

Spark Ecosystem & Internals

陈超 @CrazyJvm

开发者最佳实践日@北京3W咖啡

Show of Hands

How familiar are you with Spark?

- A. Heard of it, but haven't used it before.
- B. Kicked the tires with some basics.
- C. Worked or working on a proof-of-concept deployment.
- D. Worked or working on a production deployment.

outline

- *basis & internals*
- ecosystem

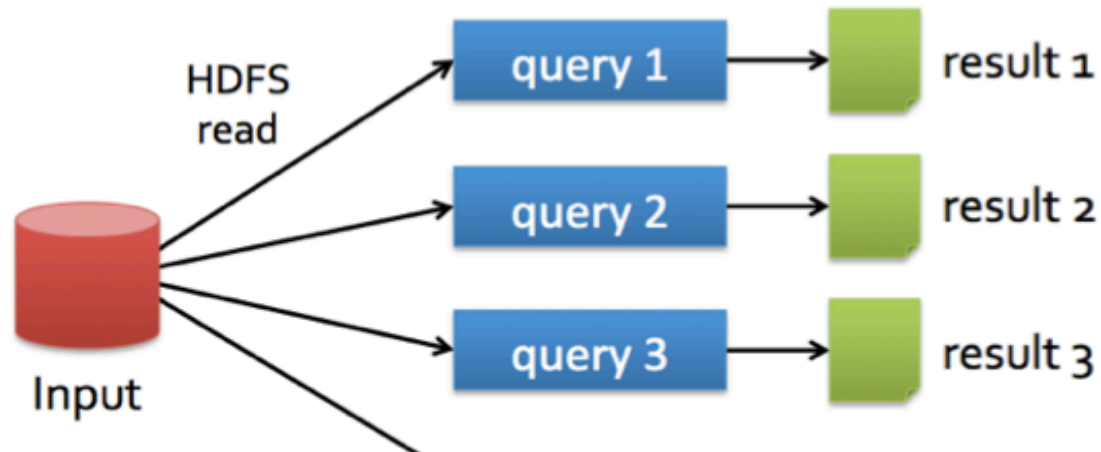
Current Major Release

- **released Spark 1.2**

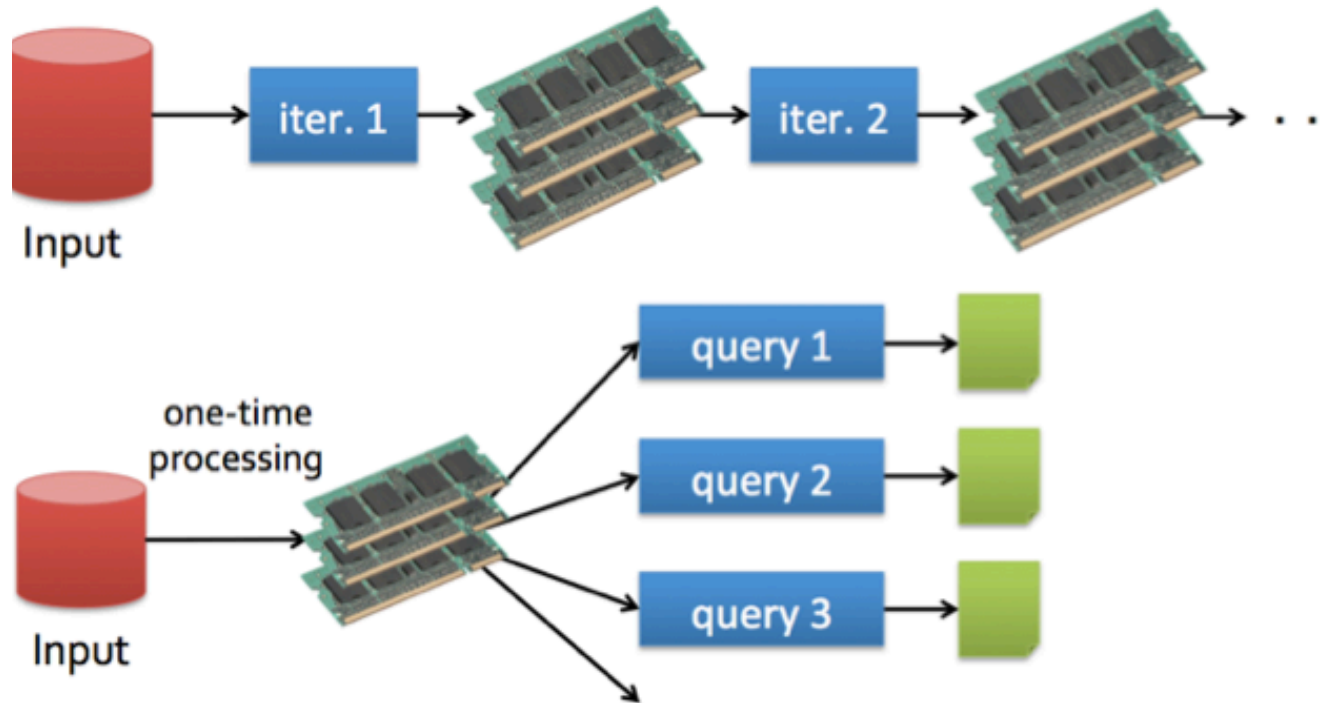
Spark : What & Why

- Apache Spark is a fast and general engine for large-scale data processing.
- Speed
- Ease of Use
- Generality
- Integrated with Hadoop

