



Spark UI, Optimizer & Internal concepts (PhysicalPlan, Jobs, Stages, Tasks)

arnaud.nauwynck@gmail.com

This document:

[https://github.com/Arnaud-Nauwynck/presentations/
/pres-bigdata/12-sparkui-internals-plan-job-stage-task](https://github.com/Arnaud-Nauwynck/presentations/pres-bigdata/12-sparkui-internals-plan-job-stage-task)

Outline

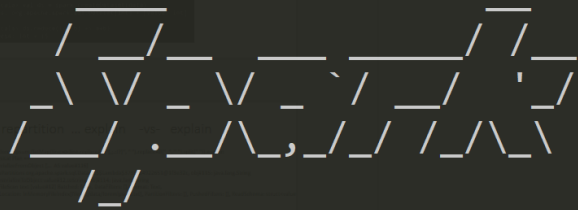
What's visible from Spark-ui

The concepts behinds Dataset/RDD execution : Job, Stage, Task

The architecture of deployment : Client, Driver, Executor

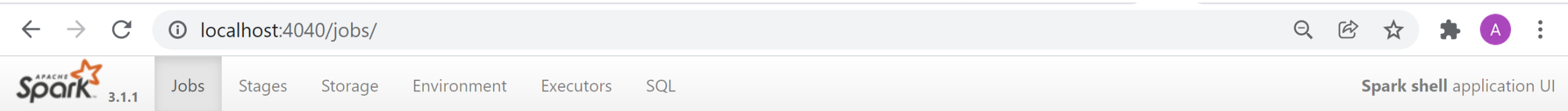
Spark-ui

```
Spark context Web UI available at http://DESKTOP-2EGCC8R:4040/
Spark context available as 'sc' (master = local[*], app id = 1)
Spark session available as 'spark'.
Welcome to
```




version 3.1.1

http://localhost:4040



spark-ui [1/6]: Jobs

 3.1.1

Jobs

Stages

Storage

Environment

Executors

SQL

Spark shell application UI

Spark Jobs (?)

User: arnaud
Total Uptime: 1,5 h
Scheduling Mode: FIFO
Completed Jobs: 11

▶ Event Timeline

▼ Completed Jobs (11)

Page:

1 Pages. Jump to . Show items in a page.

Job Id ▼	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
10	show at <console>:26 show at <console>:26	2021/12/31 19:58:07	19 ms	1/1	<div>1/1</div>
9	count at <console>:26 count at <console>:26	2021/12/31 19:57:51	66 ms	2/2	<div>2/2</div>
8	show at <pastie>:26 show at <pastie>:26	2021/12/31 19:55:51	26 ms	1/1	<div>1/1</div>
7	show at <console>:26 show at <console>:26	2021/12/31 19:54:28	15 ms	1/1	<div>1/1</div>
6	show at <console>:26	2021/12/31 19:53:59	16 ms	1/1	<div>1/1</div>

