

Hadoop & Spark Developer Hadoop & Spark Developer Hadoop & Spark Developer - Mindtree An IT professional with around 3+ years of experience in Software Development and Implementation of Big Data and Big Data related technologies. Efficient on working with Big Data and Hadoop Distributed File System (HDFS) Good working knowledge with Hive and Sqoop. Written Sqoop scripts In Order to Integrate with Hive and MYSQL Database. Expert in data ingestion, data transformations, data preprocessing, data exploration and data integration using Sqoop, Hive Scripts. Performed Hive operations on large datasets with proficiency in writing HiveQL queries using transactional and performance efficient concepts: Partitioning, Bucketing, efficient and effective Join operations. In-depth understanding of Apache Spark Architecture and performed querying using Spark (Core, SQL), RDDs, Data frames, Datasets. Experience and strong knowledge on implementation of Spark Core, Spark SQL. Implemented Spark using Scala and Spark SQL for faster testing and processing of data. Experience in Scala programming. Experience in working with Echo systems like Sqoop, Hive, Hbase. Knowledge with message broker such as Apache Kafka. Worked with Sqoop to Import and Export data from a relational database into Hadoop. Experienced with different file formats like Parquet, ORC, Avro, Sequence, CSV, XML, JSON, Text files. Scheduled jobs and automated workflows using Oozie. Experience in Agile Development and Scrum process. Work Experience Hadoop & Spark Developer Mindtree - US May 2017 to Present USA Environment: Hadoop, Hive, SGOOP, MySQL, Spark Core, Spark Sql Duration: May 2017 to till date Role: Hadoop & Spark Developer Description: HCA Data Repository is a cross platform patient medical records suite for capturing monitoring information such as measurements for hypertension or blood sugar, patient reminders, symptoms reporting, issues reporting. We will process the metadata datasets coming thousands of records of patients data from various hospitals. The purpose of publishing this dataset is to figure out the patients depending upon various constraints like type of disease, based on gender, age, seasonal diseases, on location base. Roles & Responsibilities: - Written Hive Scripts to process the HDFS data. - Created Hive tables to Store the Processed results. - Devoloped Sqoop Scripts in order to make the interaction between Hive and Mysql databases. - Implemented Spark using Scala and

Spark SQL for faster testing and Processing of data. - Involved in Conversion of SQL to HQL and My SQL procedures to Scala Code to run on Spark Engine - Creating RDD to load the Unstructured data. - Involved in gathering the requirements, designing and development. - Writing the Script files for Processing data and loading in HDFS. - Writing CLI commands using HDFS. - Creating Data frames to get tables format. - Using count and show actions after creation of Data frames. - Involved in requirement analysis phase.

Project #2: Project Name: Target E-Commerce Analysis
Client: Target Software Engineer Mindtree - Bengaluru, Karnataka March 2016 to Present
Technical Skills: Distribution: Cloudera (CDH5) Hadoop / Big Data: HDFS, Sqoop, Hive, HBase Spark-core and Spark-SQL. Operating Systems: Linux/UNIX, Windows Programming Languages: Scala RDBMS/NoSQL: SQL server, HBase Scripting: Shell Scripting IDE: Eclipse, IntelliJ Idea

Project Details: Project #1: Project Name: HCA Data Repository Education B.TECH in MECH NOVA COLLEGE OF ENGINEERING&TECHNOLOGY - Hyderabad, Telangana
Skills APACHE HADOOP HDFS (3 years), APACHE HADOOP SQOOP (3 years), Hadoop (3 years), HADOOP (3 years), HADOOP DISTRIBUTED FILE SYSTEM (3 years) Additional Information Technologies: CDH, HDFS, Hive, Sqoop, Spark Duration: May 2017 to till date Role: Hadoop & Spark Developer

Description: Target is the Second Largest E-Commerce Company in USA. Target is using Bigdata technologies to analyze their Customer transactions and draw some useful insights off it which will be useful for their business development

Roles & Responsibilities: - Used spark-sql to read the data from csv files and create tables in hive using scala API. - Used various spark transformations and actions for cleansing the input data and transformation. - Experience in Spark RDD and Spark SQL Programming using Scala. - Executed Queries on tables in hive to perform data analysis. - Responsible for writing hive queries. - Developing Hive Scripts for processing and transforming the data stored in HDFS. - Worked with Hive partition and Bucketing. - Hands on storing the data in different file formats in Hive. - Hands on with various data formats using Spark-Core and Spark-Sql.

Name: Zachary Davis

Email: bradley17@example.net

Phone: 001-671-990-3411