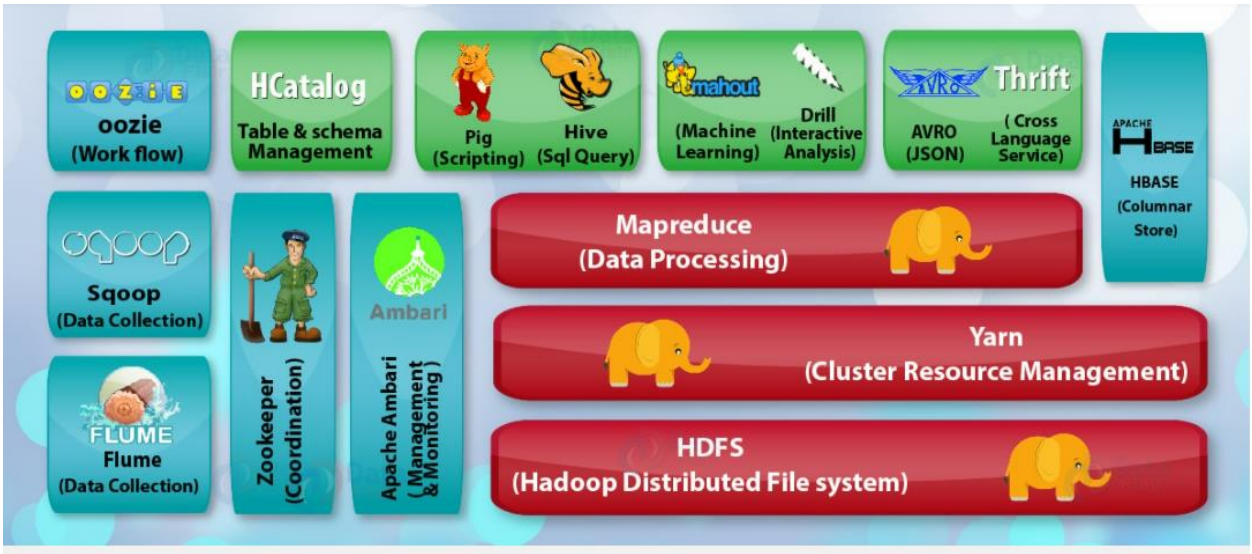
HADOOP ECOSYSTEM



In Hadoop ecosystem there are tons of components.

CORE:

1. **HDFS** – world's most popular data store
2. **YARN** – resource management layer, which manages/allocates/releases resources of the cluster
3. **MapReduce** – distributed computing model, which utilizes the power of distributed computing to process the data at lightning-fast speed.

OTHER IMPORTANT COMPONENTS:

1. **Hive –** Data warehouse on the top of Hadoop, which provides simplicity of SQL with the power of Hadoop.
2. **Pig –** top-level data processing engine, which enables users to run a script to process/parse data.
3. **HBase –** a column-oriented NoSQL DB, which handles the data with random read/write.
4. **Drill –** Schema-free SQL Query Engine, which provides faster insights without the overhead of data loading, schema creation.
5. **Mahout –** Scalable machine learning library on top of Hadoop, which provides ML algorithm at a massive scale.
6. **Flume –** Data collection system, which provides real-time collection and aggregation of Big Data.

7. **Ambari –** Installation and configuration tool, which can be used for deployment, management, maintenance & monitoring tool.
8. **Sqoop -**Sqoop imports data from external sources into compatible Hadoop Ecosystem components like HDFS, Hive, HBase etc. It also transfers data from Hadoop to other external sources.
9. **Flume -** It is a service which helps to ingest structured and semi-structured data into HDFS. Flume works on the principle of distributed processing. It aids in collection, aggregation, and movement of a huge amount of data sets.