Big Data Engineer Big Data Engineer Big Data Engineer - American Express   Having around 6 years of total IT experience with over 4years' experience in Big Data Hadoop, 2 years of experience in Development and Design of Java based enterprise applications.   Extensive working experience on Hadoop eco-system components like HDFS, MapReduce, Hive, Pig, Sqoop, Flume, Spark, Kafka, Oozie and Zookeeper. Implemented performance tuning techniques for Hive queries. Strong knowledge on Hadoop HDFS architecture, Map-Reduce (MRv1) and YARN (MRv2) framework.   Expert in working with Hive tool creating tables, data distribution by implementing partitioning and bucketing, writing and optimizing the HiveQL queries.   Strong hands on Experience in publishing the messages to various Kafka topics using Apache NIFI and consuming the message to HBase and MySql tables using Spark and Scala.   Worked on creating Spark jobs that process the true source files and successful in performing various transformations on the source data using SparkDataframe/DataSet, Spark SQL API's.   Developed Sqoop scripts to migrate data from Teradata, Oracle to Bigdata Environment.   Worked with Hue GUI in scheduling jobs with ease and File browsing, Job browsing, Metastore management.   Experience in importing and exporting the data using Sqoop from HDFS to Relational Database systems and vice-versa. Hands on experience in installation, configuration, supporting and managing Hadoop Clusters using Apache, Cloudera (CDH3, CDH4), Yarndistributions (CDH 5.X)   Performed Data scrubbing and processing with Oozie and for workflow automation and coordination.   Good knowledge in EMR (Elastic Map Reducing) to perform big data operations in AWS.Knowledge in working with Amazon Web Services (AWS) using EC2 for computing and S3 as storage mechanism.   Hands on experience in analyzing log files for Hadoop and eco-system services and finding root cause. Hands on experience on handling different file formats like AVRO, PARQUET, Sequential files, MAP Files, CSV, xml, log ORC and RC.   Experience with NoSQL Database HBase, Cassandra, MongoDB.   Experience with AIX/Linux RHEL, Unix Shell Scripting and SQL Server 2008.   Worked on data search toolElastic Search and data collection tool Logstash.   Strong knowledge in Hadoop cluster installation, capacity planning and performance tuning, benchmarking, disaster recovery plan and application deployment in production cluster.   Experience in developing stored procedures,

triggers using SQL, PL/SQL in relational databases such as MS SQL Server 2005/2008. Exposed into methodologies Scrum, Agile and Waterfall. Work Experience Big Data Engineer American Express - Phoenix, AZ November 2018 to Present RESPONSIBILITIES: Worked with extensive data sets in Big Data to uncover pattern, problem & unleash value for the Enterprise. Worked with internal and external data sources on improving data accuracy / coverage and generate recommendation on the process flow to accomplish the goal. Ingestion of various types of data feeds from SOR and use-case perspective into Cornerstone 3.0 platform. Re-engineered legacy IDN FastTrack process to get the Bloomberg data directly from source to the CS3.0. Converted legacy Shell scripts to Map-Reduce jobs in a distributed manner without performing any kind of processing on the Edgenode to eliminate the burden. Created Spark applications for data preprocessing for greater performance. Developed Spark code and Spark-SQL/streaming for faster testing and processing of data. Experience in creating spark applications using RDD, Dataframes. Worked extensively on hive to analyse the data and create reports for data quality. Implemented Partitioning, Dynamic Partitions and Buckets in HIVE for increasing performance benefit and helping in organizing data in a logical fashion. Written Hive queries for data analysis to meet the business requirements and Designed and developed User Defined Function (UDF) for Hive. Involved in creating Hive tables (Managed tables and External tables), loading and analyzing data using hive queries. Good knowledge about the configuration management tools like SVN/CVS/Github. Experience in configuring Event Engine nodes to import and export the data from Teradata to HDFS and vice-versa. Worked with source to get the history data as well as BAU data from IDN Teradata to the CornerStoneplatform and migrated also feeds from CS2.0. Expert in creating the nodes in Event Engine as per the use-case requirement to automate the process for the BAU data flow. Exported the Event Engine nodes created in the silver environment to the IDN repository in BitBucket and created DaVinci package to migrate it to Platinum. Worked with FDP team to create a secured flow to get the data from KAFKA Queue to CS3.0. Expert in creating the SFTP Connection to the internal and external source to get data in secured manner without any breakage. Handle the production Incidents assigned to our workgroup promptly and fix the bugs or

