

データブリックス ワークショップの概要

2021/6/29

データブリックスのご説明、基本的な使い方

目的：

データブリックスのデータサイエンスワークスペース上でハンズオンを実施いただき、基本操作を体験していただく。

対象：

データ分析に携わる方

アジェンダ：

データブリックスの概要説明の後、データブリックスの基本操作を体験するハンズオンを実施いたします。

- ワークスペースの基礎
- データブリックスのいいところ

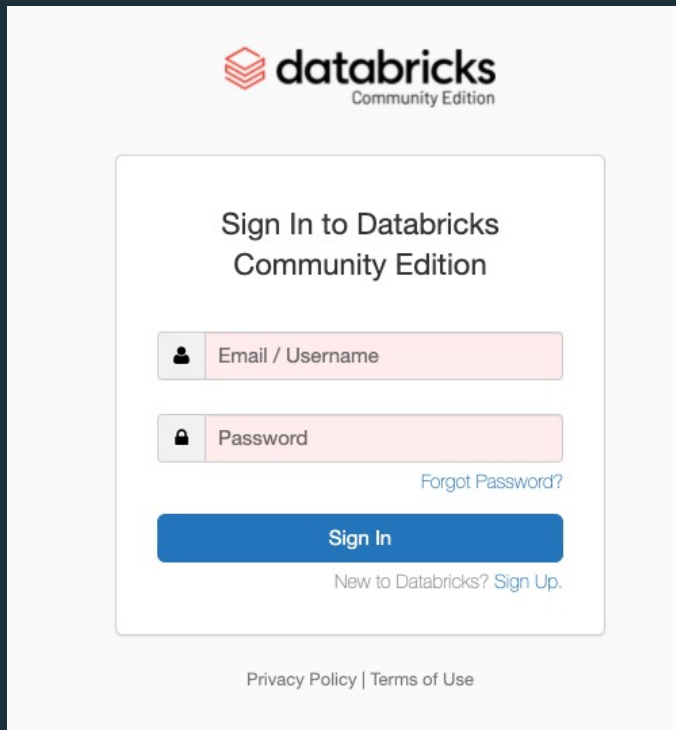
事前準備のご案内

ハンズオン参加にあたり、以下のご準備をお願いします。

- **Databricks Community Editionのユーザー登録およびログインの確認**
 - 登録手順は[Community Editionで始めるDatabricks](#)を参照ください。
- **Slackチャンネル(データブリックス/ハンズオン)への参加**
 - 当日はSlackチャンネルで質問、進捗状況の確認をさせていただきます。
 - こちらの[URL](#)をクリック頂き、チャンネルへの参加手続きをお願いします。
- **ハンズオンノートブックのインポート**
 - 当日使用するノートブックを[こちら](#)のリンクで確認ください。

ワークショップの準備 – ログイン


- <https://community.cloud.databricks.com/> にアクセスします。
- 事前準備で設定したメールアドレスとパスワードを指定してSign Inをクリックします。



The screenshot shows the login interface for Databricks Community Edition. At the top, the Databricks logo and 'Community Edition' text are displayed. The main heading is 'Sign In to Databricks Community Edition'. Below this, there are two input fields: 'Email / Username' and 'Password'. A 'Forgot Password?' link is positioned to the right of the password field. A blue 'Sign In' button is centered below the fields. At the bottom of the form area, there is a link for 'New to Databricks? Sign Up.'. At the very bottom of the page, there are links for 'Privacy Policy' and 'Terms of Use'.

ワークショップの準備 – ログイン

- ワークスペースにログインできたことを確認します。

Welcome to  databricks

Explore the Quickstart Tutorial
Spin up a cluster, run queries on preloaded data, and display results in 5 minutes.

Import & Explore Data
Quickly import data, preview its schema, create a table, and query it in a notebook.

Create a Blank Notebook
Create a notebook to start querying, visualizing, and modeling your data.

