

# AWS Summit

AWS 技术峰会 · 北京 2014

## 基于Apache Spark软件栈的 实时大数据分析

戴金权 (Jason Dai)

英特尔大数据首席架构师

2014-12-12



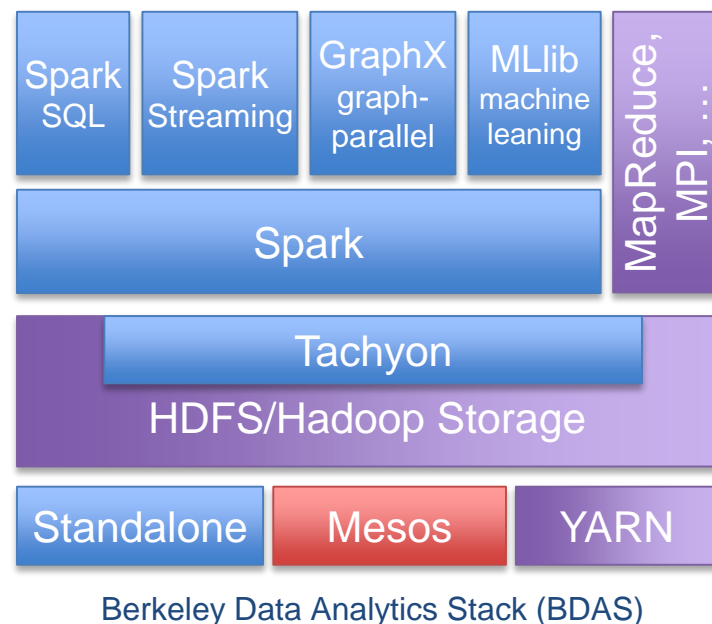
# 下一代大数据分析

- Volume
  - 海量数据 & 指数级增长
- Variety
  - 多结构化, 来自不同来源 & 不一致的数据模式(schema)
- Value
  - 简单(SQL): 描述性分析(descriptive analytics)
  - 复杂(non-SQL): 预测性分析(predictive analytics)
- Velocity
  - 交互式分析 (the speed of thought)
  - 流式分析 (drinking from the firehose)

**Apache Spark 软件栈**

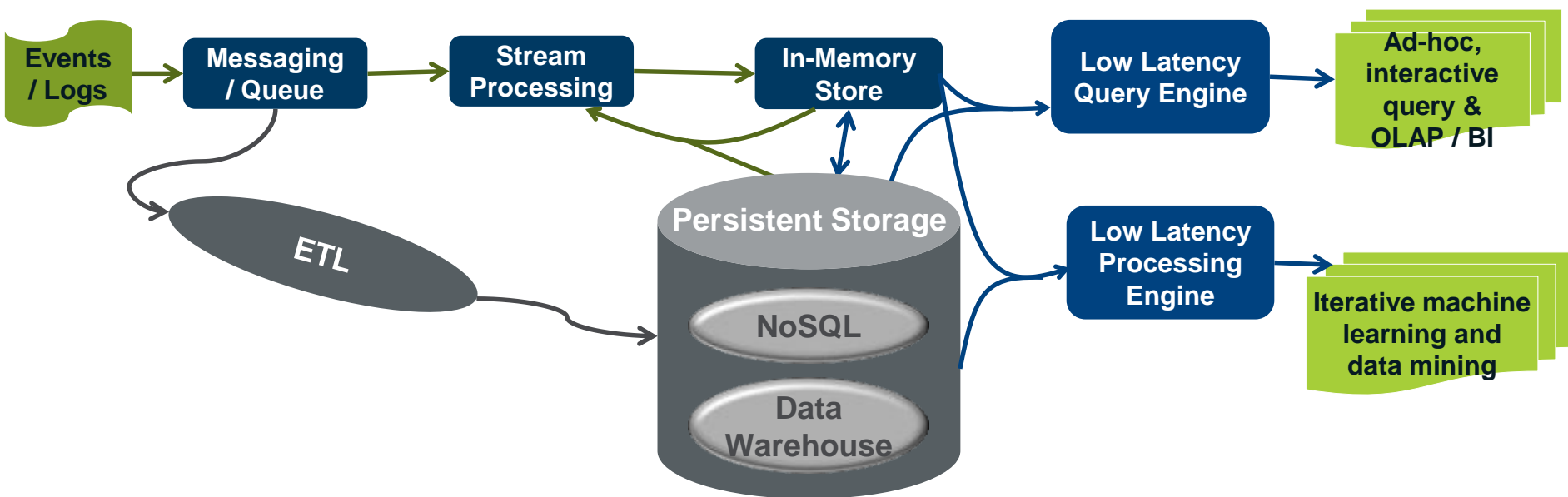
# 项目概况

- 由UC Berkeley的AMPLab发起的研究和开源项目
- Intel和AMPLab(以及开源社区)在Spark项目的开源开发上进行紧密合作
  - 合作起始于2012 (当时Spark还是一个研究项目)
  - Intel目前Spark的代码贡献量排名世界前三
    - 从Spark项目起始至今有多名committer来自Intel
- Intel和多家合作伙伴(如大型网站)进行紧密合作
  - 使用Apache Spark软件栈构建下一代大数据分析
  - 特别是实时的、基于内存的、复杂数据分析

