

Sentiment Analysis on Streaming Data using Azure Databricks

Prashant Kumar Mishra
Cloud Solution Architect



Instructions:

Please keep your microphones on mute during the entire session

Type in your questions in chat window, it will be answered towards the end of our session

Sentiment Analysis on Streaming Data using Azure Databricks

Welcome



Course Navigation

Prerequisites

Section 1

Demo

Section 2

Section 1

Create Event Hub

Create Azure Databricks workspace

Create a spark cluster in Azure Databricks

Create a Twitter app to access streaming data

Attach libraries for Event hubs and twitter API

Create a Cognitive Services account and retrieve access key

Section 2

Architecture

Send Tweets to Event Hubs

Read tweets from Event Hubs

Run Sentiment Analysis on tweets

Connect to Power BI

Create a Power BI dashboard

Sentiment Analysis on Streaming Data using Azure Databricks

Create Event Hub



Course Navigation

Prerequisites

Section 1

Create Event Hub

Create Azure Databricks workspace

Create a spark cluster in Azure Databricks

Create a Twitter app to access streaming data

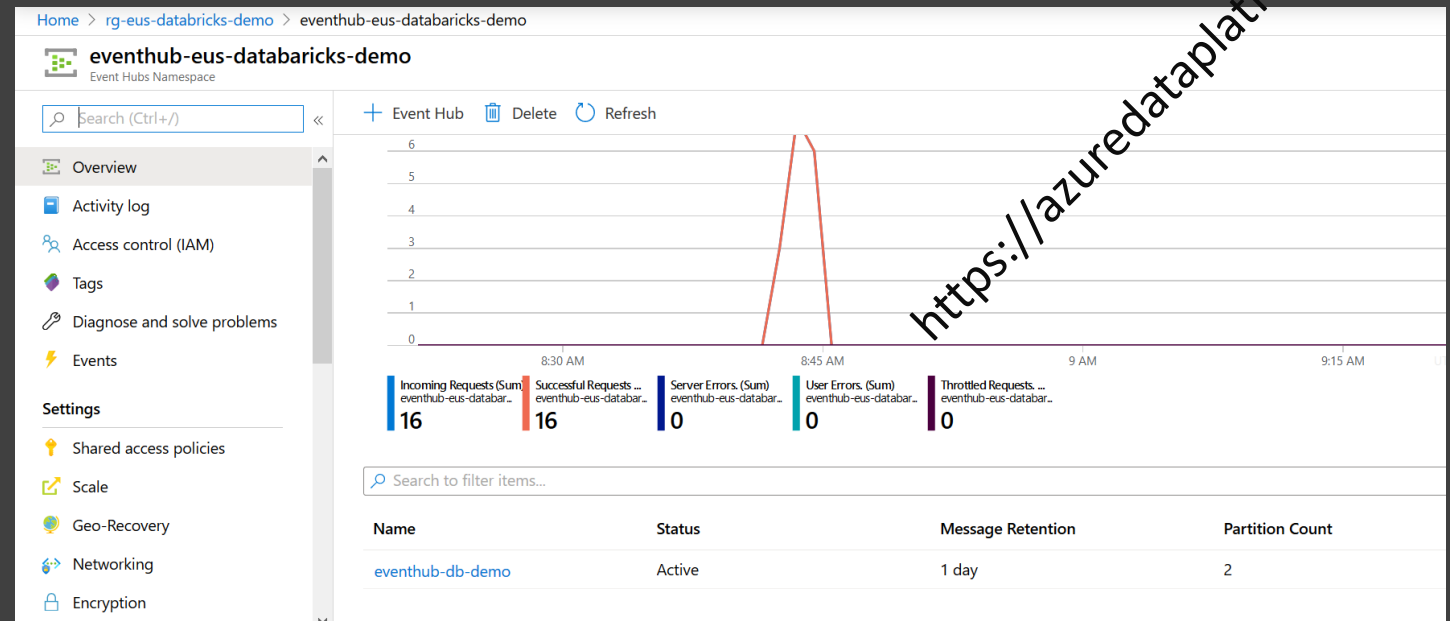
Attach libraries for Event hubs and twitter API

Create a Cognitive Services account and retrieve access key

Demo

Section 2

Azure Event Hubs is a Big Data streaming platform and event ingestion service, capable of receiving and processing millions of events per second. Event Hubs can process and store events, data, or telemetry produced by distributed software and devices.



