

AWS DynamoDB setup:

Create tables:

links_per_artist with partition key artist

links_per_title with partition key title

Example:

The screenshot displays the AWS Management Console interface for a DynamoDB table named 'links_per_title'. On the left, a sidebar shows a list of tables with 'links_per_title' selected. The main panel shows the 'Overview' tab for this table. It includes sections for 'Recent alerts' (no alarms triggered), 'Stream details' (stream disabled), and 'Table details' (provisioned capacity, creation date, and resource name).

Table details

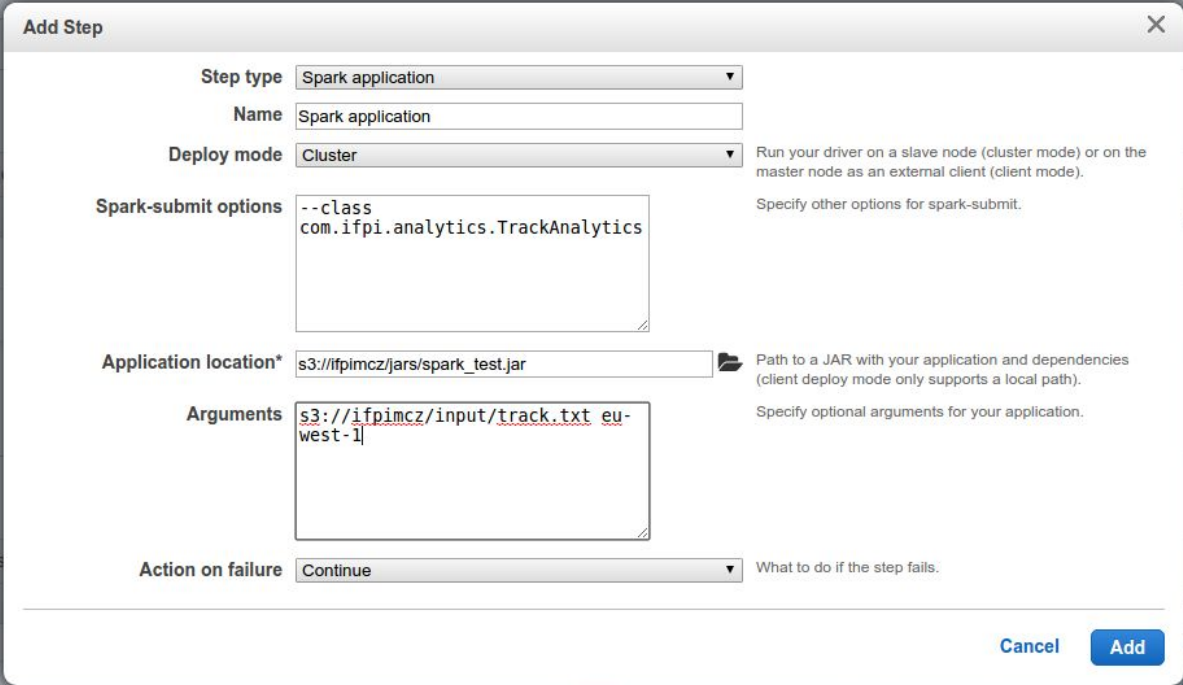
Table name	links_per_title
Primary partition key	title (String)
Primary sort key	-
Table status	Active
Creation date	July 23, 2016 at 7:39:19 PM UTC+1
Provisioned read capacity units	5
Provisioned write capacity units	5
Last decrease time	-
Last increase time	-
Storage size (in bytes)	56.00 bytes
Item count	2
Region	EU West (Ireland)
Amazon Resource Name (ARN)	arn:aws:dynamodb:eu-west-1:945222632356:table/links_per_title

Storage size and item count are not updated in real-time. They are updated periodically, roughly every six hours.

Running spark job:

1. Upload spark_test.jar fat jar to s3.
2. Add step with following configuration:
Step type: Spark application
Deploy mode: Cluster
Spark-submit options: `--class com.ifpi.analytics.TrackAnalytics`
Application location: `<path to spark_test.jar>`
Arguments: `<s3 path to input file> <dynamoDB region>`

Example:



The screenshot shows a dialog box titled "Add Step" with a close button (X) in the top right corner. The dialog contains the following fields and options:

- Step type:** A dropdown menu set to "Spark application".
- Name:** A text input field containing "Spark application".
- Deploy mode:** A dropdown menu set to "Cluster". To the right of this field is a note: "Run your driver on a slave node (cluster mode) or on the master node as an external client (client mode)."
- Spark-submit options:** A text area containing the code `--class com.ifpi.analytics.TrackAnalytics`. To the right is a note: "Specify other options for spark-submit."
- Application location*:** A text input field containing `s3://ifpimcz/jars/spark_test.jar`. To the right is a note: "Path to a JAR with your application and dependencies (client deploy mode only supports a local path)."
- Arguments:** A text area containing the code `s3://ifpimcz/input/track.txt eu-west-1`. To the right is a note: "Specify optional arguments for your application."
- Action on failure:** A dropdown menu set to "Continue". To the right is a note: "What to do if the step fails."

At the bottom right of the dialog are two buttons: "Cancel" and "Add".

Check results in DynamoDB:

Create tableActions ▼

Filter by table name X

Name ▲

☒ links_per_artist

☐ links_per_title

links_per_artistClose

OverviewItemsMetricsAlarmsCapacityIndexesTriggersAccess control

Create itemActions ▼

Scan: [Table] links_per_artist: artist ▲

Scan ▼ [Table] links_per_artist: artist ▼ ▲

⊕ Add filter

Start search

<input type="checkbox"/>	artist	count	
<input type="checkbox"/>	RIHANNA	6	

Create tableActions ▼

Filter by table name X

Name ▲

☐ links_per_artist

☒ links_per_title

links_per_titleClose

OverviewItemsMetricsAlarmsCapacityIndexesTriggersAccess control

Create itemActions ▼

Scan: [Table] links_per_title: title ▲

Scan ▼ [Table] links_per_title: title ▼ ▲

⊕ Add filter

Start search

<input type="checkbox"/>	title	count	
<input type="checkbox"/>	FOUR FIVE SE	5	
<input type="checkbox"/>	FOURFIVESEC	1	