Hadoop Data Sharing



Spark Data Sharing

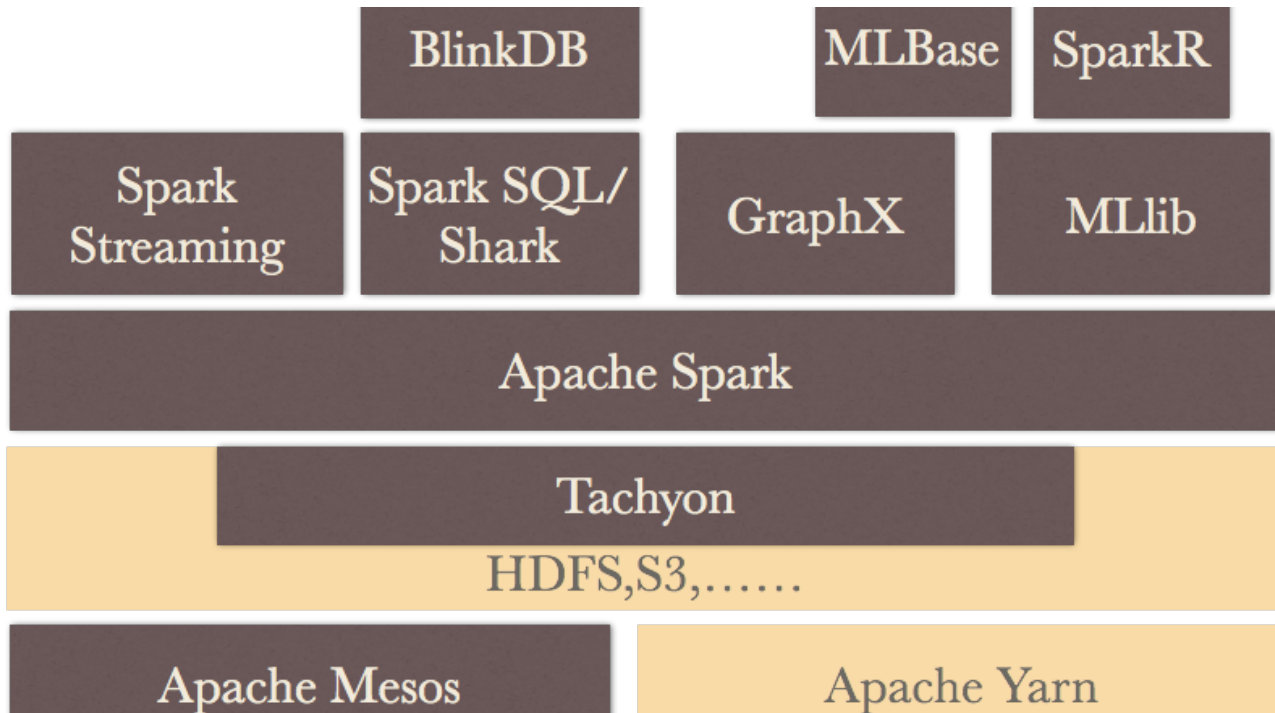


DAG & in-memory

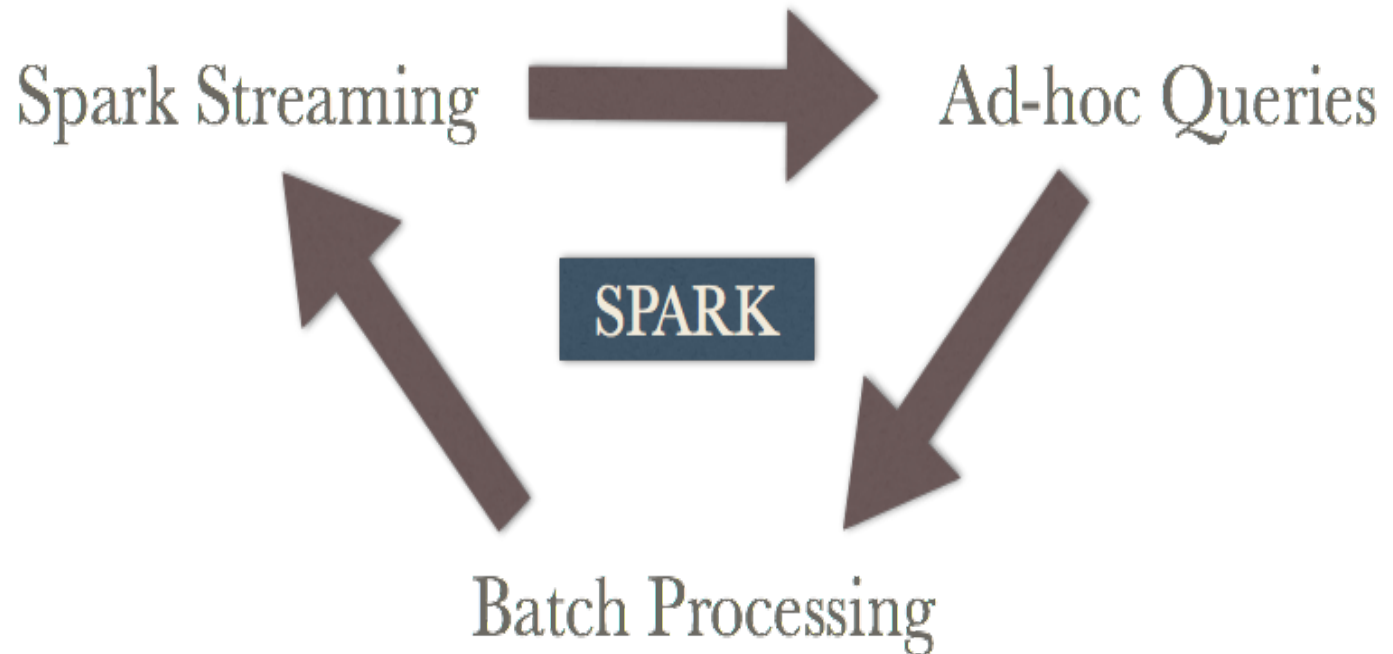
Why Spark Fast?

- Memory based computation
- DAG
- Thread Model
- Optimization(e.g. delay scheduling)

BDAS



one stack to rule them all

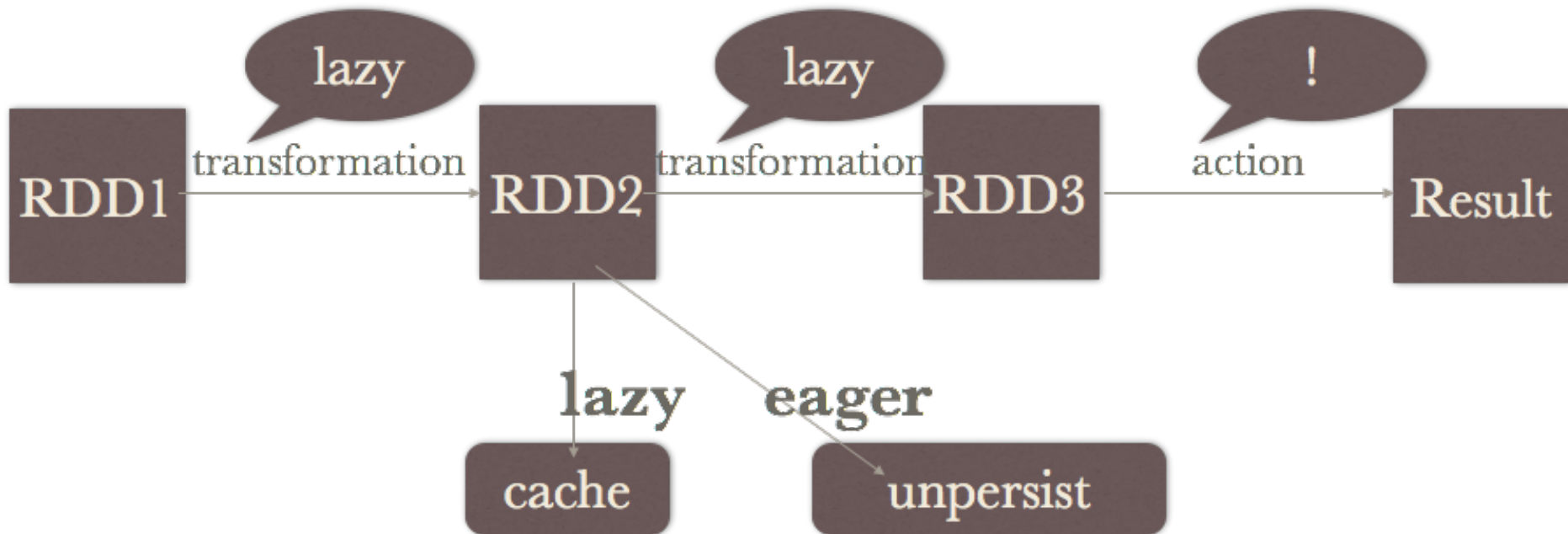


Key Concept-RDD

- A list of partitions
- A function for computing each split
- A list of dependencies on other RDDs
- Optionally, a Partitioner for key-value RDDs
- Optionally, a list of preferred locations to compute each split on

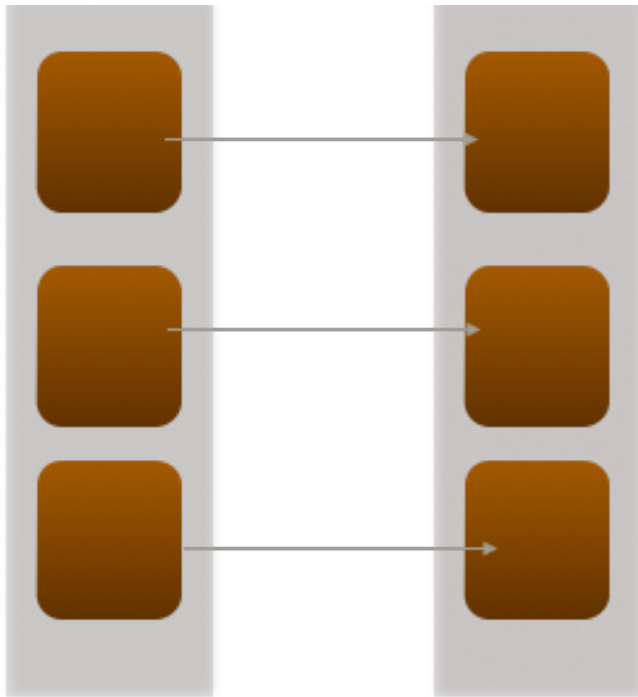
Immutable!!!