Sentiment Analysis on Streaming Data using Azure Databricks

Create Azure Databricks workspace



Course Navigation

Prerequisites

Section 1

Create Event Hub

Create Azure Databricks workspace

Create a spark cluster in Azure Databricks

Create a Twitter app to access streaming data

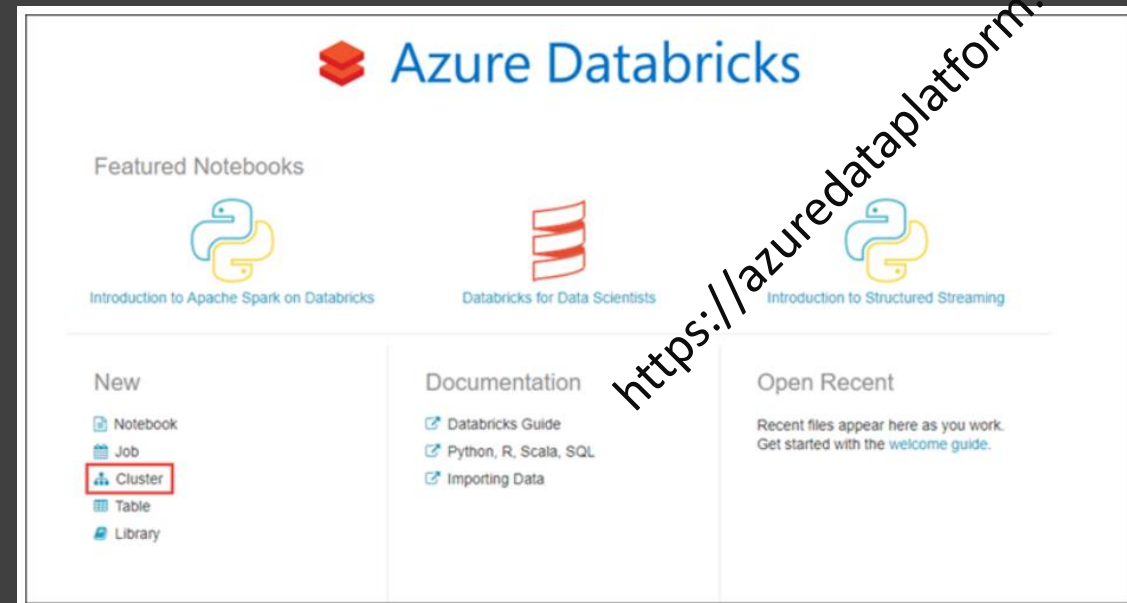
Attach libraries for Event hubs and twitter API

Create a Cognitive Services account and retrieve access key

Demo

Section 2

Azure Databricks is an Apache Spark-based analytics platform optimized for the Microsoft Azure cloud services platform. Designed with the founders of Apache Spark, Databricks is integrated with Azure to provide one-click setup, streamlined workflows, and an interactive workspace that enables collaboration between data scientists, data engineers, and business analysts.



Sentiment Analysis on Streaming Data using Azure Databricks

Create a spark cluster in Azure Databricks



Course Navigation

Prerequisites

Section 1

Create Event Hub

Create Azure Databricks workspace

Create a spark cluster in Azure Databricks

Create a Twitter app to access streaming data

Attach libraries for Event hubs and twitter API

Create a Cognitive Services account and retrieve access key

Demo

Section 2

Apache Spark in Azure Databricks

Azure Databricks builds on the capabilities of Spark by providing a zero-management cloud platform that includes:

- Fully managed Spark clusters
- An interactive workspace for exploration and visualization
- A platform for powering your favorite Spark-based applications

Microsoft Azure

Create Cluster

New Cluster Cancel Create Cluster 3 Workers: 42.0 GB Memory, 12 Cores, 2.25 DBU 1 Driver: 14.0 GB Memory, 4 Cores, 0.75 DBU Cost: \$55 per DBU

Cluster Name
mysparkcluster

Cluster Mode
Standard

Pool
None

Databricks Runtime Version
Runtime: 5.2 (Scala 2.11, Spark 2.4.0)

