



Cloudera Data Analyst Training: Instructor Guide

CONFIDENTIAL

This guide is confidential, and contains Cloudera proprietary information. It must not be made available to anyone other than Cloudera instructors and approved Partner instructors.

NOTE: The Exercise Instructions follow the course slides.

Version	Release Date	Description
190304c	04/08/2019	Maintenance revision
190304b	02/22/2019	Maintenance revision
190304a	02/05/2019	Maintenance revision
190304	01/28/2019	Major revision removing Pig and updating environment to CDH 6.1.0
171219a	02/13/2018	Maintenance revision

Suggested Course Timings

Per-Chapter Timings

	Arrivals and Registration	[15 minutes total]
1.	Introduction	[1 hour, 15 minutes total]
	<ul style="list-style-type: none"> • 60 minutes lecture • 15 minutes exercise(s) 	
2.	Apache Hadoop Fundamentals	[2 hours, 40 minutes total]
	<ul style="list-style-type: none"> • 120 minutes lecture • 40 minutes exercise(s) 	
3.	Introduction to Apache Hive and Impala	[45 minutes total]
	<ul style="list-style-type: none"> • 45 minutes lecture 	
4.	Querying with Apache Hive and Impala	[1 hour, 15 minutes total]
	<ul style="list-style-type: none"> • 45 minutes lecture • 30 minutes exercise(s) 	
5.	Common Operators and Built-In Functions	[1 hour, 45 minutes total]
	<ul style="list-style-type: none"> • 60 minutes lecture • 45 minutes exercise(s) 	
6.	Data Management	[1 hour, 45 minutes total]
	<ul style="list-style-type: none"> • 60 minutes lecture • 45 minutes exercise(s) 	
7.	Data Storage and Performance	[2 hours, 00 minutes total]
	<ul style="list-style-type: none"> • 75 minutes lecture • 45 minutes exercise(s) 	
8.	Working with Multiple Datasets	[1 hour, 45 minutes total]
	<ul style="list-style-type: none"> • 60 minutes lecture • 45 minutes exercise(s) 	
9.	Analytic Functions and Windowing	[2 hours, 35 minutes total]
	<ul style="list-style-type: none"> • 75 minutes lecture • 60 minutes exercise(s) • 20 minutes demonstration(s) 	
10.	Complex Data	[1 hour, 45 minutes total]
	<ul style="list-style-type: none"> • 60 minutes lecture 	

- 45 minutes exercise(s)
- 11. **Analyzing Text** [1 hour, 45 minutes total]
 - 60 minutes lecture
 - 45 minutes exercise(s)
- 12. **Apache Hive Optimization** [1 hour, 45 minutes total]
 - 75 minutes lecture
 - 30 minutes exercise(s)
- 13. **Apache Impala Optimization** [60 minutes total]
 - 45 minutes lecture
 - 15 minutes exercise(s)
- 14. **Extending Apache Hive and Impala** [2 hours, 15 minutes total]
 - 60 minutes lecture
 - 75 minutes exercise(s)
- 15. **Choosing the Best Tool for the Job** [1 hour, 25 minutes total]
 - 25 minutes lecture
 - 60 minutes exercise(s)
- 16. **Conclusion** [15 minutes total]
 - 15 minutes lecture
- A. **Apache Kudu** [1 hour, 45 minutes total]
 - 75 minutes lecture
 - 30 minutes exercise(s)
- Final Questions and Post-Course Survey** [15 minutes total]

Per-Day Timings

• Day 1

[Total classroom time: 6 hours, 15 minutes]

Complete all lectures and exercises for Chapters 1–4, plus lecture for Chapter 5.

• Day 2

[Total classroom time: 6 hours, 15 minutes]

Complete exercise for Chapter 5 and all lectures and exercises for Chapters 6–8.

• Day 3

[Total classroom time: 6 hours, 15 minutes]

Complete all lectures and exercises for Chapters 9–11, and most of the lecture for Chapter 12.

- **Day 4**

[Total classroom time: 6 hours, 10 minutes]

Finish lecture and exercise from Chapter 12, and complete all lectures and exercises for Chapters 13–16.

Notes on Course Timings

The timings above are a general guide, but the pace of your class will depend on several factors. These might include your presentation style, the experience level of your students, and how you manage student participation (for example, starting promptly at the scheduled time or deferring off-topic discussions until a break).

If you are running behind schedule, you might do one or more of the following:

- Skip some or all of the review questions found in the “Essential Points” slide of each chapter in the Instructor Guide.
- If only one or two students are struggling to finish an exercise, suggest that they use the `catchup.sh` script as described at the beginning of the next lab to catch up with the rest of the class. When doing this, it will be helpful to remind the students that they can later finish the exercise they had been working on at their own pace after class is over. As the step for the `catchup.sh` script at the beginning of each exercise reverts the state of data in the hands-on environment to what it would be if the exercise preceding it was successfully completed, you can start with any exercise you like so long as you follow this step. The first exercise is the only exception to this rule; the instructions for that exercise do not mention how to invoke that script since students should be starting from a new hands-on environment. An alternative to using a new copy of the environment (which would lose any code the student had written) is to call `$ADIR/scripts/advance_labs.sh cleanup` to revert all data to its original state.

If you are running ahead of schedule, you might do one or more of the following:

- Ask some or all of the review questions found in the “Essential Points” slide of each chapter in the Instructor Guide.
- Allow extra time for students to complete the bonus exercises.
- If there is widespread interest from the students, cover the material in the appendix “Apache Kudu.”