5.00 (3)

[Live] Complete Data Engineering with Apache Spark 3,Kafka, Hive, Hadoop, Azure DataBricks

ADD TO CART ~~₹17,990.00~~ **₹14,990.00**

-1:19:10

Description

Course Content :
*Linux Commands
*SQL
*Python
*Hadoop
*Hive
*Spark
*Sqoop,Flume
*SparkSQL
*Spark Streaming
*Kafka
*NoSQL
Live classroom training for Complete Data Engineering course with Big Data Hadoop and Spark. The course focuses on various aspects of Big Data frameworks like Hadoop and Spark. We will be learning about many tools in Hadoop ecosystem such as hive, sqoop, flume, spark and Kafka.

NEW BATCH WILL START ON : |
Contact(Whatsapp) : +91 9028378280 for more details

What Will I Learn?

- Job interview preparation

- Covers most of the contents for "Databricks Certified Developer For Apache Spark 3.0" Certification
- In depth understanding of Hadoop Ecosystem components.
- Resume support.
- Enhanced understanding with Hands on exercises.

Topics for this course

88 Lessons • 100h

Big Data Hadoop and Spark	
Introductory Session	1:19:10
Module 1 : Introduction to Big Data	
Introduction to Big Data (Part 1)	56:57
Introduction to Big Data (Part 2)	1:05:59
Module 2 : Introduction to Hadoop	
Introduction to Hadoop	1:05:59
High Availability and Federation	1:23:18
Module 3 : HDFS (Hadoop Distributed File System)	
File Blocks , Replication , HB , BR and Safe Mode	1:16:39
Rack awareness , Read/Write Anatomy	1:26:57
HDFS Commands	1:18:53
Module 4 : YARN (Yet Another Resource Negotiator)	
YARN Architecture	1:41:37
Mapreduce Architecture Hadoop 1.x	1:06:20
Module 5 : Setting up Pseudo Distributed Hadoop Cluster	
Document (PDF File)	
Module 6 : Map Reduce	
MapReduce Workflow	1:50:52
WordCount MR Program	2:11:12
MR and Input Splits	2:11:12
Transactions use case	1:38:21

Module 7 : Advance Map Reduce

Combiner and Partitioner	2:39:58
Reduce side join 1	1:32:32
Reduce side join 2	1:32:10
Map Side Join	1:36:28

Module 8 : Hive (SQL on top of Hadoop)

Hive Introduction and Architecture	1:00:00
Hive Basic Commands	1:31:16
Internal vs External Tables	52:48
Design level optimizations (Partitioning)	1:56:10
Design level optimizations (Bucketing)	1:55:39
Storage Level Optimizations	1:55:39
Join Optimization – I	2:15:23
Hive Project	22:29
Join Optimizations	2:15:23
CBO | Vectorization | Resource level optimization | Materialized Views	2:15:23
UDF, UDAF , UDTF	2:15:23

Module 9 : Sqoop

SQOOP Introduction	1:19:27
Sqoop Incremental Import	1:21:12

Module 10 : Flume

Flume Introduction	1:39:30
Flume Configuration	2:11:07

PreRequisite – 3 | Python for Pyspark ?

Python Session 1 | Required Setups	1:26:09

Python Session 2 | Printing Strings and Variaables	46:37
Python Session 3 | Arguments and Functions	46:37
Python Session 4 | Functions	29:23
Python Session 5 | Conditional statements and SparkSQL doubts	33:24
Python Session 6 | Python Lists	39:21
Python Session 7 | Python Tuples	33:24
python session 8 | Python Sets	33:24
Python session 9 | Dictionaries in Python	34:50
Python Session 10 | Loops in python	49:57
[Raw] Python basic exercises complete code	
Module 11 : Apache Spark	
Spark Introduction	1:04:14
Spark Architecture – I	1:07:15
Spark Architecture – II	1:49:27
Spark Deploy Modes	14:53
PySpark Installation on windows	24:26
Module 12 : Spark Programming Model	
RDD exercises Hands On | PySpark	1:26:09
RDD practice | PySpark	1:21:36
Broadcast Vraiables | PySpark	46:17
Accumulator and SparkSQL Introduction | PySpark	1:08:20
RDD creation | Scala	2:06:39
RDD Transformations | Scala	2:12:31
Broadcast Variables | Scala	2:12:31
RDD Exercise and Doubt session | Scala	1:21:25