route it to the respective teams and optimized the SLA's. Hadoop Developer Avast - Emeryville, CA September 2017 to 2018 RESPONSIBILITIES: Developed real time data processing applications by using Scala and Python and implemented Apache Spark Streaming from various streaming sources like Kafka and JMS. Experienced in writing live Real-time Processing and core jobs using Spark Streaming with Kafka as a data pipe-line system. Developed Shell, Perl and Python scripts to automate and provide Control flow to Pig scripts. Worked on Amazon AWS concepts like EMR and EC2 web services for fast and efficient processing of Big Data. Involved in loading data from Linux file systems, servers, java web services using Kafka producers and partitions. Applied Kafka custom encoders for custom input format to load data into Kafka Partitions. Implement POC with Hadoop. Extract data with Spark into HDFS. Used Spark SQL with Scala for creating data frames and performed transformations on data frames. Used Spark API over Cloudera Hadoop YARN to perform analytics on data in Hive. Developed code to read data stream from Kafka and send it to respective bolts through respective stream. Worked on Spark streaming using Apache Kafka for real time data processing. Developed Map Reduce jobs using Map Reduce Java API and HIVEQL. Developed UDF, UDAF, UDTF functions and implemented it in HIVE Queries. Developing Scripts and Batch Job to schedule a bundle (group of coordinators) which consists of various Hadoop Programs using Oozie. Experienced in optimizing Hive queries, joins to handle different data sets. Involved in ETL, Data Integration and Migration by writing pig scripts. Integrated Hadoop with Solr and implement search algorithms. Experience in Storm for handling realtime processing. Hands on Experience working in Hortonworks distribution. Worked hands on No-SQL databases like MongoDB for POC purpose in storing images and URIs. Designed and implemented MongoDB and associated RESTful web service. Involved in writing test cases and implement test classes using MRUnit and mocking frameworks. Developed Sqoop scripts to extract the data from MYSQL and load into HDFS.. Experience in processing large volume of data and skills in parallel execution of process using Talend functionality. Used Talend tool to create workflows for processing data from multiple source systems. Environment: MapReduce, HDFS, Sqoop, LINUX, Oozie, Hadoop, Pig, Hive, Solr, Spark Streaming, Kafka, Storm, Spark, Scala,

Python, MongoDB, Hadoop Cluster, Amazon Web Services, Talend.   Hadoop Developer Kellogg's Dec 2016 - Aug2017Oakbrook, IL   RESPONSIBILITIES: Experience with professional software engineering practices and best practices for the full software development life cycle including coding standards, code reviews, source control management and build processes.   Worked on analyzing Hadoop cluster and different big data analytic tools including Map Reduce, Hive.   Written multiple MapReduce programs for data extraction, transformation and aggregation from multiple file formats including XML, JSON, CSV & other compressed file formats.   Worked on Teradata parallel transport (TPT) to load data from databases and files to Teradata.   Wrote views based on user and/or reporting requirements.   Wrote Teradata Macros and used various Teradata analytic functions.   Involved in migration projects to migrate data from data warehouses on Oracle/DB2 and migrated those to Teradata.   Configured Flume source, sink and memory channel to handle streaming data from server logs and JMS sources.   Experience in working with Flume to load the log data from multiple sources directly into HDFS.   Worked in the BI team in Big Data Hadoop cluster implementation and data integration in developing large-scale system software.   Involved in source system analysis, data analysis, data modeling to ETL (Extract, Transform and Load)   Handling structured and unstructured data and applying ETL processes.   Worked extensively with Sqoop for importing and exporting the data from HDFS to Relational Database systems/mainframe and vice-versa. Loading data into HDFS.   Involved in collecting, aggregating and moving data from servers to HDFS using Flume.   Implemented logging framework - ELK stack (Elastic Search, LogStash& Kibana) on AWS.   Developed the Pig UDF'S to pre-process the data for analysis.   Coding complex Oracle stored procedures, functions, packages, and cursors for the client specific applications.   Experienced in using Java Rest API to perform CURD operations on HBase data.   Applied Hive queries to perform data analysis on HBase using Storage Handler to meet the business requirements   Writing Hive Queries to Aggregate Data that needs to be pushed to the HBase Tables.   Create/Modify shell scripts for scheduling various data cleansing scripts and ETL loading process.   Supports and assist QA Engineers in understanding, testing and troubleshooting. Environment:Hadoop, Hive, Linux, Map Reduce, Sqoop, Storm, HBase, Flume,

