

Spark Scala Developer Spark Scala Developer Spark Scala Developer - Intel Folsom, CA ? Around 10+ years of experience in Design, Analysis and Development of software application using Big Data/ Hadoop, Spark and Java/JEE Technologies. ? Knowledge in Spark Core, Spark-SQL, Spark Streaming and machine learning using Scala and Python Programming languages. ? Worked on Open Source Apache Hadoop, Cloudera Enterprise (CDH) and Hortonworks Data Platform (HDP) ? Hands on experience on major components in Hadoop Ecosystem like Hadoop Map Reduce, HDFS, HIVE, PIG, Pentaho, HBase, Zookeeper, Task Tracker, Name Node, Data Node, Sqoop, Oozie, Cassandra, Flume and Avro. ? Developed various Map Reduce applications to perform ETL workloads on terabytes of data ? Expertise in working with HIVE data warehouse infrastructure-creating tables, data distribution by implementing partitioning and bucketing, writing and optimizing the HQL queries. ? Used Spark Streaming to divide streaming data into batches as an input to Spark engine for batch processing. ? Exploring with the Spark improving the performance and optimization of the existing algorithms in Hadoop. ? Experience in working with flume to load the log data from multiple sources directly into HDFS. ? Responsible to manage data coming from different sources and involved in HDFS maintenance and loading of structured and unstructured data. ? Good understanding of RDD operations in Apache Spark like Transformations&Actions, Persistence/ Caching, Accumulators, Broadcast Variables, Optimising Broadcasts. ? Hands on experience in performing aggregations on data using Hive Query Language (HQL). ? Developed MapReduce programs in java. ? Good experience in extending the core functionality of Hive and Pig by developing user-defined functions to provide custom capabilities to these languages. ? Proficient in designing and querying the NoSQL databases like HBase and MongoDB ? Performed Importing and exporting data into HDFS and Hive using Sqoop ? Experience in scheduling time driven and data driven Oozie workflows. ? Hands on experience in working with input file formats like parquet, json, Avro. ? Knowledge on Elastic Search Concepts. ? knowledge on visualizing the NOSQL data from elastic search using Kibana ? Worked on Extraction, Transformation, and Loading (ETL) of data from multiple sources like Flat files, XML files and Databases. ? Hands-on experience in J2EE technologies such as Servlets, JSP, EJB,

JDBC and developing Web Services providers and consumers using SOAP, REST. ? Used Agile Development Methodology and Scrum for the development process ? Good Knowledge in HTML, CSS, JavaScript and web based applications. ? Excellent analytical, problem solving and interpersonal skills. Ability to learn new concepts fast consistent team player with excellent communication skills.

Work Experience Spark Scala Developer Intel - Folsom, CA June 2018 to Present

Responsibilities: ? Developed a spark data ingestion framework to ingest data from various sources like SMB servers, FTP servers and Web to HDFS. ? CDR ingestion framework cleanses the metadata files related to actual data files before proceeding for the ingestion. CDR ingestion framework will only ingest the data based on the consent from the metadata to a appropriate folder in HDFS ? Current scope of metadata collection using Rest Client by consuming the service which is created at PEGA platform. ? Worked on Storing the Output of ingestion json in csv format. ? Worked on storing the copy of metadata to HBase. ? Worked on updating the records to Kibana. ? Developed Shell script code to trigger the spark job. ? Developed file watcher using shell script to perform the spark job, as the data gets in the input path. ? Developed Error mail code in Scala, which triggers mail to specified group when error occurs in processing the json file. ? Created logs in client and cluster mode. **Environment:** HDFS, Spark, Scala, cloudera, IntelliJ, Hue, Shell script, HBase, Kibana

