

Big Data Engineer Big Data Engineer Big Data Engineer - Hadoop Houston, TX Over 10 years of professional IT experience including 5+ years of strong experience working on Apache Spark, BigData and Hadoop ecosystem. 5 years of strong end-to-end experience in Python, Java Programming involved in Design, development and implementing various web-based applications using Python and Java Technologies. ? Hands on experience in developing and deploying enterprise-based applications using major components in Hadoop ecosystem like Hadoop 2.x, YARN, Hive, Pig, Spark, Map Reduce, Impala, Kafka, Oozie, HBase, Flume,Sqoop and Zookeeper. ? Involved in all phases of Software Development Life Cycle (SDLC) and Worked on all activities related to the development, implementation, administration and support for Hadoop. ? Experience in Python programming language for framework and core java concepts ? Experience with on-prem (HortonWorks, MapR) and Google Cloud Platform. ? Experience in monitoring, tuning and administrating Hadoop cluster. ? Experience in understanding Big Data business requirements and providing them Hadoop based solutions. ? Experience in Importing and exporting data from different databases like MySQL, Oracle, Teradata into HDFS using Sqoop. ? Worked on Spark 1.6.0 for data processing using RDD's and Dataframe API. ? Experience in writing UDF'S in Hive for processing and analyzing large datasets. ? Experience in working with different file formats and compression techniques in Hadoop. ? Experience in using NFS (Network File Systems) for backing up Name node metadata. ? Experience in managing the cluster resources by implementing fair scheduler and capacity scheduler. ? Experience in developing Pig Latin scripts for data processing on HDFS. ? Excellent team player with good communication skills and effective time management. ? Understand business process management and business requirements of the customers and translate them to specific software requirements. ? In-depth understanding of Spark Architecture including Spark Core, Spark SQL, Spark Streaming. ? Experience in using Scala to convert Hive/SQL queries into RDD transformations in Spark. ? Strong knowledge of real time data analytics using Spark Streaming, Kafka & Flume. ? Proficient knowledge with kafka and spark with YARN Local & Standalone modes. ? Expertise in writing Spark RDD transformations, Actions, Case classes for input data and performing data ? transformations using Spark-Core ?

Implementing Scheduler using Azkaban, Tidal Enterprise scheduler, Crontab and Oozie. ?  
Experience in using DStreams, Broadcast Variables, RDD caching for Spark Streaming. ?  
Improving the performance and optimizing existing algorithms in Hadoop using Spark context,  
Spark-SQL, ? DataFrames, Pair RDD's & Spark YARN. ? Hands on experience with ORC, AVRO,  
Sequence and Parquet file formats. ? Experience in analyzing data using PIG Latin, HiveQL, Spark  
SQL ? Experience with Hadoop Distributions like Cloudera and Hortonworks. ? Extensive  
knowledge on designing Hive Managed/External tables, Views & Hive Analytical functions. ?  
Experience in tuning the performance of hive queries using Partitioning and Bucketing. ?  
Experience working with FLUME to handle large volume of streaming data ingestion. ? Experience  
in developing customized UDFs and UDAFs to extend core functionality of PIG and Hive. ?  
Experience in various Big Data application phases like Data Ingestion, Data analytics and Data  
visualization. ? Proficient in working with NoSQL databases such as HBase and MongoDB. ?  
Expertise in writing pig and hive queries for analyzing data to meet business requirements. ?  
Experience in design and pipeline flows with Jenkins, Tonomi and Azkaban. ? Exposed to build  
tools like MAVEN, SBT and bug tracking tool JIRA in the work environment. ? Good Knowledge in  
scheduling Job/Workflow and monitoring tools like Azkaban and Cisco Tidal Scheduler. ? Hands on  
Experience in Importing/Exporting Data from RDBS to HDFS using SQOOP. ? Excellent  
programming skills at high level abstraction using Java, Scala, Python & SQL. ? Co-ordinate patch  
upgrades, bug fixes and new releases for the application within stipulated timelines ? Performing  
Team Lead Activities and Coordination with the team members and defining time estimations for  
deliverables of change requests, patches and upgrades to the application. Authorized to work in the  
US for any employer Work Experience Big Data Engineer Hadoop - Houston, TX June 2017 to  
Present Chevron is one of the largest oil and gas companies of standard oil based in Houston, TX.  
The main goal of this project is to process and analyze the logs of data generated in chevron gas  
stations. This is a project in downstream used to analyze the logistics, usage of fuel in different gas  
stations, consumer's sentimental analysis. Data is imported into HDFS from DB2, Oracle and  
Exadata. Transformations are applied to create a single merge and flattened view, As a Hadoop

