

Spark and Hadoop Developer Spark and Hadoop Developer Spark and Hadoop Developer - Vanguard 6+ Years of professional experience in IT which includes around 4 years of comprehensive experience as Hadoop and Spark Developer, and related technologies. In depth understanding/knowledge of Hadoop Architecture and various components such as HDFS, Name Node, Data Node, Map Reduce, Spark and Spark SQL Experience in converting Hive/SQL queries into Spark transformations using Spark RDDs. Exposure on usage of Apache Kafka to develop data pipeline of logs as a stream of messages using producers and consumers. Knowledge in handling Kafka cluster and created several topologies to support real-time processing requirements.

Experience of converting various File Formats queries into Spark transformations using Data Frames and Datasets. Experience of developing SQL scripts using Spark for handling different data sets and verifying the performance over Map Reduce jobs. Experience in creating Kafka producer and Kafka consumer for Spark streaming. Exposure to Spark, Spark Streaming, Scala and Creating the Data Frames handled in Spark with Scala. In-depth understanding of Spark Architecture including Spark SQL, Data Frames, Spark Streaming. Extensive experience on importing and exporting data using stream processing platforms like Flume and Kafka. Implemented Sqoop for large dataset transfer between Hadoop and RDBMS. Hands on experience in working on Spark SQL queries, Data frames, and import data from Data sources, perform transformations, perform read/write operations, save the results to output directory into HDFS. Worked with AWS to migrate the entire Data Centers to the cloud using EC2, S3, and EMR. Experience in importing and exporting Multi Terabytes of data using Sqoop from HDFS to Relational Database Systems (RDBMS) and vice-versa Experience in writing shell scripts to dump the Shared data from MySQL, Oracle servers to HDFS. Highly experienced in importing and exporting data between HDFS and Relational Database Management systems using Sqoop. Good Experience working with Amazon AWS for setting up Hadoop cluster. Hand on experience in installing, configuring, and using Hadoop ecosystem components like Hadoop MapReduce, HDFS, HBase, Oozie, Hive, Sqoop, Zookeeper and Flume. Worked on custom Spark Transformations and variety of data formats such as JSON, Compressed CSV, ORC, AVRO etc. and reading data

from various sources like HBase and Hive. Performed map-side joins on RDD. Good understanding of Amazon web services like Elastic MapReduce (EMR), EC2. Experience in ETL operations on Hive to Spark. Gained hands on experience in writing shell scripts in UNIX. Experienced with processing different file formats like Avro, XML, JSON and Sequence file formats using Spark. Experience in implementing Spark using Scala and Spark SQL for faster analyzing and processing of data. Good understanding in configuring simple to complex work flows using Oozie. Good understanding of NoSQL databases like MongoDB and HBase. Worked on different operating systems like UNIX/Linux, Windows. Worked in managing VMs in Cloudera and Horton Works. Experience as a java Developer in client/server technologies using JSP, JDBC and SQL. Work Experience Spark and Hadoop Developer Vanguard - Malvern, PA October 2018 to Present Responsibilities:

- \* Worked on Spark streaming using Apache Kafka for real time data processing.
- \* Experienced with Kafka to ingest data into Spark Engine
- \* Configured Spark streaming to get ongoing information from the Kafka and store the stream information to HDFS.
- \* Involved in converting Hive/SQL queries into Spark transformations using Spark RDD's, Data frames, Spark SQL and Scala
- \* Performed advanced procedures like text analytics and processing, using the in-memory computing capabilities of Spark using Scala.
- \* Developed Scala and Spark SQL code to extract data from various databases.
- \* Used Spark SQL to process the huge amount of structured data and Implemented Spark RDD transformations, actions to migrate Map reduce algorithms
- \* Extensively worked with all kinds of Un-Structured, Semi-Structured and Structured data.
- \* Developed multiple Kafka Producers and Consumers from as per the software requirement specifications.
- \* Creatively communicated and presented models to business customers and executives, utilizing a variety of formats and visualization methodologies.
- \* Used Kafka extensively in gathering and moving log data files from Application Servers to a central location in Hadoop Distributed File System (HDFS)
- \* Used Spark and Spark-SQL to read the parquet data and create the tables in hive using the Scala API.
- \* Developed robust set of codes that are tested, automated, structured and efficient.
- \* Performed map-side joins on RDD, Spark SQL and Data Frames.
- \* Migrating the needed data from Oracle, MySQL in to HDFS using Sqoop and importing various

formats of flat files in to HDFS. \* Uploaded and processed more than 30 terabytes of data from various structured and unstructured sources into HDFS using Sqoop. \* Implemented the data backup strategies for the data in the HDFS cluster. \* Developed a data pipeline using Spark and Hive to ingest, transform and analyzing data. \* Developed Spark code using Scala and Spark-SQL for faster testing and data processing. \* Imported the data from relational databases into HDFS using Sqoop. \* Understanding of data storage and retrieval techniques, ETL, and databases, relational databases, tuple stores, Hadoop, HUE, MySQL and Oracle databases. Environment: Apache Spark, Cloudera (CDH 5.12), Scala, Spark SQL, Data Frames, Kafka, SBT build, HUE, Sqoop, Zookeeper, HDFS. Hadoop Developer Century link - Littleton, CO January 2018 to October 2018 Responsibilities: \* Involved in converting Map Reduce programs into Spark transformations using Spark RDD's on Scala. \* Developed analytical component using Scala and Spark \* Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs and Scala. \* Real streaming the data using Spark with Flume and store the stream data to HDFS using Scala. \* Installing configuring, troubleshooting of Hadoop with Cloudera Ecosystem, Configuring & Troubleshooting job tracker Nodes, task tracker (data) nodes, MapReduce, Spark, Hive, HBase, Sqoop and Oozie. \* Involved in complete Implementation lifecycle, specialized in writing custom Map Reduce, Pig and Hive programs. \* Extensively used Sqoop to get data from RDBMS sources like Teradata. \* Worked on reading multiple data formats on HDFS using Scala. \* Used Cloudera Manager to monitor and manage Hadoop Cluster. \* Deployed a multi-node Hadoop cluster \* Worked extensively with SGOOP for importing metadata from Oracle. \* Extensively used Hive/HQL or Hive queries to query or search for a string in Hive tables in HDFS. \* Written Hive jobs to parse the logs and structure them in tabular format to facilitate effective querying on the log data. \* Created HBase tables to store various data formats of data coming from various sources. \* Used Impala to query the Hadoop data stored in HDFS. \* Worked on streaming log data into HDFS from web servers using Flume. \* Was responsible for importing the data (Log files) from various sources into HDFS using Flume. \* Creating Hive tables on JSON data, Creating Hive tables on JSON data, Run the Hive queries in Hue and CLI. \* Implemented POC for using APACHE IMPALA for data

