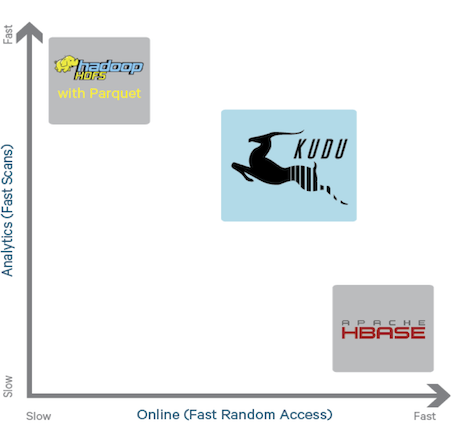
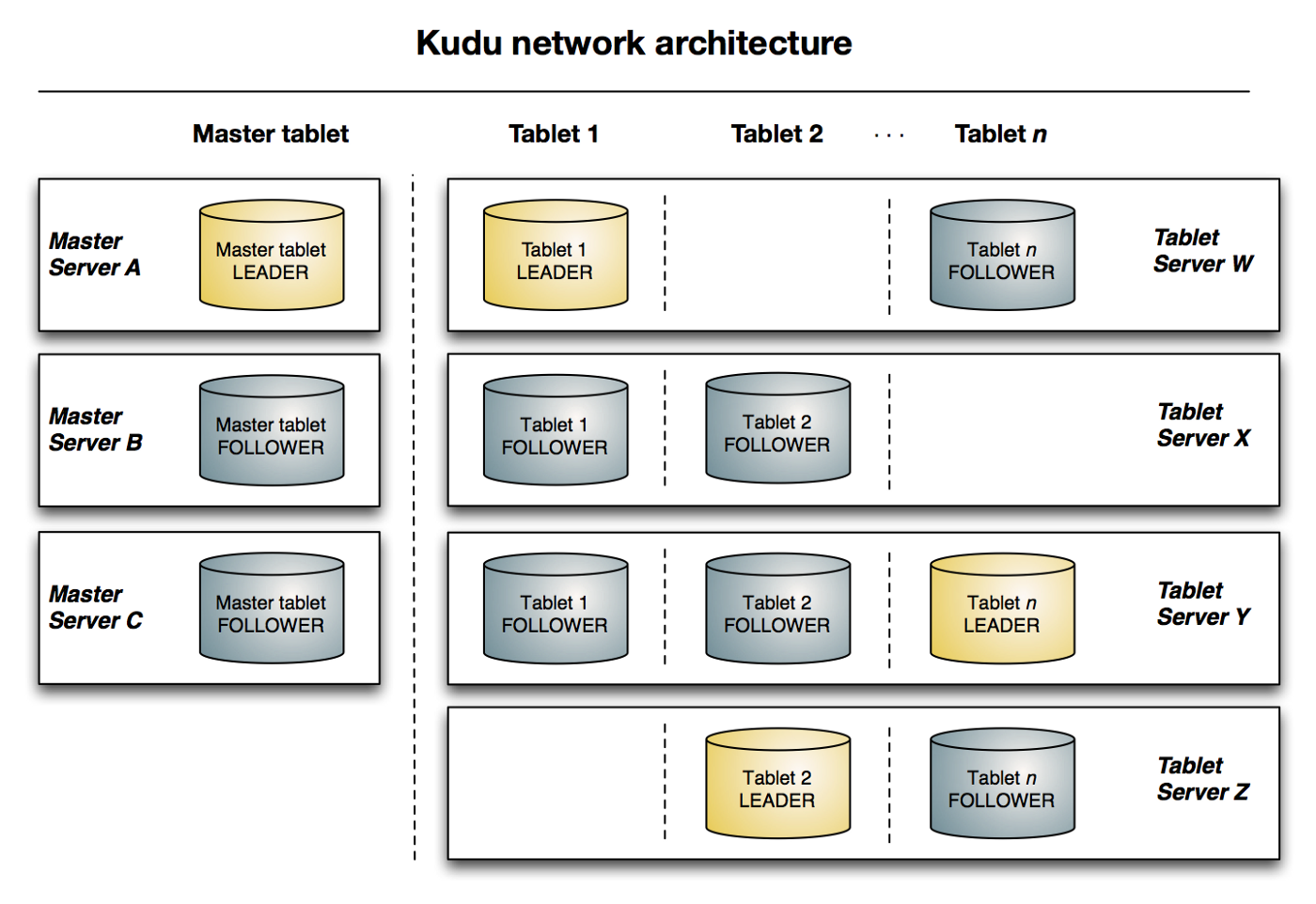
11/5 Round 2 Day 6 – **Apache** **Kudu**

* Apache Kudu is a free and open source columnar storage system developed for the Apache Hadoop
  + An engine intended for structured data that supports low-latency random access millisecond-scale access to individual rows together with great analytical access patterns
  + It sits between HBase and Impala with Parquet, attempting to remove the trade offs between quick scans and rapid random access
* Main advantages of Apache Kudu
  + Enables real-time analytics on fast data
    - Lookup for a certain value through its key
    - Lookup for a range of keys that have been sorted in a key-order
    - Carry out arbitrary queries across as many columns as needed
  + Fully distributed and fault tolerant
    - RAFT consensus algorithm
    - Horizontal scaling
  + Takes advantages of the upcoming generation of hardware
    - Optimized for SSD
  + Provides the mutability required for BI on big data
    - Slowly changing dimension (SCD)
    - Allows the user to keep track of changes inside a dimensional reference data
  + Kudu supports SQL if used with Spark or Impala
    - Connect to JDBC or ODBC



*  Resources referred:

<https://databricks.com/glossary/apache-kudu>

<https://lewisdgavin.medium.com/apache-kudu-in-5-minutes-91e9371a4f8f>

<https://kudu.apache.org/>