developer was responsible for Designing, Developing and automating several ETL workflows and analytics Use cases. Responsibilities: ? Developed Map Reduce jobs in java for data cleansing and preprocessing. ? Moving data from DB2, Oracle Exadata to HDFS and vice-versa using SQOOP. ? Collecting and aggregating large amounts of log data using Apache Flume and staging data in HDFS for further analysis ? Worked with different file formats and compression techniques to determine standards ? Developed hive queries and UDFS to analyze/transform the data in HDFS. ? Developed hive scripts for implementing control tables logic in HDFS. ? Designed and Implemented Partitioning (Static, Dynamic), Buckets in HIVE. ? Developed Pig scripts and UDF's as per the Business logic. ? Developed user defined functions in pig using Python. ? Analyzing/Transforming data with Hive and Pig. ? Developed Oozie workflows and they are scheduled through a scheduler on a monthly basis. ? Designed and developed read lock capability in HDFS. ? Implemented Hadoop Float equivalent to the DB2 Decimal. ? Involved in End to End implementation of ETL logic. ? Effective coordination with offshore team and managed project deliverable on time. ? Worked on QA support activities, test data creation and Unit testing activities.

Environment: CDH, Hadoop, HDFS, MapReduce, Hive, Sqoop, Pig, XML, ETL, DB2 and QA Big Data Engineer Cisco - San Jose, CA January 2017 to May 2018 Cisco Systems, Inc. (known as Cisco) is an American multinational technology conglomerate headquartered in San Jos , California, in the center of Silicon Valley, that develops, manufactures, and sells hardware, telecommunications, and other high-technology services and products.[4] Through its numerous acquired subsidiaries, such as OpenDNS, WebEx, and Jasper, Cisco specializes into specific tech markets, such as Internet (IoT), domain security, and energy management. Responsibilities: 1) Monitoring Activities: ? Review Tidal Enterprise Scheduler jobs, verify if all the jobs are triggered at appropriate time and completed successfully ? Investigate, RCA and fix the issue for any failed jobs. ? Develop Scripts using Python for datalake framework ? Assist project teams for their queries on DataLake tables, ? Investigate/Fix if there are any data-mismatch issues identified or raised by the project teams ? Coordinate with project teams for any datalake related changes and take initiatives like standardization of naming standards ? Any adhoc table refresh requests and

assist project teams for their testing, validating during GO Live ? We co-ordinate and help to fix the issues with data catalogue, when it goes down ? Assist project teams for the approval's and access management related queries to the tables ? If there are any PV's to be loaded with huge data, we load the data using TPT or with max\_update\_date and ensure for its success ? Coordinate with Teradata DBA team and Teradata support team for any TPT related queries/ issues ? Worked on Scala 2.10 jobs using Spark 1.6.0 for data processing using RDD's and Dataframe API ? Performance tuning of Spark and Sqoop Job

2) Ingestion Activities ? Coordinate with project teams for their ingestion requests ? Review all the metadata provided by them, database credentials, Incremental columns and uniqueness of merge keys. ? Discuss with Architect team for their approval to proceed with metadata creation ? Create metadata and get It reviewed by Architect team ? Load the tables, take the audits and share it with the teams. ? Scheduling jobs in TES and validate for their success ? Ingested data from AWS cloud to Hadoop datalake