Common Tasks

- New Notebook
- Create Table
- New Cluster
- New Job
- New MLflow Experiment
- Import Library
- Read Documentation

Recents

- データブリックスを使ってみよう
- Databricksでデータ分析のAtoZをご紹介
- 1-はじめてのDatabricksノートブック

What's new in v3.48

- Databricks Status
- View latest release notes

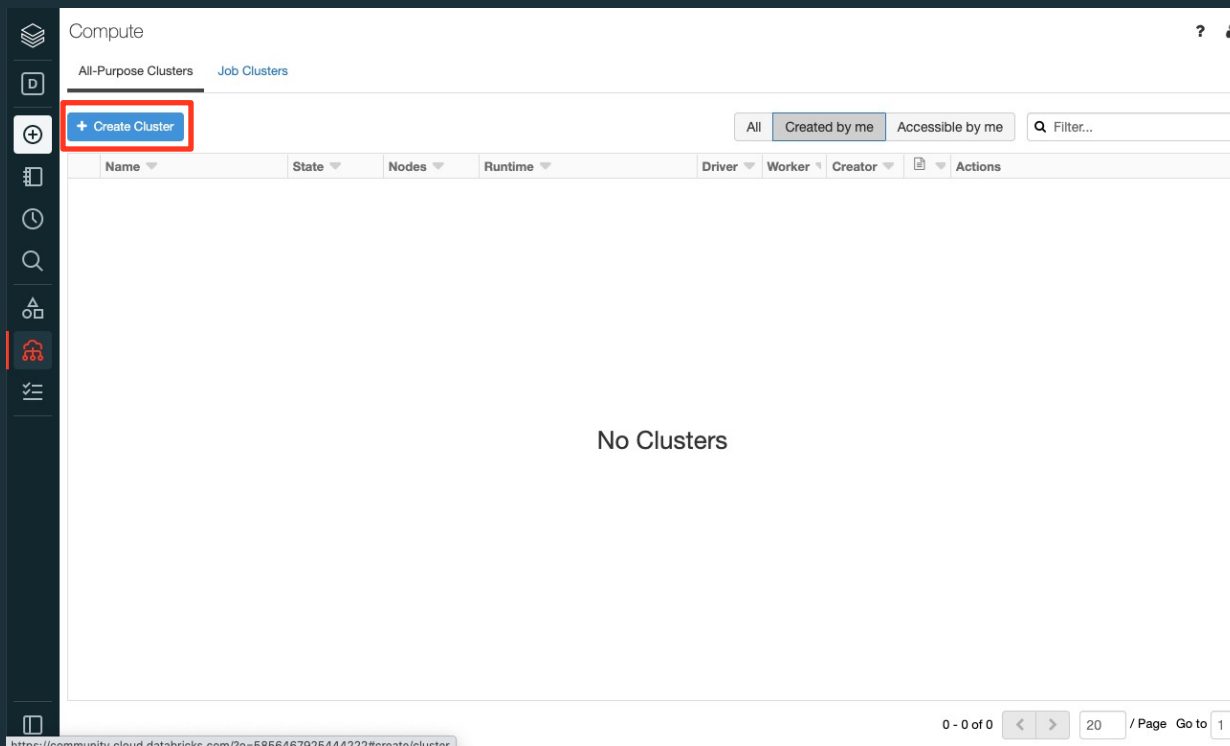
ワークショップの準備 – クラスターの作成

- サイドバーにマウスカーソルを移動し展開します。
- 「Compute」をクリックします。

The screenshot displays the Databricks workspace interface. On the left, a dark sidebar contains navigation options: 'Data Science & E...', 'Create', 'Workspace', 'Recents', 'Search', 'Data', 'Compute', and 'Jobs'. The 'Compute' option is highlighted with a red rectangular box. The main workspace area features a 'Welcome to databricks' header and three primary action cards: 'Follow the Quickstart Tutorial', 'Import & Explore Data', and 'Create a Blank Notebook'. Below these cards, there are sections for 'Recents' and 'What's new in v3.48'. The 'Recents' section lists recent documents, including 'データブリックスを使ってみよう', 'Databricksでデータ分析のAtoZをご紹介', and '1-はじめてのDatabricksノートブック'. The 'What's new in v3.48' section includes links for 'Databricks Status' and 'View latest release notes'. At the bottom of the sidebar, there is a 'Menu options' section with a link to 'Documentation'.