Spark-ui [1/6]: Job details DAG

Details for Job 4

Status: SUCCEEDED

Submitted: 2022/01/02 18:55:23

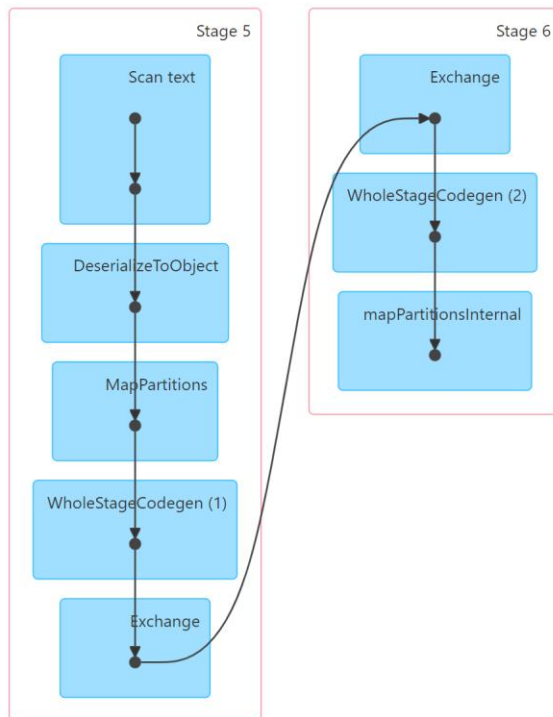
Duration: 0,2 s

Associated SQL Query: 2

Completed Stages: 2

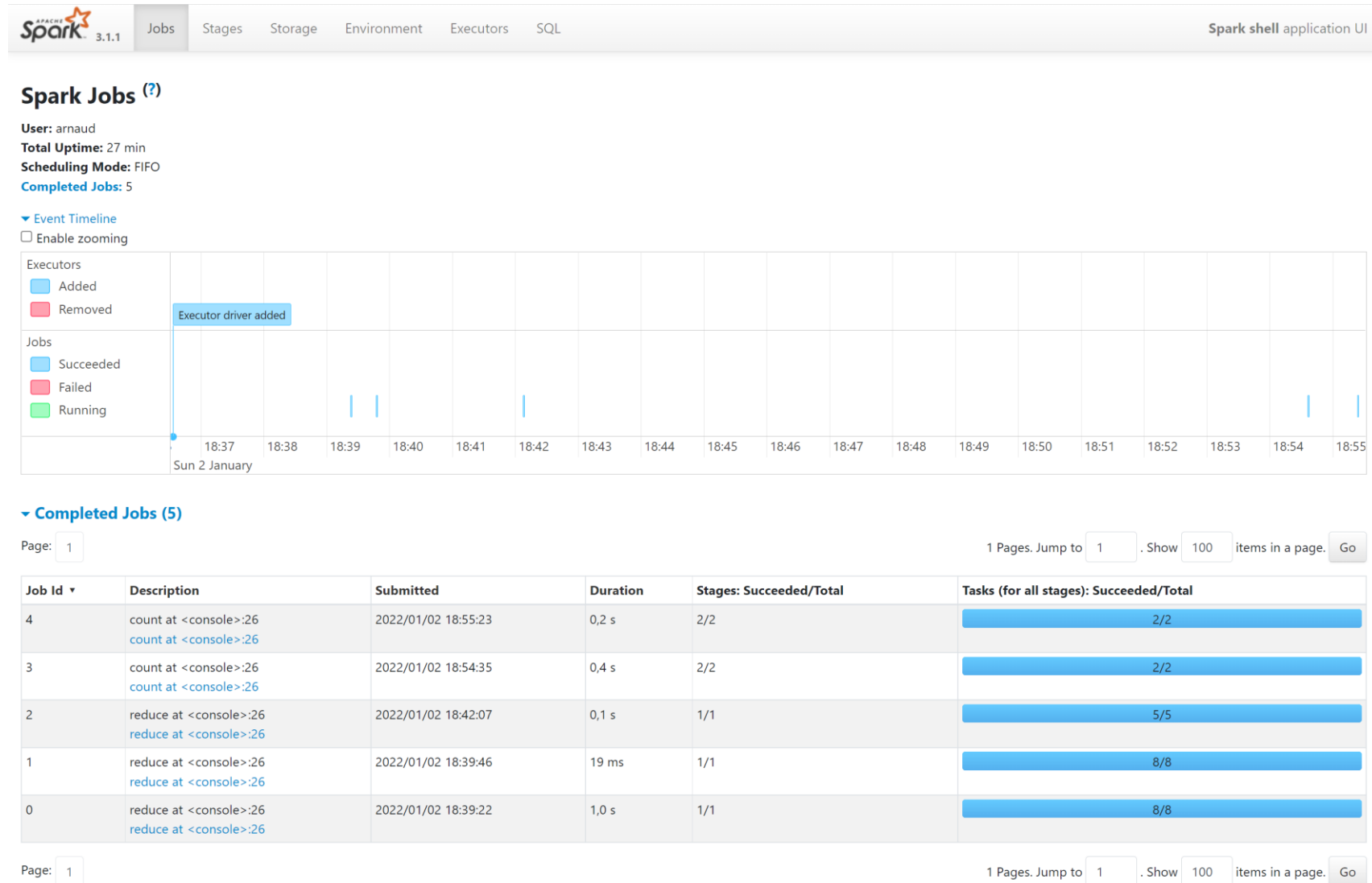
▶ Event Timeline

▼ DAG Visualization



▼ Completed Stages (2)

Spark-ui [1/6] detail Job timeline + statistics



Spark-ui [2/6]: detail Job > Stage > Task

Details for Stage 12 (Attempt 0)

Resource Profile Id: 0

Total Time Across All Tasks: 4 ms

Locality Level Summary: Process local: 1

Input Size / Records: 450.0 B / 1

Associated Job Ids: 10

- ▶ [DAG Visualization](#)
- ▶ [Show Additional Metrics](#)
- ▶ [Event Timeline](#)

Summary Metrics for 1 Completed Tasks

Metric	Min	25th percentile	Median	75th percentile	Max
Duration	4.0 ms	4.0 ms	4.0 ms	4.0 ms	4.0 ms
GC Time	0.0 ms	0.0 ms	0.0 ms	0.0 ms	0.0 ms
Input Size / Records	450 B / 1	450 B / 1	450 B / 1	450 B / 1	450 B / 1

Showing 1 to 3 of 3 entries

▶ [Aggregated Metrics by Executor](#)

Tasks (1)

Show 20 entries

Search:

Index	Task ID	Attempt	Status	Locality level	Executor ID	Host	Logs	Launch Time	Duration	GC Time	Input Size / Records	Errors
0	12	0	SUCCESS	PROCESS_LOCAL	driver	DESKTOP-2EGCC8R		2021-12-31 19:58:07	4.0 ms		450 B / 1	

Showing 1 to 1 of 1 entries

Previous 1 Next

Spark-ui [3/6] Storage (cache, persist)

[Jobs](#)[Stages](#)[Storage](#)[Environment](#)[Executors](#)[SQL](#)

Spark shell application UI

Storage

▼ RDDs

ID	RDD Name	Storage Level	Cached Partitions	Fraction Cached	Size in Memory	Size on Disk
28	<code>*(1) SerializeFromObject [staticinvoke(class org.apache.spark.unsafe.types.UTF8String, StringType, fromString, input[0, java.lang.String, true], true, false) AS value#30] +- MapPartitions org.apache.spark.sql.Dataset\$\$Lambda\$4450/17435982@68a3d4, obj#29: java.lang.String +- DeserializeToObject value#21.toString, obj#28: java.lang.String +- FileScan text [value#21] Batched: false, DataFilters: [], Format: Text, Location: InMemoryFileIndex[file:/c:/data/loremIpsum.txt], PartitionFilters: [], PushedFilters: [], ReadSchema: struct<value:string></code>	Disk Memory Deserialized 1x Replicated	1	100%	952.0 B	0.0 B
40	<code>FileScan text [value#21] Batched: false, DataFilters: [], Format: Text, Location: InMemoryFileIndex[file:/c:/data/loremIpsum.txt], PartitionFilters: [], PushedFilters: [], ReadSchema: struct<value:string></code>	Disk Memory Deserialized 1x Replicated	1	100%	960.0 B	0.0 B

Spark-ui [4/6] Environment: jvm+spark+hadoop+..

[Jobs](#)[Stages](#)[Storage](#)[Environment](#)[Executors](#)[SQL](#)[Spark shell application UI](#)

Environment

- ▶ [Runtime Information](#)
- ▶ [Spark Properties](#)
- ▶ [Resource Profiles](#)
- ▶ [Hadoop Properties](#)
- ▶ [System Properties](#)
- ▶ [Classpath Entries](#)

▼ Spark Properties















Name	Value
spark.app.id	local-1640975191762
spark.app.name	Spark shell
spark.app.startTime	1640975189843
spark.driver.host	DESKTOP-2EGCC8R
spark.driver.port	57122
spark.executor.id	driver
spark.home	C:\apps\hadoop\spark-3.1.1
spark.jars	
spark.master	local[*]
spark.repl.class.outputDir	C:\Users\arnaud\AppData\Local\Temp\spark-169a28ae-4479-4585-ab47-1e354d2d1347\repl-ff2cda14-38ec-4b1e-9c63-c4f02e84bf14
spark.repl.class.uri	spark://DESKTOP-2EGCC8R:57122/classes
spark.scheduler.mode	FIFO
spark.sql.catalogImplementation	hive
spark.submit.deployMode	client
spark.submit.pyFiles	
spark.ui.showConsoleProgress	true

Spark-ui [5/6]: Executors / Driver

Executors

[Show Additional Metrics](#)

