



For this Azure Databricks Unified Analytics Workshop, please download the following files so you can follow along with your Databricks Solution Architect with these notebooks. You can run these notebooks on your own after the workshop by using Databricks.

If you do not have a Databricks account, you can try Azure Databricks by going to <https://docs.azuredatabricks.net/getting-started/try-databricks.html>.

To following along, you will need to download the following files:

- a. Data Engineering: [Data Preparation with Delta Lake](#)
- b. Data Sciences: [Evaluate Risk on Loan Approvals](#)

If you would like to dive further, you can also download the [Evaluate Risk on Loan Approvals \(XGBoost 0.81\)](#) and try using XGBoost (included as part of [Databricks Runtime for Machine Learning](#)).

For more information on the Evaluate Risk on Loan Approvals scenario, refer to the blog post [Loan Risk Analysis with XGBoost and Databricks Runtime for Machine Learning](#).

Refer to the following pages for additional details

Import Data	2
Data Engineering	3
Data Sciences	4

Import Data

To download the data, please uncomment the earlier cell in the notebook which will download the files using `wget` into the Databricks File System `tmp` folder (i.e. `/dbfs/tmp`) as noted in the following screenshot.

```
1 %sh wget -P /dbfs/tmp/azure-workshop/ https://sscsa.blob.core.windows.net/azure-databricks-workshop/loanstats_2012_2017_pq.parquet

wget: /databricks/python/lib/libcrypto.so.1.0.0: no version information available (required by wget)
wget: /databricks/python/lib/libssl.so.1.0.0: no version information available (required by wget)
wget: /databricks/python/lib/libssl.so.1.0.0: no version information available (required by wget)
--2019-06-05 16:23:45-- https://sscsa.blob.core.windows.net/azure-databricks-workshop/loanstats_2012_2017_pq.parquet
Resolving sscsa.blob.core.windows.net (sscsa.blob.core.windows.net)... 52.239.152.74
Connecting to sscsa.blob.core.windows.net (sscsa.blob.core.windows.net)|52.239.152.74|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 162431066 (155M) [application/octet-stream]
Saving to: '/dbfs/tmp/azure-workshop/loanstats_2012_2017_pq.parquet'

 0K ..... 0% 5.87M 26s
 50K ..... 0% 5.72M 27s
100K ..... 0% 5.18M 28s
150K ..... 0% 4.60M 29s
200K ..... 0% 5.78M 29s
250K ..... 0% 5.84M 28s
300K ..... 0% 5.60M 28s
350K ..... 0% 4.99M 29s
400K ..... 0% 5.71M 28s
450K ..... 0% 5.85M 28s
500K ..... 0% 5.47M 28s
550K ..... 0% 4.85M 28s
600K ..... 0% 5.71M 28s

Command took 24.90 seconds -- by denny.lee@databricks.com at 6/5/2019, 9:23:43 AM on rocinante
```

Once downloaded, you can see the file via the following command.

```
1 %fs ls /tmp/azure-workshop/
```

