



CONNECT

Data Import

Workflow

Warehouse Explorer

Experimentation

Administration >

Monitoring >

DISCOVER

ACT

WORKFLOW

New job

Import job

Activity	Name	Description	Creation date	Actions
	AMI-DEMO	Sample Spark stages in Scala and Java with metrics and logs	05/20/2015 13:11:07	
	EnronEmails Tachyon	Ingest emails from the public enron dataset, convert them in parquet format in Tachyon, create a SQL table for querying	05/06/2015 15:33:35	
	EnronEmails Hdfs	Ingest emails from the public enron dataset, convert them in parquet format in HDFS, create a SQL table for querying	05/06/2015 11:30:52	
	MovieLens (recommender system)	Recommendation system based on collaborative filtering using MovieLens dataset.	05/06/2015 10:26:07	
	AMI-DEMO-old	Sample Spark stages in Scala and Java with metrics and logs	05/05/2015 18:38:51	

CONNECT

Data Import

Workflow

Warehouse Explorer

Experimentation

Administration >

Monitoring >

DISCOVER

ACT

WORKFLOW



Job name: AMI-DEMO

Run this job

Created: 05/20/2015 13:11:07

[Edit job properties](#)

Job runs

Scheduler

Designer

SPARK

SPARK
SQL

HIVE

IMPORT
DATA

SHELL

SSH

JAVA

EMAIL





Job name: AMI-DEMO

Run this job



Job runs



Scheduler



Designer

SPARK

SPARK
SQL

HIVE

IMPORT
DATA

SHELL

SSH

JAVA

EMAIL



Created: 05/20/2015 13:11:07

[Edit job properties](#)

Spark Stage



Configure Spark stage

Name *	<input type="text" value="scala"/>		
Description *	<input type="text" value="demo"/>		
Parameters ?	<input type="text" value='input="test param"'/>	+ Add	
Jar name ?	<input type="text" value="xpatterns-spark-api-4-ami.jar"/>		
Main class *	<input type="text" value="com.xpatterns.ami.sample.SampleScalaSparkJob"/>		
Parallelism	<input type="radio"/> Unlimited	<input checked="" type="radio"/> Limited to:	<input type="text" value="2"/> / 2 cores
Memory per node	<input type="range" value="2.0"/>		2.0 / 2.0 GB
Stage resources			
File to upload	<input type="button" value="Choose File"/> No file chosen		
Relative path	<input type="text"/>		
Resource description	<input type="text"/>		
			<input type="button" value="Upload"/>

[Show 1 resource](#)

OK

Cancel

Job name: AMI-DEMO Run this job

Created: 05/20/2015 13:11:07
[Edit job properties](#)

🔗 Job runs Scheduler Designer

SPARK

SPARK
SQL

HIVE

IMPORT
DATA

SHELL

SSH

JAVA

EMAIL



Spark Stage



Configure Spark stage

Name *

Description *

Parameters ? + Add 🗑️ ℹ️

Jar name ?

Main class *

Parallelism ☐ Unlimited ☒ Limited to: / 2 cores

Memory per node 2.0 / 2.0 GB

Stage resources

File to upload Choose File No file chosen

Relative path

Resource description

Upload

[Hide 1 resource](#)

xpatterns-spark-api-4-ami.jar 👁️ ℹ️ ✖️

OK

Cancel

CONNECT

Data Import

Workflow

Warehouse Explorer

Experimentation

Administration

Monitoring

DISCOVER

ACT

WORKFLOW

Start date

End date

Status

Duration

Trigger

05/20/2015 13:16:40

05/20/2015 13:18:23

✓

SUCCEEDED

1 min 43 secs

UI

✓

JOB RUN

⏮

✕

⏸

▶

✓

scala

✓

java

Logs

Metrics

○

Live

●

History

First page

◀ Previous page

Next page ▶

Last page

Filter by:

No filter

(181 / 181)

⛶

05/20/2015 13:17:13 INFO [bridge] *SparkLog* [SID[2] overall tasks [60 of 128] overall Shuffle [read [0.0 B] write [0.0 B]]

05/20/2015 13:17:13 INFO [bridge] *SparkLog* [SID[2] overall Tasks [72 of 128] overall Shuffle [read [0.0 B] write [0.0 B]]

