



Spark Summit Europe 2017

Deep Learning, Keras, TensorFlow, + Spark

- Intro + Artificial Neural Networks
- TensorFlow Basics
- Convolutional Networks
- Recurrent Nets, LSTM
- Intro to Reinforcement Learning
- Building Deep Learning Models with Spark
- Using Spark for Inference (Prediction)
- Operations in the Real World

Instructor: Adam Breindel

LinkedIn: <https://www.linkedin.com/in/adbreind>

Email: adbreind@gmail.com



- ~20 years building systems for startups and large enterprises
- 10 years teaching front- and back-end technology
- Fun big data projects...
 - Streaming neural net + decision tree fraud scoring (debit cards)
 - Realtime & offline analytics for banking
 - Music synchronization and licensing for networked jukeboxes
- Industries
 - Finance / Insurance
 - Travel
 - Media / Entertainment

Today's Schedule

10:30 a.m. – 11 a.m. Coffee Break

12 p.m. – 1 p.m. Lunch

3 p.m. – 3:30 p.m. Afternoon Break

5 p.m. End of Class

Today's Class – Informal Survey

- Today is my first day trying machine learning
- I've done some tutorials, know the ideas, built a few toy models
- I've been working with machine learning for < 1 year in my job
- I'm a data scientist and I've been working with predictive analytics and machine learning for years

Files and Resources

Documents

- These slides: <http://tinyurl.com/dublin-start>

Databricks

- You should have a working Databricks account
 - If not, sign up for **free Community Edition** now at <http://tinyurl.com/databricks-ce>
 - Use a laptop with **Firefox** or **Chrome** (Internet Explorer / MS Edge not supported)