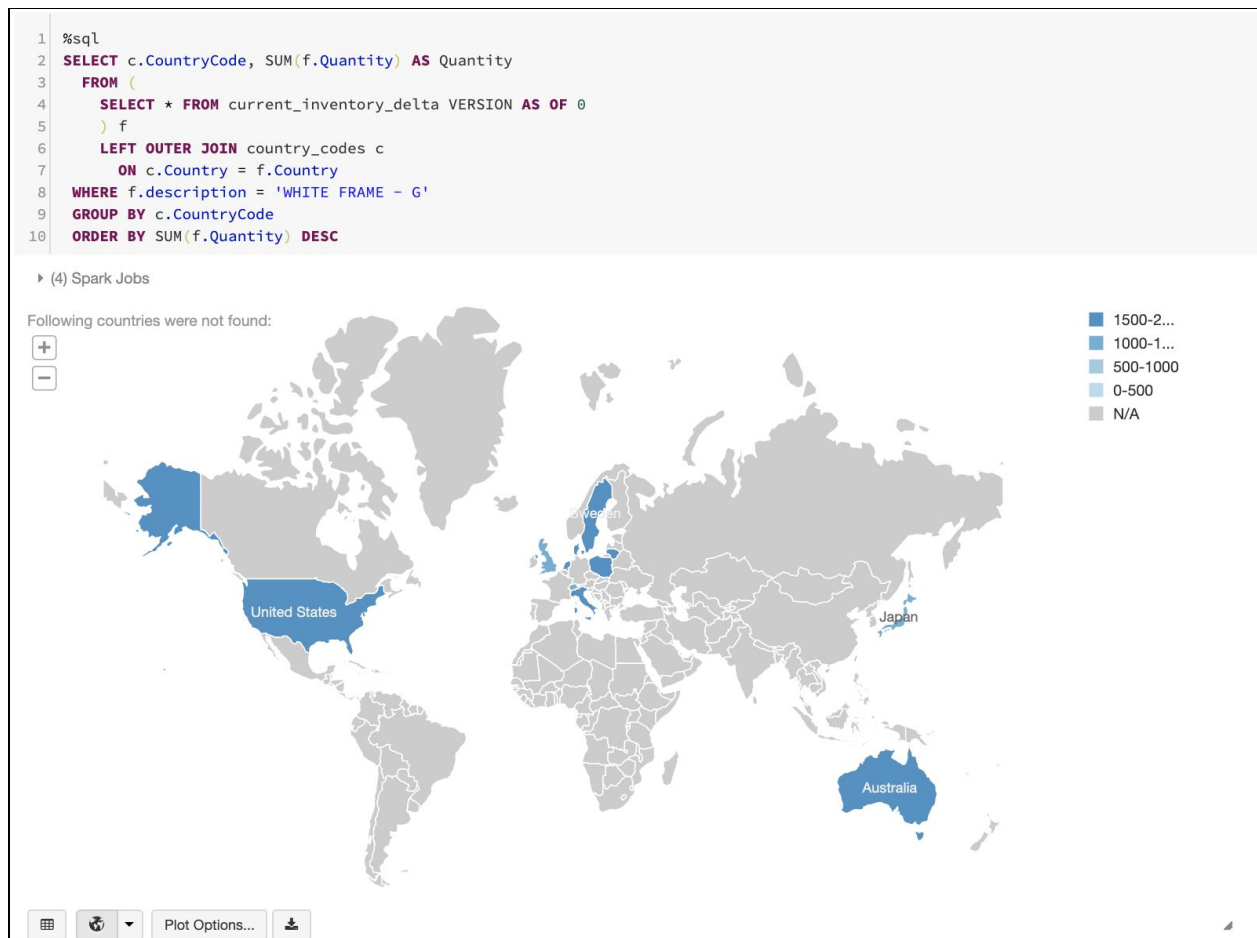
path
dbfs:/tmp/azure-workshop/loanstats_2012_2017_pq.parquet



Data Engineering

This demo showcases batch/streaming sync, ACID transactions, and time travel with Delta Lake. Download the notebook: [Data Preparation with Delta Lake](#),

As well, click on the image below to watch a < 2 min video showcasing this scenario. You can access the video directly at <https://dbricks.co/dl-db>.



Data Sciences

This demo showcases running Machine Learning with visualizations within a Databricks workspace. Download the following notebooks and resources: [Evaluate Risk on Loan Approvals](#).

If you would like to dive further, you can also download the [Evaluate Risk on Loan Approvals \(XGBoost 0.81\)](#) and try using XGBoost (included as part of [Databricks Runtime for Machine Learning](#)).

As well, click on the image below to watch a < 2min video showing this scenario. You can access the video directly at <https://dbricks.co/lra-video>

The screenshot displays a Databricks workspace interface. On the left is a sidebar with navigation icons for Home, Workspace, Recents, Data, Clusters, Jobs, and Search. The main area shows a notebook titled "Loan Risk Analysis (Python)". The notebook content includes a section "Evaluating Risk for Loan Approvals" with a subsection "Business Value". A yellow callout box with green text says "Let's start with the Databricks workspace for our loan risk analysis". Below the text is a flowchart showing the process: Data Exploration -> Machine Learning Models -> Evaluate Results -> Score New Loans. The video player interface at the bottom shows the title "The Data" and a progress bar at 0:13 / 1:18.