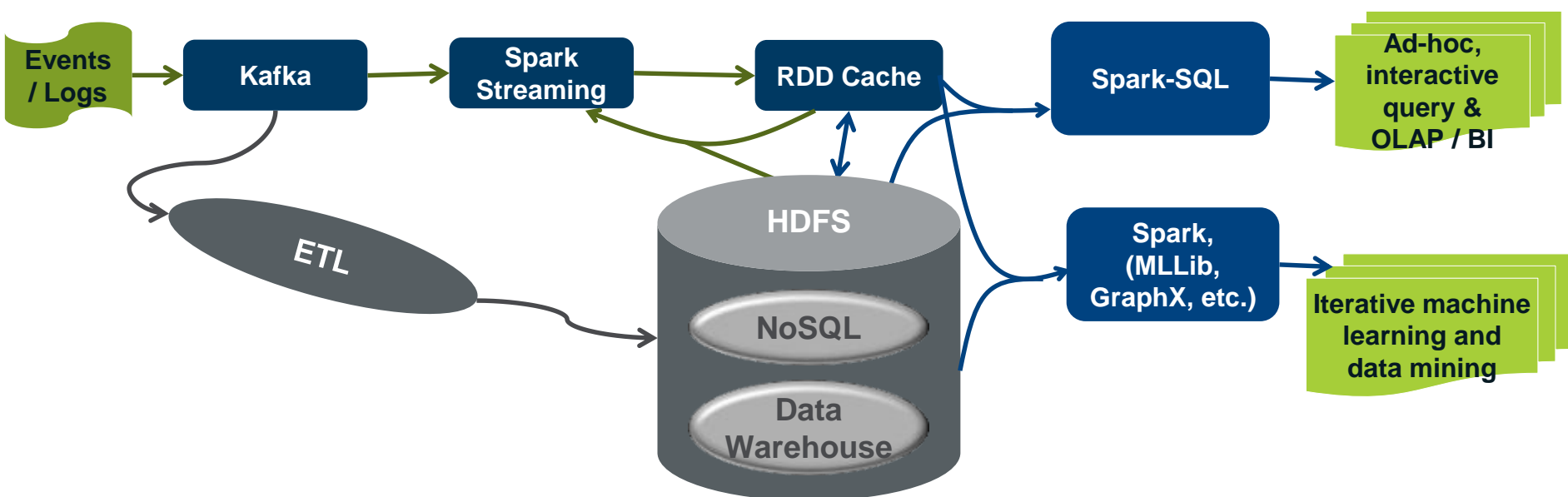


# 实时大数据分析处理

- 下一代实时大数据分析架构
  - Data captured & processed in a (semi) real-time / streaming fashion
  - Data mined using SQL queries as well as complex machine learning & graph analysis
  - Iterative and/or interactive analysis leveraging distributed in-memory data store



