Introduction to big data architecture

Why big data

- Volume
- Variety
- Velocity
- Extensibility (Schemaless)

Use case

- A company wishes to provide its employees medical coverage. So, they create medical plans tailored to the employees needs. Each plan consists of large number of covered services, e.g. acupuncture, physical, well-baby visits, emergency room visits, and so on. Additionally, each plan specifies the cost associated with that plan. For example, the co-pay for the various visits, and any deductible that should be met before the patient is reimbursed.
- the company has created a website for its employees where they can view each medical plan and the covered services associated with the plans. Additionally, the website is also used by the plan administrators to create new plans and modify existing plans.
- Is this a use case for big data?

How can we tell?

- Start by asking a few questions:
  what is the data size of a medical plan?
  What does a medical plan look like? That is, how can we model a medical plan?

  Other factors?
  How many people are viewing the website? E.g. throughput rates, latency requirements?

  Is there a need to batch import/export plans from the system?

Use case continued

- The analyst responsible for plan creation wants to quickly modify any plans that he created with additional attributes. For example, the analyst may want to remove services and add services to the plan. The analyst may also wish to extend any plan with additional attributes that may not have been foreseen during the design of the system.
- The question is: how can we extend the definition of a plan?

Technical requirements so far

- Need for data modeling
- Need for CRUD APIs
- Need for batch APIs
- Need for data extensibility
- Need for data validation

Use case continued

- While an analyst editing a plan, this plan must not be visible to employees. Furthermore, other analysts may view, but not edit, this plan
- Hence, the need to secure the system with authentication, and authorization support

Use case continued

- An employee using the system may find the medical plan that best fit his/her needs by using the search box. A user may search on any attribute

- Technical requirement:
  - need for search

Assignments

- Json schema: http://json-schema.org/
- Json: HTTP://json.org
- JSON Parser; JSON Simple
- Jsonpath
- Springboot
- marwansabbouh@gmail.com