Summary

	 RDD Blocks 	 Storage Memory 	 Disk Used 	 Cores 	 Active Tasks 	 Failed Tasks 	 Complete Tasks 	Total Tasks	Task Time (GC Time)	Input	Shuffle Read	Shuffle Write	Excluded
Active(1)	0	0.0 B / 413.9 MiB	0.0 B	8	0	0	13	13	0.8 s (18.0 ms)	4.6 KiB	118 B	118 B	0
Dead(0)	0	0.0 B / 0.0 B	0.0 B	0	0	0	0	0	0.0 ms (0.0 ms)	0.0 B	0.0 B	0.0 B	0
Total(1)	0	0.0 B / 413.9 MiB	0.0 B	8	0	0	13	13	0.8 s (18.0 ms)	4.6 KiB	118 B	118 B	0

















Executors

Show

20

 entries

Search:

Executor ID 	Address 	Status 	 RDD Blocks 	 Storage Memory 	 Disk Used 	 Cores 	 Active Tasks 	 Failed Tasks 	 Complete Tasks	Total Tasks	Task Time (GC Time)	Input	Shuffle Read	Shuffle Write	Thread Dump
driver	DESKTOP-2EGCC8R:57178	Active	0	0.0 B / 413.9 MiB	0.0 B	8	0	0	13	13	0.8 s (18.0 ms)	4.6 KiB	118 B	118 B	Thread Dump

Showing 1 to 1 of 1 entries

Previous

1

Next

Spark-ui [6/6] SQL

SQL

Completed Queries: 3

▼ Completed Queries (3)

Page: 1

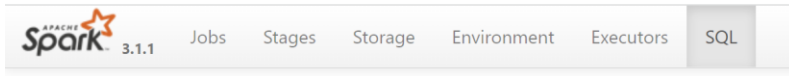
1 Pages. Jump to 1. Show 100 items in a page. Go

ID ▼	Description	Submitted	Duration	Job IDs
2	count at <console>:26 +details	2022/01/02 18:55:23	0,2 s	[4]
1	count at <console>:26 +details	2022/01/02 18:54:35	0,7 s	[3]
0	reduce at <console>:26 +details	2022/01/02 18:42:06	0,2 s	[2]

Page: 1

1 Pages. Jump to 1. Show 100 items in a page. Go

Spark-ui [6/6] Sql detail for query



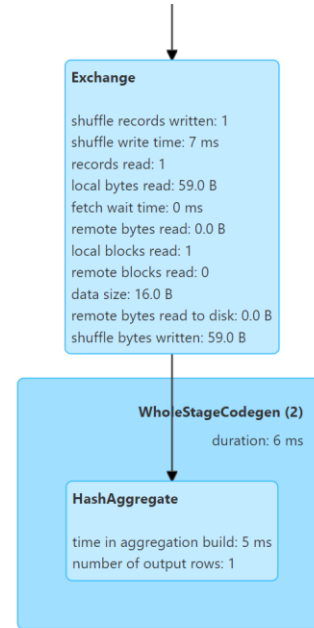
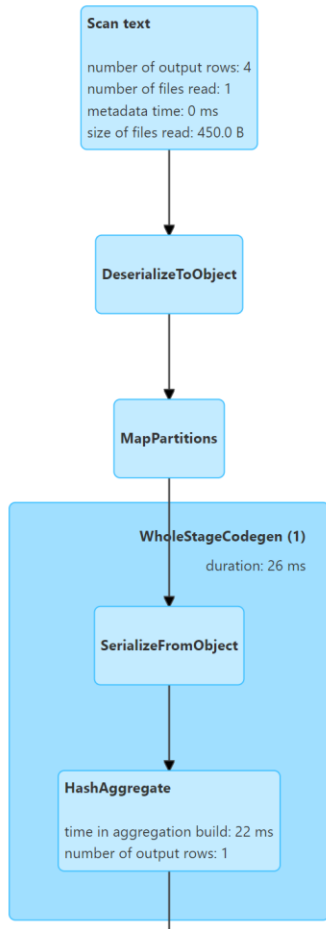
Details for Query 2

Submitted Time: 2022/01/02 18:55:23

Duration: 0,2 s

Succeeded Jobs: 4

☐ Show the Stage ID and Task ID that corresponds to the max metric



Details

```
== Physical Plan ==
* HashAggregate (7)
+- Exchange (6)
  +- * HashAggregate (5)
    +- * SerializeFromObject (4)
      +- MapPartitions (3)
        +- DeserializeToObject (2)
          +- Scan text (1)

(1) Scan text
Output [1]: [value#21]
Batched: false
Location: InMemoryFileIndex [file:/c:/data/loremIpsum.txt]
ReadSchema: struct<value:string>

(2) DeserializeToObject
Input [1]: [value#21]
Arguments: value#21.toString, obj#28: java.lang.String

(3) MapPartitions
Input [1]: [obj#28]
Arguments: org.apache.spark.sql.Dataset$$Lambda$4450/17435982@68a3d4, obj#29: java.lang.String
```

```
(4) SerializeFromObject [codegen id : 1]
Input [1]: [obj#29]

(5) HashAggregate [codegen id : 1]
Input: []
Keys: []
Functions [1]: [partial_count(1)]
Aggregate Attributes [1]: [count#35L]
Results [1]: [count#36L]

(6) Exchange
Input [1]: [count#36L]
Arguments: SinglePartition, ENSURE_REQUIREMENTS, [id=#55]

(7) HashAggregate [codegen id : 2]
Input [1]: [count#36L]
Keys: []
Functions [1]: [count(1)]
Aggregate Attributes [1]: [count(1)#32L]
Results [1]: [count(1)#32L AS count#33L]
```

