데이터 엔지니어링하둡에코 시스템 하둡에고 시스템 하답에고 시스템의 이해

Park Suhyuk



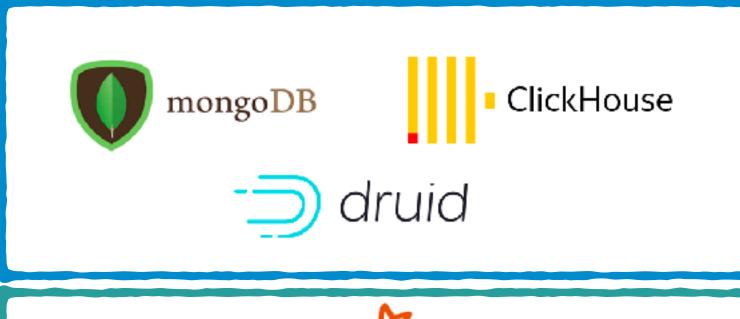




Core Ecosystem

데이터 저장 및 처리 시스템

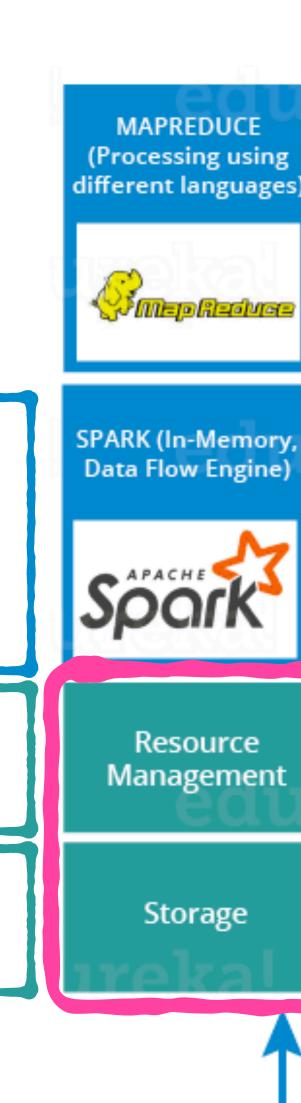
Hadoop Ecosystem



Spark Streaming

ထို kafka











MAHOUT &

SPARK MLlib





SOLR & LUCENE

(Searching

& Indexing)



HBASE

(NoSQL Database)









ZOOKEEPER

& AMBARI

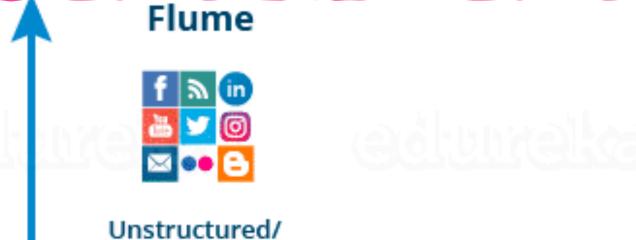
(Management

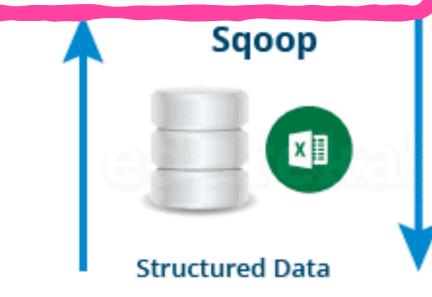
& Coordination)