Key Concept-Lineage

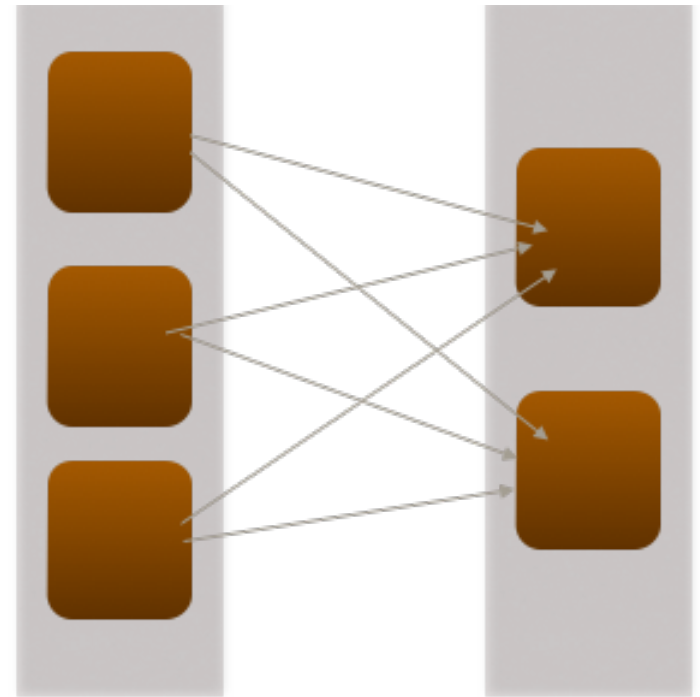


unroll partition safely when caching

Key Concept-Dependency



Narrow
Dependency

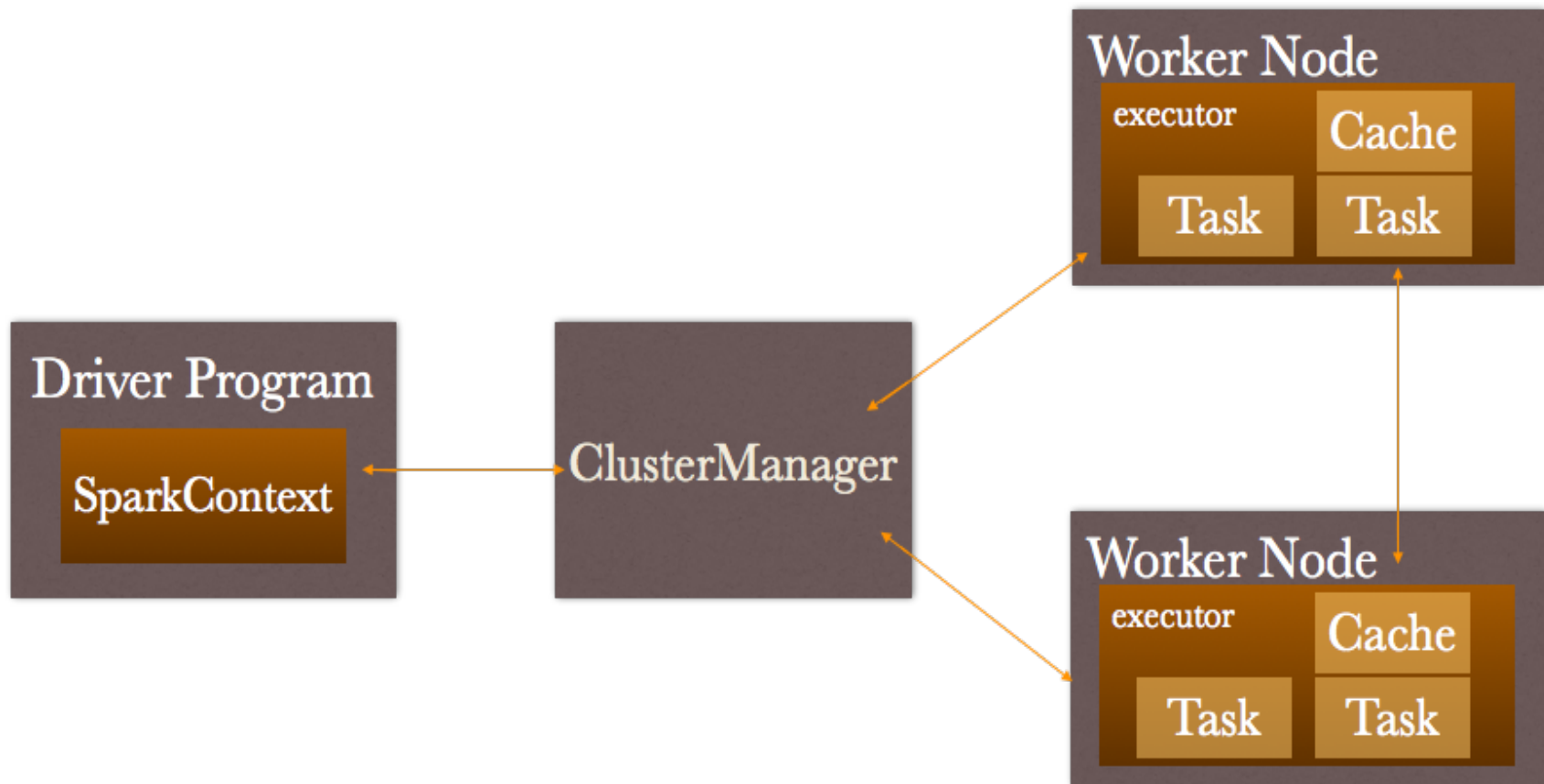


Wide
Dependency

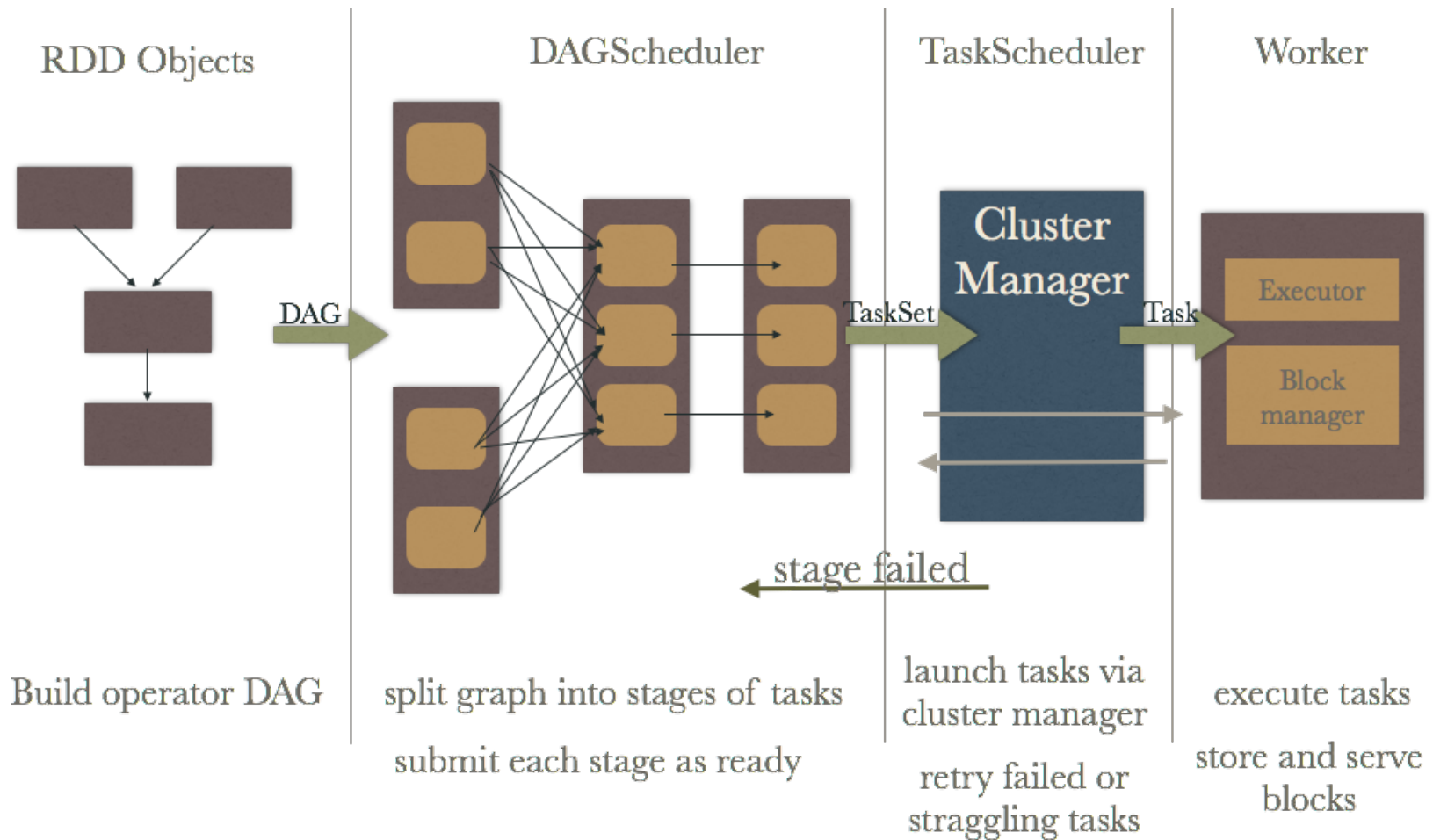
Key Concept-ClusterManager

- Local
- Standalone
- Yarn
- Mesos

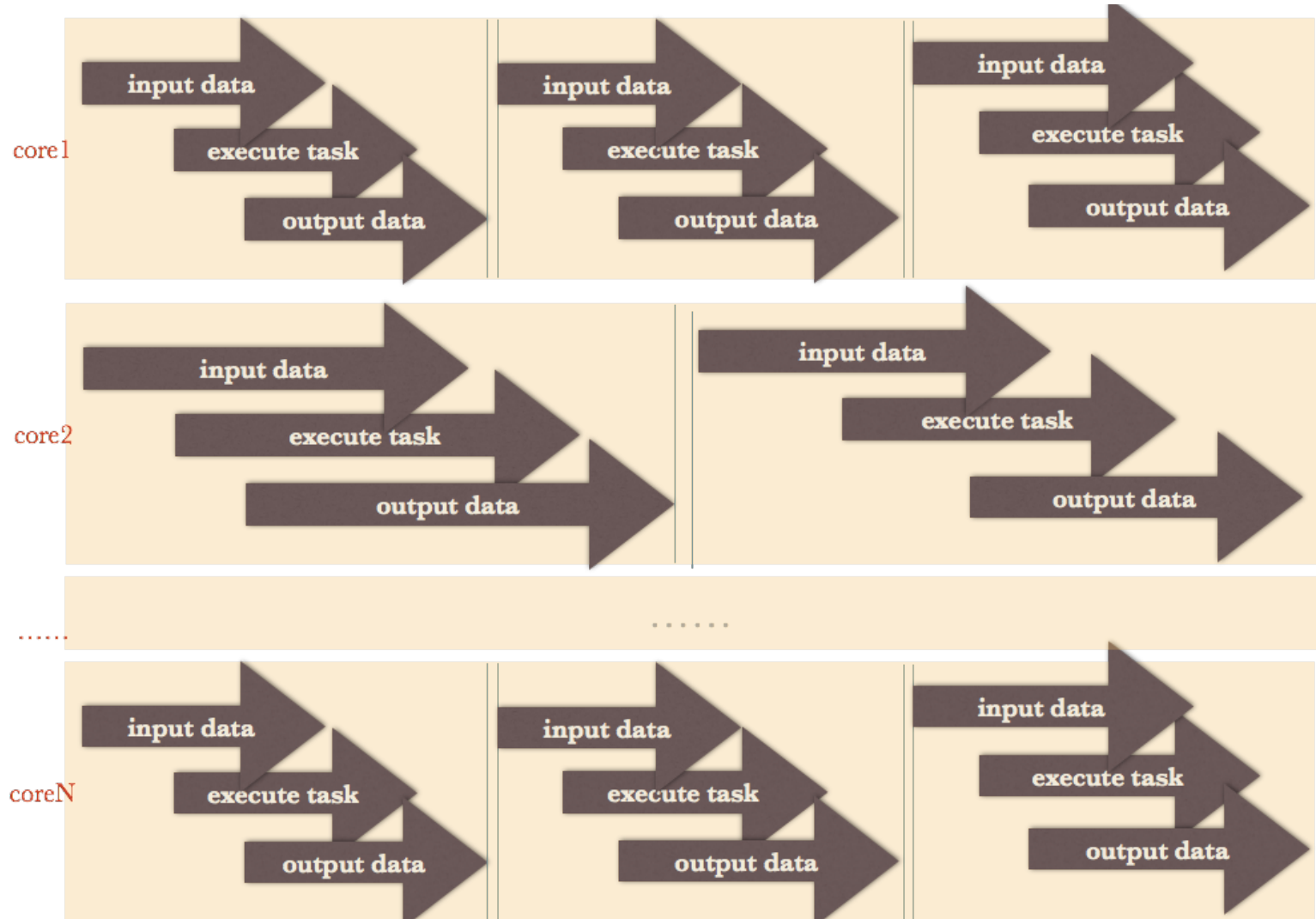
