Homework 1

1

> In class, we mentioned the 3 (or 4) V's of Big Data, two of which are volume and velocity
> What is the third V (Googling is ok)? Why is it an important part of Big Data?

Variety
Because in Big data, many data is unstructured data or semi-structured data. These data can be different types and in different forms including images, sounds, texts and videos. So we need different approach as well as different techniques to store all raw data.
Reference

> Choose a Big Data use case that is of interest to you
> Analyze its requirements around each of the V's?

Facebook to store their social media content.
Volume: 1.13 billion people daily active users are posting to facebook everyday which geneate a large volume of data.
Velocity: Every 60 seconds, 293,000 statuses are updated on facebook which need the system to store and handle data very quick.
Variety: Social media content including comments, photos, videos and more.

3

> Come up with some simple characterizations of what applications the lambda

> architecture is good for and bad for

Lambda architecture is good for Big Data system which has massive quantities of data and has requirements on fault-tolerant, latency and throughput.
Lambda architecture is not good for applications not has so much data, or applications require very high accuracy like financial system, or applications with simple structured data.