Semi-structured Data







Batch Processing

배치 처리 프레임워크 및 도구 - Sqoop, Flume, Spark, Hive, Pig

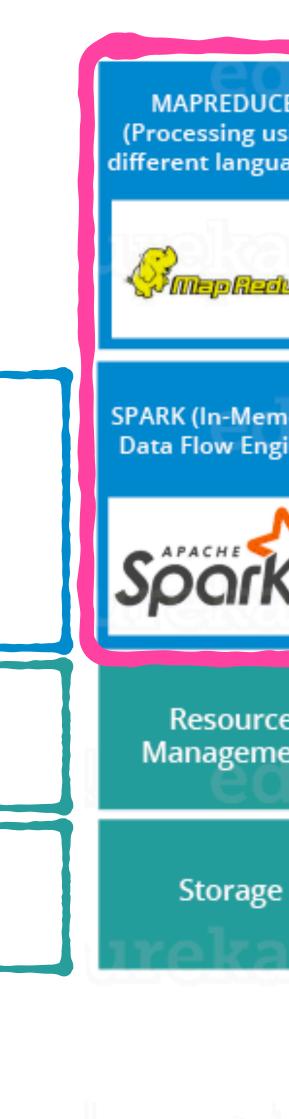
Hadoop Ecosystem

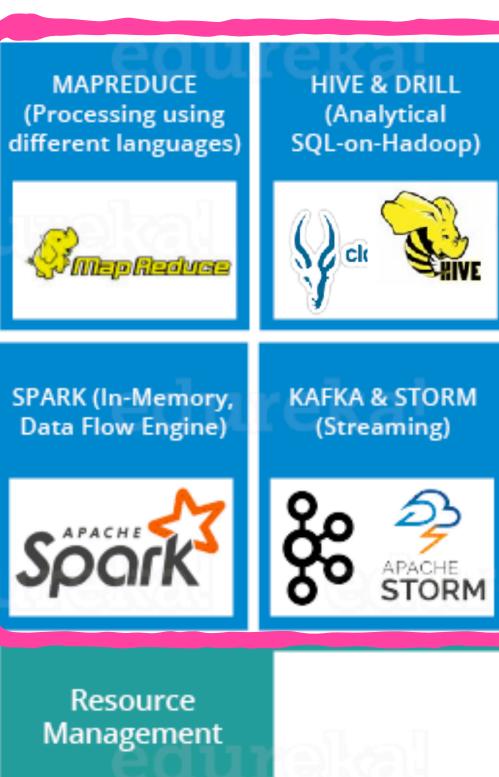


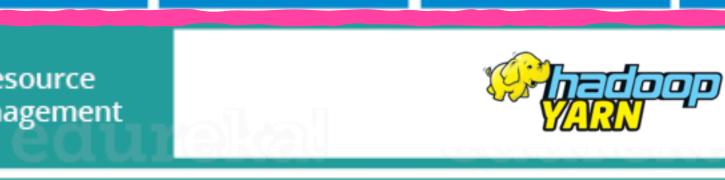
Spark Streaming

ထို kafka

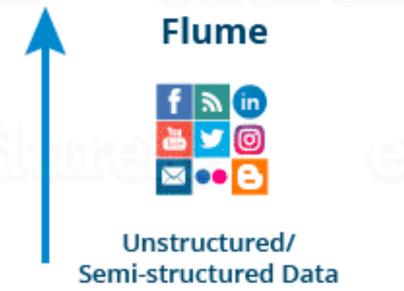








itorage





MAHOUT &

SPARK MLlib

(Machine learning)

mahout

Spark

MLlib







HBASE

(NoSQL Database)

HBASE





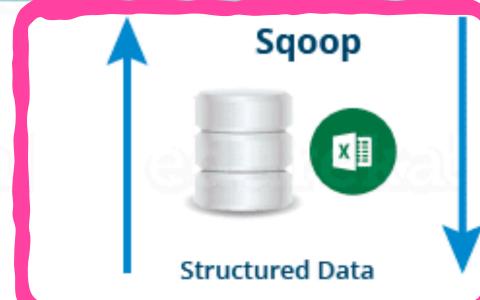


ZOOKEEPER

& AMBARI

(Management

& Coordination)