Eclipse, Maven, Junit, agile methodologies. Java Developer M-ize Software Solutions 2014 to December 2014 July 2016Hyderabad, India.   RESPONSIBILITIES: Review the system requirements and attending requirements meetings with analysts and users.   Involved in the life cycle of the project from documentation to unit testing making development as priority.   Developed web pages using Struts framework, JSP, XML, JavaScript, HTML/ DHTML and CSS, configure struts application, use tag library.   Used Apache Struts framework includes the integrated AJAX. Played major role in designing & developing JSP pages and XML reports.   Developed Servlets and custom tags for JSP pages.   Developed few module Web pages using Springs IOC and Hibernate.   Designed and developed dynamic pages using HTML, CSS- layout techniques, Java script.   Took the various challenges in the enhancement and completed them on time.   Extensive Used Exception handling and Multi-threading for the optimum performance of the application.   Involved in design and implemented (SOA, SOAP) next generation system on distributed platform. Extensively used XSL as a XML parsing mechanism for showing Dynamic Web Pages in HTML format.   Implemented SOAP protocol to get the requests from the outside System.   Used CVS as a source control for code changes.   Used ANT scripts to build the project and JUnit to develop unit test cases.   Developed coding using SQL, PL/SQL, Queries, Joins, Views, Procedures/Functions, Triggers and Packages.   Provided development support for System Testing, Product Testing, User Acceptance Testing, Data Conversion Testing, Load Testing, and Production.   Environment: Java 1.5, J2EE, AJAX, Servlets, JSP, RUP, Eclipse 3.1, Struts, Spring 2.0, Hibernate, XML, CVS, Java Script, JQuery, ANT, SOAP, Log4J, DB2, Web Sphere server, UNIX, IBM Web Sphere Portal Server Java/web Developer Highgate Systems - Hyderabad, Telangana May 2013 to September 2014 Hyderabad, India.   RESPONSIBILITIES: Collecting and understanding the User requirements and Functional specifications   Development of GUI Using HTML, CSS, JSP and JavaScript. Creating components for isolated business logic.   Deployment of application in J2EE Architecture.   Implemented Session Facade Pattern using Session and Entity Beans   Developed message driven beans to listen to JMS.   Developed the Web Interface using Servlets, Java Server Pages, HTML and CSS.   Used WebLogic to deploy applications on local and development environments of

the application. Extensively used the JDBC Prepared Statement to embed the SQL queries into the java code. Developed DAO (Data Access Objects) using Spring Framework 3. Developed Web applications with Rich Internet applications using Java applets, Silverlight, Java. Used JavaScript to perform client side validations and Struts-Validator Framework for server-side validation. Environment: Java, J2EE, JDBC, JSP, Struts, JMS, Spring, SQL, MS-Access, JavaScript, HTML Education Master of Science in Computer Science Western Illinois University 2017 Bachelor of Technology in Information Technology JNTU - Hyderabad, Telangana 2013 Skills SQL (2 years), JAVA (1 year), CSS (1 year), HTML (1 year), JAVASCRIPT (1 year) Additional Information TECHNICAL SKILLS: Programming Languages Java, Python, SQL, Scala, and C/C++ Big Data Ecosystem Hadoop, MapReduce, Kafka, Spark, Pig, Hive, YARN, Flume, Sqoop, Oozie, Zookeeper, Talend. Hadoop Distributions Cloudera Enterprise, Horton Works, EMC Pivotal. Databases Oracle, SQL Server, PostgreSQL. Web Technologies HTML, XML, JQuery, Ajax, CSS, JavaScript, JSON. Streaming Tools Kafka, RabbitMQ Testing Hadoop Testing, Hive Testing, MRUnit. Operating Systems Linux Red Hat/Ubuntu/CentOS, Windows 10/8.1/7/XP. Cloud AWS S3, Redshift Cluster Technologies and Tools Servlets, JSP, Spring (Boot, MVC, Batch, Security), Web Services, Hibernate, Maven, GitHub. Application Servers Tomcat, JBoss. IDE's Eclipse, Net Beans, IntelliJ.

Name: Miss Jeanne Villegas

Email: smithsteven@example.org

Phone: 8203641272