05/20/2015 13:17:13 INFO [bridge] *SparkLog* [SID[2] overall Tasks [84 of 128] overall Shuffle [read [0.0 B] write [0.0 B]]

05/20/2015 13:17:13 INFO [bridge] *SparkLog* [SID[2] overall Tasks [96 of 128] overall Shuffle [read [0.0 B] write [0.0 B]]

05/20/2015 13:17:13 INFO [bridge] *SparkLog* [SID[2] overall Tasks [108 of 128] overall Shuffle [read [0.0 B] write [0.0 B]]

05/20/2015 13:17:13 INFO [bridge] *SparkLog* [SID[2] overall Tasks [120 of 128] overall Shuffle [read [0.0 B] write [0.0 B]]

05/20/2015 13:17:13 INFO [bridge] *SparkLog* [SID[2] Name[collect at SampleScalaSparkJob.scala:51] for JID[2] finished in 0.807 s ! Shuffle read[0.0 B] Shuffle write[0.0 B]]

05/20/2015 13:17:13 INFO *SCALA* [altered RDD]: 100

05/20/2015 13:17:13 INFO [bridge] *SparkLog* [JID[2] finished in 0.821 s with result JobSucceeded]

05/20/2015 13:17:13 INFO *SCALA* [altered RDD]: 200

05/20/2015 13:17:13 INFO *SCALA* [altered RDD]: 300

05/20/2015 13:17:13 INFO *SCALA* [altered RDD]: 400

05/20/2015 13:17:13 INFO *SCALA* [altered RDD]: 500

05/20/2015 13:17:13 INFO *SCALA* [altered RDD]: 600

05/20/2015 13:17:13 INFO ----- SampleScalaSparkJob done ... final results :..-----

05/20/2015 13:17:13 INFO [bridge] ===== XPatternsSparkBridge [SCALA branch - runJob results] =====

05/20/2015 13:17:13 INFO [bridge] ===== XPatternsSparkBridge [done :)] =====

05/20/2015 13:17:13 INFO [bridge] ===== XPatternsSparkBridge [SCALA branch - finished] =====

05/20/2015 13:17:13 INFO [bridge] *SCALA* XPatternsSparkBridge - Spark Job is finished !!!

05/20/2015 13:17:14 [launcher] Spark with jobId:84288ff9-e112-4a20-a825-5c04c19448fc finished with status: Finished

05/20/2015 13:17:14 [launcher] Delete context tcomponent-ubuntu-554-561-localhost-coarse-2-2048M if exists!

05/20/2015 13:17:14 [launcher] Validate delete context response!

05/20/2015 13:17:14 [launcher] Context deleted.

05/20/2015 13:17:14 INFO [bridge] *SparkLog* [Finished application in [1432117034918] ===

05/20/2015 13:17:14 INFO [bridge] =====

05/20/2015 13:17:14 INFO [bridge] *SparkLog* [closing all loggers (Cassandra, RabbitMq)

05/20/2015 13:17:19 [launcher] Deleted spark context: tcomponent-ubuntu-554-561-localhost-coarse-2-2048M

05/20/2015 13:17:19 [launcher] Spark execution for stage 554 finished successfully

