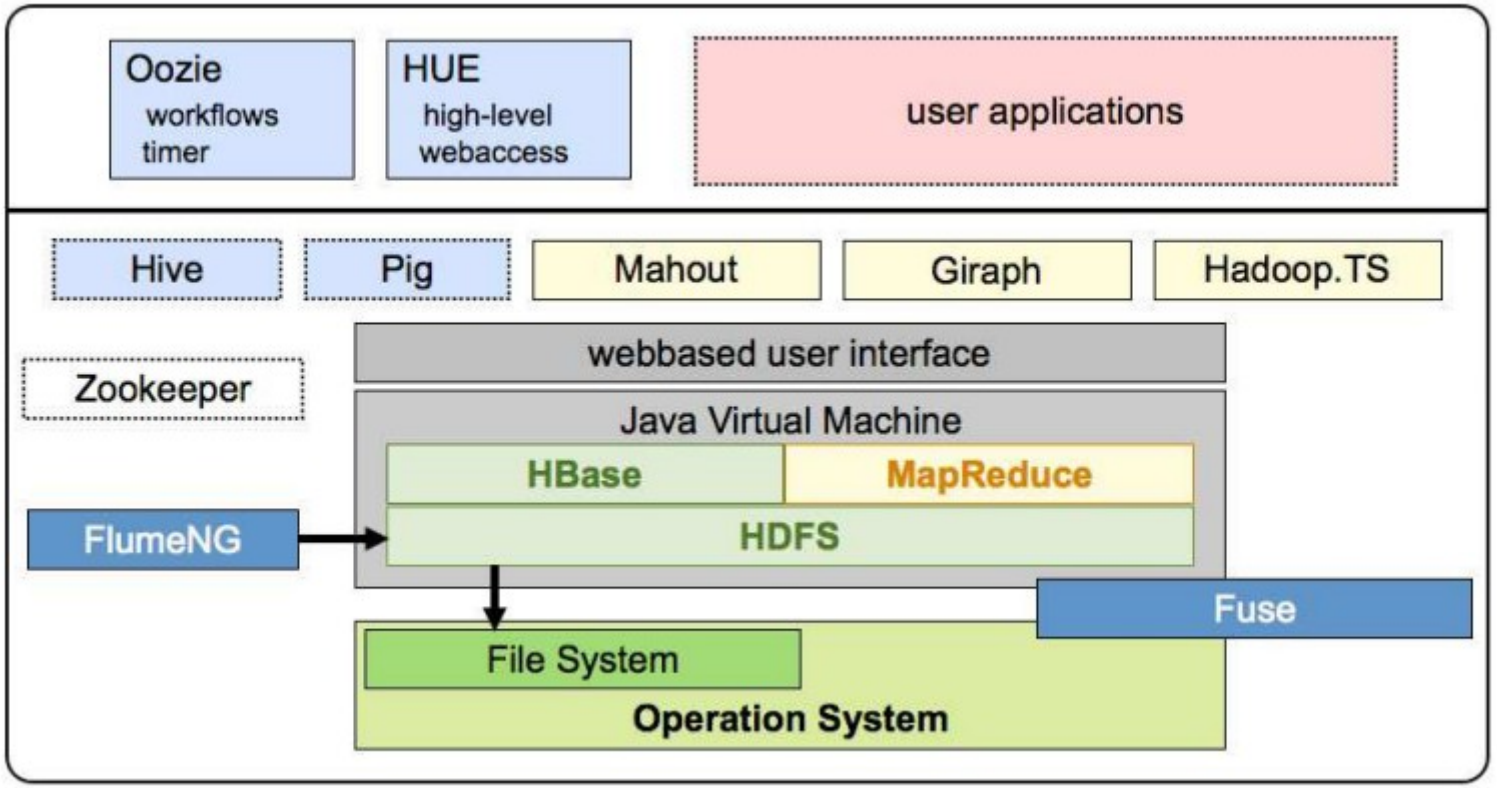
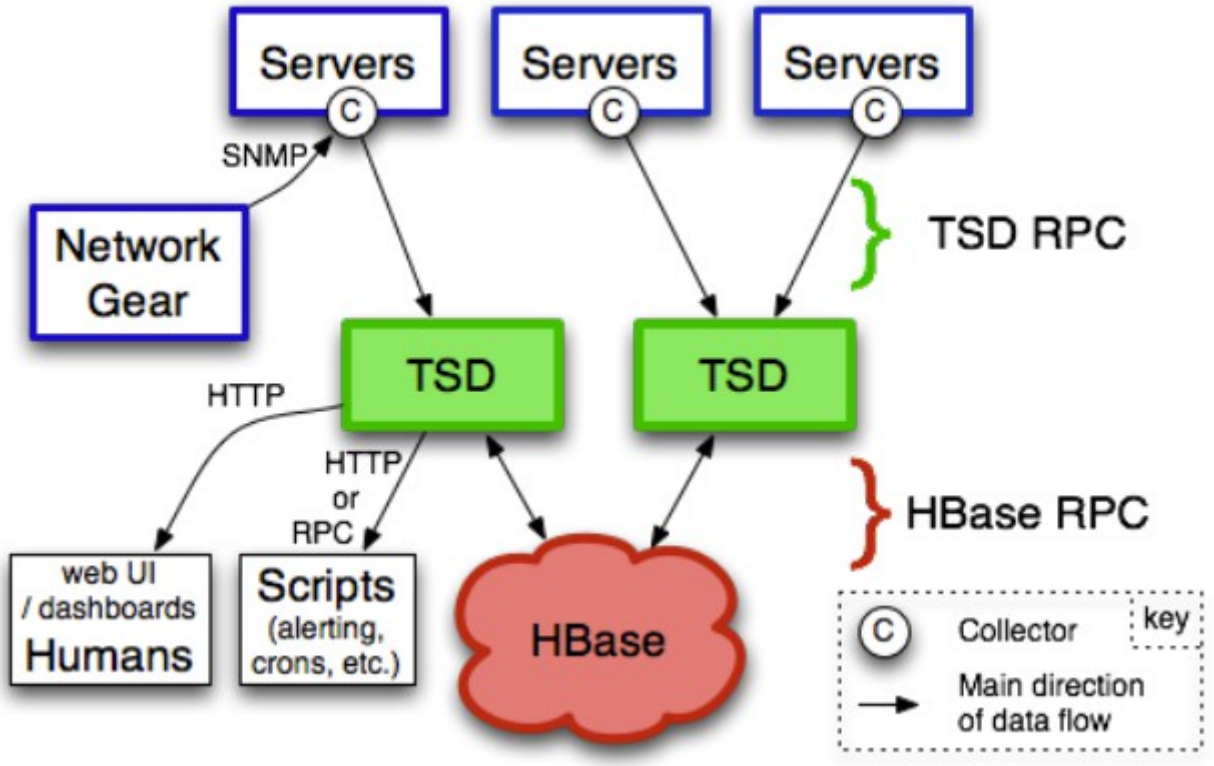
HadoopTS