ワークショップの準備 – クラスターの作成

- 「+ Create Cluster」をクリックします。



The screenshot displays the Databricks Compute interface. At the top, the title 'Compute' is visible. Below it, there are two tabs: 'All-Purpose Clusters' (selected) and 'Job Clusters'. A prominent blue button labeled '+ Create Cluster' is highlighted with a red rectangular box. To the right of this button, there are filter buttons for 'All', 'Created by me', and 'Accessible by me', along with a search bar labeled 'Filter...'. Below the filters is a table header with columns: Name, State, Nodes, Runtime, Driver, Worker, Creator, and Actions. The main content area of the table is empty and contains the text 'No Clusters'. At the bottom right, there is a pagination control showing '0 - 0 of 0' and 'Page Go to 1'. The URL at the bottom of the browser window is 'https://community.cloud.databricks.com/?o=5856487925444222/#create/cluster'.

ワークショップの準備 – クラスターの作成

- Cluster Nameにクラスター名を入力します。
- 「Create Cluster」をクリックします。

Create Cluster

New Cluster Cancel Create Cluster DBU / hour: 1 ? 0 Workers: 0.0 GB Memory, 0 Cores
1 Driver: 15.3 GB Memory, 2 Cores

Cluster Name

takaaki

Databricks Runtime Version ?

Runtime: 8.2 (Scala 2.12, Spark 3.1.1) | v

Note Databricks Runtime 8.x uses Delta Lake as the default table format. [Learn more](#)

Instance

Free 15GB Memory: As a Community Edition user, your cluster will automatically terminate after an idle period of two hours. For [more configuration options](#), please [upgrade your Databricks subscription](#).

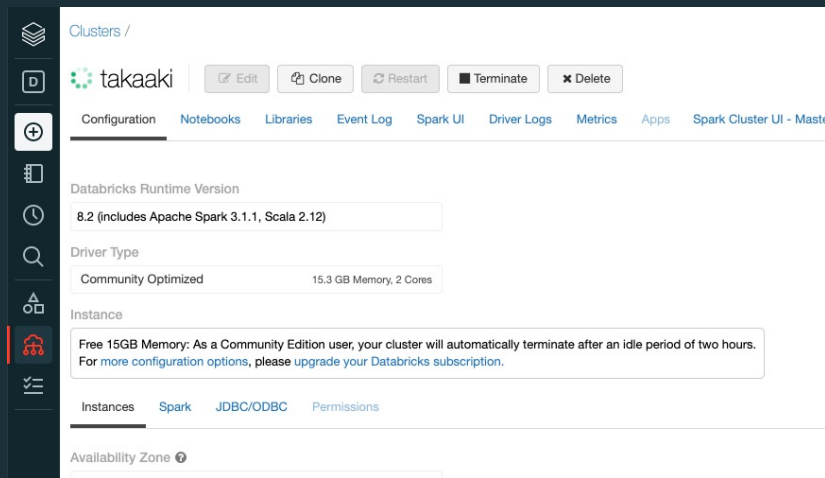
Instances Spark

Availability Zone ?

auto | v

ワークショップの準備 – クラスターの作成

- ・ クラスター名の左のインジケータが●(起動状態)になるまで待ちます。



Clusters /

takaaki [Edit] [Clone] [Restart] [Terminate] [Delete]

Configuration Notebooks Libraries Event Log Spark UI Driver Logs Metrics Apps Spark Cluster UI - Master