3) Enterprise data and Analytics - In the process of automating Hadoop ? Involved in requirement gathering, discussions with project teams, Joint Application Design meetings, Build/update user stories ? Validate requirements, Review screen design prototypes, test the functionalities, participate in User Acceptance Testing and actively involved in GO live activities

Environment: CDH, Hadoop, HDFS, MapReduce, Hive, Sqoop, Pig, Jira, XML, DB2 and QA. Big Data Engineer Kohl's South Digital Center - Milpitas, CA April 2015 to December 2016 Kohl's Digital Center in Milpitas, California, puts Kohl's in the fast-paced technology industry and in close proximity to key technology partners. The creative incubator is dedicated to seek and implement the latest and greatest technology to revolutionize the customer experience. Kohls is one of the leading retail chain markets. Develop, Support and handle fixes related to project tracks in on-prem and GCP. Involve in Business interactions, client management and coordination with the team. Responsibilities: ? Being in Retail domain - Online and stores data - Need to be Agile, Aggressive and focused team player in monitoring, debugging and Issue fix. Any delay may impact propensities for top trending ? As a center of Excellence team, Involve in any of the application issues, triage/investigate them, build and fix the issues ? Using GCP Console, monitor dataproc cluster and jobs. Stack Driver to

monitor Dashboards and do a performance tuning and optimization of jobs which are memory intensive and provide L3 support for the applications in production environment ? DAS - Data as a Service, in GCP - google storage buckets. ? Monitor Azkaban jobs in on-prem (Hortonworks distribution) and GCP (Google Cloud Platform). ? Investigate, RCA and fix the issue for any failed jobs. ? As part of Production Engineering team, keep the environment healthy 24/7. ? Monitor Azkaban, Ambari, splunk and Net Diagnostics for Hbase Timeouts, Propensities. ? GCP - Stackdriver for monitoring, logging, compute engine and dataproc. ? Involve in discussion with project teams, understand the issue, Investigate, RCA and fix the issue ? Handle production issues assigned in JIRA tasks, Incidents / Requests through Remedy. ? Involve in discussions, take initiatives and work with Application teams for the smooth flow of projects. ? Involve in all the Technical Discussions and scrum meetings. ? Reviews project documents) received from other Technical Specialists, Business Technical Specialists, and Project Managers to ensure quality, completeness, and adherence to documentation standard. Environment: Hadoop, Java, HDFS, Jira,Azkaban, MapReduce, Hive, Sqoop, Pig, XML, ETL, DB2 and QA Java/Hadoop Developer Samsung America Inc - San Diego, CA February 2013 to March 2015 Samsung's solutions build on the company's 35 years' experience in developing and commercializing some of the world's most popular communication services. Based on its wide range of field-experience in any technology, Samsung has already successfully partnered with dozens of mobile service providers around the world. Responsibilities Installation & configuration of a Hadoop cluster along with Hive. Developed Map Reduce application using Hadoop Map reduce programming, a framework for processing. Large data sets in parallel across the Hadoop cluster for pre-processing. Developed the code for Importing and exporting data into HDFS and Hive using Sqoop. Responsible for writing Hive Queries for analyzing data in Hive warehouse using Hive Query Language (HQL). Responsible for designing Front end system using HTML, CSS, JSP, Servlets and Ajax. Transformed web application into compatible Mobile & Tablet application by designing responsive designs using HTML & CSS. Used LDAP for user Authentication and authorization. Created Stored Procedures, Views, Cursors and functions to support application. Involved in defining job

