

Sr. Big Data Developer Sr. Big Data Developer Sr. Big Data Developer - Nordstrom, Inc Seattle, WA

8 years of working experience in IT Industry including Big Data and Java/J2EE Worked in various domains included E-commerce, telecommunication, finance and investment Strong experience in Hadoop 2.3 Eco system technologies such as in HDFS, YARN, Map Reduce, Spark 1.3+, Hive 1.0+, Pig 0.14+, HBase 0.98, Zookeeper 3.4+, Oozie 3.0+, Flume 1.3+, Sqoop 1.3+, Impala 1.2+ and Kafka 1.2+ Experienced in building, maintaining multiple Hadoop clusters of different sizes and configuration and setting up the rack topology for large clusters also in Hadoop Administration/Architecture/ Developer Developed Java 6+ and Scala 2.10+ applications on Hadoop and Spark Streaming for high-volume and real-time data processing Expertise in Spark Streaming and Spark SQL with Scala Exceptional skills with NoSQL databases such as HBase and Cassandra Experienced in writing Sqoop, HiveQL and Pig scripts as well as the UDFs to do ELT process Utilized Kafka, RabbitMQ and Flume to gain real-time data stream and save it in HDFS and HBase from the different data sources Experience other Hadoop ecosystem tools in jobs such as ZooKeeper, Oozie, Implala Experience in all the phases of Data warehouse life cycle involving requirement analysis, design, coding, testing, and deployment Strong in core Java, data structure and Java components like Collections Framework, Exception handling, I/O system, and Multithreading Earned designation with J2EE development by using frameworks such as Spring MVC and Hibernate 3 & 4, as well as using Web services such as SOAP and REST Familiar in Agile/Scrum Development and Test Driven Development (TDD) Extensive Experienced in Unit Testing with JUnit, MRUnit, Pytest Worked in development environment tools such as Git, JIRA, Jenkins, Agile/Scrum and Waterfall Self-motivated, teamwork, working under high pressure, working in several projects simultaneously, dynamic problem-solving, adept at working with minimal to supervision, highly-caliber teams of professionals Authorized to work in the US for any employer

Work Experience Sr. Big Data Developer Nordstrom, Inc - Seattle, WA February 2017 to Present

Project: Customer Behaviors Analysis System (CBAS) Nordstrom, Inc. is a leading fashion specialty retailer offering compelling clothing, shoes and accessories for men, women and children. Nordstrom needs to maximize profits by reducing the overstock products. Therefore, the company

decides to analyze the historical shopping transactions from the customers. The purpose of this project is to process and analyze the transactions (etc. user's shopping histories), and eventually for data science team to generate reports. Our team is responsible to process the data by using big data technologies, with data modeling, data transformation, data analyzing and data visualization.

Responsibilities: Design and build scalable infrastructure and platform for very large amounts of data ingestion, aggregation, integration and advanced analytics in Hadoop, including Map Reduce, Spark, Hive, HBase, Pig Design the HBase schemes based on the requirements and HBase data migration and validation Design and implement the HBase query APIs in Java for BI teams Apply Spark Streaming to receive data from Kafka to do the continually data cleaning and aggregating, then store the data in HBase Work on the core and Spark SQL modules of Spark Streaming extensively Write customized Spark SQL UDFs in Scala and Java Load the data from different sources such as HDFS or HBase into Spark RDD and implement in memory data computation to generate the output response Expertise at designing tables in Hive, MySQL using Sqoop and processing data like importing and exporting of databases to the HDFS ETL: Data extraction, managing, aggregation and loading into HBase Involve in converting Hive/SQL queries into Spark transformations using Spark RDDs and Scala to improve the performance Develop multiple POCs using Scala and deploy on the Yarn cluster, compared the performance of Spark, with Hive and SQL Develop predictive analytic using Apache Spark Scala APIs HBase, Spark Streaming performance tuning Configure Zookeeper to coordinate and support Kafka, Spark, Spark Streaming, HBase and HDFS Environment: Hadoop 2.6, HDFS, YARN, Spark 1.5, Spark SQL, Spark Streaming, Zookeeper 3.5, HBase 0.98, Cassandra, Hive 1.2, Kafka 1.3, RabbitMQ 3.4.4, Oozie 4.1, Eclipse, Scala, UNIX Shell Scripting. Sr. Big Data Developer Avvo, Inc - Seattle, WA March 2016 to January 2017 Project: sWIG3 Avvo was founded in Seattle, Washington by tech-savvy lawyer Mark Britton to make legal easier and help people find a lawyer. Avvo need to enhance the existing metric table because the requirements from business are updated. In this project, we need to load the raw data from MySQL to do delta difference. Then we save the historical data in to source tables which will be used into dimension layer. Finally, the tables by

