livelessons
video instruction from technology experts

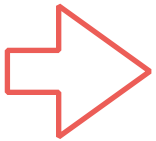
Lesson 4: Spark Internals

4.9 Spark Performance: Monitoring and Optimization

Performance Tuning

Monitoring

- Web UI (application)
- History Server (application)
- Ganglia (infrastructure)
- jstack, jmap, jstat (JVM)



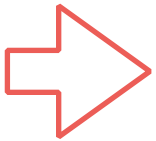
Optimization

- Data (serialization + locality)
- Application
  - data structures
  - caching
  - broadcasting
  - shuffle
- Framework (parallelism + memory + GC)

Performance Tuning

Monitoring

- Web UI (application)
- History Server (application)
- Ganglia (infrastructure)
- jstack, jmap, jstat (JVM)



Optimization

- Data (serialization + locality)
- Application
  - data structures
  - caching
  - broadcasting
  - shuffle
- Framework (parallelism + memory + GC)

Spark Web UI

Monitoring: Every `SparkContext` launches a web UI on the Driver machine defaulting to port 4040.

Note: The Spark Standalone cluster manager launches its own web UI on the Master node

Spark 1.4.1 **Spark Master at spark://127.0.0.1:7077**

URL: spark://127.0.0.1:7077
REST URL: spark://127.0.0.1:6066 *(cluster mode)*
Workers: 2
Cores: 2 Total, 0 Used
Memory: 2.0 GB Total, 0.0 B Used
Applications: 0 Running, 0 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

Workers

Worker Id	Address	State	Cores	Memory
worker-20150812221322-10.0.1.156-64391	10.0.1.156:64391	ALIVE	1 (0 Used)	1024.0 MB (0.0 B Used)
worker-20150812221322-10.0.1.156-64392	10.0.1.156:64392	ALIVE	1 (0 Used)	1024.0 MB (0.0 B Used)

Spark 1.4.1 Spark Master at spark://127.0.0.1:7077

URL: spark://127.0.0.1:7077
REST URL: spark://127.0.0.1:6066 *(cluster mode)*
Workers: 2
Cores: 2 Total, 2 Used
Memory: 2.0 GB Total, 1024.0 MB Used
Applications: 1 Running, 1 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

Workers

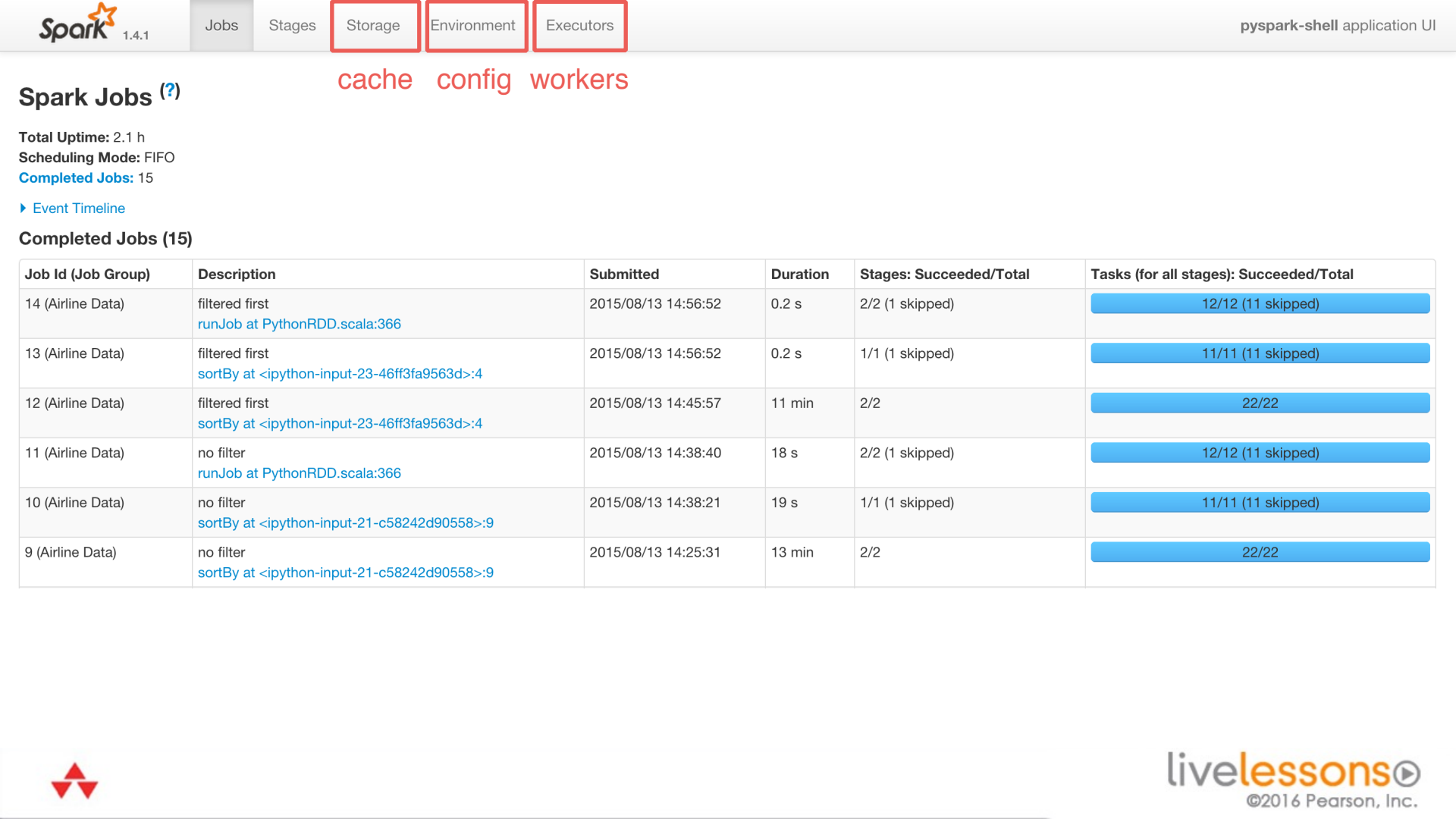
Worker Id	Address	State	Cores	Memory
worker-20150812221322-10.0.1.156-64391	10.0.1.156:64391	ALIVE	1 (1 Used)	1024.0 MB (512.0 MB Used)
worker-20150812221322-10.0.1.156-64392	10.0.1.156:64392	ALIVE	1 (1 Used)	1024.0 MB (512.0 MB Used)