05/20/2015 13:17:19 [launcher] ##### Stopping Launcher (System.exit(0)) #####

05/20/2015 13:17:23 Stage scala of job 551 changed status to SUCCEEDED

05/20/2015 13:17:23 Time to process stage 554 of job 551 was 0 min, 42 sec

Start date	End date	Status	Duration	Trigger
05/20/2015 13:16:40	05/20/2015 13:18:23	✔ SUCCEEDED	1 min 43 secs	UI
<div>✔ JOB RUN</div> <div><div>⏮</div><div>✖</div><div>⏸</div><div>▶</div></div> <div><div>✔ scala</div><div>✔ java</div></div>				
<div><div>Logs</div><div>Metrics</div></div> <div><div><div><div>○ Live</div><div>● History</div><div>First page</div><div>◀ Previous page</div><div>Next page ▶</div><div>Last page</div></div><div>Filter by: <div>No filter</div> (171 / 171)</div></div><div><div>05/20/2015 13:17:55 INFO [bridge] *SparkLog* [SID[1] progress [1 / 1] - TID[1][SUCCESS] in 15091 ms on ip-10-0-2-168.ec2.internal read[0.0 B]]</div><div>05/20/2015 13:17:55 INFO [bridge] *SparkLog* [SID[1] Name[count at SampleJavaSparkJob.java:74] for JID[1] finished in 15.092 s ! Shuffle read[0.0 B] Shuffle write[0.0 B]]</div><div>05/20/2015 13:17:55 INFO [bridge] *SparkLog* [JID[1] finished in 15.09 s with result JobSucceeded]</div><div>05/20/2015 13:17:55 INFO [bridge] *SparkLog* [JID[2] with SIDs[2] has been started</div><div>05/20/2015 13:17:55 INFO [bridge] =====</div><div>05/20/2015 13:17:55 INFO [bridge] *SparkLog* [JID[2] ->spark.jobInstanceId = 1</div><div>05/20/2015 13:17:55 INFO [bridge] *SparkLog* [JID[2] ->spark.xpatterns.jobId = 551</div><div>05/20/2015 13:17:55 INFO [bridge] *SparkLog* [JID[2] ->spark.xpatterns.jobInstanceId = 561</div><div>05/20/2015 13:17:55 INFO [bridge] *SparkLog* [JID[2] ->spark.xpatterns.stageId = 552</div><div>05/20/2015 13:17:55 INFO [bridge] *SparkLog* [SID[2] Name[saveAsTextFile at SampleJavaSparkJob.java:76] NumberOfTasks[1] has been started for JID[2]</div><div>05/20/2015 13:18:11 INFO [bridge] *SparkLog* [SID[2] progress [1 / 1] - TID[2][SUCCESS] in 15936 ms on ip-10-0-2-168.ec2.internal read[0.0 B]]</div><div>05/20/2015 13:18:11 INFO [bridge] *SparkLog* [SID[2] Name[saveAsTextFile at SampleJavaSparkJob.java:76] for JID[2] finished in 15.937 s ! Shuffle read[0.0 B] Shuffle write[0.0 B]]</div><div>05/20/2015 13:18:11 INFO [bridge] *SparkLog* [JID[2] finished in 15.958 s with result JobSucceeded]</div><div>05/20/2015 13:18:12 INFO [bridge] ----- SampleJavaSparkJob done -----</div><div>05/20/2015 13:18:12 INFO [bridge] ===== XPatternsSparkBridge [JAVA branch - runJob results:] =====</div><div>05/20/2015 13:18:12 INFO [bridge] ===== XPatternsSparkBridge [6] =====</div><div>05/20/2015 13:18:12 INFO [bridge] ===== XPatternsSparkBridge [JAVA branch - finished] =====</div><div>05/20/2015 13:18:12 INFO [bridge] *SCALA* XPatternsSparkBridge - Spark Job is finished !!!</div><div>05/20/2015 13:18:16 [launcher] Spark with jobId:d465c8aa-f9ff-4ba9-a4c3-95719486ec37 finished with status: Finished</div><div>05/20/2015 13:18:16 [launcher] Delete context tcomponent-ubuntu-552-561-localhost-coarse-2-2048M if exists!</div><div>05/20/2015 13:18:16 [launcher] Validate delete context response!</div><div>05/20/2015 13:18:16 [launcher] Context deleted.</div><div>05/20/2015 13:18:16 INFO [bridge] *SparkLog* [Finished application in [1432117096188] ==</div><div>05/20/2015 13:18:16 INFO [bridge] =====</div><div>05/20/2015 13:18:16 INFO [bridge] *SparkLog* [closing all loggers (Cassandra, RabbitMq)</div><div>05/20/2015 13:18:21 [launcher] Deleted spark context: tcomponent-ubuntu-552-561-localhost-coarse-2-2048M</div><div>05/20/2015 13:18:21 [launcher] Spark execution for stage 552 finished successfully</div><div>05/20/2015 13:18:21 [launcher] ##### Stopping Launcher (System.exit(0)) #####</div><div>05/20/2015 13:18:23 Stage java of job 551 changed status to SUCCEEDED</div><div>05/20/2015 13:18:23 Time to process stage 552 of job 551 was 1 min, 1 sec</div></div></div>				
05/20/2015 13:13:08	05/20/2015 13:14:51	✔ SUCCEEDED	1 min 43 secs	UI

✓ JOB RUN Logs Metrics Spark: line.process|Throughput

