

Spark Developer Spark Developer Spark Developer - Highmark Health Solutions 3.6 Years of IT experience in Application development in Python and Hadoop. 2.6 years of experience in deployment of Hadoop Ecosystems like Hadoop(HDFS), Hive, Pig, HBase, Sqoop, Flume, Spark and Scala. Good working experience on Hive and Pig. Having experience on developing Apache Spark programs using scala for large-dataset processing and using the in-memory computing capabilities for faster data processing with spark core and spark-SQL. Expertise in developing spark RDD transformation, actions, Dataframes, case classes for the required input data and performed the data transformation using spark-core Very good understanding of partition, bucketing concepts in hive and designed both external and internal tables in hive optimize performance. Involved in creating Hive tables, loading with data and writing hive queries which will run internally in map reduce way. Experience in writing Pig Scripts to build the source to target mapping. Experience in migrating the data using Sqoop from HDFS to relational data base system and vice-versa according to the other clients. Good experience working with NoSQL technologies using HBase. Having good knowledge on UNIX commands. I have knowledge on PySpark.

Work Experience Spark Developer Highmark Health Solutions October 2017 to Present Project Summary: Highmark is national, diversified health care partner serving members through its business in health insurance, dental insurance, vision care and reinsurance. Highmark provides information, guidance, and operational services necessary for hospitals and health plans to transform their organizations into high-performing accountable delivery systems. Highmark operate health care plans in Pennsylvania, Delaware and West Virginia that serve 5.3 million members.

Roles & Responsibilities: Responsible for processing scalable distributed data solutions using Hadoop. Developing Data pipe line using spark and hive to ingest data into hadoop cluster for analysis. Used various transformation and actions for cleaning the input. Loaded data into spark RDD and Data Frame to do in memory data computation to generate the output response. Used SparkSQL to process the huge amount of structured data to perform SQL operations. Exploring with the spark improving the performance and Optimizing of existing algorithms in Hadoop using Spark Context, Spark-SQL, Data frames and pair RDDs. Involved in creating hive tables, loading

and analyzing data using hive queries. Worked extensively with sqoop for importing data from RDBMS. Environment: HDFS, Spark, Scala, Hive, Sqoop, Kafka, HBase, Oozie, Flume, Parquet, Linux, IntelliJ. Project#2: Software Engineer InfoGain - Bangalore, KARNATAKA, IN August 2016 to Present Hadoop Developer Prepaid and Merchant solutions July 2016 to September 2017 Project Summary: TSYS make it possible for millions of buyers and sellers to move money around the world with trust and confidence supports Issuing Services, Acquiring services, Prepaid and Merchant solutions. This project is responsible to get the merchant sales and the credit card transactions. Then, it splits the transactions based on the card type and generates the settlement files for the respective providers. The project is critical as it deals with millions of dollars on a daily basis. The platform is built on Hadoop ecosystem with HDFS/HBase being the primary data storage.

Roles & Responsibilities: 1. Involved in end to end data processing like ingestion, processing, quality checks and splitting. 2. Bringing the data into Big Data Lake using Pig, Sqoop and Hive. 3. Refined terabytes of data from different sources and created hive tables. 4. Importing and exporting data into HDFS and HIVE from MySql database using Sqoop. 5. Responsible to manage data coming from different sources. 6. Responsible for loading data from UNIX file systems into HDFS. 7. Wrote Pig scripts to process unstructured data and create structure data for use with Hive. Environment: Sqoop, HDFS, Pig, Hive, Map Reduce, Java, Oozie, Eclipse, Linux, Oracle, Teradata.

Project#3 IMI Mobile - Hyderabad, ANDHRA PRADESH, IN July 2015 to July 2016 Python Developer IMI Mobile July 2015 to June 2016 Project Summary: The main idea for implementing this project is to replace existing manual tender system with computerized system. Online tendering in its simplest form is described as the electronic publishing, communicating, accessing, receiving and submitting of all tender-related information and documentation via the internet, there by replacing the traditional processes, and achieving a more efficient and effective business process for all government PWD's and Contractors. Roles & Responsibilities: Preparation of functional requirement specification as per client requirement. Implemented code specified by client. Developing/Enhancing the code based on the client requirements. Involved in development, enhancement and support activities coordinating. Involved in creating different

models using python required for the project. Environment: python, Django, MySQL db. Additional Information Technical Skills: Big Data Technologies: Hadoop, HDFS, MapReduce, Hive, Pig, Hbase, Sqoop, Flume, Oozie, Spark Data Bases: Oracle, MySQL, Sql. Programming Languages: C, C++, Python, Scala Operating Systems: Windows, Linux (centos and Ubuntu)

Name: Christopher Taylor

Email: nicholasrodriguez@example.net

Phone: +1-800-751-7191x97533