary cloud, provide reasons for your choice of cloud services for the following:
 - a. Data Ingestion
 - b. Data Storage and Data Processing
 - c. Data Warehousing and BI
 - d. Business Intelligence and Analytics
 - e. Data Science - ML/AI
 - f. Applications – Webapps, Mobile
5. Data Sizing : estimate the total data size requirements of all the patient data and addition
6. Capacity Sizing : number of required EC2 virtual machines and other compute infrastructure
7. You will need to get the IT funding approval for the project. Use a pricing calculator ([AWS](#)) to estimate monthly costs of the end to end solution.
8. Design a customized [system architecture diagram](#) for this big data solution. This could be used by the IT systems engineers to build out the infrastructure to meet your needs.