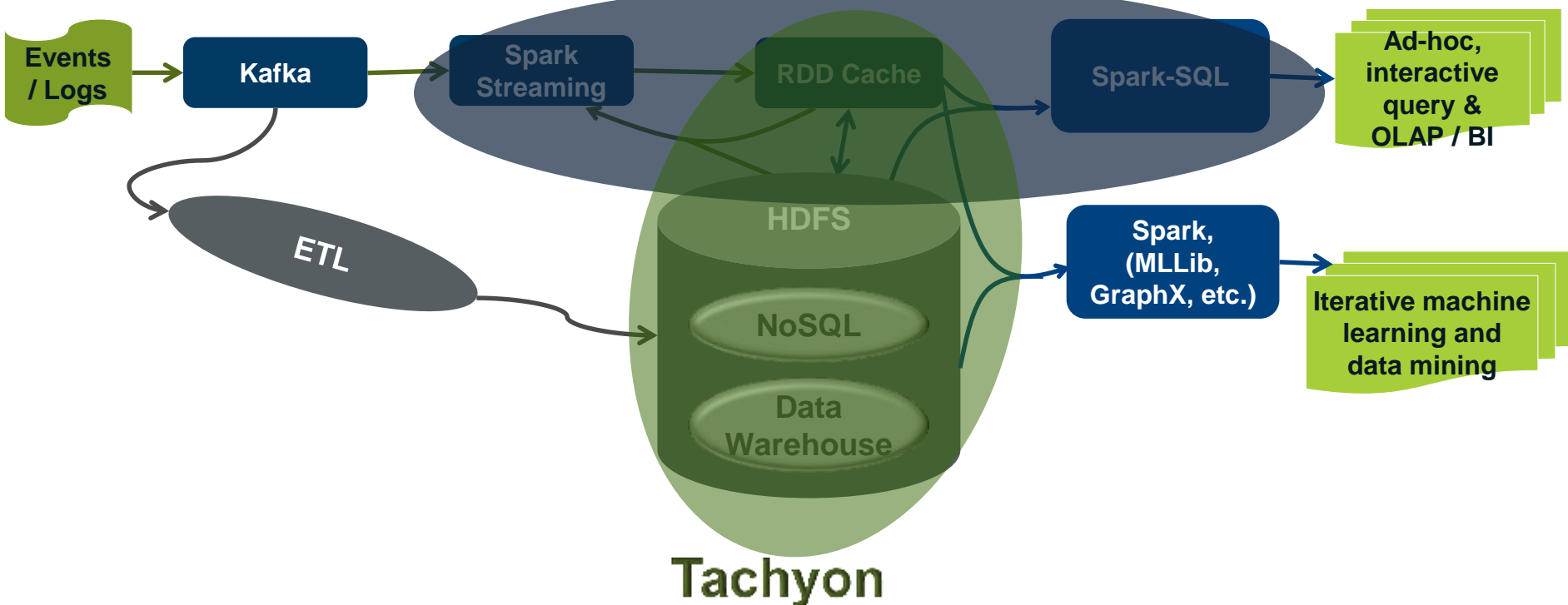
# 基于Apache Spark软件栈的实时大数据分析



# 基于Apache Spark软件栈的实时大数据分析

## Stream-SQL

使用SQL查询，对输入数据流(包括结合历史数据、参考数据)进行处理分析



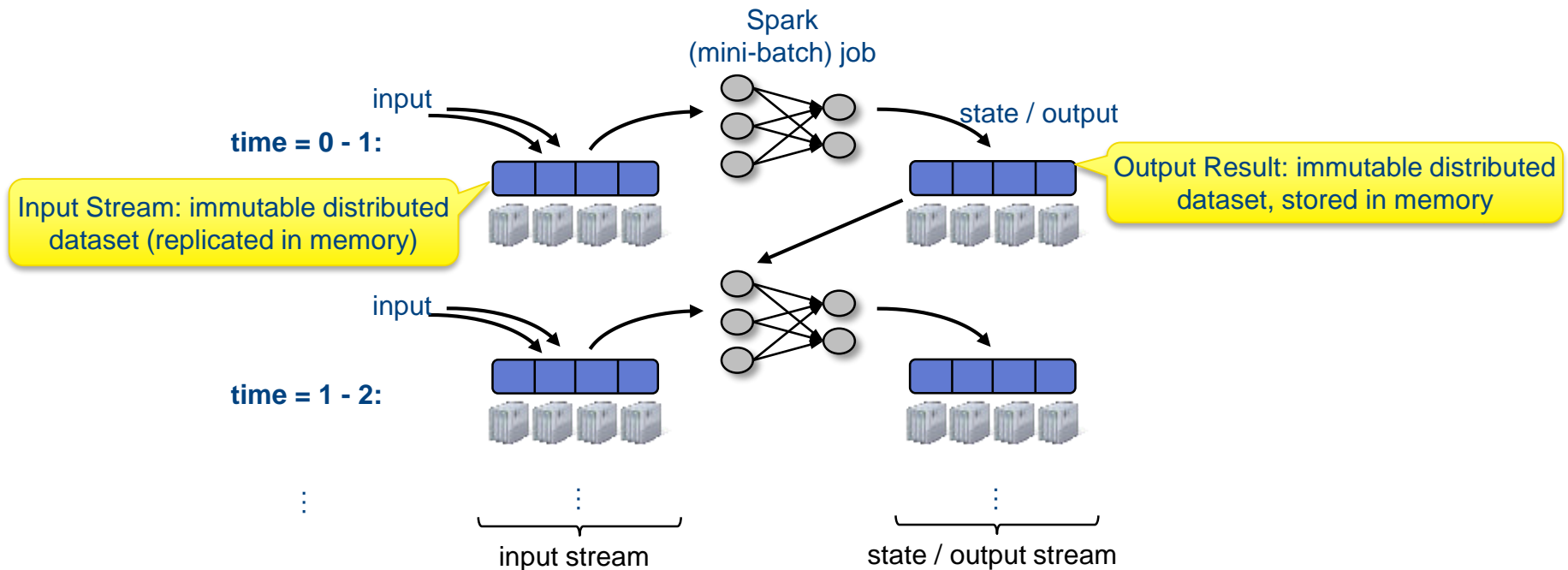
可靠的、分布式内存文件系统，支持多种不同的底层存储系统

# Spark Stream-SQL: 流式处理 + SQL查询

- 支持使用SQL查询，对输入数据流(包括结合历史数据、参考数据)进行处理分析
- 构建于Spark Streaming和Spark SQL框架之上

# Spark Streaming概述

- Discrete Stream (DStream)概念
  - Run streaming computation as a series of very small, deterministic (mini-batch) Spark jobs
    - As frequent as  $\sim 1/2$  second
  - Better fault tolerance, straggler handling & state consistency

