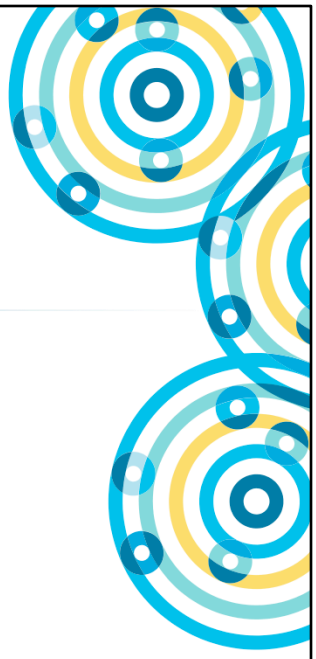
cloudera®



Conclusion

Chapter 18

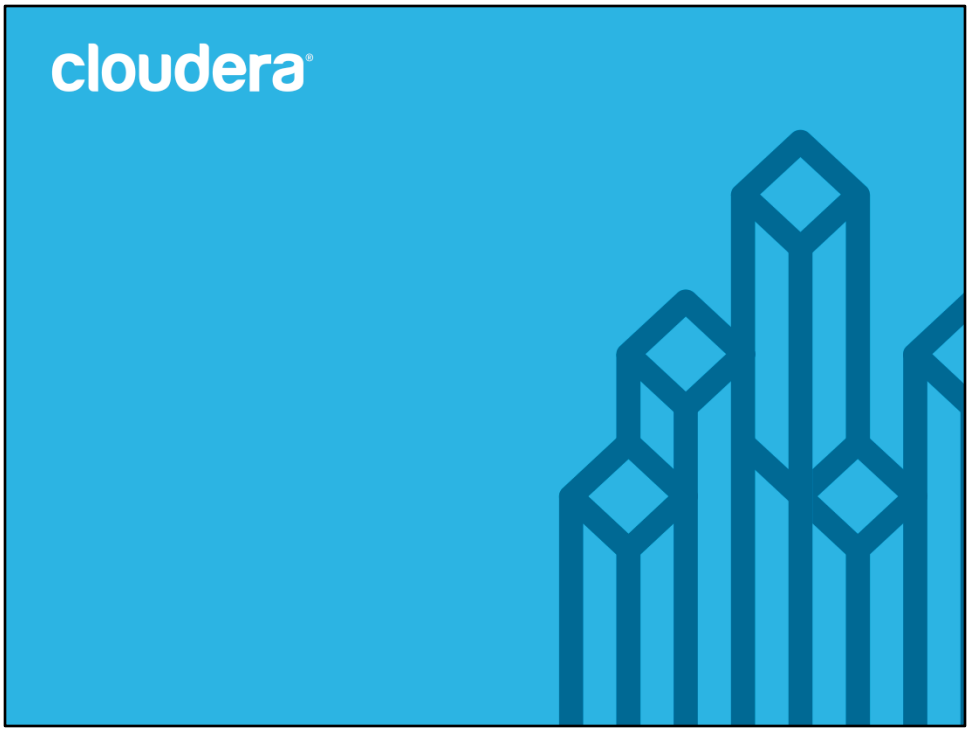
Course Chapters

1	Introduction	Course Introduction
2 3	Introduction to Hadoop and the Hadoop Ecosystem Hadoop Architecture and HDFS	Introduction to Hadoop
4 5 6 7 8	Importing Relational Data with Apache Sqoop Introduction to Impala and Hive Modeling and Managing Data with Impala and Hive Data Formats Data Partitioning	Importing and Modeling Structured Data
9	Capturing Data with Apache Flume	Ingesting Streaming Data
10 11 12 13 14 15 16 17	Spark Basics Working with RDDs in Spark Aggregating Data with Pair RDDs Writing and Deploying Spark Applications Parallel Processing in Spark Spark RDD Persistence Common Patterns in Spark Data Processing Spark SQL and DataFrames	Distributed Data Processing with Spark
18	**Conclusion**	**Course Conclusion**

Course Objectives

During this course, you have learned

- **How the Hadoop Ecosystem fits in with the data processing lifecycle**
- **How data is distributed, stored and processed in a Hadoop cluster**
- **How to use Sqoop and Flume to ingest data**
- **How to model structured data as tables in Impala and Hive**
- **Best practices for data storage**
- **How to choose a data storage format for your data usage patterns**
- **How to process distributed data with Spark**


cloudera