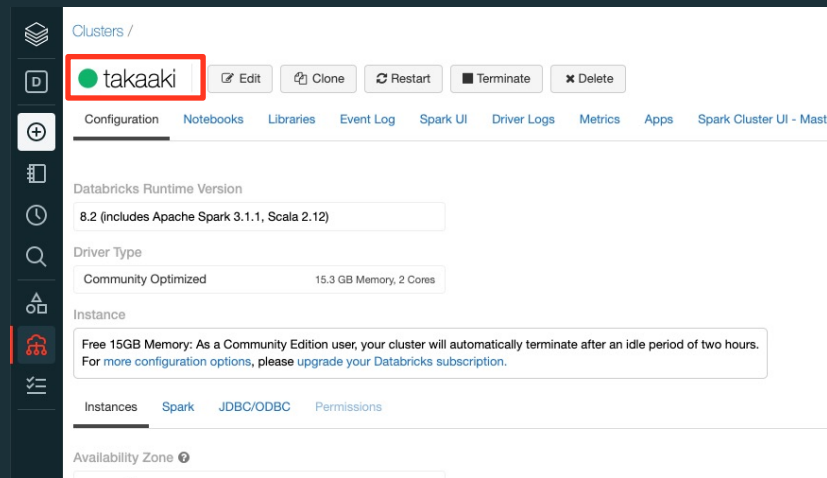
Databricks Runtime Version
8.2 (includes Apache Spark 3.1.1, Scala 2.12)

Driver Type
Community Optimized 15.3 GB Memory, 2 Cores

Instance
Free 15GB Memory: As a Community Edition user, your cluster will automatically terminate after an idle period of two hours. For more configuration options, please upgrade your Databricks subscription.

Instances Spark JDBC/ODBC Permissions

Availability Zone



Clusters /

takaaki [Edit] [Clone] [Restart] [Terminate] [Delete]

Configuration Notebooks Libraries Event Log Spark UI Driver Logs Metrics Apps Spark Cluster UI - Master

Databricks Runtime Version
8.2 (includes Apache Spark 3.1.1, Scala 2.12)

Driver Type
Community Optimized 15.3 GB Memory, 2 Cores

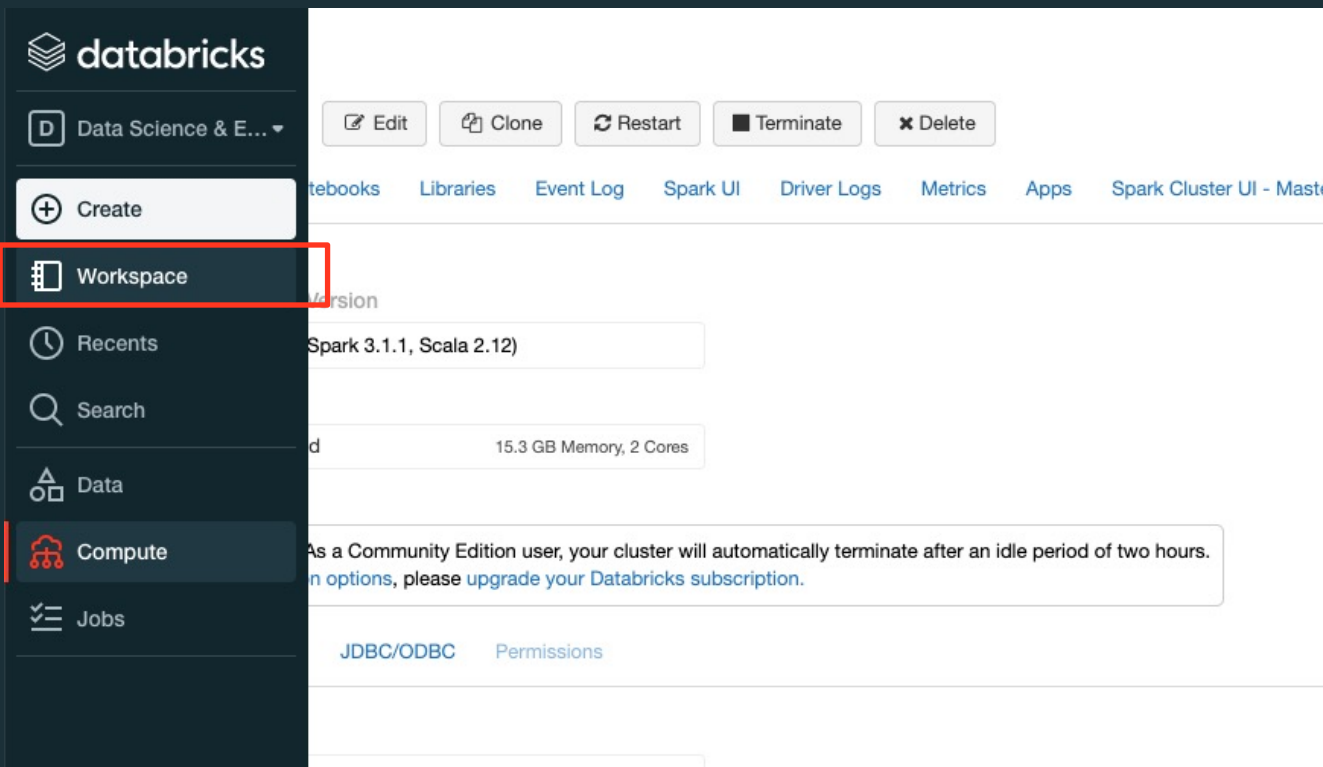
Instance
Free 15GB Memory: As a Community Edition user, your cluster will automatically terminate after an idle period of two hours. For more configuration options, please upgrade your Databricks subscription.