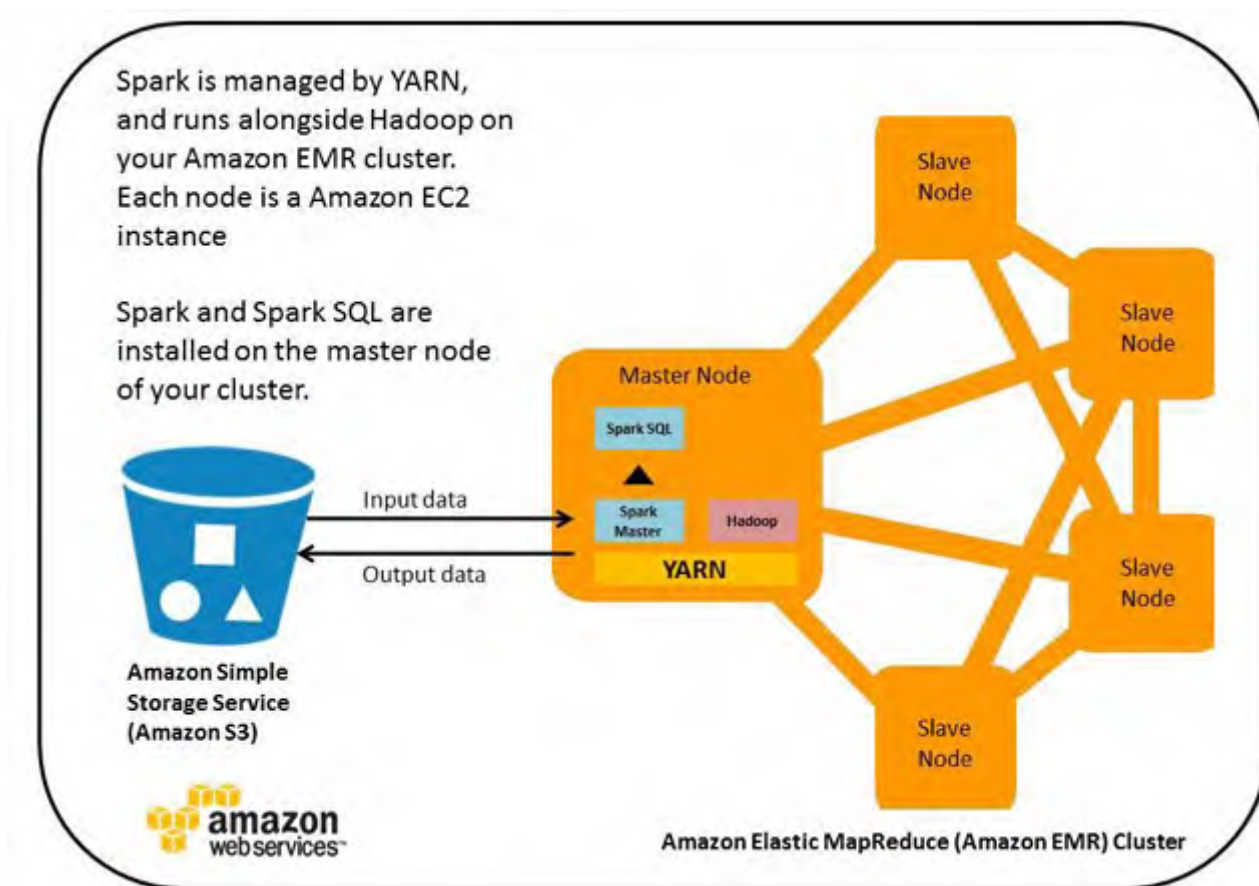




# Spark SQL概述

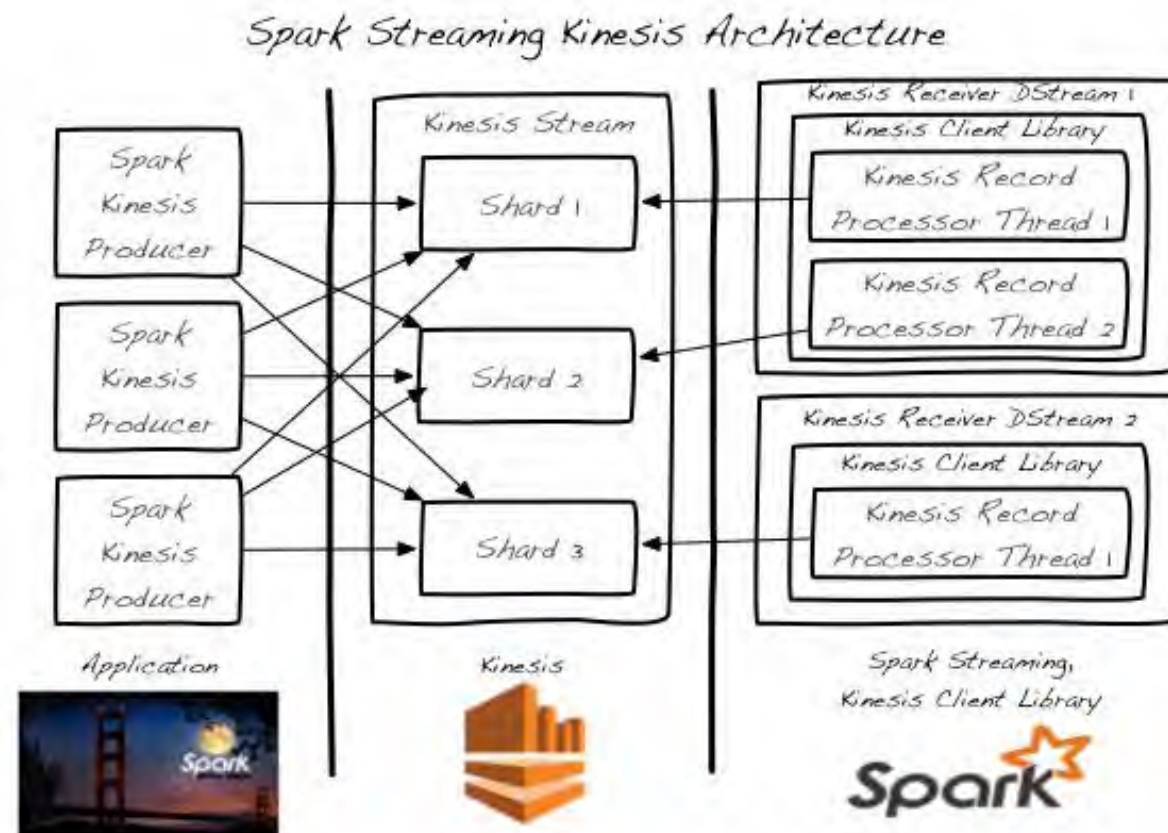
- 在Spark框架上支持SQL查询
  - Structured data analysis using SQL queries on Spark
    - Hive tables, Parquet files, etc.
  - Integration with analytics pipelines
- Hive兼容性
  - Directly reading data stored in Hive
  - Writing queries in HiveQL

# 在EMR上运行Spark / Spark-SQL



(Source: <https://aws.amazon.com/articles/4926593393724923>)

# Spark Streaming + Kinesis



(Source: <https://spark.apache.org/docs/latest/streaming-kinesis-integration.html>)

# Spark Stream-SQL: 流式SQL分析框架

- 用户使用Stream-SQL查询，对输入数据流进行处理分析
- 框架自动将Stream-SQL查询编译成Discretized Stream
- 生成的Discretized Stream在每一个“batch”运行一个Spark作业
  - Conceptually, each job runs the same Spark-SQL query as the Stream-SQL query (with the input “Stream” replaced by an input table)
  - The input table will contain the data received over that stream during this “batch” (or data received in the “current” window)

# Stream-SQL查询

```
CREATE STREAM IF NOT EXISTS
    people_stream1 (name STRING, age INT)
STORED AS LOCATION 'kafka://...';

CREATE STREAM IF NOT EXISTS
    people_stream2 (name STRING, zipcode INT)
STORED AS LOCATION 'kafka://...';

SELECT count(*)
FROM people_stream1
WHERE age >= 10 && age <= 19;

SELECT zipcode, AVG(age)
FROM people_stream1 JOIN people_stream2
ON people_stream1.name = people_stream2.name
GROUPBY zipcode;
```

# Spark Stream-SQL和Hive的兼容性

- Hive: Hadoop平台上的数据仓库系统
- Stream-SQL将Hive扩展为一个构建在Spark上的数据流管理系统
  - Support writing queries in HiveQL for Stream
  - Stream created & registered in Hive MetaStore (just as normal Hive tables)
  - Query both input data stream and (history/reference) data table stored in Hive

