Connect Jupyter Notebook with CDP Data Lake over AWS Glue

- Working with Notebooks on the AWS Glue Console
  - Amazon SageMaker Notebooks on the AWS Glue Console
  - Zeppelin Notebook Servers on the AWS Glue Console

Working with Notebooks on the AWS Glue Console

A *development endpoint* is an environment that you can use to develop and test your AWS Glue scripts. A *notebook* enables interactive development and testing of your ETL (extract, transform, and load) scripts on a development endpoint.

AWS Glue provides an interface to Amazon SageMaker notebooks and Apache Zeppelin notebook servers. On the AWS Glue notebooks page, you can create Amazon SageMaker notebooks and attach them to a development endpoint. You can also manage Zeppelin notebook servers that you created and attached to a development endpoint. To create a Zeppelin notebook server, see Creating a Notebook Server Hosted on Amazon EC2.

The **Notebooks** page on the AWS Glue console lists all the Amazon SageMaker notebooks and Zeppelin notebook servers in your AWS Glue environment. You can use the console to perform several actions on your notebooks. To display details for a notebook or notebook server, choose the notebook in the list. Notebook details include the information that you defined when you created it using the **Create SageMaker notebook** or **Create Zeppelin Notebook server** wizard.

Amazon SageMaker Notebooks on the AWS Glue Console

The following are some of the properties for Amazon SageMaker notebooks. The console displays some of these properties when you view the details of a notebook.

> Important
>
> AWS Glue only manages Amazon SageMaker notebooks in certain AWS Regions. For more information, see Managing Notebooks.

Ensure you have permissions to manage Amazon SageMaker notebooks in the AWS Glue console. For more information, see **AWSGlueConsole SageMakerNotebookFullAccess** in Step 3: Attach a Policy to IAM Users That Access AWS Glue.

Notebook nameThe unique name of the Amazon SageMaker notebook.Development endpointThe name of the development endpoint that this notebook is attached to.ImportantThis development endpoint must have been created after August 15, 2018.StatusThe provisioning status of the notebook and whether it is Ready, Failed, Starting, Stopping, or Stopped.Failure reasonIf the status is Failed, the reason for the notebook failure. Instance typeThe type of the instance used by the notebook.IAM roleThe IAM role that was used to create the Amazon SageMaker notebook. This role has a trust relationship to Amazon SageMaker. You create this role in the IAM console. When creating the role, choose Amazon SageMaker, and attach a policy for the notebook, such as AWSGlueServiceSageMakerNotebookRoleDefault. For more information, see Step 7: Create an IAM Role for Amazon SageMaker Notebooks.

Zeppelin Notebook Servers on the AWS Glue Console

The following are some of the properties for Zeppelin notebook servers. The console displays some of these properties when you view the details of a notebook.

Notebook server nameThe unique name of the Zeppelin notebook server.Development endpointThe unique name that you give the endpoint when you create it.Provisioning statusDescribes whether the notebook server is CREATE_COMPLETE or ROLLBACK_COMPLETE.Failure reasonIf the status is Failed, the reason for the notebook failure.CloudFormation stackThe name of the AWS CloudFormation stack that was used to create the notebook server.EC2 instanceThe name of Amazon EC2 instance that is created to host your notebook. This links to the Amazon EC2 console (https://console.aws.amazon.com/ec2/) where the instance is tagged with the key aws-glue-dev-endpoint and value of the name of the development endpoint.SSH to EC2 server commandEnter this command in a terminal window to connect to the Amazon EC2 instance that is running your notebook server. The Amazon EC2 address shown in this command is either public or private, depending on whether you chose to Attach a public IP to the notebook server EC2 instance.Copy certificateExample scp command to copy the keystore that is required to set up the Zeppelin notebook server to the Amazon EC2 instance that hosts the notebook server. Run the command from a terminal window in the directory where the Amazon EC2 private key is located. The key to access the Amazon EC2 instance is the parameter to the -i option. You provide the path-to-keystore-file. The remaining part of the command is the location where the development endpoint private SSH key on the Amazon EC2 server is located.HTTPS URLAfter setting up a notebook server, enter this URL in a browser to connect to your notebook using HTTPS.