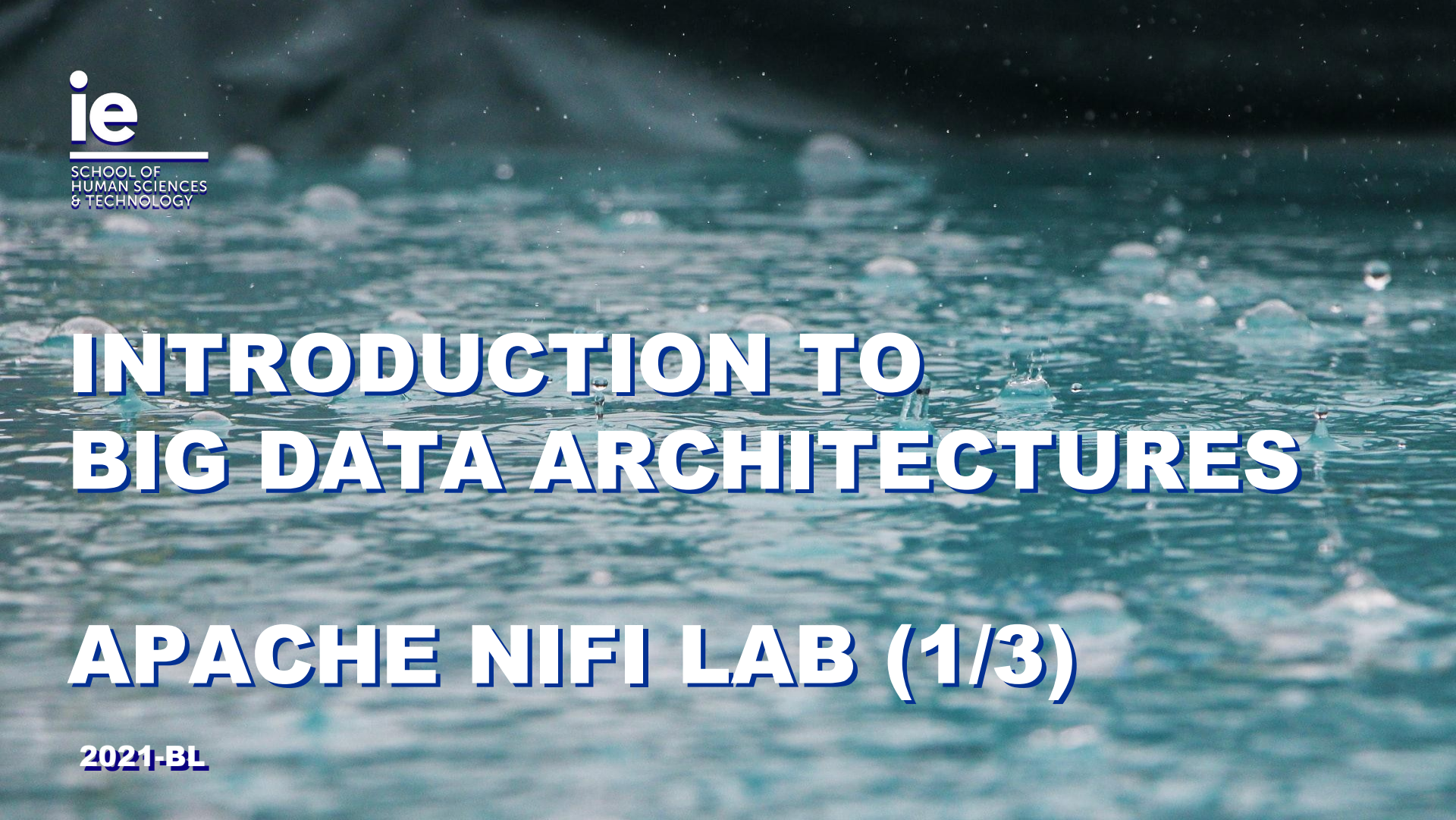

ie
SCHOOL OF
HUMAN SCIENCES
& TECHNOLOGY
INTRODUCTION TO
BIG DATA ARCHITECTURES
APACHE NIFI LAB (1/3)
2021-BL

Agenda

- Lab context
- Setting up Twitter

1.

LAB CONTEXT

Goals

- To create a complete NiFi flow
- To store streaming data in a distributed file system (HDFS).

Description

The lab is all about creating a data flow with NiFi to ingest tweets into HDFS.

Given that the default processor doesn't work as expected, we'll use a Python library, Tweepy, to ingest tweets from Twitter.

NiFi data flow

sources	ingestion	storage	processing

2.