# Stream-SQL查询

```
CREATE TABLE IF NOT EXISTS  
    city_table (zipcode INT, city_name STRING);
```

```
CREATE STREAM IF NOT EXISTS  
    people_stream (name STRING, zipcode INT)  
STORED AS LOCATION 'kafka://...';
```

...

```
SELECT cityname, count(*)  
FROM people_stream JOIN city_table  
ON people_stream.zipcode = city_table.zipcode  
GROUPBY city_table.zipcode;
```

# Spark Stream-SQL开发状态

- 基于Apache 2.0协议开源
  - <https://github.com/intel-spark/stream-sql>
  - Developer preview (based on Spark 1.0) available
- 目前正处于积极开发中
  - An update based on latest Spark version will be available soon
  - Many more features & optimizations are being added
  - Plan to contribute back to the main Spark project

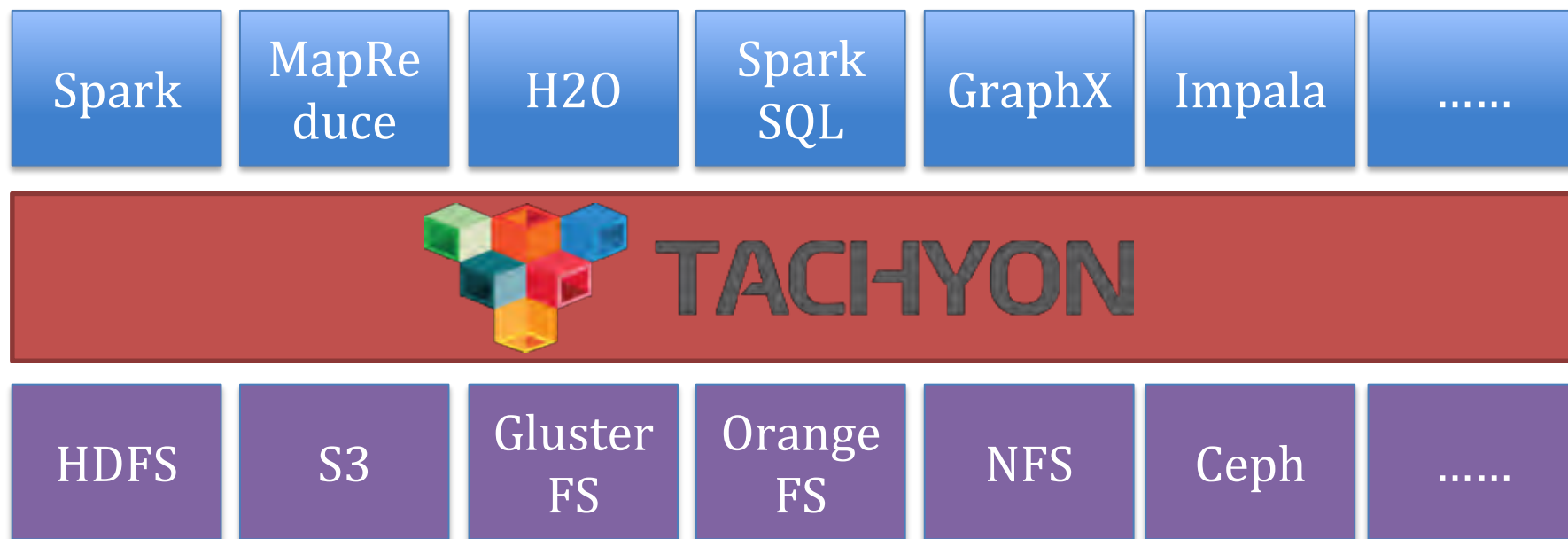
**Welcome Collaboration!**



# Tachyon概述

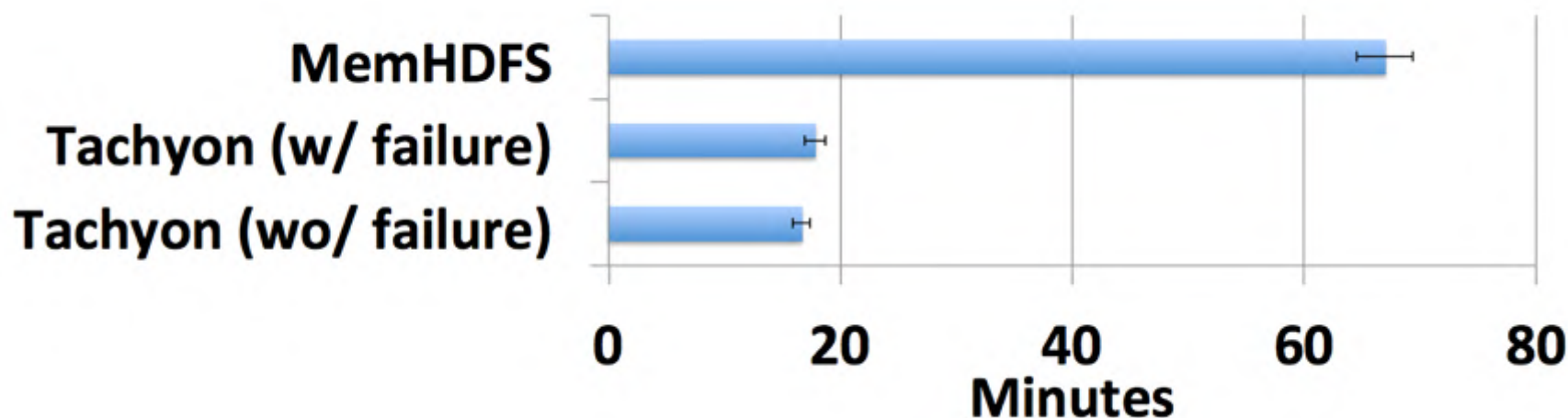
- 可靠的、分布式内存文件系统，支持多种不同的底层存储系统
- 在不同的集群计算框架和作业之间，提供可靠的、内存级读写速度的数据共享

# 支持多种框架的分布式内存文件系统



(Source: <http://www.slideshare.net/haoyuanli/tachyon20141121ampcamp5-41881671>)

# 应用性能改进



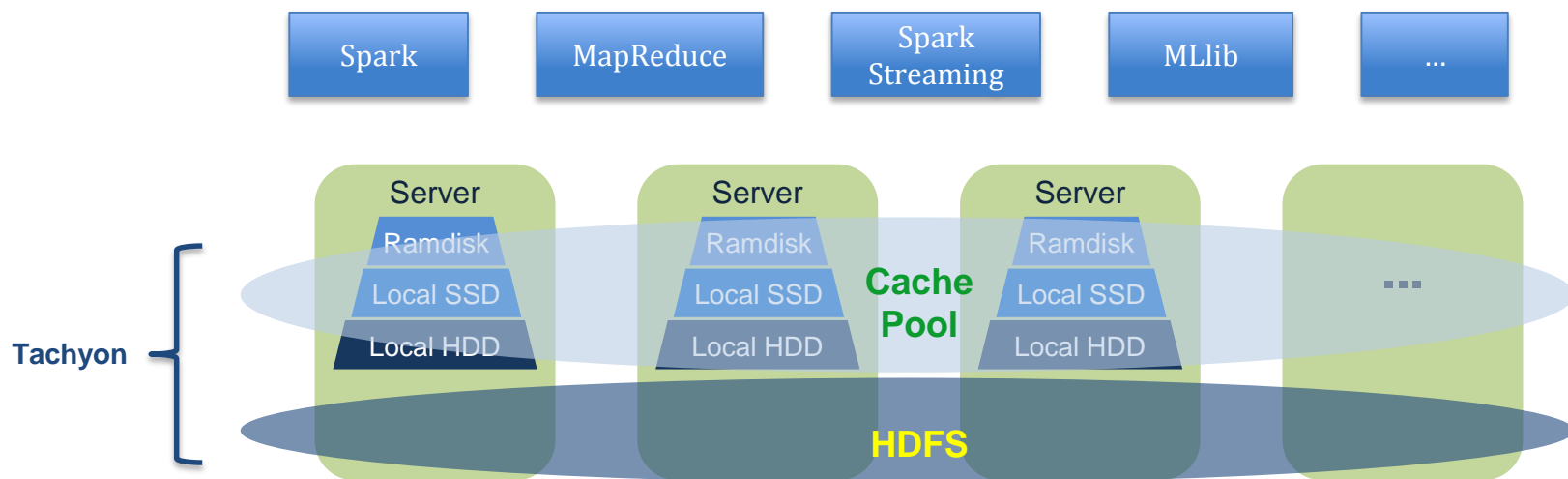
Performance comparison for realistic workflow. The workflow ran 4x faster on Tachyon than on MemHDFS. In case of node failure, applications in Tachyon still finishes 3.8x faster.

