

Intro to Apache Spark

<http://databricks.com/>

download slides:

http://cdn.liber118.com/workshop/itas_workshop.pdf



00: Getting Started

Introduction

installs + intros, while people arrive: 20 min

Intro: *Online Course Materials*

Best to download the slides to your laptop:

cdn.liberl18.com/workshop/itas_workshop.pdf

Be sure to complete the course survey:

<http://goo.gl/QpBSnR>

In addition to these slides, all of the code samples are available on GitHub gists:

- gist.github.com/ceteri/f2c3486062c9610eac1d
- gist.github.com/ceteri/8ae5b9509a08c08a1132
- gist.github.com/ceteri/11381941

Intro: *Success Criteria*

By end of day, participants will be comfortable with the following:

- open a Spark Shell
- use of some ML algorithms
- explore data sets loaded from HDFS, etc.
- review Spark SQL, Spark Streaming, Shark
- review advanced topics and BDAS projects
- follow-up courses and certification
- developer community resources, events, etc.
- return to workplace and demo use of Spark!

Intro: *Preliminaries*

- intros – what is your background?
- who needs to use AWS instead of laptops?
- PEM key, if needed? See tutorial:
Connect to Your Amazon EC2 Instance from Windows Using PuTTY

[Click here to download full PDF material](#)