SETTING UP TWITTER

Setting up Twitter

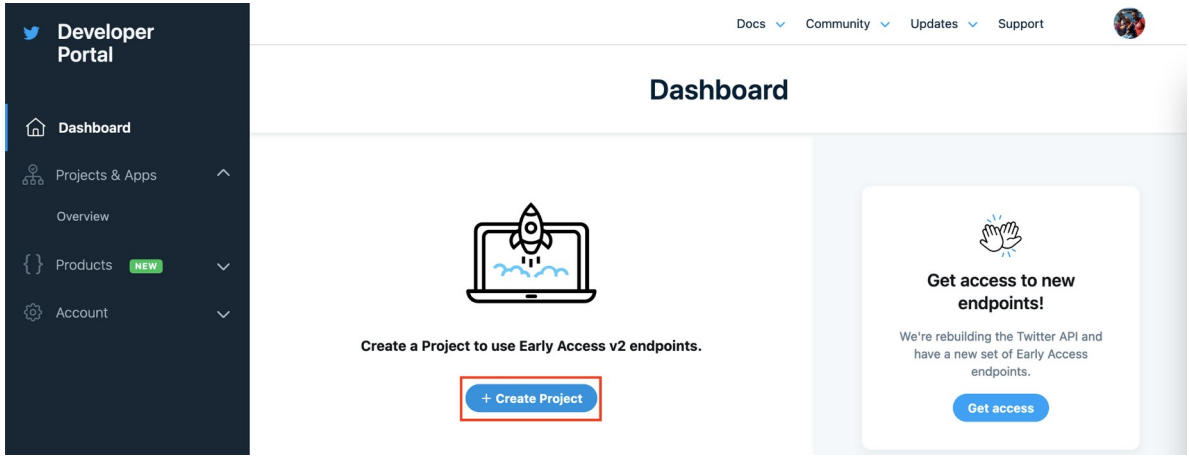
Before anything, you need to create a “Twitter App” by following these steps:

1. Visit Projects & Apps section in the Developer Portal
2. Sign into your account*
3. Click the *“Create an app”* button.
4. Fill-in the form (*at least required fields*).
5. Grab the details to setup the **ingestion script** later on.

*** You might need to create a new one; follow the wizard and provide the required information**
(feel free to mention that is going to be used for academic purposes and share my name/email if needed)