Python Version
3

Autopilot Options
☐ Enable autoscaling
☒ Terminate after 120 minutes of inactivity

Worker Type
Standard_DS3_v2 14.0 GB Memory, 4 Cores, 0.75 DBU 3 Workers

Driver Type
Same as worker 14.0 GB Memory, 4 Cores, 0.75 DBU

Advanced Options

Sentiment Analysis on Streaming Data using Azure Databricks

Create a Twitter app to access streaming data



Course Navigation

Prerequisites

Section 1

Create Event Hub

Create Azure Databricks workspace

Create a spark cluster in Azure Databricks

Create a Twitter app to access streaming data

Attach libraries for Event hubs and twitter API

Create a Cognitive Services account and retrieve access key

Demo

Section 2

The screenshot shows the Twitter Developer portal. At the top, it says 'Account Application Approved' with a blue checkmark. Below this, a message states: 'Your Twitter developer account application has been approved! Thanks for applying for access. We've completed our review of your application, and are excited to share that your request has been approved.' It then prompts the user to 'Sign in to your developer account to get started.' Below this, there's a section titled 'No apps here. You'll need an app and API key in order to authenticate and integrate with most Twitter developer products. Create an app to get your API key.' The 'Create an app' button is highlighted with a red box and a blue circle with the number 1. The user is then taken to the 'Create a new app' form. The form has several fields: 'App name (required)' with the value 'SparkEventHubsSample', 'Application description (required)' with the value 'Get tweets with certain keywords and filter them!', 'Website URL (required)' with the value 'https://sample.com', 'Callback URL' with the value 'https:// or scheme://', 'Terms of Service URL', 'Privacy policy URL', 'Organization name', and 'Organization website URL'. The 'Create' button is highlighted with a red box and a blue circle with the number 3. After clicking 'Create', the user is taken to the 'Keys and tokens' page. This page shows the 'Consumer API keys' section with a 'Regenerate' button. The 'Access token & access token secret' section is also visible, with a 'Create' button highlighted with a red box and a blue circle with the number 4. A watermark 'https://azuredataplatform.com' is visible across the bottom of the screenshot.