Oozie
workflows
timer
HUE
high-level
webaccess
user applications
Hive
Pig
Mahout
Giraph
Hadoop.TS
webbased user interface
Zookeeper
Java Virtual Machine
HBase
MapReduce
FlumeNG
HDFS
Fuse
File System
Operation System

HadoopTS

- Use HBase as the underlying data store
- TSBucket: time-series bucket saves the times series data of equal resolution and length
- TSProcessor: Controls the analysis flow for all individual opeations on all records within TSBucket.
- UDF: Integrate HadoopTS with other systems, such as Hive, Pig.
- Time series analysis:
  - Auto correlation
  - Detrended fluctuation analysis (DFS)
  - Return Interval statistics
  - Cross correlation
  - Event synchronization (ES)

OpenTSDB


Servers
Servers
Servers
C
C
C
SNMP
Network Gear
TSD RPC
TSD
TSD
HTTP
HTTP or RPC
HBase RPC
web UI / dashboards
Humans
Scripts (alerting, crons, etc.)
HBase
C
Collector
key
Main direction of data flow

Hbase Tables

- tsdb - Data point table. Massive
- tsdb-uid - Name to UID and UID to name mappings
- tsdb-meta - Time series index and meta-data (new in 2.0)
- tsdb-tree - Config and index for heirarchical naming schema

HBase Tables

- Integer UIDs assigned to each value per type (metric, tagk, tagv) in *tsdb-uid* table
- 64 bit integers in row *\x00* reflect last used UID

CF:Qualifier	Row Key	UID
id:metric	sys.cpu.user	1
id:tagk	host	1
id:tagv	web01	1
id:tagk	cpu	2
id:tagv	0	2

Data table schema

- Row key is a concatenation of UIDs and time:
  - metric + timestamp + tagk1 + tagv1… + tagkN + tagvN
- sys.cpu.user 1234567890 42 host=web01 cpu=0
  \x00\x00\x01\x49\x95\xFB\x70\x00\x00\x01\x00\x00\x01\x00\x00\x02\x00\x00\x02
- Timestamp normalized on 1 hour boundaries
- All data points for an hour are stored in one row
- Enables fast scans of all time series for a metric
- …or pass a row key regex filter for specific time series with particular tags