Volume

Variety

Velocity

Veracity

Value

Visualization

Viscosity

Virality

Excel – structured

XML, json – semi-structured (self-describing)

CSV – quasi-structured (formatted with effort)

PDF – unstructured

For Mid Term

* - meaning – probable questions in exam
* Format for application of big data concepts for **use cases** on **all 8 Vs** (will be asked in midterm for **5 marks**)
* Volume: Quantification of data (gb/tb) (xx\*yy\*zz format also OK)
* Velocity: Comparison (e.g., data generated Last yr vs this yr)
* Variety: Variety of data (e.g., Text, videos, images for amazon)
* HDFS Architecture
* HDFS Read
* HDFS Write
* How MapReduce Works
* MRv1
* MRv2
* Map-reduce Failure Recovery
* (Pig) Group by, union, split

10 subjective, 10 programming, 5 MCQ – 45 minutes

For End Term

Pig + Hive

Theory + Practical

Theory = 20 marks

Practical = 20 marks (5 queries to fire)

Scoring the marks