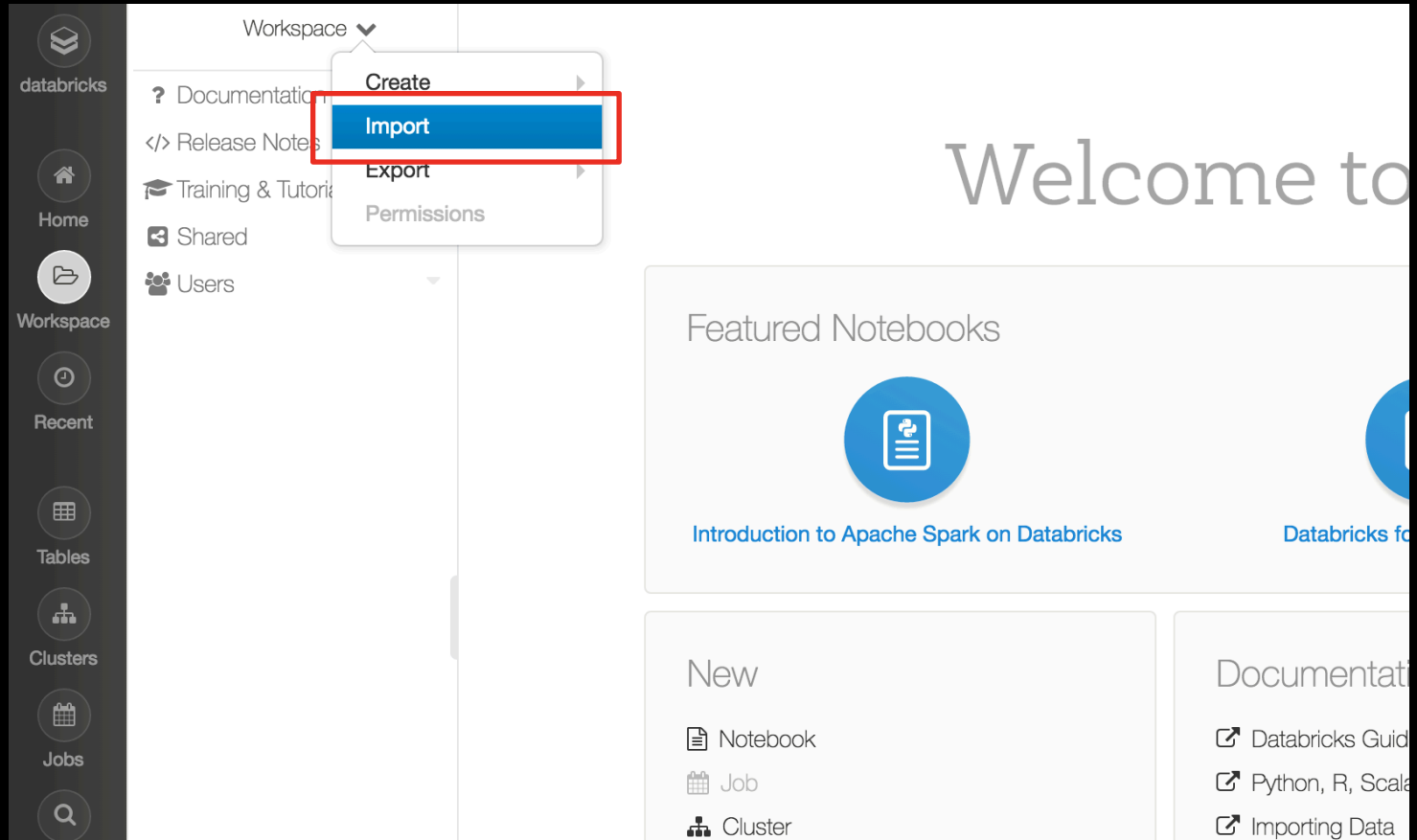
Getting Started

- Import
<http://files.training.databricks.com/events/summit-2017-10/deep-learning/Labs.dbc>
directly into your Databricks account per the instructions (and live demo) on the following slides.
- Create a cluster: choose **Databricks Runtime 3.2**

Log in to Databricks

The screenshot shows the Databricks Community Edition (2.54) interface. The browser address bar displays the URL `https://community.databricks.com/?o=2347565288706789`. The sidebar on the left contains navigation icons for Home, Workspace, Recent, Data, Clusters, Jobs, and Search. The main content area is titled "Welcome to databricks™" and includes a "Featured Notebooks" section with three featured notebooks: "Introduction to Apache Spark on Databricks", "Databricks for Data Scientists", and "Introduction to Structured Streaming". Below this, there are three columns: "New" (containing Notebook, Job, Cluster, Table, and Library), "Documentation" (containing Databricks Guide, Python, R, Scala, SQL, and Importing Data), and "What's new?" (containing R CRAN library management, Bitbucket Cloud (Private Preview), and Historical snapshots of cluster metrics). A "Latest release notes" link is also present. At the bottom right, there is a "Send Feedback" button.

Import Notebook for Today...



Import Notebook for Today...

Import Notebooks

Import from: ☐ File ☒ URL

<http://files.training.databricks.com/events/summit-2017-10/deep-learning/Labs.dbc>

Accepted formats: .dbc, .scala, .py, .sql, .r, .ipynb, .html
(To import a library, such as a jar or egg, [click here](#))

Cancel

Import

Introduction to Apache Spark on Databricks

Databricks for Data Scientists

Create a Cluster

1 →

2 →

data
Clusters

data
+ Create Cluster

▼ Interactive Clusters

Name	State	Nodes	Driver	Worker	Runtime	Created
● test2	Terminated ?	-	Community ...	Community ...	3.2 (includes ...	adam
● test (clone)	Terminated ?	-	Community ...	Community ...	3.2 (includes ...	adam

▼ Job Clusters

No clusters found

data
Jobs

data
data

Create a Cluster

The screenshot shows the Databricks 'Create Cluster' page. On the left sidebar, the 'Clusters' icon is highlighted with a red box and the number 3. Below it, the 'Workspace' icon is highlighted with a red box and the number 4. In the main content area, the 'Create Cluster' button is highlighted with a red box and the number 5. The 'Cluster Name' field contains 'demo'. The 'Databricks Runtime Version' dropdown is set to '3.2 (includes Apache Spark 2.2.0, Scala 2.11)'. The 'Instance' section shows a warning message about memory and idle time. The 'Availability Zone' dropdown is set to 'us-west-2c'.