Running Applications

Application ID		Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20150812221733-0001	(kill)	Performance Tuning	2	512.0 MB	2015/08/12 22:17:33	jonathandinu	RUNNING	12 s

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20150812221435-0000	pyspark-shell	2	512.0 MB	2015/08/12 22:14:35	jonathandinu	FINISHED	2.6 min



Spark 1.4.1 | Jobs | Stages | Storage | Environment | Executors | **pyspark-shell** application UI

Spark Jobs (?)

Total Uptime: 2.1 h
Scheduling Mode: FIFO
Completed Jobs: 15

▸ Event Timeline

Completed Jobs (15)

Job Id (Job Group)	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
14 (Airline Data)	filtered first runJob at PythonRDD.scala:366	2015/08/13 14:56:52	0.2 s	2/2 (1 skipped)	12/12 (11 skipped)
13 (Airline Data)	filtered first sortBy at <ipython-input-23-46ff3fa9563d>:4	2015/08/13 14:56:52	0.2 s	1/1 (1 skipped)	11/11 (11 skipped)
12 (Airline Data)	filtered first sortBy at <ipython-input-23-46ff3fa9563d>:4	2015/08/13 14:45:57	11 min	2/2	22/22
11 (Airline Data)	no filter runJob at PythonRDD.scala:366	2015/08/13 14:38:40	18 s	2/2 (1 skipped)	12/12 (11 skipped)
10 (Airline Data)	no filter sortBy at <ipython-input-21-c58242d90558>:9	2015/08/13 14:38:21	19 s	1/1 (1 skipped)	11/11 (11 skipped)
9 (Airline Data)	no filter sortBy at <ipython-input-21-c58242d90558>:9	2015/08/13 14:25:31	13 min	2/2	22/22

Actions initiate jobs

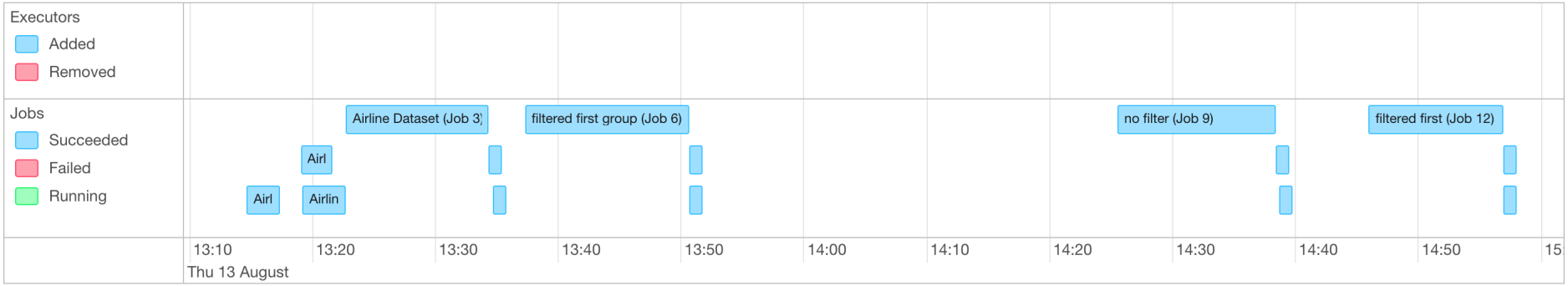
Spark 1.4.1 | Jobs | Stages | Storage | Environment | Executors | **pyspark-shell** application UI

Spark Jobs (?)

Total Uptime: 2.2 h
Scheduling Mode: FIFO
Completed Jobs: 15

▼ Event Timeline

☐ Enable zooming



Completed Jobs (15)

Job Id (Job Group)	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
14 (Airline Data)	filtered first runJob at PythonRDD.scala:366	2015/08/13 14:56:52	0.2 s	2/2 (1 skipped)	12/12 (11 skipped)

Actions initiate jobs