Cluster Overview



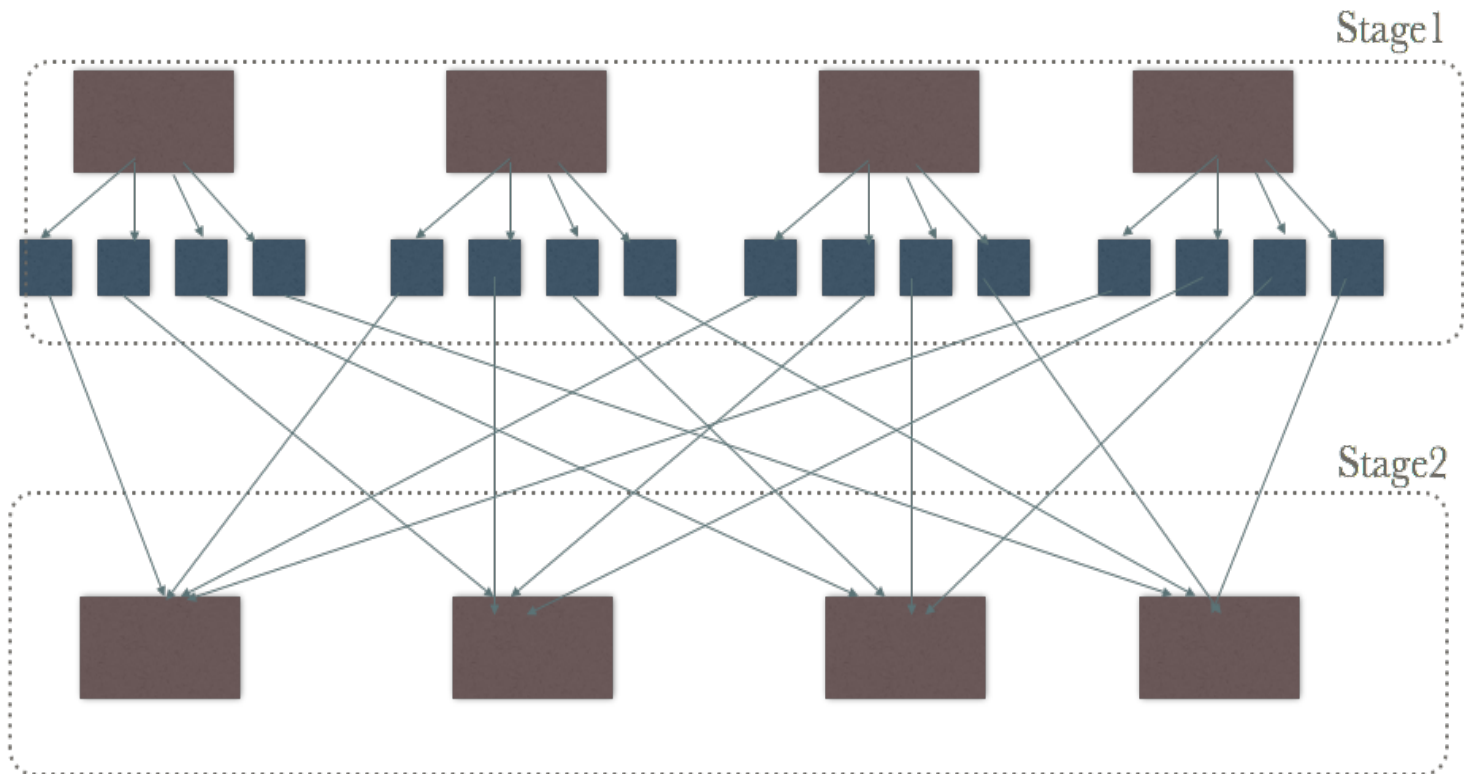
Schedule



Executor



Shuffle



Sort-based shuffle supported

Shuffle

- Pull-based (not push-based)
- Write intermediate files to disk
- Build hash map within each partition
- Can spill across keys
- A single key-value pair must fit in memory