processing on top of HIVE. \* Migrated ETL jobs to Spark to do Transformations, even joins and some pre-aggregations before storing the data onto HDFS. \* Implement Flume, Spark, Stream framework for real time data processing. Environment: UNIX, HDFS, Hive, Spark, Scala, Flume, Sqoop, HBase, Zookeeper, CDH 5.4. Hadoop Developer ArcelorMittal - Burns Harbor, IN December 2015 to December 2017 Responsibilities: \* Build and maintain scalable data pipelines using the Hadoop ecosystem and other open source components like Hive and HBase. \* Handle the data exchange between HDFS and RDBMS using Sqoop. \* Responsible for building scalable distributed data solutions using Hadoop Cloudera Distribution. \* Developed several advanced Map Reduce programs to process data files received. \* Developed Hive Scripts, Hive UDFs to load data files into Hadoop. \* Close monitoring and analysis of the Map Reduce job executions on cluster at task level. \* Import the data from various sources like HDFS/HBase into Hive. \* Extensively worked on Hive for ETL Transformations and optimized Hive Queries. \* Used Flume to collect, aggregate, and store the web log data from various sources like web servers, mobile and network Devices and pushed to HDFS. \* Created tables in Impala, Created Partition tables, run the hive queries in hue, created tables HBase, Created folders in HDFS. \* Importing and Exporting Data from MySQL/Oracle to HiveQL. \* Importing and Exporting Data from MySQL/Oracle to HDFS. \* Write the Impala queries. \* Installed Oozie workflow engine to run multiple Hive jobs. Environment: CDH5.9, HDFS, Hive, hue, Sqoop, Impala, HBase, Oozie, Hive, Flume, Linux) Java developer Integra Micro Systems February 2015 to November 2015 Responsibilities: \* Involved in Analysis, Design, Development, Integration and Testing of application modules. \* Involved in developing a custom framework like Spring Framework, with more features to meet the business needs. \* Performed requirement analysis, design, coding and implementation, team co-ordination, code review, testing, and Installation. \* Developed server side utilities using JAVA technologies Servlets, JSP. \* Developed presentation layers using JSP custom tags and JavaScript. \* Implemented design patterns - Business Delegate, Singleton, Flow Controller, DAO and Value Object patterns. \* Developed Role Based Access Control to restrict the users to access specific modules based on their roles. \* Used Oracle as the back-end application and used Hibernate Framework for ORM. \* Deployed the

application on WebSphere server using Eclipse as the IDE. \* Used Tomcat server 5.5 and configured it with Eclipse IDE. \* Performed extensive Unit Testing for the application. \* Responsible for Design and development of Web pages using HTML, CSS including Ajax controls and XML.

Environment: WebSphere 5.1, Tomcat 5.0, Oracle 9i, Hibernate3.0, Eclipse 3.2, JSP, Java Script, Servlets, XML, Eclipse, Junit, Spring, plug-ins (Tomcat) Jr Java Developer PRIME TECHNOLOGY GROUP May 2013 to February 2015 Responsibilities: \* Involved in the analysis, design, implementation, and testing of the project. \* Implemented the presentation layer with HTML and JavaScript. \* Developed web components using JSP, Servlets and JDBC. \* Implemented database using SQL Server. \* Designed tables and indexes. \* Wrote complex SQL and stored procedures. \* Involved in fixing bugs and unit testing with test cases using JUnit. \* Developed user and technical documentation.

Environment: Java, JSP, Servlets, JDBC, HTML, JavaScript, MySQL, JUnit, Eclipse IDE.

Education Bachelor's Skills MYSQL (4 years), ORACLE (4 years), Hadoop (3 years), HADOOP (3 years), APACHE HADOOP HDFS (3 years)

Additional Information TECHNICAL SKILLS: Big data/Hadoop Ecosystem HDFS, Map Reduce, Hive, HBase, Sqoop, Flume, Oozie, Spark, Hue, Impala, Kafka, Spark Data Frames, Spark SQL. Hadoop Distribution Platforms Cloudera (CDH4/CDH5), Horton works. Programming Languages Core Java, Scala, SQL, Linux. Application Servers Tomcat, WebLogic Databases MySQL, Oracle, No-SQL Database (HBase) IDE Tools Eclipse, IntelliJ, Putty, MobaXterm, Tableau. Operating System Centos, Windows (7, 8, 10), Ubuntu, UNIX

Name: Joshua Payne

Email: edwardsanchez@example.net

Phone: (735)808-9419x5152