flows using Oozie for scheduling jobs to manage apache Hadoop jobs by directed. Acyclic graph (DAG) of actions with control flows. Developing Hive User Defined Functions in java, compiling them into jars and adding them to the HDFS and executing them with Hive Queries. Experienced in managing and reviewing Hadoop log files. Responsible to manage data coming from different sources. Assisted in monitoring the Hadoop cluster using Ganglia tool. Dealing with high volume of data in the cluster. Tested and reported defects in an Agile Methodology perspective. Consolidate all defects, report it to PM/Leads for prompt fixes by development teams and drive it to closure. Installed and configured Hadoop Cluster (3 Node Cluster) in fully distributed mode. Installed hadoop ecosystems(Hive, Pig, Sqoop, HBase, Oozie) on top of hadoop cluster Importing data from Oracle to HDFS & Hive for analytical purpose. Analyzing imported data in HDFS & Hive using HiveQL and custom Map Reduce programs in Java Environment: Java, CDH, Hadoop, HDFS, Map Reduce, Hive and Sqoop JAVA Developer Value Labs July 2009 to December 2012 Project Description: E-order management project is part of initiatives to harness and leverage on the advancement of IT for better Health. It will enable medical practitioners to electronically order supplies, prescribe medicine and requisite services such as laboratory, radiology, referral and other clinical and therapy services. Responsibilities: ? Responsible and active in the analysis, design, implementation and deployment of full software development ? lifecycle (SDLC) of the project. ? Designed and developed user interface using JSP, HTML and JavaScript. ? Developed JSP Custom Tag Libraries for Tree Structure and Grid using Pagination Logics. ? Worked extensively with JSP's and Servlets to accommodate all presentation customizations on the front end.Used Building tools like Maven to build, package, test and deploy application in the application server. ? Developed Struts action classes, action forms and performed action mapping using Struts framework and ? Performed data validation in form beans and action classes. ? Extensively used Struts framework as the controller to handle subsequent client requests and invoke the ? Model based upon user requests. ? Defined the search criteria and pulled out the record of the customer from the database. ? Make the required changes and save the updated record back to the database. ? Validated the fields of user registration screen and login screen by writing JavaScript

validations. ? Involved in developing and coding the Interfaces and classes required for the application and created ? appropriate relationships between the system classes and the interfaces provided. ? Developed build and deployed scripts using Apache ANT to customize WAR and EAR files. ? Used DAO and JDBC for database access. ? Developed stored procedures and triggers using PL/SQL to calculate and update the tables to implement ? business logic. ? Used SVN to maintain source and version management. ? Using JIRA to manage the issues/project work flow. ? Involved in peer code reviews and performed integration testing of the modules. Followed coding and ? documentation standards. ? Involved in post-production support and maintenance of the application. Environment: Oracle, Java, Struts, Servlets, HTML, XML, SQL, J2EE, JUnit, Tomcat, PL/SQL, JIRA, SVN. Education Master's Skills Ambari, Hdfs, Sqoop, Hbase, Apache kafka, Kafka, Flume, Hadoop, Teradata, Apache spark, C++, Hadoop, Hbase, Hive, Javascript, Json, Perl, Pig, Scripting, Zookeeper Additional Information Technical Skills: Hadoop Ecosystem Development HDFS, Map-Reduce, Hive, Pig, TES, HBase, Sqoop, Zookeeper, Spark, MCS Azkaban, Ambari Hadoop Distribution Framework MapR , HortonWorks Cloud Technologies Google Cloud Platform Languages Java, C , C++ Technologies Scripting Shell Script, Perl, JavaScript and PowerShell Hadoop Ingestion Apache Sqoop, Apache Kafka, Apache spark, Apache Flume, Storm. Database Mongo Database, Oracle 10g, Oracle 11g, MySQL, Teradata, Hbase, Netezza. Operating Systems Linux, Unix, Windows Development Tools Eclipse, Putty, Tectia Java Technologies JSON,JDBC,AJAX

Name: Tammy Stuart

Email: mary88@example.net

Phone: 225.792.2435