

业界动态

2016.04.20

韩岳岐

hanyueqi@cnic.cn

<https://github.com/sens2010/bigdata>

Apache Sentry项目

- Cloudera发布，经两年多孵化变为定级项目
- Hadoop开源组件，细粒度级、基于角色的授权以及多租户的管理模式
- Hive/Hcatalog、Solr、Impala集成
- 加强Hadoop中数据保护策略
- <http://sentry.apache.org/>

APACHE SENTRY

Apache Sentry is a system for enforcing fine grained role based authorization to data and metadata stored on a Hadoop cluster.

Apache Sentry has successfully [graduated](#) from the Incubator in March of 2016 and is now a Top-Level Apache project.

How to get started:

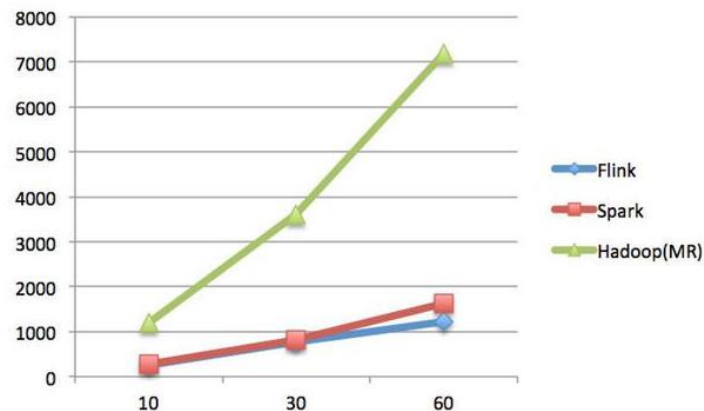
- Read a [tutorial](#)
- [Contribute](#) a patch
- Reach out on the [mailing lists](#)



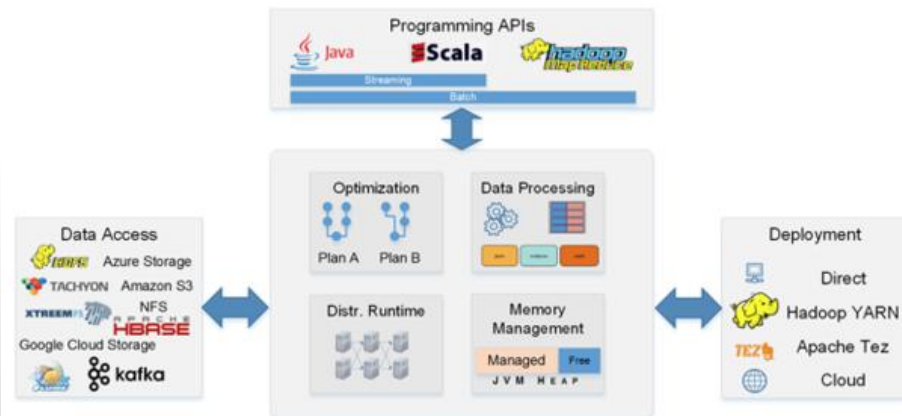
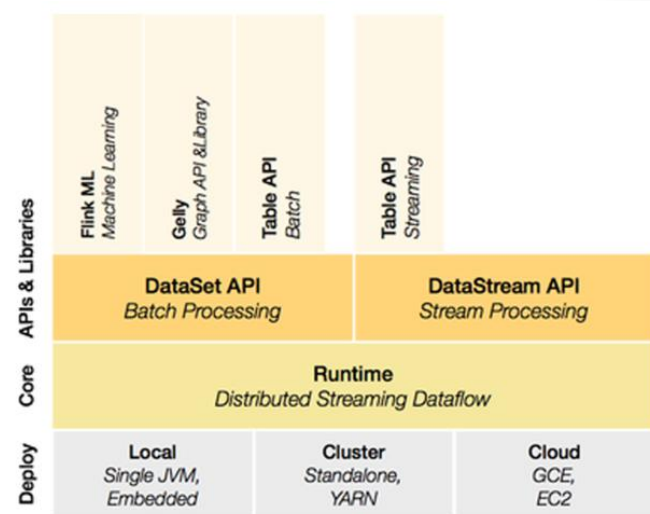
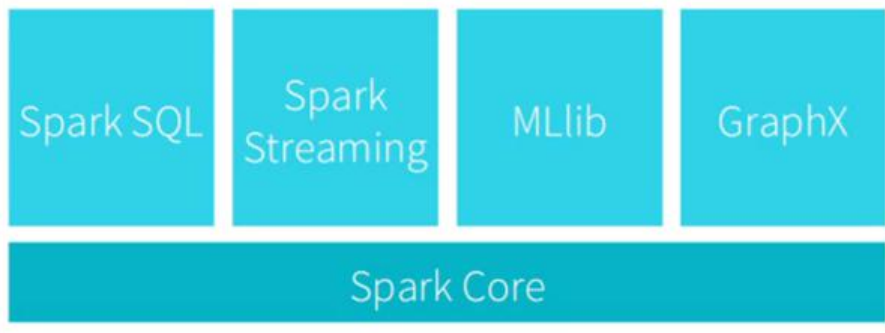
<https://github.com/sens2010/bigdata>