Accumulator | Scala	42:24
Module 13 : Spark SQL	
DF vs DS, catalyst Optimizer | PySpark	1:01:13
SC vs SS | TempView vs GlobalTempView | PySpark	1:27:34
DF Structured Transformations | PySpark	29:55
Hands on | Creating DFs | PySpark	43:06
SparkSQL – I | Scala	43:06
SparkSQL – II | Scala	1:46:55
SparkSQL -III | Scala	1:32:29
Spark hive Integration | Scala	1:11:46
Deploy modes and Resource calculations | Scala	1:34:47
PySpark DF Scenarios and Databricks Certification Practice	
Session : 1	1:07:53
Session : 2	1:02:28
Session : 3	1:16:19
Session : 4	51:23
Session : 5	47:50
Session : 6	35:49
Session : 7	40:39
Session : 8	29:37
SparkSQL Advanced	
Executor Memory Architecture	1:16:56
Adaptive Query Execution	57:57
Cache vs Persist | Repartition vs Coalesce	1:25:50
Resource Calculations for Spark Application	22:06

Module 14 : Spark Streaming	
Spark streaming – I	47:13
Spark streaming – II	14:02
Spark Streaming – III	14:02
Module 15 : Kafka Basics	
Kafka introductory	1:35:19
kafka storage architecture	1:30:56
flume to kafka	1:30:57
Module 16 : Hbase NoSQL database	
Hbase Introductory	1:37:59
Hbase Commands	2:19:01
Module 17 : Airflow	

Instructor

Sandeep Patil

5.00 ratings (5) 3 Courses 88 students

Student Feedback

5.0

Total 3 Ratings

Harpreet Kaur Sodhi

1 month ago

The way of teaching is really amazing!! Joining this course is totally worth it. Anyone who is looking to have a successful career in Big data Engineering must enroll for the course!

A

ajit_sharma

3 months ago

★★★★★

Sandeep sir You are great. We easily understand what you taught. We are learning more from you. Thanks sir.

SK

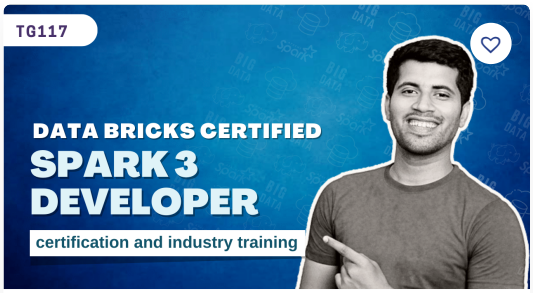
Sonali Kakde

3 months ago

★★★★★

Good Teaching and everyone must join if they need to have good knowledge of Big Data.

Similar Courses



Sandeep Patil

[Hindi] DataBricks Certified Spark 3 Developer

25 Lessons 27 5.00(1)

INTERMEDIATE ~~₹7,990.00~~ ₹3,990.00

Sandeep Patil

[Hindi] Data Engineering Using Hadoop 3 and Apache Hive

41 Lessons 29

INTERMEDIATE

Sandeep Patil

[Live] Complete D Engineering with 3,Kafka, Hive, Ha DataBricks

88 Lessons 32

ALL LEVELS ~~₹17,990.0~~

Technological Geeks

The main objective of "TG" family is to provide guidance about the recent technologies to students , beginners and IT professionals .

Quick Links

Courses
Contact Us
Terms and Conditions
Refund and Return Policy
Privacy Policy

Contact Us

sdp117@gmail.com
Contact No: 9028378280
New Vishal Nagar, Gajanan Mandir, Garkheda, Aurangabad

Follow Us