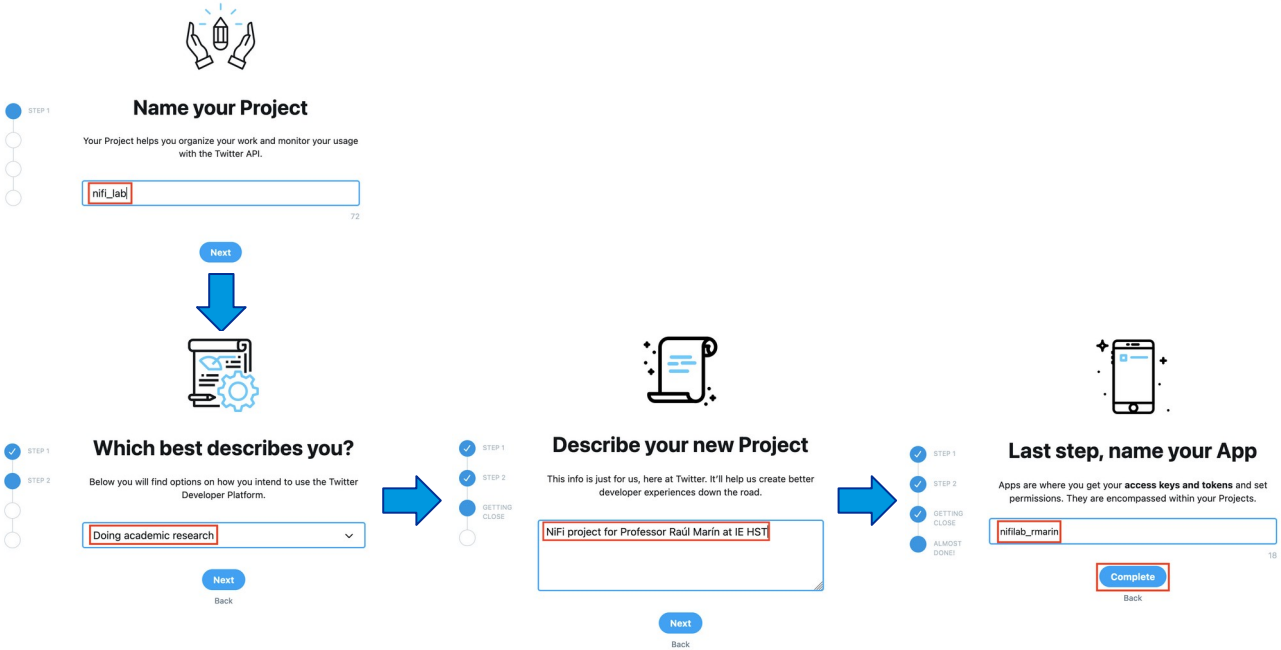
Setting up Twitter - App Creation

1. Once you're in, click the **"Create an app"** button to start the process:



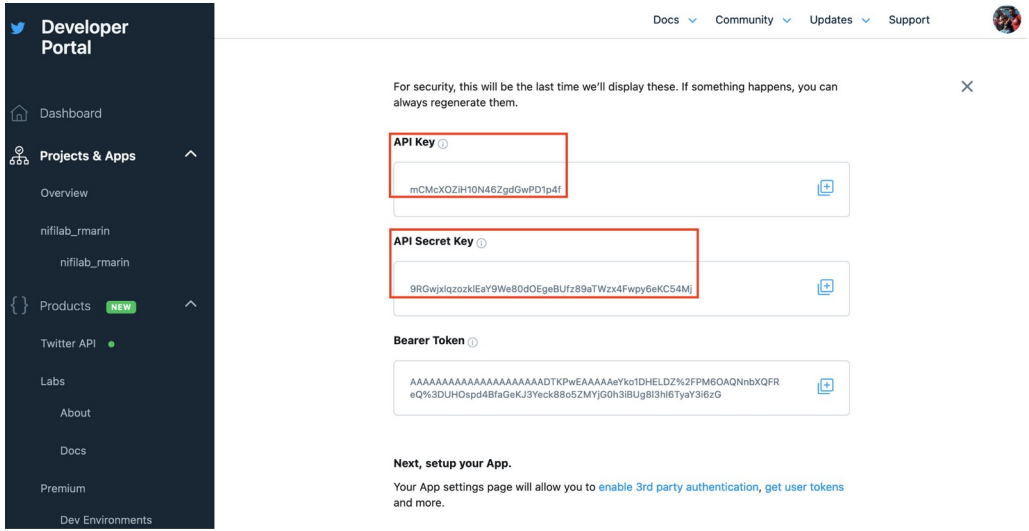
Setting up Twitter - App Creation

2. Let's fill-in **required fields**:



Setting up Twitter - App Creation

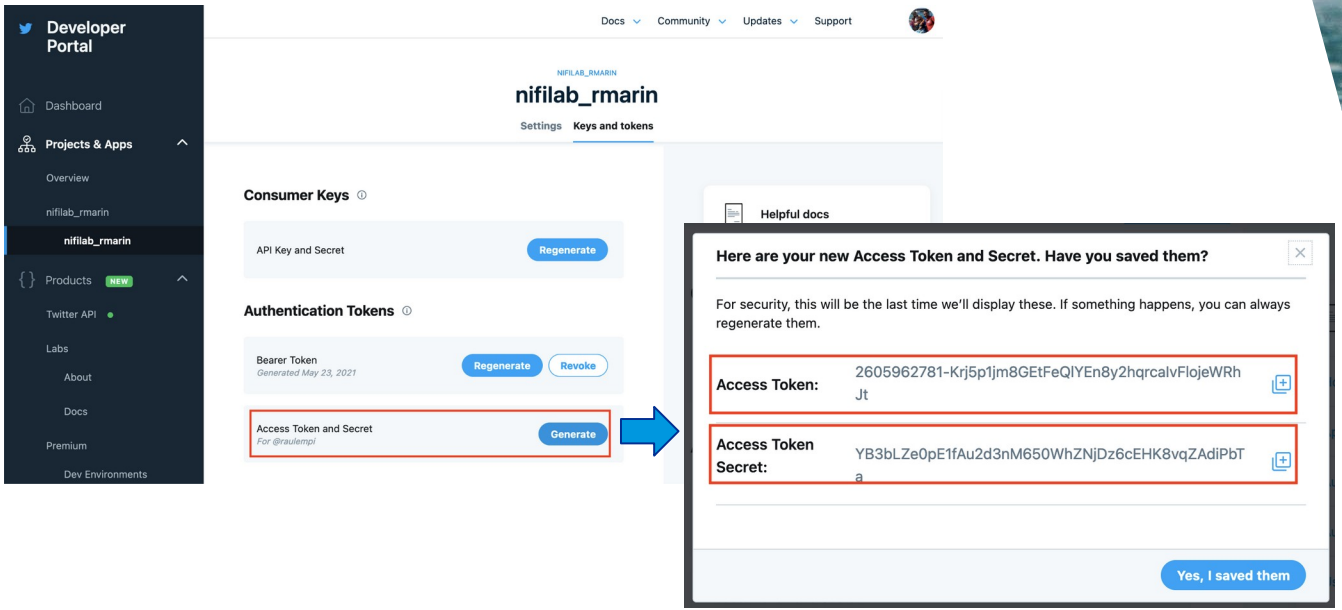
3. If all goes well, your **Twitter App should be created** and the **API Key**, **API Secret Key** will show up*:



*** Write them down for the following steps.**

Setting up Twitter - App Creation

4. The **Access Token** and **Access Token Secret** will be also needed:



*** Write them down for the following steps.**