(Source: <http://www.slideshare.net/haoyuanli/tachyon20141121ampcamp5-41881671>)

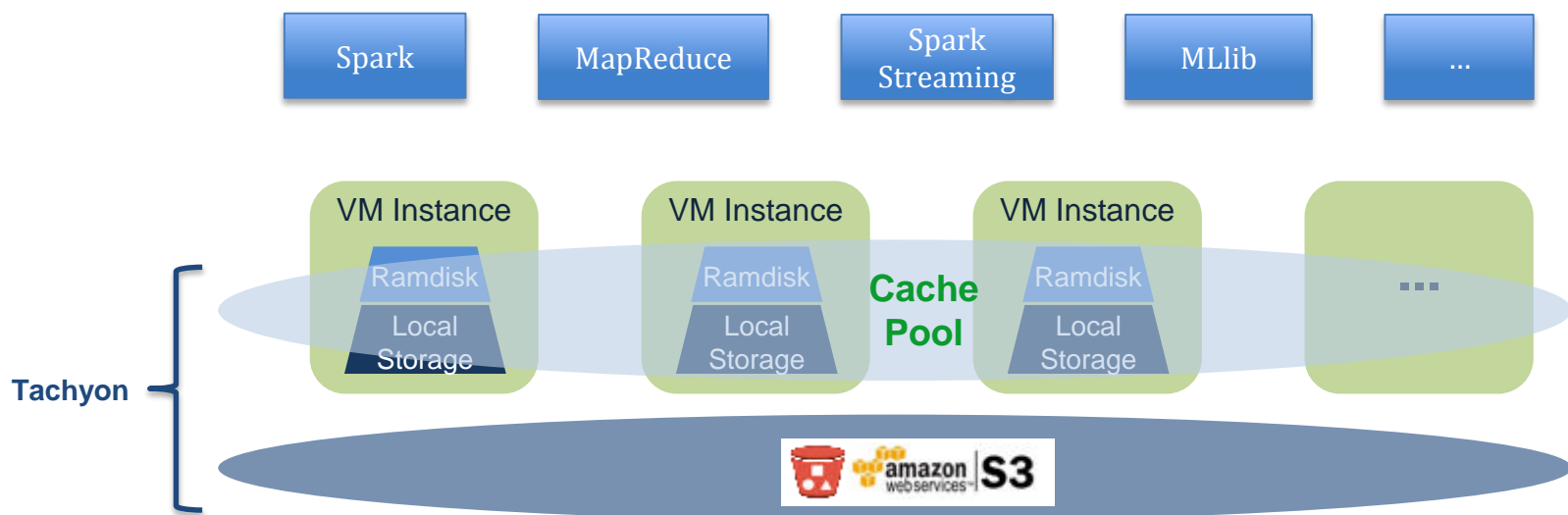
# Tachyon分级存储管理

- 当前Tachyon中的2级存储架构
  - Memory across different servers in the cluster are organized as a **cache pool** to provide *memory-speed* data sharing
  - All data are reliably persisted in the **underlying file system**
- Tachyon中新的分级存储管理
  - The data cache pool manages **multiple storage tiers** (for different types of storage) to provide *memory-speed* data sharing
  - Provide efficient support for new storage media (e.g., flash) and/or computing environments (e.g., cloud, HPC)

# Tachyon分级存储管理(闪存SSD案例)



# Tachyon分级存储管理(Amazon S3案例)



# Tachyon分级存储管理开发状态

- 目前正处于积极开发中
  - <https://tachyon.atlassian.net/browse/TACHYON-33>
- 将于Tachyon 0.6 release中发布
  - <https://github.com/amplab/tachyon>

# 总结

- 使用Apache Spark软件栈构建下一代、实时大数据分析
- 通过开源社区、共同合作开发
  - Spark Stream-SQL: 使用SQL查询，对输入数据流进行处理分析
  - Tachyon分级存储管理: 支持多种不同的分级存储系统，提供分布式、可靠的、内存级的数据共享

**Welcome Collaboration!**



# Notices and Disclaimers

- Copyright © 2014 Intel Corporation.
- Intel, the Intel logo are trademarks of Intel Corporation in the U.S. and/or other countries. \*Other names and brands may be claimed as the property of others.  
See Trademarks on [intel.com](http://intel.com) for full list of Intel trademarks.
- All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest Intel product specifications and roadmaps
- Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors.  
Performance tests are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.
- For more complete information about performance and benchmark results, visit [www.intel.com/benchmarks](http://www.intel.com/benchmarks).
- Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.
- Results have been estimated or simulated using internal Intel analysis or architecture simulation or modeling, and provided to you for informational purposes. Any differences in your system hardware, software or configuration may affect your actual performance.
- Intel technologies may require enabled hardware, specific software, or services activation. Check with your system manufacturer or retailer.
- No computer system can be absolutely secure. Intel does not assume any liability for lost or stolen data or systems or any damages resulting from such losses.
- You may not use or facilitate the use of this document in connection with any infringement or other legal analysis concerning Intel products described herein. You agree to grant Intel a non-exclusive, royalty-free license to any patent claim thereafter drafted which includes subject matter disclosed herein.
- No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.
- The products described may contain design defects or errors known as errata which may cause the product to deviate from publish.

# AWS Summit

AWS 技术峰会 · 北京 2014