Spark Developer Charter Communications September 2017 to May 2018

Responsibilities: ? Developed Apache spark jobs using Scala in test environment for faster data processing and used spark SQL for querying. ? Developed Scala code using specific monad pattern for different calculations based on the requirement. ? Developed and executed shell scripts to automate the jobs ? Written complex hive queries and automated using Azkaban for analyzing hourly calculations. ? Analyzed large amounts of data sets using Pig scripts and Hive scripts. ? Worked with Hue manager for developing hive queries and checking data in both development and production environments. ? Developed Pig Latin scripts for extracting data. ? Used Pig for data loading, filtering and storing the data. ? Worked on Data Integration from different source systems. ? Used Mongo DB for storing the data. ? Worked on retrieving data from Amazon kinesis streams and Amazon s3. ? Worked on creating topics in Kafka and integrating kinesis streams to Kafka and

storing the data in the HDFS using gobblin. ? Scheduled/Automate jobs using Azkaban. ? Developed and executed shell scripts to automate the jobs. ? Developed python code for creating fields in mongo. ? Used Jenkins for continuous automation of code. ? Complete end-to-end design of Apache NiFi to get connected to AWS and store the final output in HDFS. Environment: HDFS, Hive, Pig, Spark RDD, Spark-SQL, Kafka, Spark, Scala, Python, MongoDB, Horton works, IntelliJ, Azkaban, Ambari/Hue, Jenkins, Apache NiFi, Agile. Spark Developer T. Rowe Price - Owings Mills, MD September 2016 to August 2017 Responsibilities: ? Import the data from different sources like HDFS/HBase into Spark RDD. ? Importing and exporting data into HDFS and HIVE using Sqoop. ? Involved in gathering the requirements, designing, development and testing. ? Developed PIG scripts for source data validation and transformation. ? Designing and developing tables in HBase and storing aggregating data from Hive. ? Developed Spark scripts by using Scala shell commands as per the requirement. ? Developed Spark core and Spark-SQL scripts using Scala for faster data processing. ? Involved in code review and bug fixing for improving the performance. ? Implemented the workflows using Apache Oozie framework to automate tasks. ? Developing design documents considering all possible approaches and identifying best of them. ? Implemented Partitioning, Bucketing in Hive for better organization of the data ? Optimized Hive queries for performance tuning. ? Used Spark-Streaming APIs to perform necessary transformations and actions on the fly for building the common learner data model which gets the data from Kafka in near real time and Persists into Cassandra. ? Responsible for developing data pipeline with Amazon AWS to extract the data from weblogs and store in HDFS. ? Populated HDFS and Cassandra with huge data using Apache Kafka ? Basic knowledge on Machine Learning and Predictive Analysis. ? Data analysis using Spark with Scala. ? Analyze and report the data using Tableau. ? Create dashboards in Tableau. Environment: HDFS, Hive, Pig, Spark RDD, Spark Streaming, Spark-SQL HBase, Sqoop, Oozie, Kafka, Cassandra, Scala, Tableau,Agile. Hadoop Developer CareFirst - Owings Mills, MD January 2015 to September 2016 Responsibilities: ? Worked on importing data from various sources and performed transformations using MapReduce, hive to load data into HDFS. ? Worked on compression mechanisms to optimize MapReduce Jobs.

? Developed Big Data Solutions that enabled the business and technology teams to make data-driven decisions on the best ways to acquire customers and provide them business solutions.

? Created scripts to automate the process of Data Ingestion. ? Performed joins, group by and other operations in MapReduce by using Java and PIG. ? Configured Sqoop jobs to import data from RDBMS into HDFS using Oozie workflows. ? Worked on setting up Pig, Hive and HBase on multiple nodes and developed using Pig, Hive, HBase and MapReduce ? Worked on the conversion of existing MapReduce batch applications for better performance. ? Created HBase tables to store variable data formats coming from different portfolios ? Performed real time analytics on HBase using Java API and Rest API ? Implemented HBase Co-processors to notify Support team when inserting data into HBase Tables ? Worked on compression mechanisms to optimize MapReduce Jobs ? Analyzed the customer behavior by performing click stream analysis and to ingest the data used flume ? Experienced with working on Avro Data files using Avro Serialization system ? Implemented business logic by writing UDF's in Java and used various UDF's from Piggybanks and other sources ? Continuous monitoring and managing the Hadoop cluster using Cloudera Manager ? Exported the analyzed data to the relational databases using Sqoop for visualization and to generate reports for the BI team

Environment: HDFS, Hive, MapReduce, Pig, Sqoop, RDBMS, HBase, Java API, REST API, Cloudera, AVRO, Flume.

