Background:

**HBase**: open-source, non-relational, distributed database modeled after Google’s Bigtable

1. Highlight:
2. Wiki: “HBase is a column-oriented key-value data store and has been idolized widely because of its lineage with Hadoop and HDFS”
3. Wiki: “HBase runs on top of HDFS and is well-suited for faster read and write operations on large datasets with high throughput and low input/output latency”

**Cassandra**: free and open-source, distributed, wide column store, NoSQL database management system

1. Highlight:
   1. Reliability: “treats failures as the norm rather than the exception”
   2. Efficiency
   3. Scalable
2. Facebook released Cassandra as an open-source project on Google code in July 2008
3. Why are they created at the first place?
   1. **HBase**: Apache HBase began as a project by the company *Powerset* out of a need to process massive amounts of data for the purpose of natural language search
   2. **Cassandra**: Facebook designed Cassandra to fulfill the storage needs of the Inbox Search problem
      1. Inbox Search enables users to search through their Facebook Inbox. The system was required to handle a very high write throughout, billions of writes per day, and also scale with the number of users
4. Commons between HBase and Cassandra:
   1. Open-source
   2. High throughput
5. s