Apache Flink项目

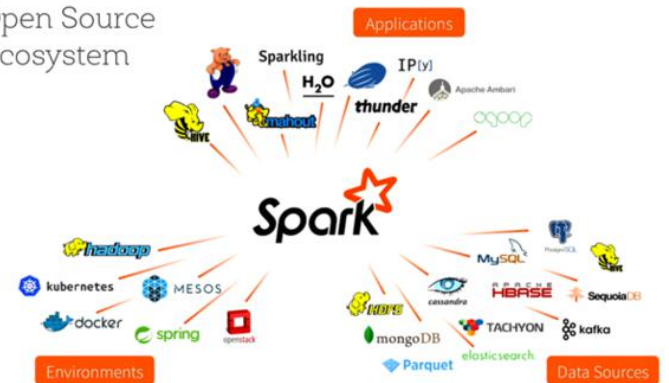
- 类Spark
- DataSet API 支持Java、Scala和Python。
- DataStream API支持Java and Scala。
- Table API支持类SQL。
- 有机器学习和图处理（**Gelly**）的各种库。
- 有自动优化迭代的功能，如有增量迭代。
- 支持高效序列化和反序列化，非常便利。
- 与Hadoop兼容性很好。



Apache Flink项目



Open Source Ecosystem



FaceBook在2015年的5大开源项目

- Presto
 - 2013年，新型分布式SQL引擎，对GB到PB数据源进行交互式查询，多数据源合并查询分析，支持Hive、Cassandra、关系型及专有数据存储，直接从HDFS读数据，号称比Hive快好几个数量级
- HipHop虚拟机（HHVM）
 - 2013年，Hack和PHP语言，JIT，目标取代Apache+mod_php
- RocksDB
 - 2013年，嵌入式持久化键值数据库，MongoDB存储引擎，Percona提供企业级支持
- React
 - 2013年，shadow dom，完全组件化开发
- ReactNative
 - 2013年，移动版React，快速开发手机应用

谷歌的海量数据排序实验史

- 2007, (1PB, 12.13小时, 1.37TB/分钟, 2.9 MB/秒/worker)
- 2008, (1PB, 6.03小时, 2.76TB/分钟, 11.5 MB/秒/worker)
- 2010, (1PB, 2.95小时, 5.65TB/分钟, 11.8 MB/秒/worker)
- 2011, (1PB, 0.55小时, 30.3TB/分钟, 63.1 MB/秒/worker)
- 2012, (50PB, 23小时, 36.2TB/分钟, 50 MB/秒/worker)

The 2016 Big Data Landscape

Big Data Landscape 2016 (Version 3.0)



Last Updated 3/23/2016

© Matt Turck (@mattturck), Jim Hao (@jimhao), & FirstMark Capital (@firstmarkcap)

FIRSTMARK