Better Metrics System

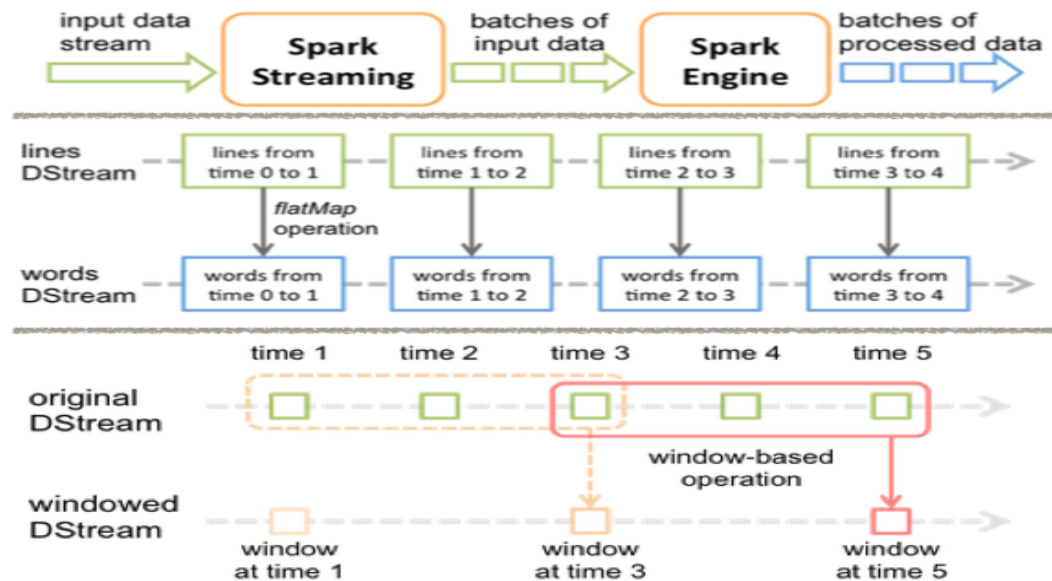
- Previously: only collect after task completed
- Now : report when task is still running

outline

- basis & internals
- *ecosystem*

Spark Streaming

- Mini-batch



rate limiting supported!

Streaming + MLlib

Enhanced HA

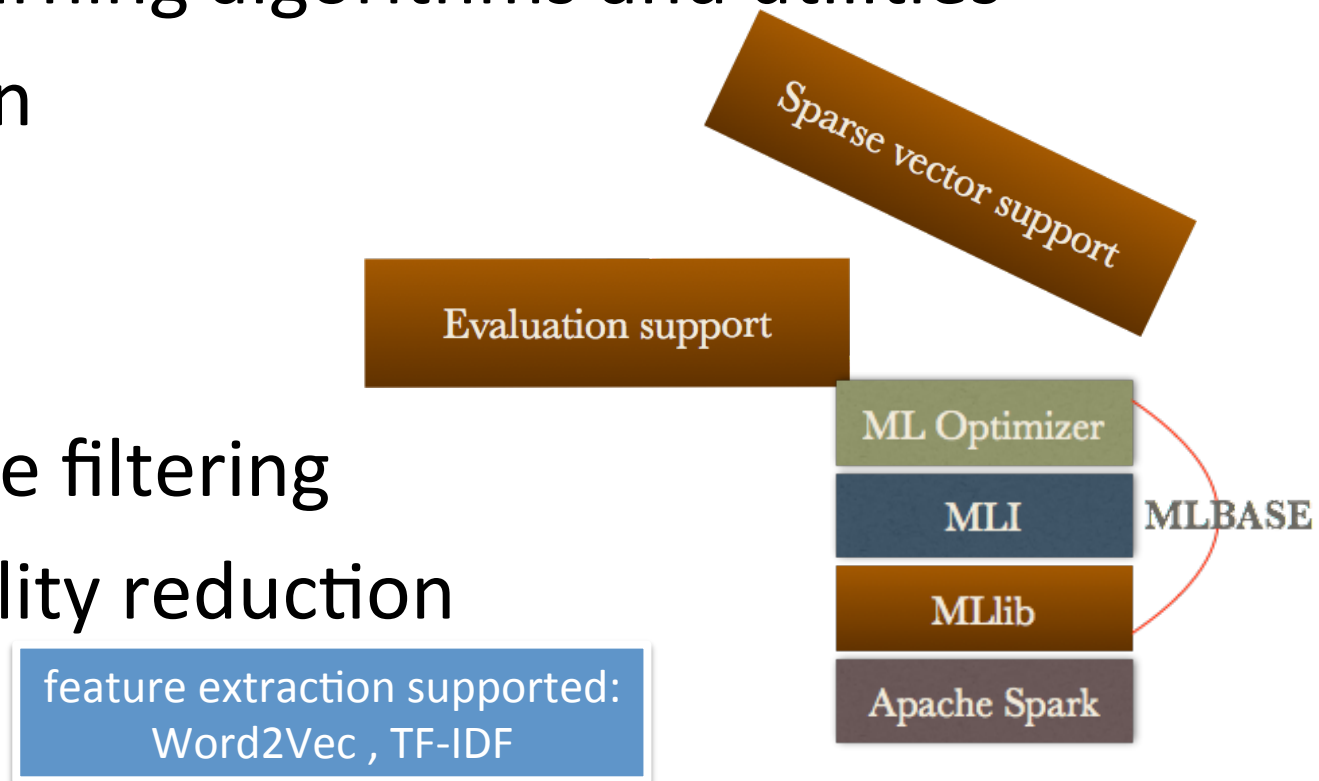
Data Lose?