applied business logics are save in the dimension layer. Responsibilities: Develop big data applications in CDH platform and maintain the data processing workflows Design the Hive schemes based on the requirements and Hive data migration and validation Apply Sqoop to extract data from MySQL to do the continually data cleaning and aggregating, then store the data in HDFS Work on creating HiveQL scripts to process and load data into Hive tables Experienced performing data quality investigations writing interactive ad-hoc queries using Spark shell Expertise at designing tables in Hive, MySQL using Sqoop and processing data like importing and exporting of databases to the HDFS ETL: Data extraction, managing, aggregation and loading into Hive Extensive experience in data monitoring and integrating existing monitoring to maintain data workflows via DataDog Develop POC using Hive and deploy on the Test cluster, compared the performance of the old data architecture Develop predictive analytic using Impala and check the data consistency Develop workflow automation in Oozie and define coordinator to connect the upstreaming workflow and generate trigger files for downstreaming workflow Environment: Hadoop 2.6, MySQL, HDFS, YARN, Impala, DataDog, Spark 1.6, Spark SQL, Zookeeper 3.5, Hive 1.2, Kafka 1.3, Oozie 4.1, Eclipse, Scala, Python, UNIX Shell Scripting. Senior Data Engineer Groupon - Seattle, WA January 2015 to February 2016 Project: Shopping Discount System (SDS) Groupon is an American worldwide e-commerce marketplace connecting millions of subscribers with local merchants by offering activities, travel, goods and services. The Groupon is providing services, which connects the suppliers and users with discount coupons of different products. This project focused on analyzing the feedbacks from the suppliers based on the customers shopping behaviors, and provides the best discount ranges of different coupons. Our team worked on data ingestions from varieties data sources, processing data with Hadoop related solutions, and worked for designing statistical models for analyzing data with data science team. Responsibilities: Analyzed large data sets by running custom MapReduce and Hive queries Designed the HBase schemes based on the requirements Designed Cassandra schemes and connect it by Spark HBase data migration and validation Assisted in exporting analyzed data to relational databases using Sqoop Utilized Kafka and RabbitMQ to capture the data stream Processed data by using

Spark Streaming and Kafka then stored the results in HBase    Worked on the core and Spark SQL modules of Spark extensively    Loaded the data from different source such as HDFS or HBase into Spark RDD and do in memory data computation to generate the output response    Implemented POC to migrate MapReduce jobs into Spark RDD transformation using Scala IDE for IntelliJ    Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs and Scala to improve the performance    Developed predictive analytic using Apache Spark Scala APIs    Scheduled workflow with Oozie    Environment: Hadoop 2.6, HDFS, Spark 1.5, Spark SQL, Spark Streaming, Zookeeper 3.5, HBase 0.98, Hive 1.2, Kafka 1.3, RabbitMQ 3.4.4, Oozie 4.1, IntelliJ, Scala, UNIX Shell Scripting. Senior Data Engineer Disney ABC Television Group - Seattle, WA September 2013 to December 2014 The Disney/ABC Television Group is composed of The Walt Disney Company's global entertainment and news television properties, owned television stations group, as well as radio and publishing businesses. The Disney/ABC User Tracking Platform focused on development for the Hadoop related solutions and services to achieve better advertisement strategy. We provide data processing and analytic solutions including streaming data ingestion, RDBMS and log integration, data transformation/cleaning and data modeling.

Responsibilities:    Worked on Amazon EMR 4.x with Agile methodology    Developed Kafka consumer to receive and store real time data from Kafka to Amazon S3    Used Flume to collect, aggregate, and store web log data from different sources    Used Sqoop to transfer data between RDBMS and HBase    Extracted data from MongoDB through MongoDB Connector for Hadoop    Involved in migrating of MapReduce programs into Spark using Spark and Scala    Scheduled workflow with Oozie    Use Spark with Scala and Spark SQL for testing and processing of data    Worked with analytics team to build statistical model with MLlib and PySpark    Worked with analytics team to prepare and visualize tables in Tableau for reporting    Performed unit testing using JUnit and Pytest    Used Git for version control, JIRA for project tracking and Jenkins for continuous integration    Environment: Hadoop 2.6, Cloudera CDH 5.4, HDFS, MapReduce, HBase, Sqoop, Flume, Zookeeper, MongoDB, Spark 1.4, Spark SQL, Pyspark, MLlib, Tableau 9.2, JUnit, Pytest Jr.