Instances Spark JDBC/ODBC Permissions

Availability Zone

ワークショップの準備 – ノートブックのインポート

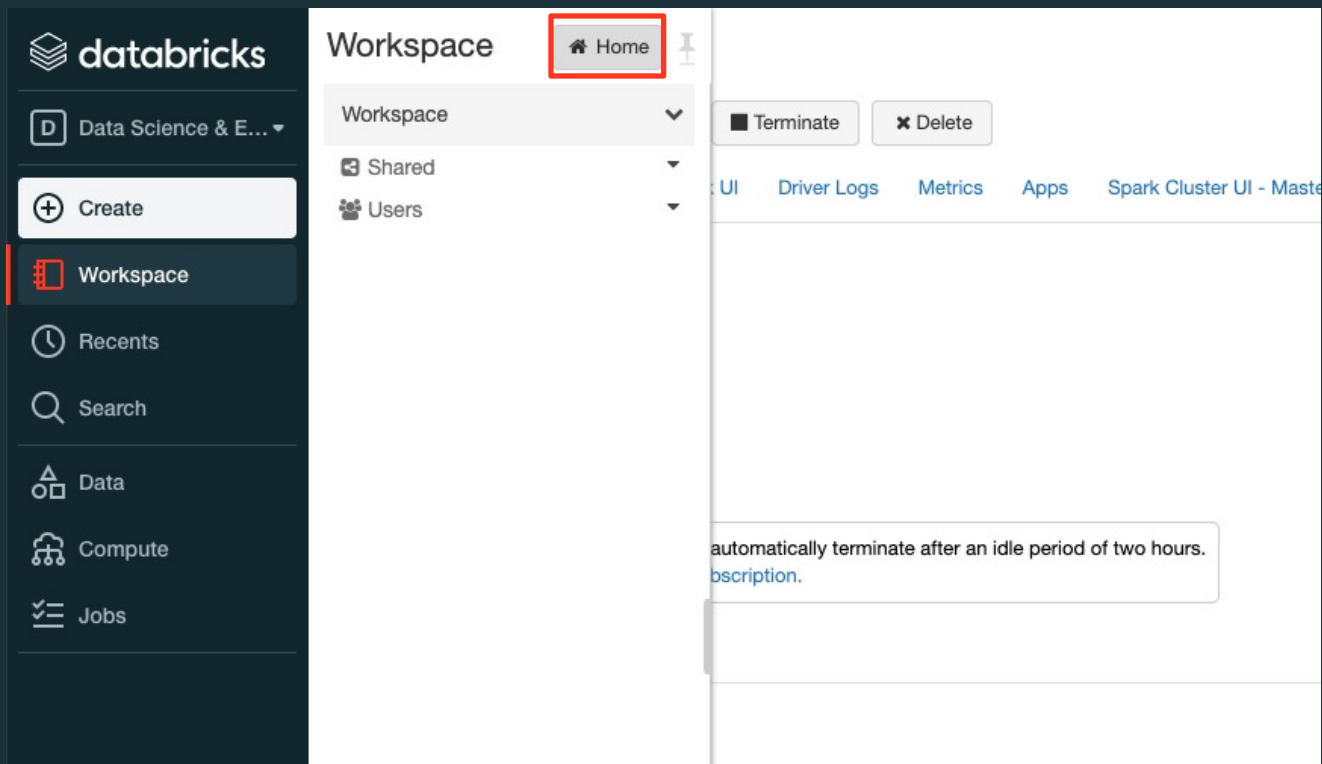
- サイドバーのWorkspaceをクリックします。ワークスペースのホームディレクトリ一覧が表示されます。



