



- Several alternative title slide layouts are available toward the end of this template deck.

Who I Am

Carl Howe
Principal

carl.howe@thinkbiganalytics.com



Course Agenda



Day 1	Day 2	Day 3
Intro to Big Data	SparkSQL and DataFrames	Spark Streaming
Cluster Architecture		
SQL on Hadoop		Spark in Python
Big Data Bottlenecks		
Spark Architecture	Spark Machine Learning Techniques	Using R In Spark
Spark In Scala		

3

Pull up <https://github.com/HortonworksUniversity/Essentials> to show that all demo instructions are available

There are optional/time-permitting labs on the github project for the Ingesting Data and Parallel Processing modules.
There is a "Backup and Recovery" section as hidden slides after the YARN section should time and/or interest suggest it be used.

Introductions



- Your name
- Your job role and responsibilities
- Your Big Data and/or Hadoop experience, if any
- Your expectations for this course

Class Logistics



- Stopping time for the day
- Breaks and lunch
- Facility information (if applicable)
 - Exits, restrooms, break room...
- Courseware
 - Do you have it?
 - Where do you get it?
- Technical information
- Wireless access, cloud access, virtual machine information...



- Closing slide for all presentations