Spark Internals

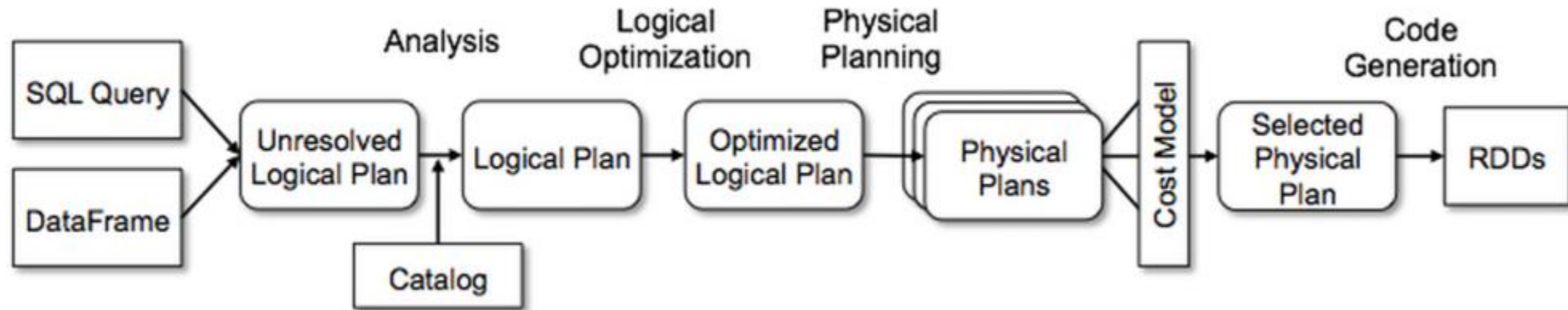
Logical/PhysicalPlan

Jobs - Stage - Tasks

Sql -> parsed AST -> Dataset + expression API

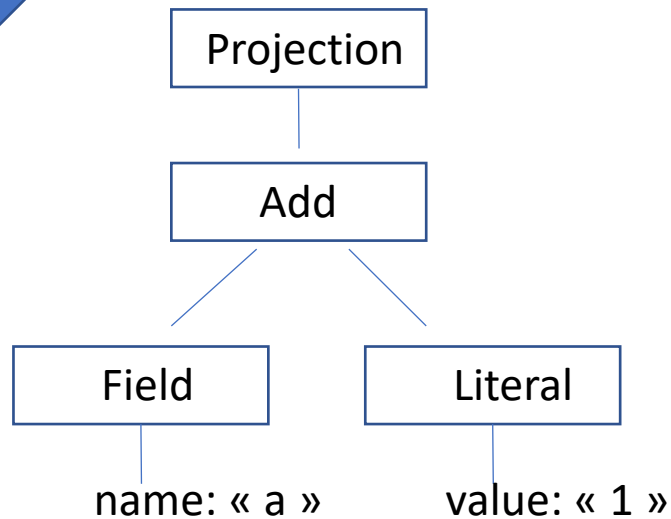
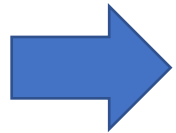
Optimisation: re-write
from LogicalPlan to PhysicalPlan

SQL / DataSet -> AST -> Catalyst -> Code + RDD



Compiler chain: Parser -> AST -> ..

Sql> select **(a+1)** from ...



```
110 case class Add(left: Expression, right: Expression) extends BinaryArithmetic {
111   override def symbol: String = "+"
112
113   lazy val numeric = dataType match {
114     case n: NumericType => n.numeric.asInstanceOf[Numeric[Any]]
115     case other => sys.error(s"Type $other does not support numeric operations")
116   }
117
118   override def eval(input: Row): Any = {
119     val evalE1 = left.eval(input)
120     if(evalE1 == null) {
121       null
122     } else {
123       val evalE2 = right.eval(input)
124       if (evalE2 == null) {
125         null
126       } else {
127         numeric.plus(evalE1, evalE2)
128       }
129     }
130   }
131 }
```


Catalyst AST -> CodeGenerator

📄 [apache / spark](#) Public

👁 Watch 2.1k ▾

🍴 Fork 25k

☆ Star 31.7k ▾

<> Code

🔗 Pull requests 241

🎬 Actions

📁 Projects

🛡 Security

📈 Insights

🔗 fedbfc7074 ▾

[spark](#) / [sql](#) / [catalyst](#) / [src](#) / [main](#) / [scala](#) / [org](#) / [apache](#) / [spark](#) / [sql](#) / [catalyst](#) /
[expressions](#) / [codegen](#) / **CodeGenerator.scala**

Go to file

...

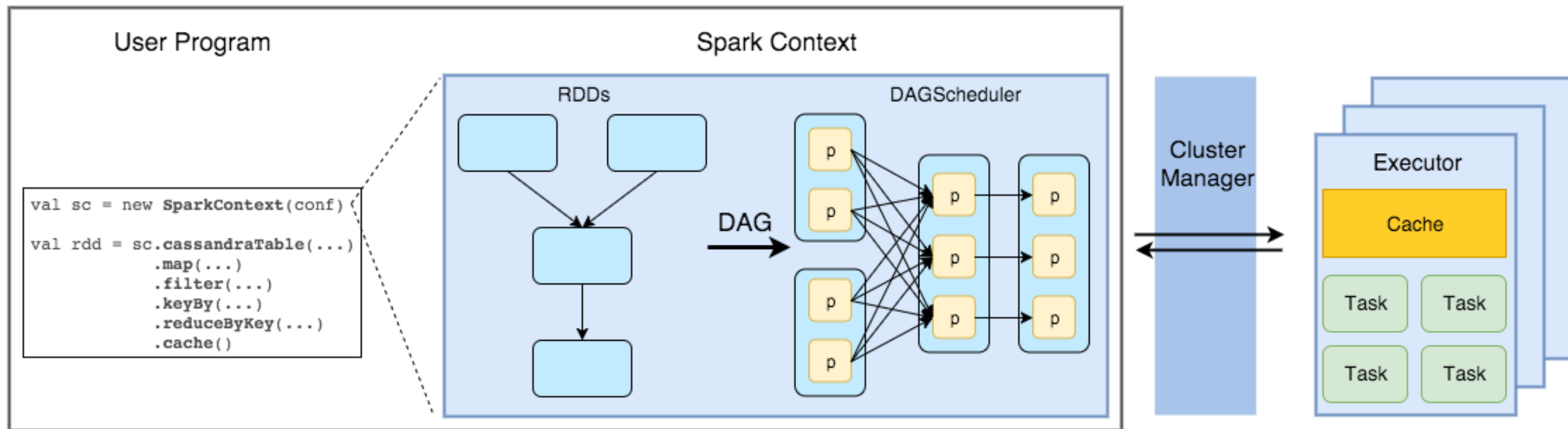
```
33  /**
34   * A base class for generators of byte code to perform expression evaluation. Includes a set of
35   * helpers for referring to Catalyst types and building trees that perform evaluation of individual
36   * expressions.
37   */
38  abstract class CodeGenerator[InType <: AnyRef, OutType <: AnyRef] extends Logging {
39
363      case Add(e1, e2) => (e1, e2) evaluate { case (eval1, eval2) => q"$eval1 + $eval2" }
364      case Subtract(e1, e2) => (e1, e2) evaluate { case (eval1, eval2) => q"$eval1 - $eval2" }
365      case Multiply(e1, e2) => (e1, e2) evaluate { case (eval1, eval2) => q"$eval1 * $eval2" }
366
640  }
641  }
```