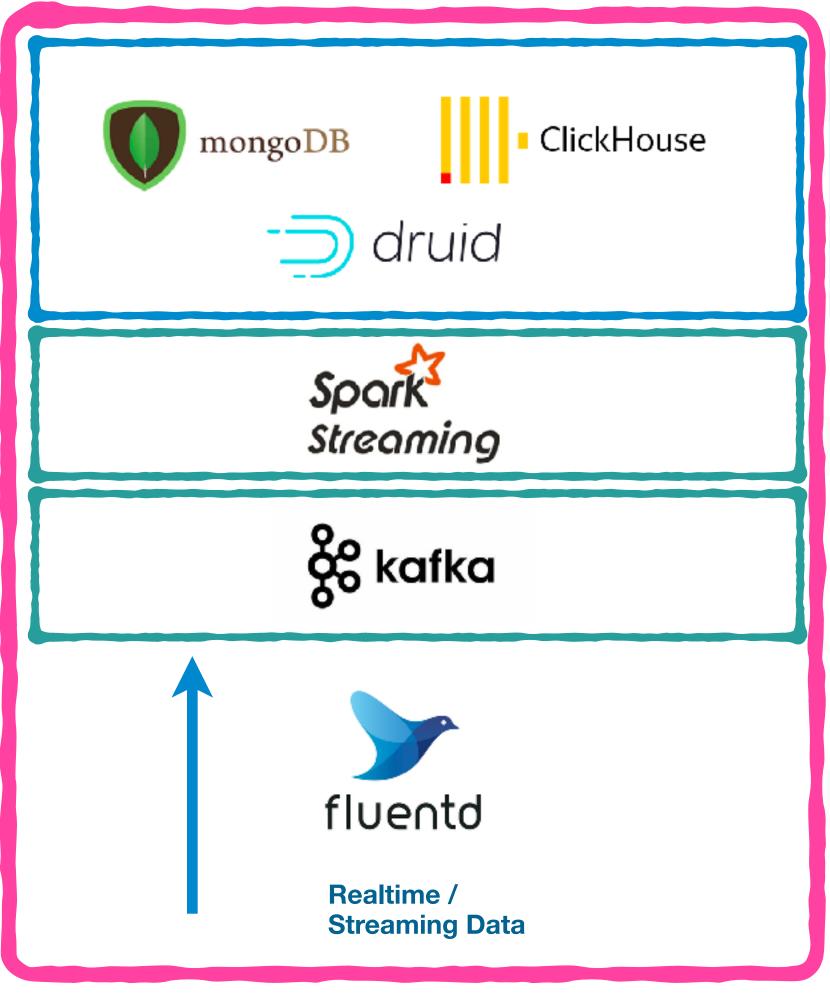


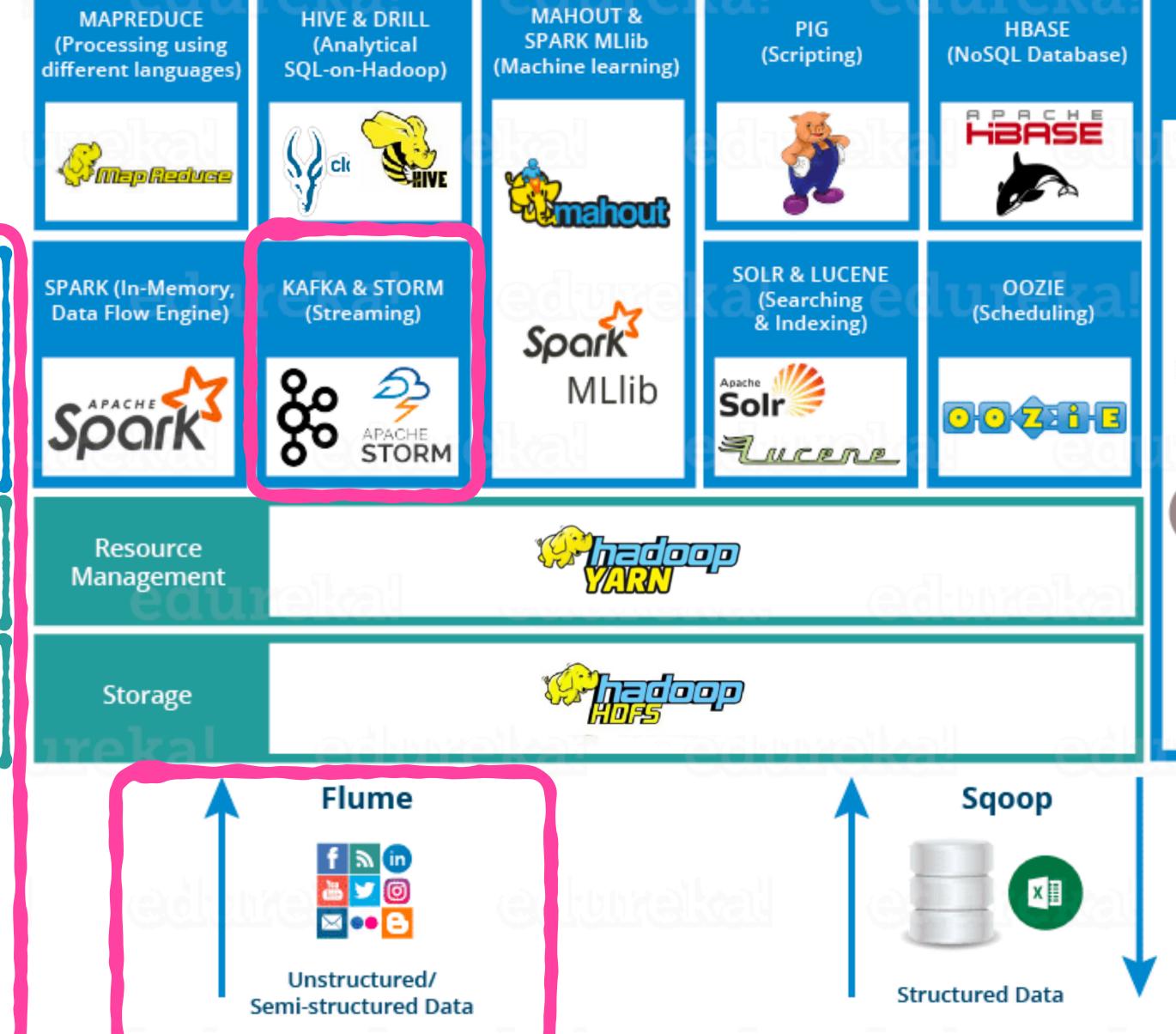


Stream Processing

스트림 처리 프레임워크 및 도구 - Fluentd, Kafka, Spark Streaming, Druid

Hadoop Ecosystem





ZOOKEEPER

& AMBARI

(Management

& Coordination)