Data Engineer Starbucks - Seattle, WA June 2012 to August 2013 Starbucks Corporation is an

American coffee company and coffeehouse chain. The team contributes to detect, mitigate, and prevent loss and quickly and helps to easily identify fraudulent merchants to reduce risk. Greater precision in search results and prioritization of the results is helping customers more quickly and easily evaluate a merchant, without the cost of executing additional searches or further investigations. Responsibilities: Responsible for building scalable distributed data solutions using Hadoop Handled importing of data from various data sources, performed transformations using Hive, MapReduce, loaded data into HDFS and Extracted the data from MySQL into HDFS using Sqoop Installed and configured Hive, Pig, Sqoop, Flume and Oozie on the Hadoop cluster Developed simple to complex Map/Reduce jobs using Java, and scripts using Hive and Pig Analyzed the data by performing Hive queries (HiveQL) and running Pig scripts (Pig Latin) for data ingestion and egress Implemented business logic by writing UDFs in Java and used various UDFs from other sources Experienced on loading and transforming of large sets of structured and semi structured data Managing and Reviewing Hadoop Log Files, deploy and Maintaining Hadoop Cluster Export filtered data into HBase for fast query Environment: Hadoop, HBase, Hive, Pig, Map Reduce, Sqoop, Oozie, Eclipse, Java Jr. Java/J2EE/Big Data Developer UHG(Optum) - Minnetonka, MN July 2011 to May 2012 UnitedHealth Group Inc. is an American for-profit managed health care company based in Minnetonka, Minnesota. It is sixth in the United States on the Fortune 500. UnitedHealth Group offers health care products and insurance services. Responsibilities: Developed Map Reduce programs in Java to help the analysis team to read, write, delete, update and analyze the big data Developed a series of Map Reduce algorithms to increase the performance Developed Web pages using Struts view component JSP, JavaScript, HTML, jQuery, AJAX, to create the user interface views migration 3rd party applications Java Hibernate is used to access the database within the system Environment: Eclipse, Java, Map Reduce, Algorithm, Hibernate, Struts, JavaScript, JQuery, CSS, HTML, MySQL and XML Jr. Java Developer BBVA Compass - Birmingham, AL April 2009 to June 2011 This is an intranet application for stock screening. This application is used by the customers and other client based users to build their query and send the result to the clients they cover. The application has alerting capability, which will

alert users according to the criteria and alert level set for a stock in the form chosen. The result set built is saved in downloadable, viewable formats as preferred by the clients. Responsibilities: Involved in system design, which is based on Spring Struts Hibernate framework. Implemented the business logic in standalone Java classes using core Java. Developed database (MySQL) applications. Worked in Hibernate Template to access the MySQL database. Involved in Unit testing of the components and created unit test cases and did unit test review. Environment: Eclipse, MySQL Client 4.1, Spring, HTML, JavaScript, Hibernate, JSF, Junit, SDLC: Agile/Scrum Education Bachelor's Skills APACHE HADOOP OOZIE (6 years), APACHE HADOOP SQOOP (6 years), Hadoop (6 years), HADOOP (6 years), OOZIE (6 years) Additional Information TECHNICAL SKILLS: Hadoop Eco Systems \ Web Technologies\ Hadoop 2.7.0+, MapReduce, HBase 0.98\ SOAP, REST, JSP 2.0, JavaScript 1.8\ Spark 1.6+, Hive 1.0+, Pig 0.14+, Kafka 1.2+\ Servlet 3.0, HTML 5, CSS 3\ Sqoop 1.3+, Flume 1.3+, Impala 1.2+, Oozie \ 3.0+, Zookeeper 3.4+\ NoSQL\ Java Frameworks\ HBase 0.98, Cassandra 2, MangoDB 3\ spring MVC, Hibernate 3 & 4\ Programming Languages\ Others Java, Scala, SQL, SparkSQL\ JSON, AVRO, XML \ HiveQL, Pig-Latin, C++, C, Python\ RabbitMQ 3.0+\

Name: Jesus Jones

Email: westrada@example.net

Phone: 301.424.5843x86319