Runtime Execution:

Dataset -> RDD -> Dag Scheduler
-> Jobs, Stages, Tasks

Spark Application

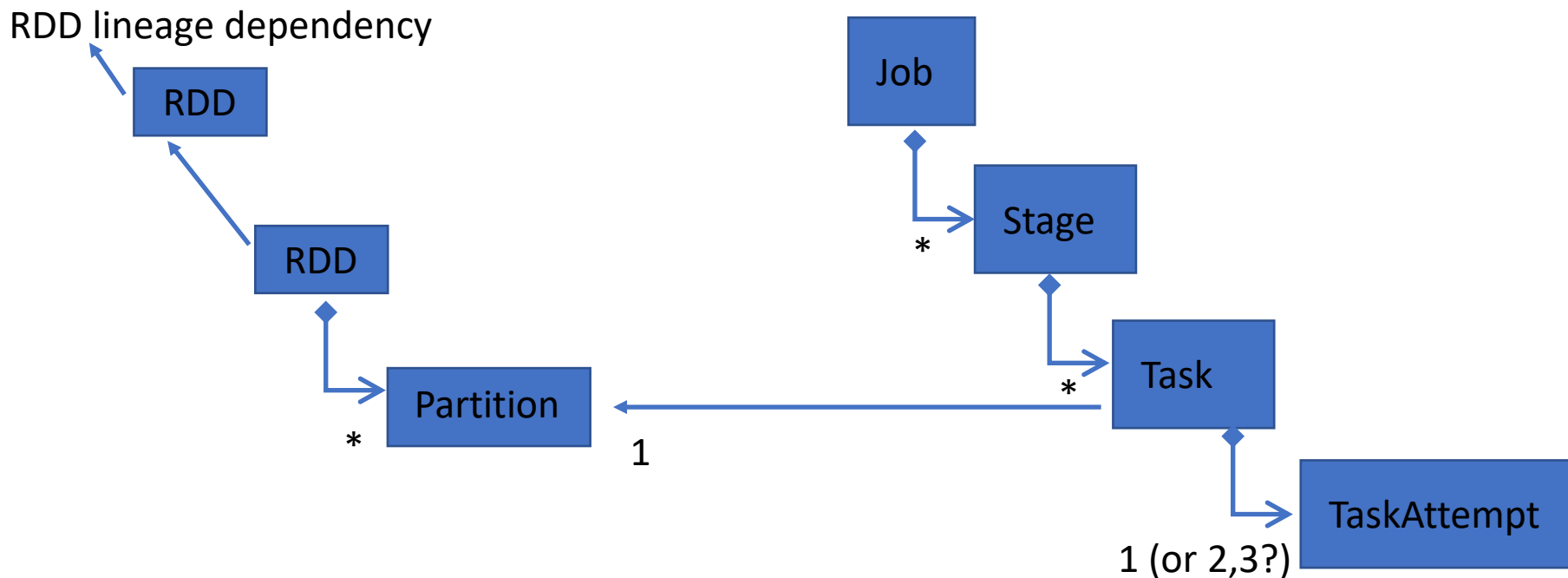
Workers



Spark-core Internal Concepts...

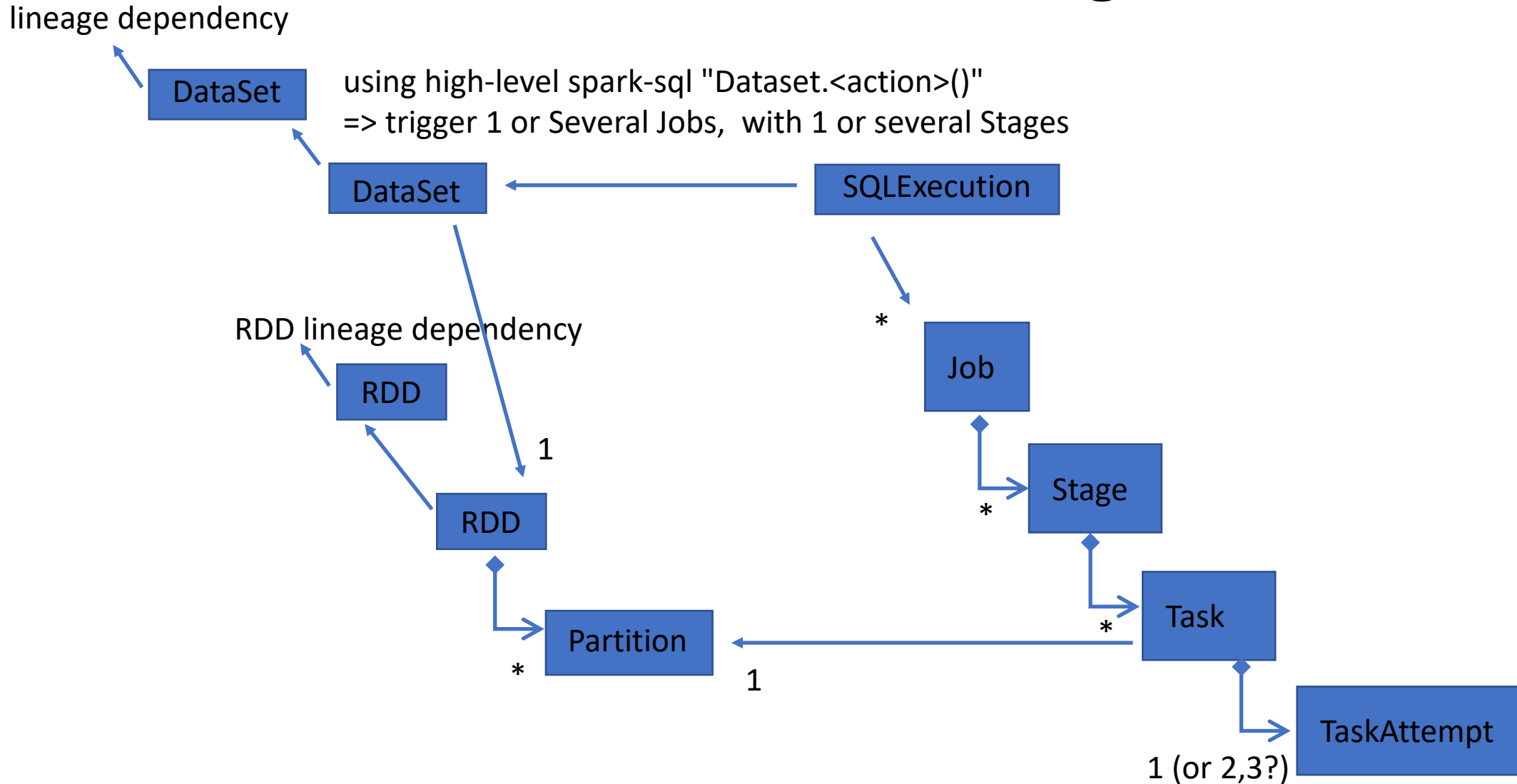
RDD, Job, Stage, Task

using only low-level spark-core "RDD.eval()"
=> trigger 1 Job, with 1 or several Stages