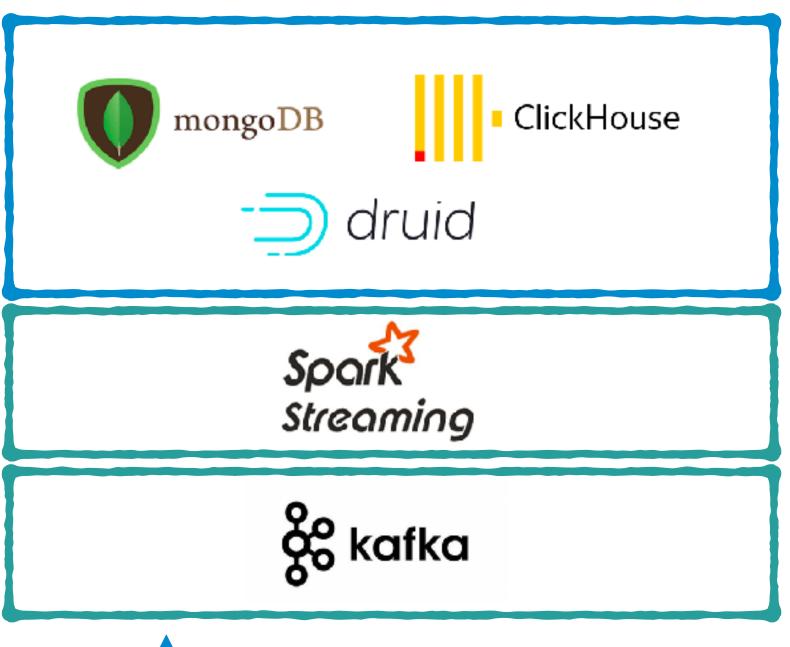
Apache

Ambari

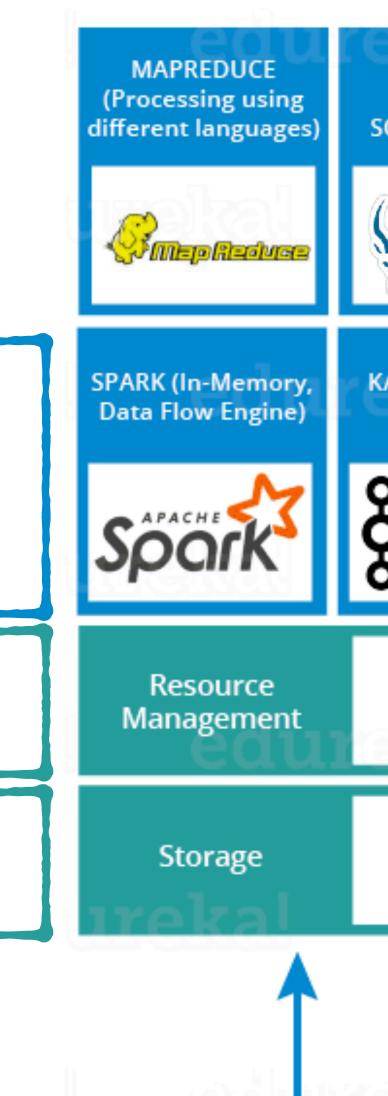
Others

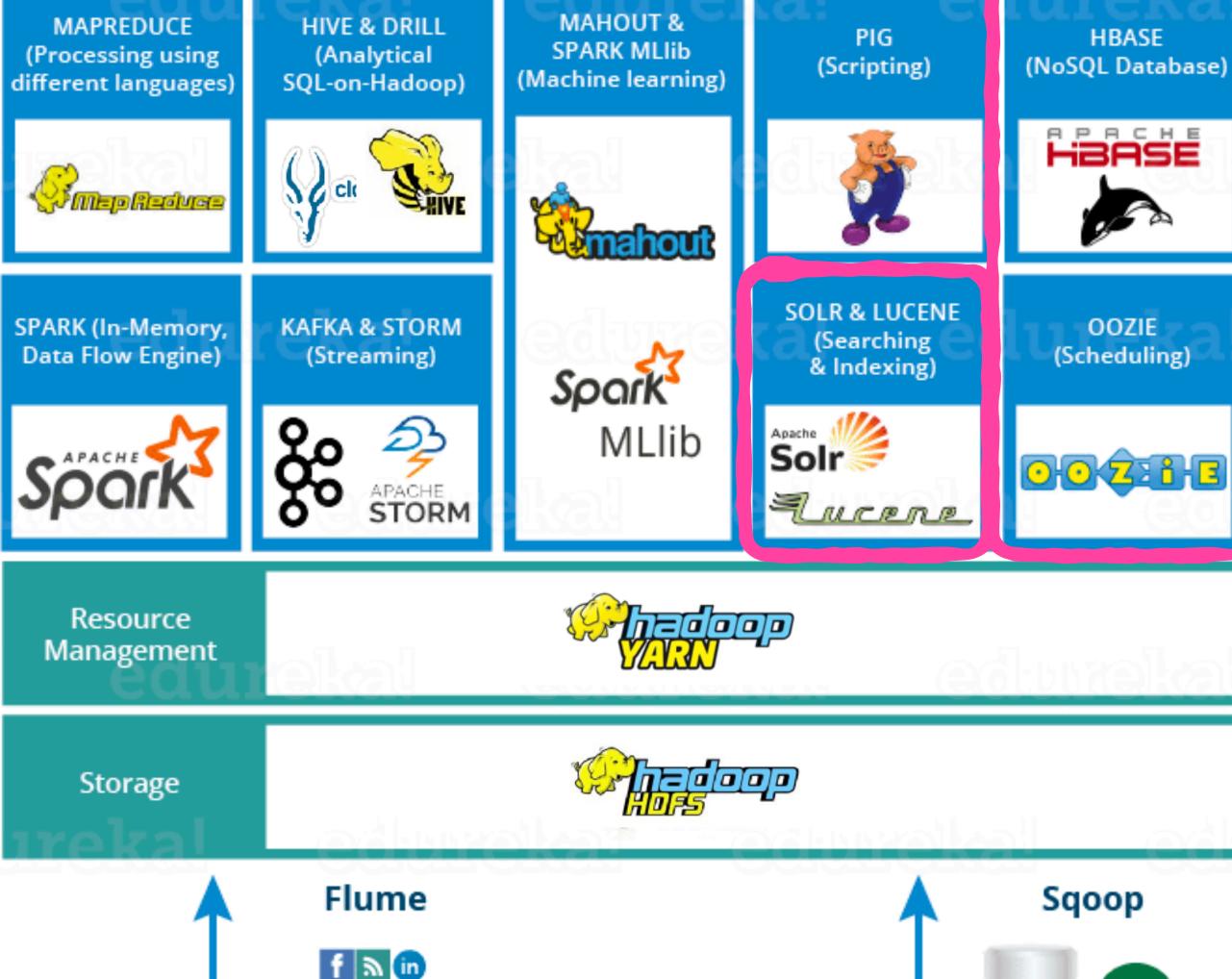
코디네이팅, 데이터 조회, 배포, 스케줄링 및 보안

Hadoop Ecosystem











ZOOKEEPER

& AMBARI

(Management

& Coordination)

Apache

Ambari