3

4

5

Create Cluster

New Cluster

Cancel

Create Cluster

Cluster Name

demo

Databricks Runtime Version ?

3.2 (includes Apache Spark 2.2.0, Scala 2.11)

Instance

Free 6GB 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 ?

us-west-2c

0 Workers: 0.0 GB Memory, 0 Cores, 0 DBU

1 Driver: 6.0 GB Memory, 0.88 Cores, 1 DBU ?

databricks

Home

Workspace

Recent

Data

Clusters

Jobs

databricks

Install TensorFlow Library

The screenshot shows the Databricks web interface. On the left sidebar, the 'Workspace' icon (a folder) is highlighted with a red box and a blue arrow labeled '1'. In the main area, the 'Workspace' tab is active, showing a 'Connecting...' status. A 'Create' dropdown menu is open, and the 'Library' option is highlighted with a red box and a blue arrow labeled '2'. Below the 'Library' option, the 'Folder' option is visible. The 'Featured Notebooks' section at the bottom shows a Python logo and the title 'Introduction to Apache Spark on Databricks'.

1 →

2 →

Workspace

Connecting...

Create

Library

Folder

Featured Notebooks

Introduction to Apache Spark on Databricks

Install TensorFlow

The screenshot shows the 'Create Library' interface in Databricks. A sidebar on the left contains navigation icons and labels: 'databricks', 'Home', 'Workspace', 'Recent', and 'Clusters'. The main content area is titled 'New Library' and includes the following elements:

- Language:** A dropdown menu with the option 'Upload Python Egg or PyPI' selected. A red box highlights this dropdown, and a blue arrow labeled '3' points to it.
- Install PyPi Package:** A section with the text 'You can specify a package name with an optional [version specification](#)'. Below this is a text input field for the 'PyPi Name' containing the text 'tensorflow==1.3.0'. A red box highlights this input field, and a blue arrow labeled '4' points to it.
- Install Library:** A large grey button at the bottom of the form. A purple box highlights this button, and a blue arrow labeled '5' points to it.

Install Keras Library

The screenshot shows the Databricks web interface. On the left sidebar, the 'Workspace' icon (a folder) is highlighted with a red box and a blue arrow labeled '1'. In the main area, the 'Workspace' tab is active, and a context menu is open over a folder. The 'Create' option in the menu is highlighted with a blue bar, and its sub-menu is displayed. In the sub-menu, the 'Library' option is highlighted with a red box and a blue arrow labeled '2'. The background shows the 'Workspace' section with links to Documentation, Release Notes, Training & Tutorials, Shared, and Users. Below this, there is a 'Connecting...' status bar and a 'Featured Notebooks' section with a Python logo and the title 'Introduction to Apache Spark on Databricks'.

1 →

2 →

Install Keras Library

The screenshot shows the 'Create Library' interface in Databricks. A sidebar on the left contains navigation icons and labels: 'databricks', 'Home', 'Workspace', 'Recent', and 'Clusters'. The main content area is titled 'Create Library' and 'New Library'. Step 3 points to the 'Language' dropdown menu, which is set to 'Upload Python Egg or PyPI'. Step 4 points to the 'PyPi Name' input field, which contains the text 'keras==2.0.8'. Step 5 points to the 'Install Library' button. The interface also includes a section for 'Install PyPi Package' with a hint: 'You can specify a package name with an optional [version specification](#)'.

3 Language Upload Python Egg or PyPI

4 PyPi Name `keras==2.0.8`

5 Install Library



Let's Get Started!