Introduction

...

Welcome to the Course

Learning Spark2 with Python

A Little bit about myself

- Soumendra Dhanee
  - Head of Data Science and Senior Data Scientist in a number of startups
  - Currently consulting Head of Data Science in a few startups
  - Building my own products now

- soumendra@gmail.com
- https://github.com/soumendra
- https://twitter.com/dataBiryani

A Little bit about myself

Worked mostly in

- Machine Learning
- Deep Learning (Computer Vision)
- Recommender Systems
- Natural Language Processing
- Data Science Engineering

In

- Educational Technologies
- Transportation
- HR

Why should you learn Spark?

One of the coolest and most impactful technologies right now

If you hope to work for a company that deals with massive amount of data, you'll most probably need Spark.

What we'll be learning in this course

Mission Statement

We’ll evaluate if we have met our objective or not at the end of the course.

By the end of the course

- we'll be analysing and visualizing datasets with millions of records/ratings,
- using clusters of computers that we'll spin up in the cloud, and
- we’ll build predictive models using state-of-the-art algorithms.

Eco-system Goals

We'll learn the tools necessary to accomplish data science tasks

- Jupyter and Zeppelin notebooks to create reproducible workflows
- Visualizations and Reporting with Python
- Deployment with AWS EMR
- Databricks

The Approach

What we'll be learning in this tutorial

- We'll start from the simplest concepts and move towards complex applications
- We'll do a lot of hands-on exercises
- We'll focus on building our profiles to signal our experience on Spark

Takeaway

Apart from all the skills

By the end,

- we'll have 3 Spark projects that you can play with and learn from
- you can build on top of them to create portfolio projects

Capstone Project

A Recommender System

We'll take a business problem, solve it with Spark and then deploy that solution in the cloud.

What you need to bring to the table

- Go slow to go fast
- Do the in-class assignments

Evaluation Scheme

- 80% for in-class assignments
- 20% for the capstone project