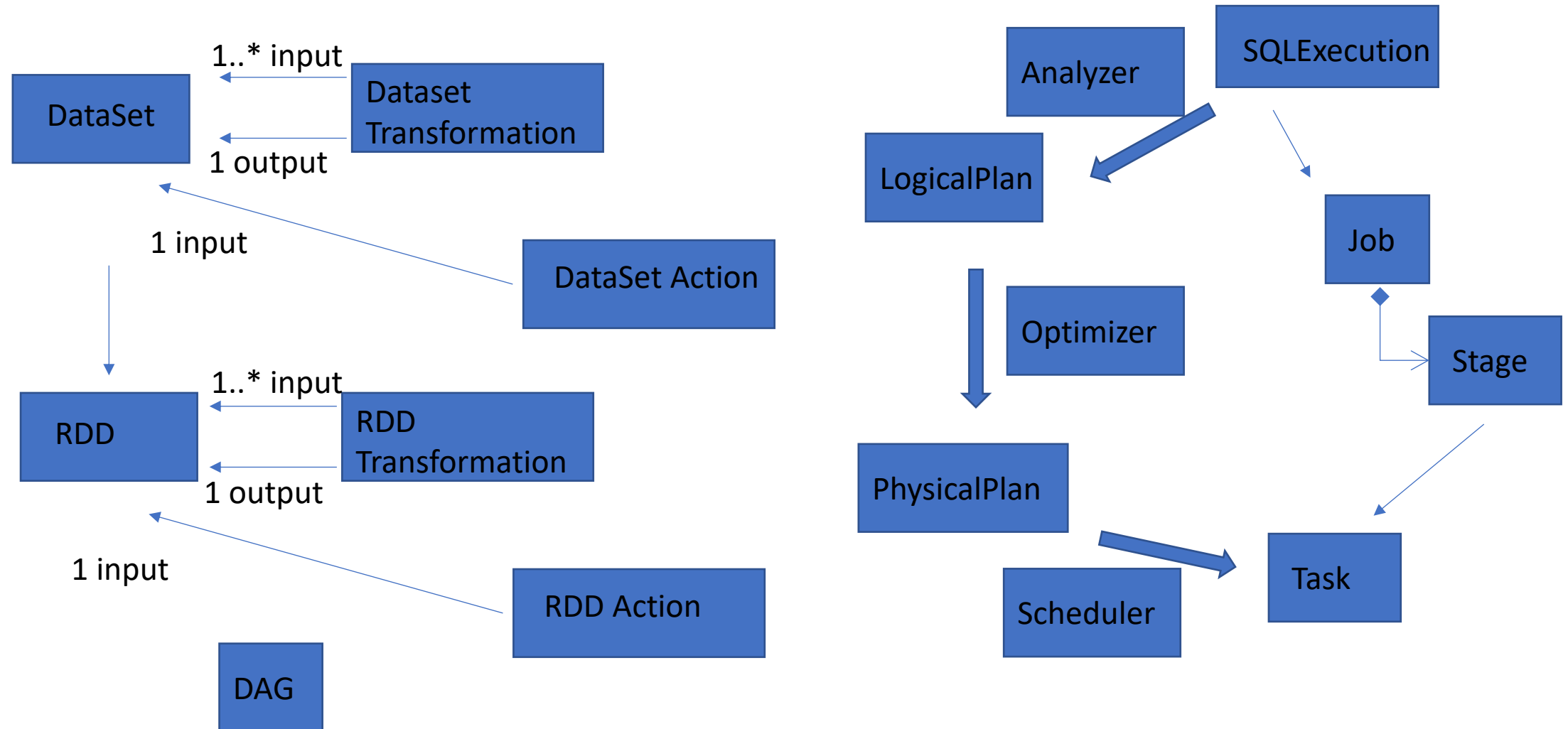
Spark-sql Internal Concepts...