The screenshot displays the Databricks user interface. On the left is a dark sidebar with the 'databricks' logo at the top. Below the logo are several menu items: 'Data Science & E...', 'Create', 'Workspace', 'Recents', 'Search', 'Data', 'Compute', and 'Jobs'. The 'Workspace' menu item is highlighted with a red rectangular box. To the right of the sidebar, the main content area shows a workspace view with buttons for 'Edit', 'Clone', 'Restart', 'Terminate', and 'Delete'. Below these buttons are navigation links for 'Notebooks', 'Libraries', 'Event Log', 'Spark UI', 'Driver Logs', 'Metrics', 'Apps', and 'Spark Cluster UI - Master'. The main area also displays technical details for a workspace, including 'Spark 3.1.1, Scala 2.12' and '15.3 GB Memory, 2 Cores'. A warning message at the bottom states: 'As a Community Edition user, your cluster will automatically terminate after an idle period of two hours. For more options, please upgrade your Databricks subscription.' At the bottom of the main area, there are links for 'JDBC/ODBC' and 'Permissions'.

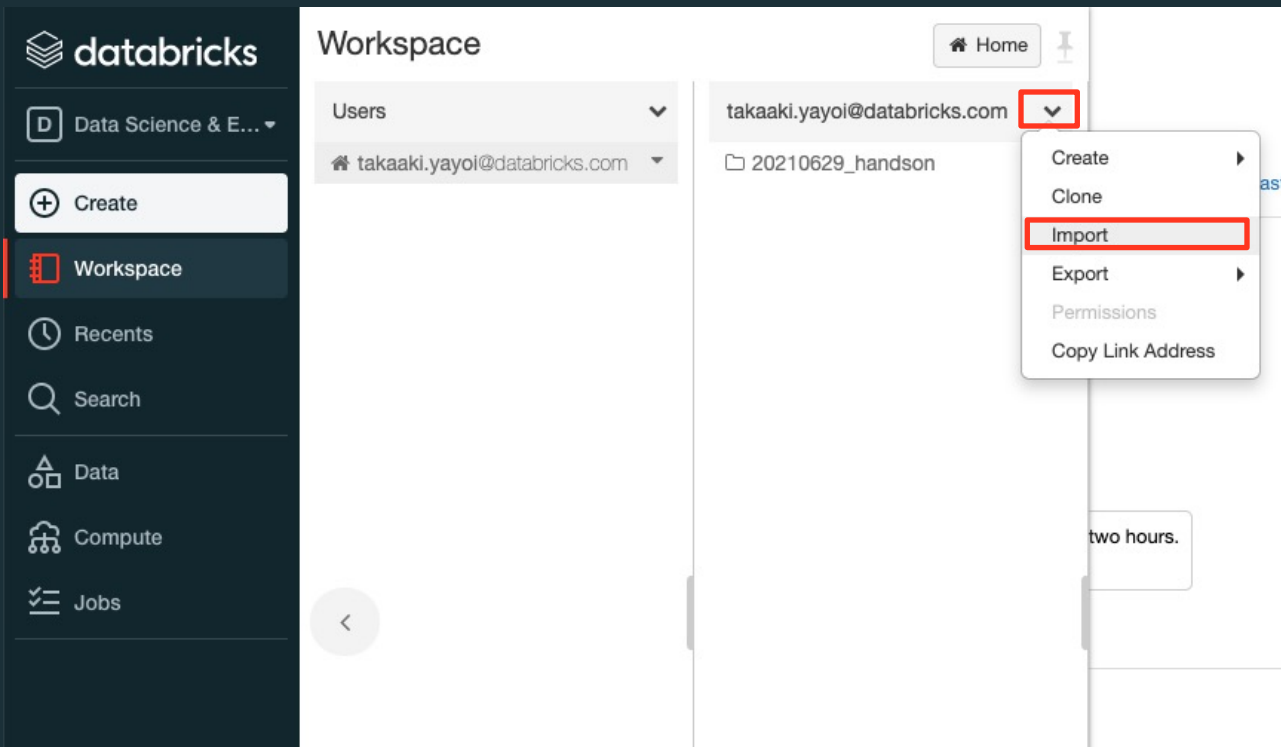
ワークショップの準備 – ノートブックのインポート

- Homeボタンをクリックして自分のホームディレクトリに移動します。



ワークショップの準備 – ノートブックのインポート

- 自分のメールアドレスの右にある下向き矢印をクリックしてメニューを展開します。
- Importを選択します。



ワークショップの準備 – ノートブックのインポート

- 自分のメールアドレスの右にある下向き矢印をクリックしてメニューを展開します。
- Importを選択するとダイアログが表示されますので「URL」を選択します。

Import Notebooks

