**Redis** (remote dictionary server)

redis.io/commands

In-memory db, i.e. write to the memory and cache, also have options to write to the disk

Redis is a Key-Value store (Value can be any data structure), NO-SQL db

String

Lists

Sets

Sorted Sets

Hashes

Bitmaps

Hyperloglogs

Geospatialinexes

|  |
| --- |
| SET “name” “john”  Get “name” |

Ad: fast, simple & flexible (no need for table), durable

A close up of a device

Description automatically generated

Master-slave feature

A screenshot of a cell phone

Description automatically generated

**Hadoop**

Run on a cluster, for distributed storage and distributed processing of very large data sets

**Hive**

Hive provides SQL intellect, so users can write SWL like queries called HQL (hive query language) to extract the data from Hadoop



Can use:

Hive can be used for OLAP (online analytic) processing

Hive is a platform for the SQL users to interact with the large datasets that reside on HDFS filesystem

Cannot use:

A picture containing text

Description automatically generated

Features: 1. Supports different file formats: sequence file, text file, avro file format, ORC

file, RC file

2. Metadata is in RDBMS

3. Compression tech

4. HQL

5. UDF (user define function)

6. Specialized join

A screenshot of a cell phone

Description automatically generated

**Apache Spark**

Spark is a fast and general-purpose cluster computing system for large scale data processing.

Has high-level APIs in Java, Scala, Python and R

A screenshot of a cell phone

Description automatically generated