SQLExecution, Job, Stage, Task

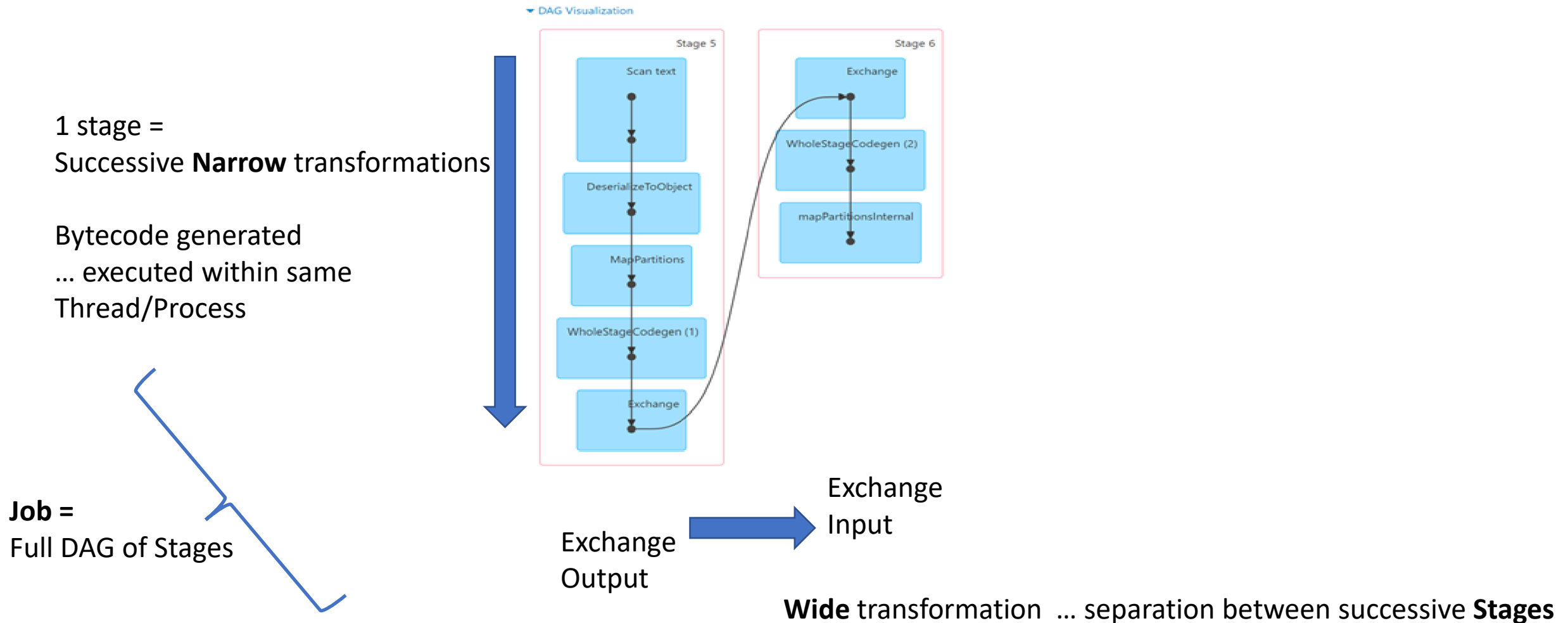


Spark Internal Concepts...

SQL, Job, Stage, Task

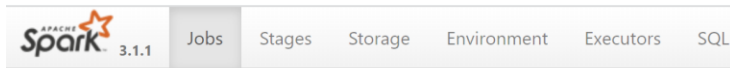


DAG Job = list of Stages
each Stage = N "narrow" transformations
between Stages = Exchange Shuffle ("Wide" transfo.)



example Word Count DAGs

simple -vs- repartition()



Details for Job 10

Status: SUCCEEDED

Submitted: 2021/12/31 19:58:07

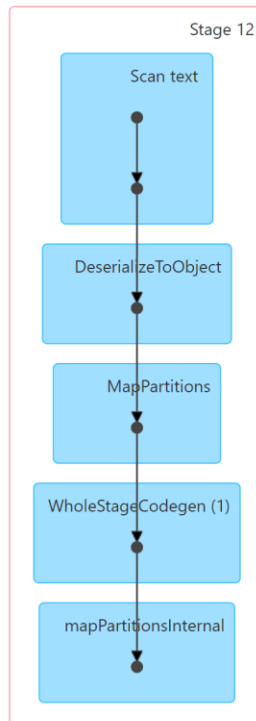
Duration: 19 ms

Associated SQL Query: 10

Completed Stages: 1

▶ Event Timeline

▼ DAG Visualization



▼ Completed Stages (1)



Details for Job 9

Status: SUCCEEDED

Submitted: 2021/12/31 19:57:51

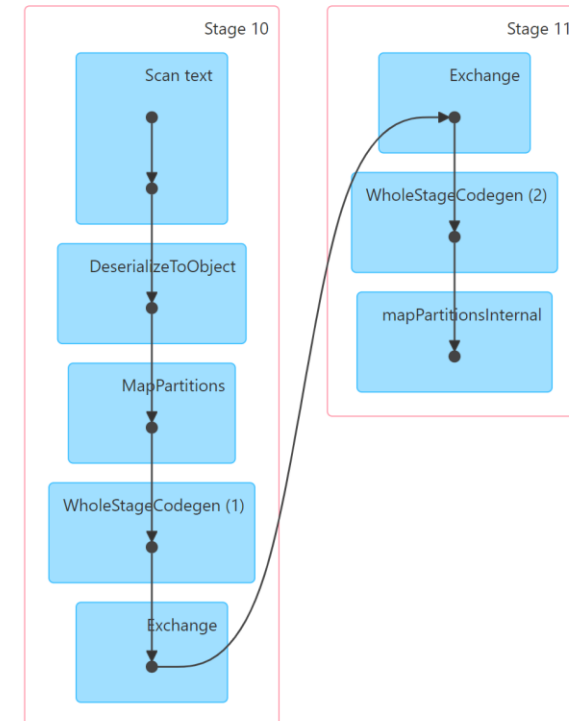
Duration: 66 ms

Associated SQL Query: 9

Completed Stages: 2

▶ Event Timeline

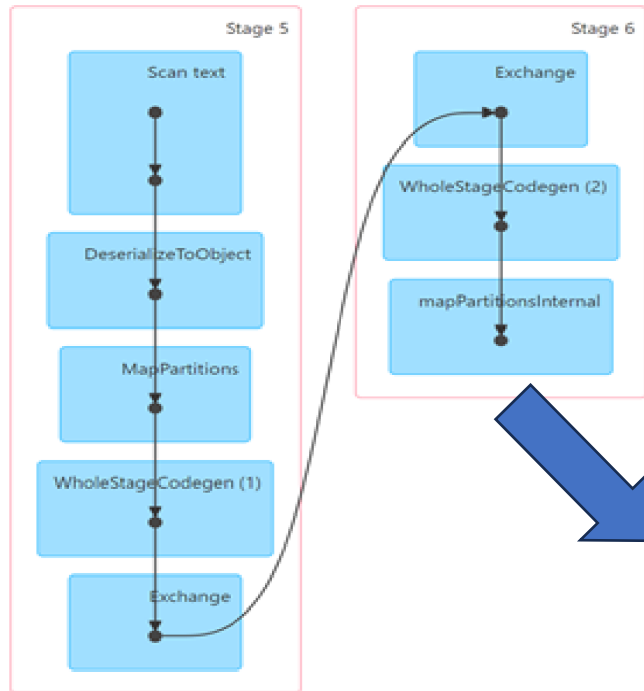
▼ DAG Visualization



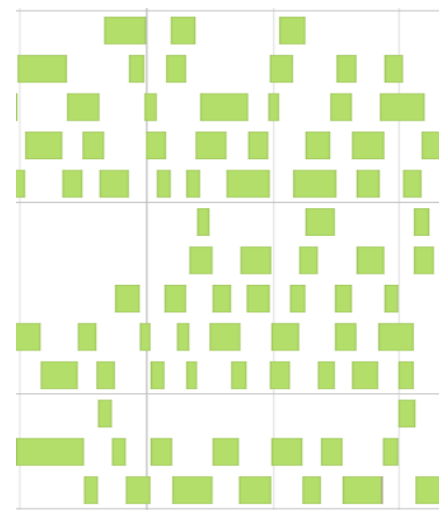
▼ Completed Stages (2)