- Worker Failure
- Driver Failure
- Reliable Receiver

Deployment Scenario	Worker Failure	Driver Failure
Spark 1.1 or earlier, or Spark 1.2 without write ahead log	Buffered data lost with unreliable receivers Zero data loss with reliable receivers and files	Buffered data lost with unreliable receivers Past data lost with all receivers Zero data loss with files
Spark 1.2 with write ahead log	Zero data loss with reliable receivers and files	Zero data loss with reliable receivers and files

MLlib

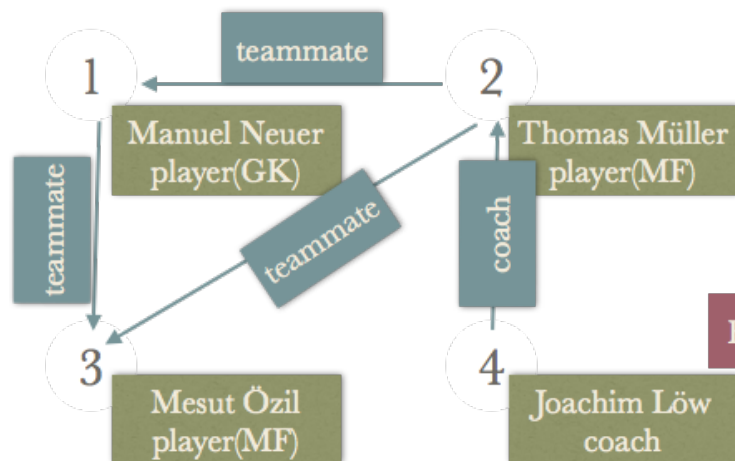
- Spark implementation of some common machine learning algorithms and utilities
- classification
- regression
- clustering
- collaborative filtering
- dimensionality reduction



ML Pipeline

- ML Dataset
- Transformer
- Estimator
- Pipeline

GraphX



Vertex Table

id	
1	(Manuel Neuer,player)
2	(Thomas Müller,player)
3	(Mesut Özil,player)
4	(Joachim Löw,coach)

Edge Table

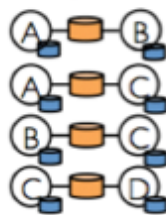
SrcId	DstId	Property(E)
2	1	teammate
2	3	teammate
1	3	teammate
4	2	coach

GraphX

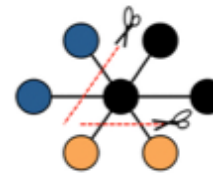
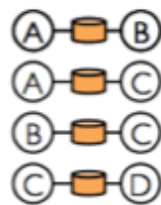
Vertices