Import from: File URL

|

Accepted formats: .dbc, .scala, .py, .sql, .r, .ipynb, .Rmd, .html

(To import a library, such as a jar or egg, [click here](#))

Cancel Import

ワークショップの準備 – ノートブックのインポート

- ブラウザで別のタブ、ウィンドウを開きます。
- こちらの[URL](https://sajpstorage.blob.core.windows.net/handson20210629/%E3%83%86%E3%82%99%E3%83%BC%E3%82%BF%E3%83%95%E3%82%99%E3%83%AA%E3%83%83%E3%82%AF%E3%82%B9%E3%82%92%E4%BD%BF%E3%81%A3%E3%81%A6%E3%81%BF%E3%82%88%E3%81%86.html)をクリックするか、以下のURLをアドレスバーに貼り付けます。
- <https://sajpstorage.blob.core.windows.net/handson20210629/%E3%83%86%E3%82%99%E3%83%BC%E3%82%BF%E3%83%95%E3%82%99%E3%83%AA%E3%83%83%E3%82%AF%E3%82%B9%E3%82%92%E4%BD%BF%E3%81%A3%E3%81%A6%E3%81%BF%E3%82%88%E3%81%86.html>



ワークショップの準備 – ノートブックのインポート

- ノートブックが表示されるので、右上の「Import Notebook」をクリックします。



The screenshot shows the Databricks web interface. At the top left is the Databricks logo. In the center, the text reads 'データブリックスを使ってみよう (Python)'. On the top right, there is a button labeled 'Import Notebook' which is highlighted with a red rectangular border. Below the header, the main content area features the heading 'データブリックスを使ってみよう' and a sub-heading '準備中です。当日までに最新版に差し替えます。'

ワークショップの準備 – ノートブックのインポート

- 表示されるURLをコピーします。

Import Notebook

You can edit and run this notebook by importing it into your Databricks account. Select **Import** from any folder's menu and paste the URL below.