Sentiment Analysis on Streaming Data using Azure Databricks

Attach libraries for Event hubs and twitter API



Course Navigation

Prerequisites

Section 1

Create Event Hub

Create Azure Databricks workspace

Create a spark cluster in Azure Databricks

Create a Twitter app to access streaming data

Attach libraries for Event hubs and twitter API

Create a Cognitive Services account and retrieve access key

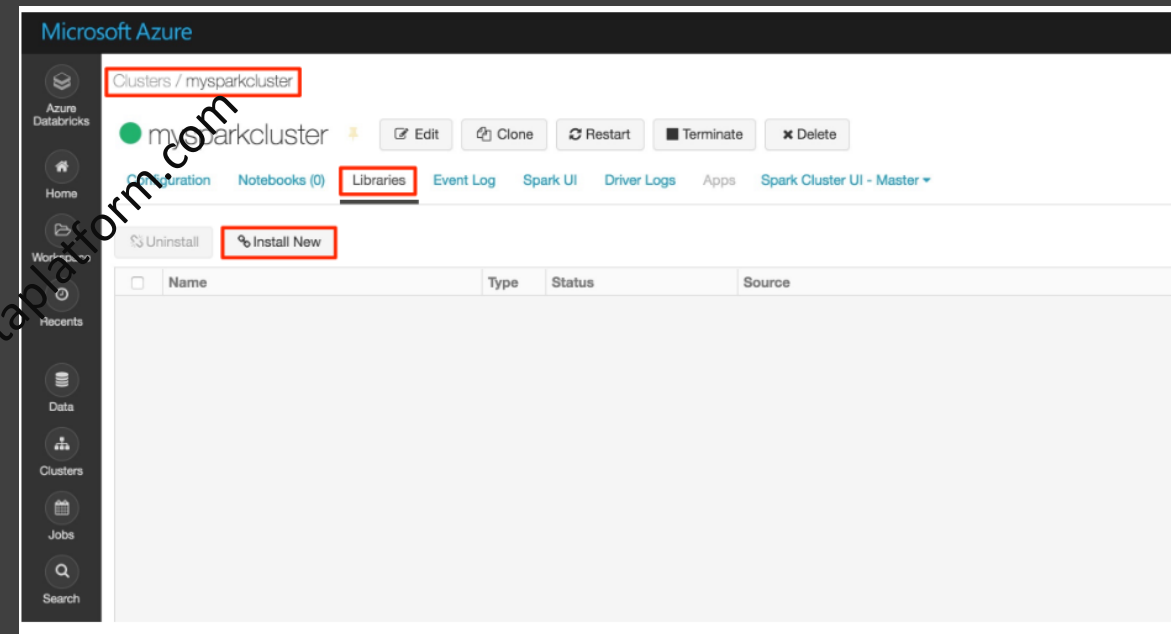
Demo

Section 2

In the New Library page, for Source select Maven. For Coordinate, click Search Packages for the package you want to add. Here is the Maven coordinates for the libraries used in this tutorial:

Spark Event Hubs connector - `com.microsoft.azure:azure-eventhubs-spark_2.11:2.3.10`

Twitter API - `org.twitter4j:twitter4j-core:4.0.7`



Sentiment Analysis on Streaming Data using Azure Databricks

Create a Cognitive Services account and retrieve access key



Course Navigation

Prerequisites

Section 1

Create Event Hub

Create Azure Databricks workspace

Create a spark cluster in Azure Databricks

Create a Twitter app to access streaming data

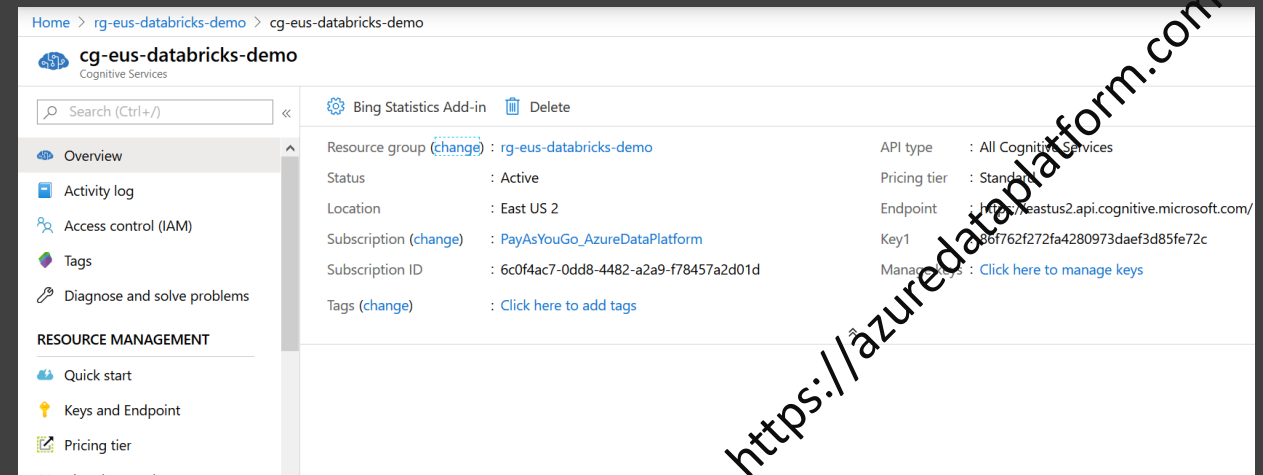
Attach libraries for Event hubs and twitter API

Create a Cognitive Services account and retrieve access key

Demo

Section 2

Azure Cognitive Services are APIs, SDKs, and services available to help developers build intelligent applications without having direct AI or data science skills or knowledge. Azure Cognitive Services enable developers to easily add cognitive features into their applications. The goal of Azure Cognitive Services is to help developers create applications that can see, hear, speak, understand, and even begin to reason. The catalog of services within Azure Cognitive Services can be categorized into five main pillars - Vision, Speech, Language, Web Search, and Decision.