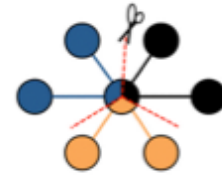
Triplets



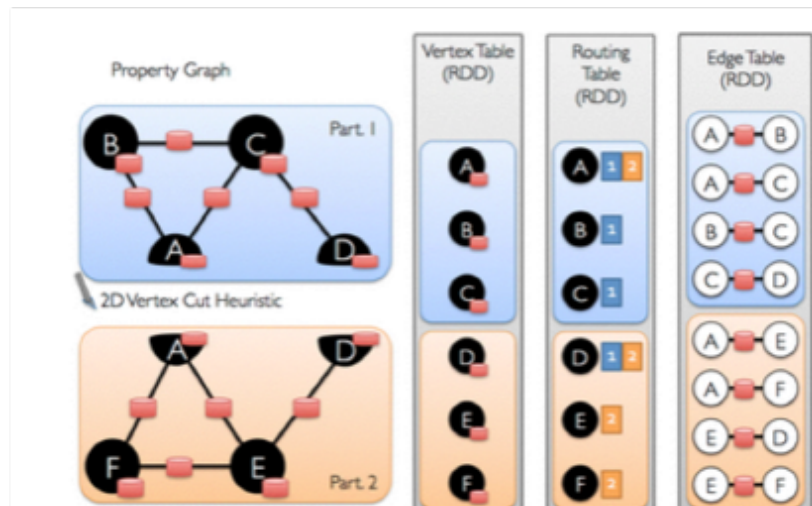
Edges



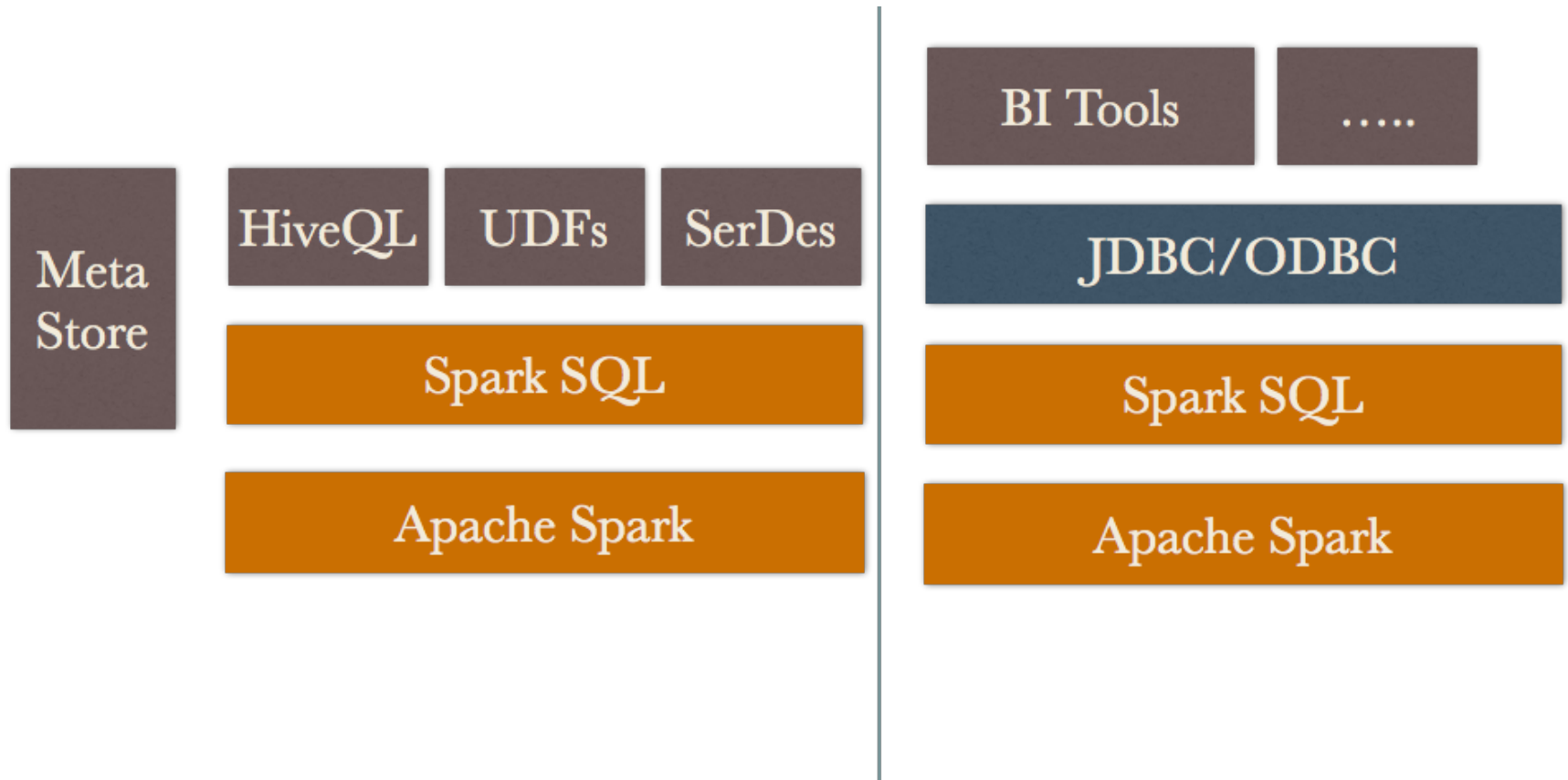
Edge Cut



Vertex Cut



Spark SQL



Spark SQL

- Data Sources
- RDDs/Parquet Files/JSON Datasets/Hive Table
- DSL
- JDBC Server

Programmatically Specifying the Schema

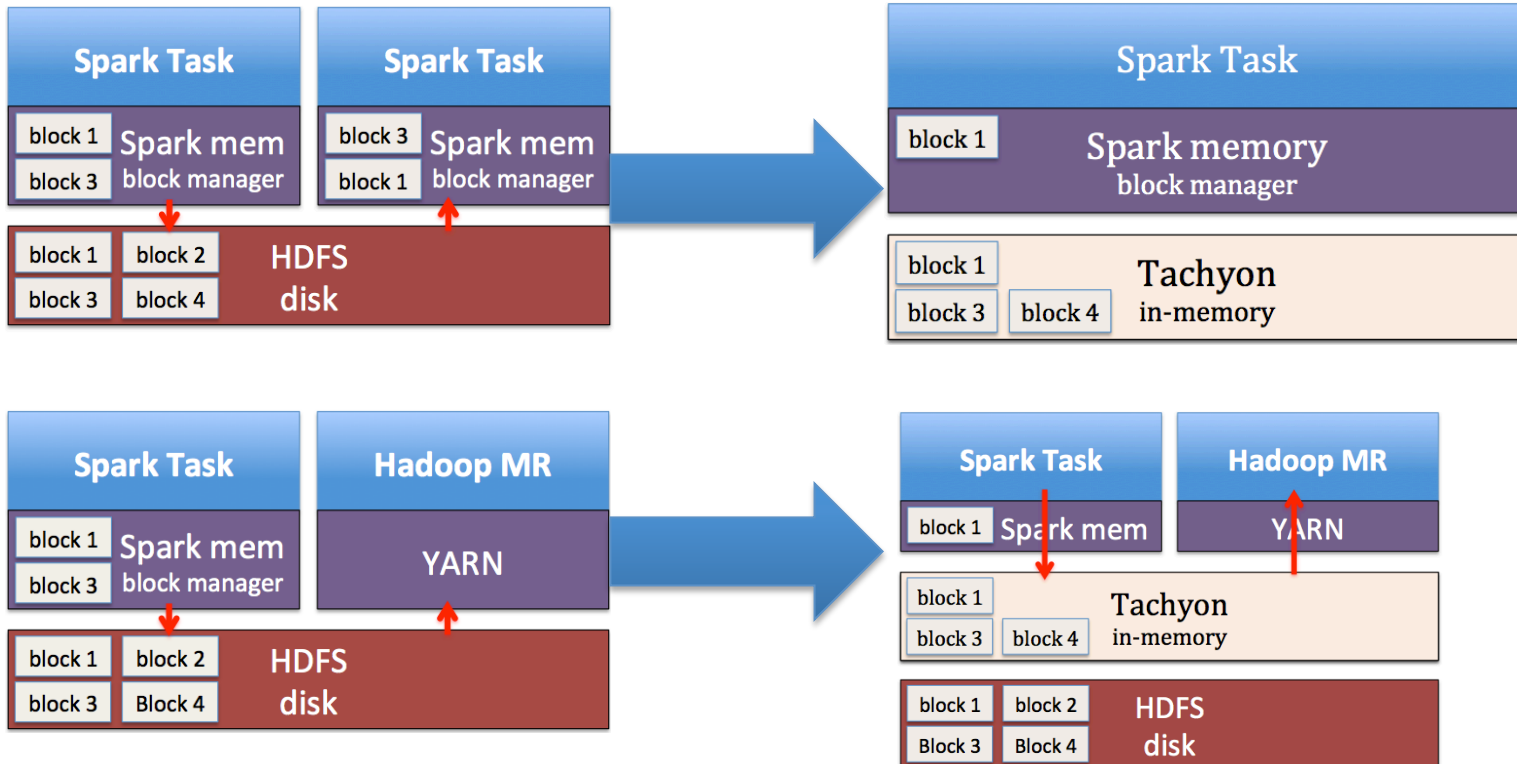
So... what is Hive on Spark?

Shark

- Development in Shark has been ended and subsumed by Spark SQL
- Mission completed !!!