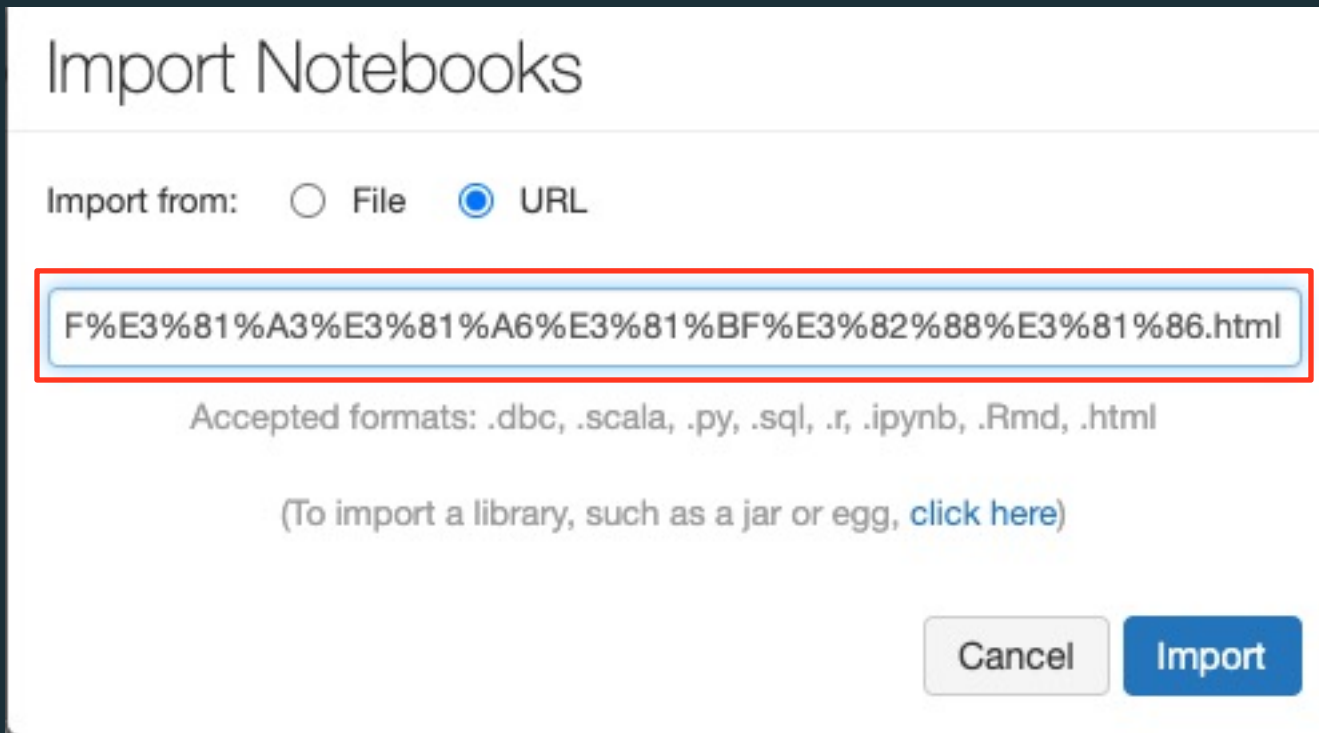
`https://sajpstorage.blob.core.windows.net/handson20210629/%E3%83%86%`

New to Databricks? [Try it now.](#)

Done

ワークショップの準備 – ノートブックのインポート

- ワークスペースに戻り、コピーしたURLを貼り付けます。
- Importをクリックします。



Import Notebooks

Import from: File URL

`F%E3%81%A3%E3%81%A6%E3%81%BF%E3%82%88%E3%81%86.html`

Accepted formats: .dbc, .scala, .py, .sql, .r, .ipynb, .Rmd, .html

(To import a library, such as a jar or egg, [click here](#))

Cancel Import

ワークショップの準備 – ノートブックのインポート

- ワークスペースにノートブックがインポートされます。
- 公開されている他のデータブリックスノートブックについても同様の手順でインポートすることができます。



データブリックスを使ってみよう (Python)

Detached

データブリックスを使っ... Cmd 1

END

データブリックスを使ってみよう

準備中です。当日までに最新版に差し替えます。

作者	Databricks Japan
日付	2021/06/03
バージョン	1.4
クラスター	8.3ML

 databricks

ワークショップの準備 – クラスターへのアタッチ

- データブリックスでノートブックを実行するには、先ほど作成したクラスターにノートブックを関連づける(アタッチ)する必要があります。
- 左上のドロップダウンを展開して、先ほど作成したクラスターを選択します。



データブリックスを使ってみよう (Python)

Detached

Attach:

- takaaki
15.25 GB | 2 Cores | DBR 8.2 | Spark 3.1.1 | Scala 2.12

使ってみよう

準備中です。当日までに最新版に差し替えます。

作者	Databricks Japan
日付	2021/06/03
バージョン	1.4
クラスター	8.3ML

 databricks

ワークショップの準備 – クラスターへのアタッチ

- これで準備は完了です！



データブリックスを使ってみよう (Python)

takaaki

データブリックスを使っ... Child 1

END

データブリックスを使ってみよう

準備中です。当日までに最新版に差し替えます。

作者	Databricks Japan
日付	2021/06/03
バージョン	1.4
クラスター	8.3ML

 databricks