Stage = list of Tasks : 1 per Partition
... each scheduled on an executor

▼ DAG Visualization



Tasks "**eventTimeline**" for stage

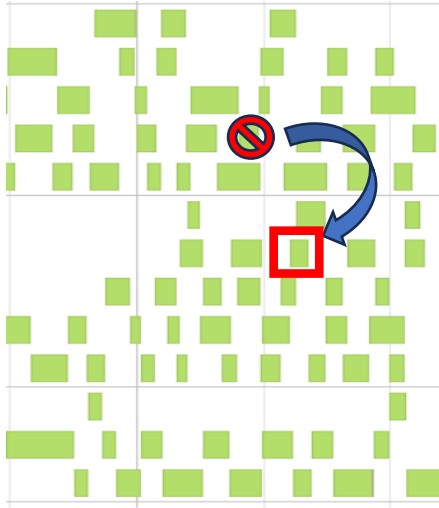


Tasks executed on executor 1

Tasks executed on executor 2

Tasks executed on executor 3

Tasks may fail, Retry individually



Tasks errors are automatically retried
(rescheduled)

By default,

`spark.task.maxFailures = 4`

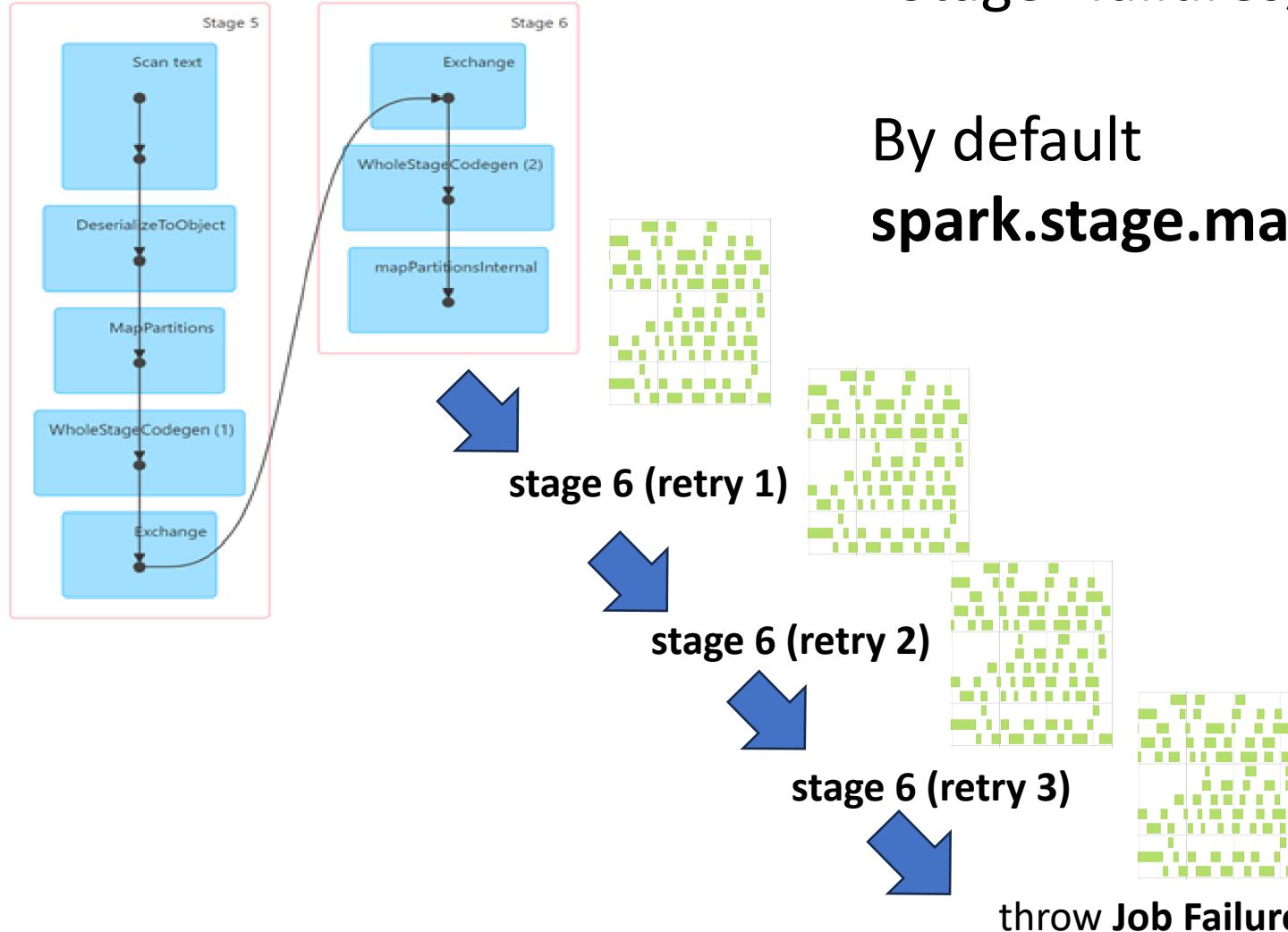
Stage Retry

"Stage" failures, are also retried

By default

`spark.stage.maxConsecutiveAttempts = 4`

▼ DAG Visualization



Architecture Components

Client-Cluster ... Driver & Executors

program/library/servers

Cluster Client -> Driver (-> SparkContext) -> Executor

Case 1/ Launch directly java.. main(String[] args) {
 SparkContext.getOrCreate()..
}

Case 2/ spark-submit -mode **client**

Case 3/
Yarn launch spark -mode **cluster**
+ relaunch if failed

Java spark-submit
-master yarn
-mode cluster

OS

Spark Driver

Java spark-submit

-master yarn

-mode client

SparkContext

SparkSession

ClusterManager

OS WorkerNode
(+ YarnNodeManager)

Spark Executor

Java CoarseGrainExecutor

Task

Cache

OS WorkerNode
(+ YarnNodeManager)

Spark Executor

Java CoarseGrainExecutor

Task

Cache

OS WorkerNode
(+ YarnNodeManager)

Questions ?

... Next Steps ?