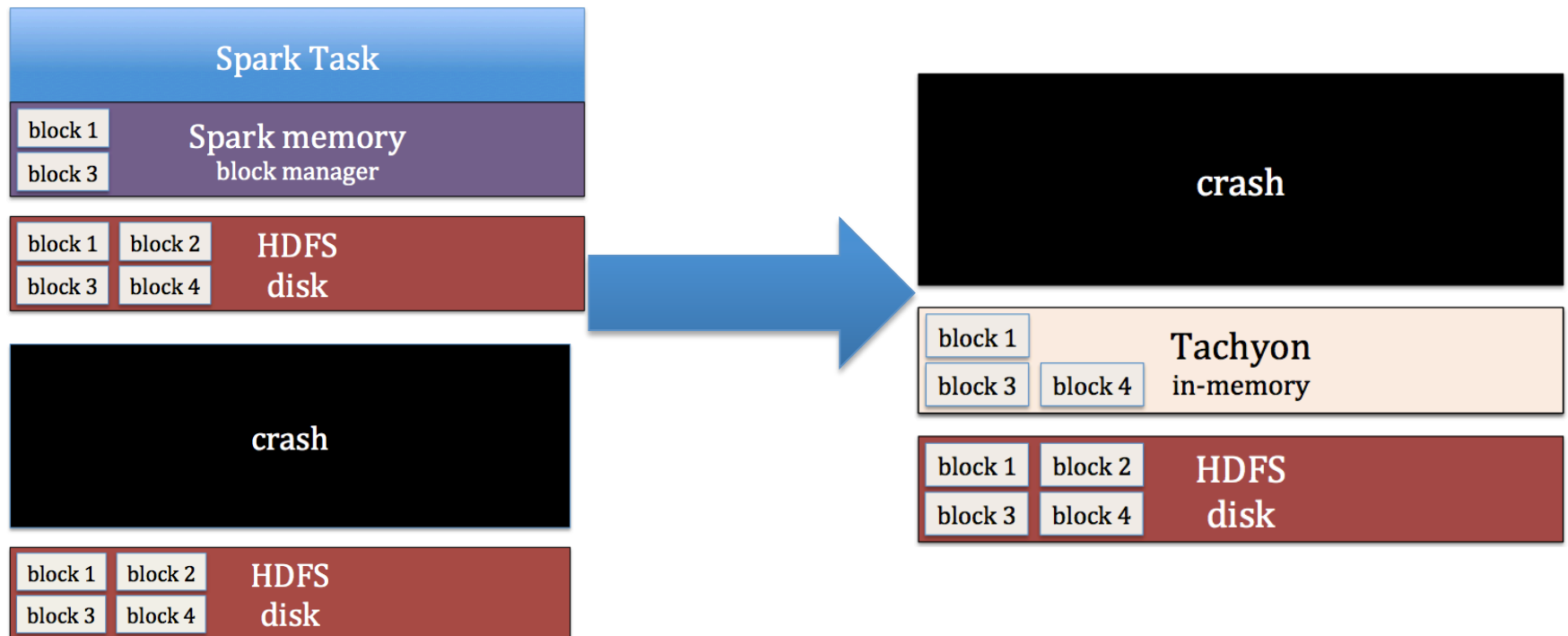
Tachyon

- Data sharing between different jobs or frameworks



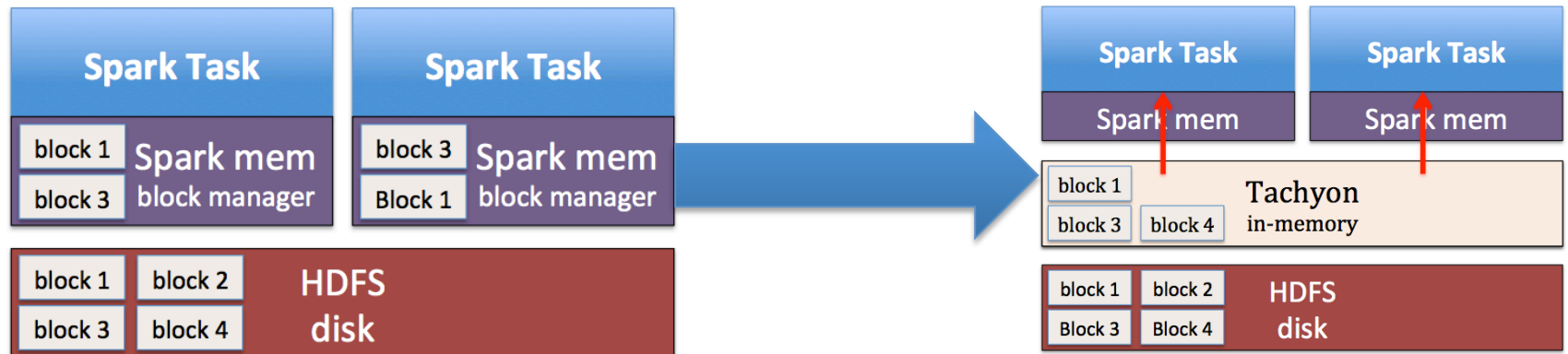
Tachyon

- Process crash



Tachyon

- Duplicated memory and GC



Tachyon



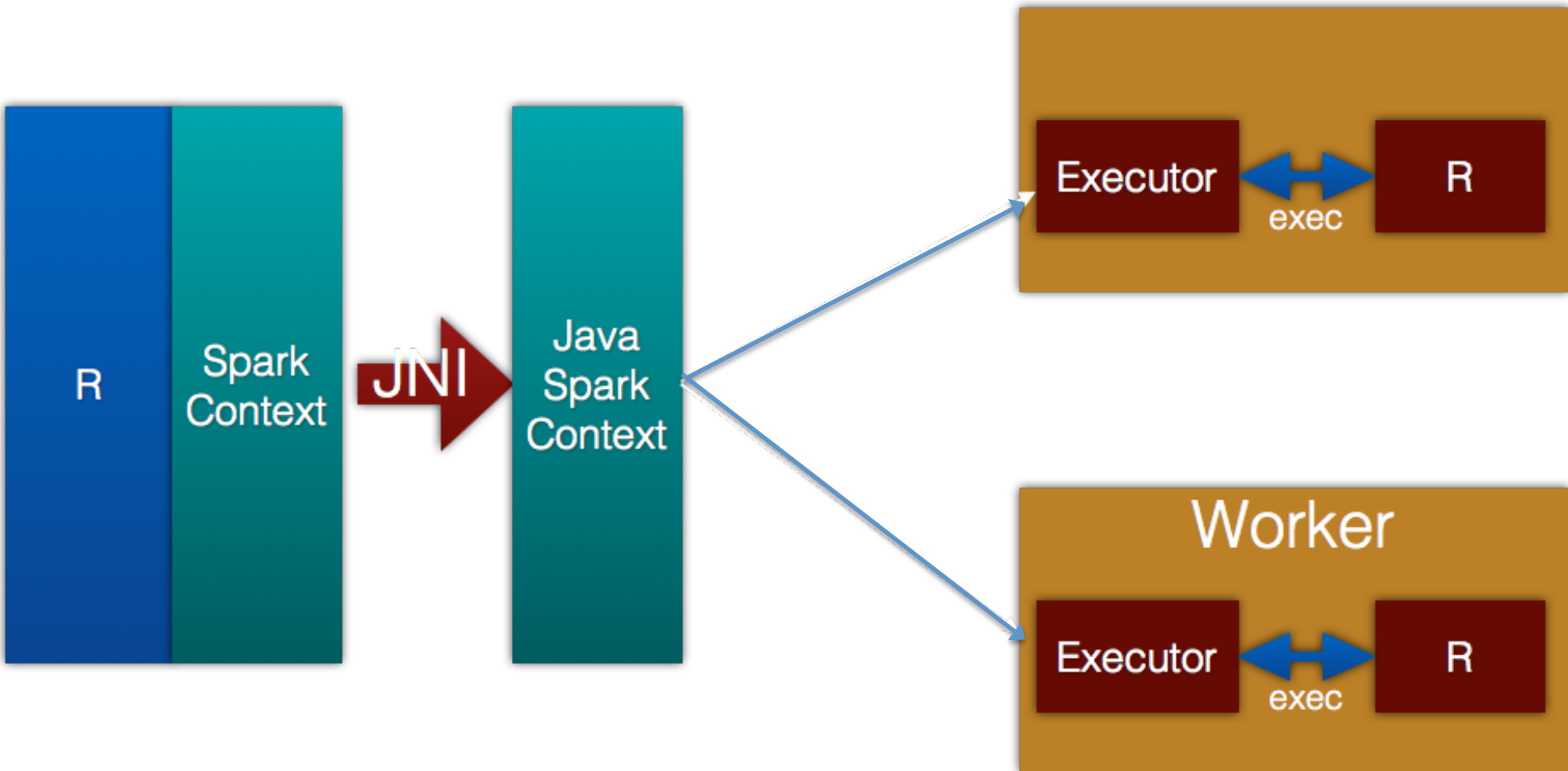
SparkR



RDDs as Distributed Lists

```
sc <- sparkR.init("local")  
lines <- textFile(sc, "hdfs://data.txt")  
wordsPerLine <- lapply(lines, function(line) { length(unlist(strsplit(line, " ")) }) })
```

SparkR



BlinkDB

- Queries with Bounded Errors and Bounded Response Times on Very Large Data

```
SELECT avg(sessionTime)
FROM Table
WHERE city='San Francisco'
WITHIN 2 SECONDS
```

Queries with Time Bounds

```
SELECT avg(sessionTime)
FROM Table
WHERE city='San Francisco'
ERROR 0.1 CONFIDENCE 95.0%
```

Queries with Error Bounds

JobServer

- provides a RESTful interface for submitting and managing Apache Spark jobs, jars, and job contexts

Spark Packages

- [http:// spark-packages.org](http://spark-packages.org)
- A community index of packages for Apache Spark

Q & A

weibo:@CrazyJvm

wechat public account : ChinaScala

Thank you all !