Hadoop Developer EOG

Resources - Houston, TX June 2013 to December 2014 Responsibilities:

? Installed and configured Apache Hadoop to test the maintenance of log files in Hadoop cluster ? Installed and configured Hive, Pig, Sqoop, and Oozie on the Hadoop cluster ? Installed Oozie Workflow engine to run multiple Hive and Pig Jobs ? Developed multiple MapReduce jobs in Java for data cleansing and preprocessing ? Developed Simple to complex Map/Reduce Jobs using Hive and Pig ? Involved in loading data from UNIX file system to HDFS ? Evaluated business requirements and prepared detailed specifications that follow project guidelines required to develop written programs ? Analyzed large amounts of data sets to determine optimal way to aggregate and report on it ? Provided quick response to ad hoc internal and external client requests for data and experienced in creating ad hoc reports ? Responsible for building scalable distributed data solutions using Hadoop

? Migration of ETL processes from Oracle to Hive to test the easy data manipulation ? Performed optimization on Pig scripts and Hive queries increase efficiency and add new features to existing code ? Developed PIG Latin scripts for the analysis of semi structured data ? Developed Hive queries to process the data ? Used Hive and created Hive tables and involved in data loading and writing Hive UDFs ? Used Sqoop to import data into HDFS and Hive from other data systems ? Installed Oozie workflow engine to run multiple Hive ? Continuous monitoring and managing the Hadoop cluster using Cloudera Manager ? Conducted some unit testing for the development team within the sandbox environment. Environment: Hadoop Cluster, Hive, Pig, Sqoop, Oozie, Oracle, Cloudera Manager, UNIX, ETL, Mongo DB, Agile. Hadoop Developer Merck, Bluebell, PA April 2012 to May 2013 Responsibilities: ? Installed and configured Hadoop MapReduce, HDFS, developed multiple MapReduce jobs in Java for data cleaning and preprocessing. ? Good understanding and related experience with Hadoop stack - internals, Hive, Pig and Map/Reduce ? The system was initially developed using Java. The Java filtering program was restructured to have business rule engine in a jar that can be called from both java and Hadoop. ? Wrote MapReduce jobs to discover trends in data usage by users. ? Involved in defining job flows ? Involved in managing and reviewing Hadoop log files ? Load and transform large sets of structured, semi structured and unstructured data ? Responsible to manage data coming from different sources ? Supported Map Reduce Programs those are running on the cluster ? Involved in loading data from UNIX file system to HDFS. ? Responsible to manage data coming from different sources. ? Installed and configured Hive and developed Hive UDFs to extend core functionality of hive ? Involved in creating Hive tables, loading with data and writing hive queries which will run internally in map reduce way ? Implemented Partitioning, Dynamic Partitions, Buckets in HIVE. ? Monitor System health and logs and respond accordingly to any warning or failure conditions. Environment: Apache Hadoop, HDFS, Map Reduce, Pig, Hive tables, Hive UDFs, UNIX, Java, ETL, Eclipse. Java Developer Gimmel - New York, NY December 2010 to March 2012 Responsibilities: ? Designed the application using Agile Methodology. ? Developed Maven based project structure having data layer, ORM, and Web module. ? Developed MVC framework based website using JSF and spring. ?