Sentiment Analysis on Streaming Data using Azure Databricks

Architecture



Course Navigation

Prerequisites

Section 1

Demo

Section 2

Architecture

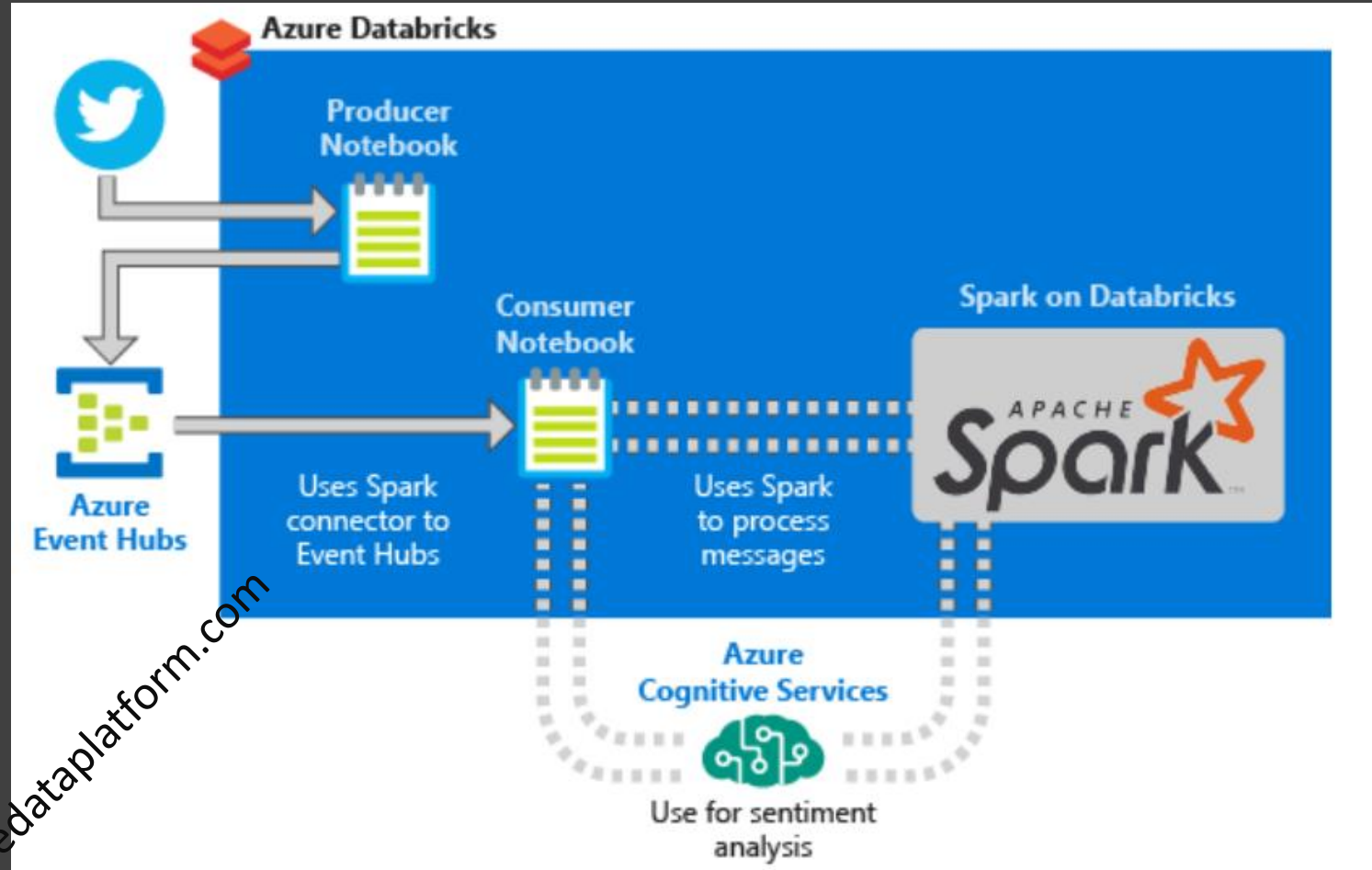
Send Tweets to Event Hubs

Read tweets from Event Hubs

Run Sentiment Analysis on tweets

Connect to Power BI

Create a Power BI dashboard



Sentiment Analysis on Streaming Data using Azure Databricks

Demo



Course Navigation

Prerequisites

Section 1

Demo

Section 2

Architecture

Send Tweets to Event Hubs

Read tweets from Event Hubs

Run Sentiment Analysis on tweets

Connect to Power BI

Create a Power BI dashboard

Demo

<https://azuredatapatform.com>