

1. **Ingestion – 1-2 man?**
   * Import US stocks data (5 days timeframe) from Yahoo finance (possibly via Python and other tools?) on Hadoop
   * Propose a data structure/frame to ingest the data (e.g. date, stock, daily returns as variables)
2. **Storage – 1 man?**
   * Write the data imported into file on HDFS
   * Ensures python libraries supported on Hadoop and writing into HDFS.
   * Can work with sample csv first while someone else works on ingestion
3. **Big Data processing – 2 man?**
   * Read data from HDFS
   * Write Spark computing job
   * Write basic code to compute correlations (derived variables) between stock returns
4. **Graph analytics – 1 man?**
   * Write code to analyse stock returns (to plot stocks as nodes and correlations (> 0.7) as arcs?) in terms of graphs
   * Ensures appropriate libraries (R/Python) supported on Hadoop platform