Designed and Developed HTML pages and JSP pages. ? Developed Business components using spring framework and database connections using JDBC. ? Responsible for creating tables of client's information in and writing Hibernate mapping files to manage one-to-one and one-to-many mapping relationships. ? Implemented data reading, saving and modification by stored procedures in MySQL database and Hibernate criteria. ? Developed Graphical User Interfaces by using JSF, JSP, HTML, CSS, and JavaScript. ? Installation and configuration of Development Environment using Eclipse with WebLogic Application server. ? On the server side, post the access to the application and provided result on the network using RESTfulweb service. ? Developed the XMLGateway to help the ordering process system communicate with the Order Execution Tool and different online tools such as Line Qualification, Billing Information and Credit Card Validation Systems. ? Used NodeJS to develop scalable web application. ? Development (TDD Approach) environment using Agile methodologies. ? Used JUnit to test, debugged and implemented the application. ? Implemented payment gateway using PayPal. ? Developed build script using MAVEN to build, package, test and deploy application in application server. ? Auditing tool is implemented by using log4j. Environment: MVC, HTML, JSP, XML, Maven, JavaScript, Node.js, TDD, Junit, log4j, JDBC, MYSQL, Hibernate, WebLogic, JSF, Spring, Eclipse. Java Developer Metlife - Bridgewater, NJ October 2009 to November 2010 Responsibilities: ? Designing the Use Case Diagrams, Class Model, Sequence diagrams, for SDLC process of the application. ? Implemented GUI pages by using JavaScript, HTML, JSP, CSS, and AJAX. ? Designed and developed UI components using JSP, JMS, JSTL. ? Deployed project on Web Sphere application server in Linux environment. ? Implemented the online application by using Web Services (SOAP), JSP, Servlets, JDBC, and Core Java. ? Implemented Singleton,DAO Design Patterns, factory design pattern based on the application requirements. ? Used DOM and SAX parsers to parse the raw XML documents. ? Tested the web services with SOAP UI tool. ? Developed back end interfaces using PL/SQL packages, stored procedures, Functions, Procedure, Anonymous PL/SQL programs, Cursor management, Exception Handling in PL/SQL programs. ? Tuning complex database queries and joining the tables to improve the performance of the application. ? Used

Eclipse as Development IDE for web applications. Environment: JDBC, HTML, CSS, JSP, AJAX, XML, SOAP, DOM, SAX, PL/SQL, Eclipse, SOAP, Servlets. Java Developer H-Line Soft - Hyderabad, Telangana June 2008 to September 2009 Responsibilities: ? Involved in the complete SDLC software development life cycle of the application from requirement gathering and analysis to testing and maintenance. ? Worked with the business community to define business requirements and analyze the possible technical solutions. ? Requirement gathering, Business Process flow, Business Process Modeling and Business Analysis. ? Implemented the User Login logic using Spring MVC framework encouraging application architectures based on the Model View Controller design paradigm ? Used various Java, J2EE APIs including JDBC, XML, Servlets, and JSP. ? Generated Hibernate Mapping files and created the data model using mapping files ? Developed UI using JavaScript, JSP, HTML and CSS for interactive cross browser functionality and complex user interface ? Developed action classes and form beans and configured the struts-config.xml ? Provided client-side validations using Struts Validator framework and JavaScript ? Created business logic using servlets and session beans and deployed them on Apache Tomcat server ? Created complex SQL Queries, PL/SQL Stored procedures and functions for back end ? Prepared the functional, design and test case specifications ? Performed unit testing, system testing and integration testing ? Used JUnit for unit testing of the application ? Provided Technical support for production environments resolving the issues, analyzing the defects, providing and implementing the solution defects. Resolved more priority defects as per the schedule Environment: SDLC, Spring MVC, JSP, Servlets, JavaScript, SQL, HTML, CSS, PL/SQL, Hibernate, Junit. Education Bachelors in Computer Science in Computer Science JNTU 2008 Skills HDFS (6 years), Hive (6 years), Pig (6 years), JAVA (8 years), APACHE (5 years), Drupal, Servicenow, Hadoop, AEM, Cobol Additional Information TECHNICAL SKILLS Hadoop/Big Data Hadoop MapReduce, HDFS, Hive, Pig, Sqoop, Flume, Oozie, Zookeeper, Spark, Kafka Languages and Web Technologies C, C++, C#, Scala, XML, HTML, CSS, JavaScript, J2EE, Java, Python, JSP Frameworks Spring , struts, Hibernate, Servlets Web Services REST, SOAP Databases MY SQL, Oracle, DB2, MongoDB, Spark-SQL, HBase Tools Tableau, Weka Application servers Apache Tomcat, WebLogic 8.0 IDE/Modelling

Tools Eclipse, IntelliJ Development Methodologies Agile, Waterfall Logging tools Log4j Operating Systems Windows 7/8/10, LINUX

Name: Joseph Novak

Email: brandonmcdaniel@example.com

Phone: 4409367783