Data and Analytics Case Study

Slalom New York City Data and Analytics Practice

Jan 2020

slalom

What we typically do in D&A

Our Data and Analytics experts are called upon to service our clients in a variety of ways. These can include:

1. Serving as a Data Analysts for a complex data mapping exercise creating Traceability matrix, Data Flows, Lineage etc.
2. Acting as a Data Engineer building Data Engineering code base that integrates data from source to target application/s
3. Acting as a Visualization Engineer providing data insights and metrics that can be used for operations and analytics
4. Serving as Data Scientist to obtain remarkable data insights to help marketing KPIs, customer segmentation etc. leveraging machine learning techniques
5. Defining end to end architectures, best practices and guidelines for the customer
6. Design and implement Data engineering platforms to execute the best in class solutions using various platforms, tools, services and languages
7. Serving as a leader and mentor to the team

Case Study Introduction

PROBLEM STATEMENT: We are building an **E-commerce** enterprise data platform in Cloud for Customer 'Slalom' to consolidate their segregated platforms into one environment in cloud that reduces their operations cost and helps perform an efficient analytics which they are unable to perform today.

SOLUTIONS: Please present your story using **one** of the below **scenarios**:

Analyze, Act and React!

Build a platform to enable machine learning analytics against their data, to influence segmentation and targeting

Derive data insights!

Build insights to provide key information about the customers and products, which helps increase sales and revenue

Modern Data Platform!

Build a platform to consolidate data into an efficient model within a cloud platform & database, that enables efficient data delivery to end consumers and applications for analytics

slalom

About sample data

For any scenario that you have picked, you can use the below sample data for mockups / prototyping.

The key dataset is transactional consumer data related to purchases and incentives. These datasets can be download HERE. It would be consumed in your solution using vendor provided APIs on daily basis. Data volume – 5 GB/Day.

In future:

- Platform will also receive social media datasets in semi-structured format using both Push (vendor pushes to cloud storage) and pull (using API) methodologies
- Integrate with on premises Master databases that provides customer demographics & product catalog information (no sample data available). Data volume – 1 TB total. Preferred data access method: pull

When solutioning, please keep in mind that:

- There are multiple ways to solve this and that there is no one right answer! We are looking for your understanding of the scope and approach to the solution, not just the answer
- Your case study interview will last 60 minutes and there will be Q/A with the interviewer(s), so plan your time accordingly
- You will have at least 7 days to put this content together. The effort estimate for preparation is about 8-12 hours
- Please send the deck and any code base to your Slalom Recruiter before your scheduled interview and present during the interview

Deliverables

Please choose any 4 deliverables from the below to solution the scenario that you have selected.
Notice that some deliverables belong to a specialty area, eg: #3 for Data Engineering, #5 for Visualization, #7/8 for Advanced Analytics, while others are more generic. We want you to choose the right deliverables that makes your case study robust.
Given that you have exposure on multiple areas, we would like to understand that in your solution, otherwise it is acceptable if your solution is focused on one specialty area.

1. Technical Rationalization - Which platform, technologies, tools, services and language would you choose and why?
2. Conceptual Data model for Data Lake, ODS and DWH layers describing how you will structure the data to fulfill all requirements
3. Data Integration approach to manage smooth hand offs between data feeds and interfaces. Explain your design for a robust error handling. Explain your design for Audits and troubleshoots. Please explain your strategy and approach, coding is not expected.
4. Architecture diagram describing the services you have identified and how you envision an end to end operational platform.
5. Pease visualize the data by creating an application within Tableau, QlikView or QlikSense, or Power BI. Feel free to highlight your skills by using the highest level of sophistication.
6. Describe approach towards productionization of your solution
7. Use the sample datasets to predict the customer loyalty and influence targeting customers for future promotions, using Python / (Tableau, QlikView or QlikSense, or Power BI)
8. Design a model to influence customers with similar products recommendations based on their purchases, using Python / (Tableau, QlikView or QlikSense, or Power BI)
9. Think out of the box: any additional perspectives about the data that can make your case study look more appealing?

- Be perspective in describing the solution approach with key assumptions and Highlight best practices leveraged in your solution
- We want to understand your approach and thinking abilities. Please use Power Point to represent your case study, with any accompanying code base
- You can integrate other technologies or supplemental datasets to tell your story with the data

Tips

- Case study samples:
  - Tropical Smoothie Story: https://www.slalom.com/work/customer-story/tropical-smoothie-data
  - LiveRamp Marketing Story: https://www.slalom.com/work/customer-story/liveramp-marketing-attribution
  - Modern Data Load Accelerator Story: https://medium.com/@sharmaric/https-medium-com-sharmaric-slalom-accelerator-for-snowflake-data-ingestions-1eb62a5dc549