Sr. Big Data/Hadoop Developer Sr. Big Data/Hadoop Developer Sr. Big Data/Hadoop Developer - VF Corporation, Greensboro, NV Greensboro, NC   Over 9+ years of working experience as a Big Data/Hadoop Developer in designed and developed various applications like big data, Hadoop, Java/J2EE open-source technologies.   Strong development skills in Hadoop, HDFS, Map Reduce, Hive, Sqoop, HBase with solid understanding of Hadoop internals.   Experience in Programming and Development of java modules for an existing web portal based in Java using technologies like JSP, Servlets, JavaScript and HTML, SOA with MVC architecture.   Expertise in ingesting real time/near real time data using Flume, Kafka, Storm   Good knowledge of NO SQL databases like Mongo DB, Cassandra and HBase.   Excellent knowledge of Hadoop Architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node and MRA and MRv2 (YARN).   Expertise in writing Hadoop Jobs to analyze data using MapReduce, Apache Crunch, Hive, Pig and SOLR, Splunk.   Hands on experience in installing, configuring and using Apache Hadoop ecosystem components like Hadoop Distributed File System (HDFS), MapReduce, Pig, Hive, HBase, Apache Crunch, Zookeeper, Scoop, Hue, Scala, AVRO.   Strong Programming Skills in designing and implementing of multi-tier applications using Java, J2EE, JDBC, JSP, JSTL, HTML, CSS, JSF, Struts, JavaScript, Servlets, POJO, EJB, XSLT, JAXB.   Extensive experience in SOA-based solutions - Web Services, Web API, WCF, SOAP including Restful APIs services Good Knowledge in Amazon Web Service (AWS) concepts like EMR and EC2 web services which provides fast and efficient processing of Teradata Big Data Analytics.   Experience working on EC2 (Elastic Compute Cloud) cluster instances, setup data buckets on S3 (Simple Storage Service), set EMR (Elastic MapReduce).   Experienced in collection of Log Data and JSON data into HDFS using Flume and processed the data using Hive/Pig.   Expertise in developing a simple web based application using J2EE technologies like JSP, Servlet, and JDBC.   Work Extensively in Core Java, Struts2, JSF2.2, Spring3.1, Hibernate, Servlets, JSP and Hands-on experience with PL/SQL, XML and SOAP.   In depth understanding/knowledge of Hadoop Architecture and various components such as HDFS, Job Tracker, Task Tracker, NameNode, DataNode.   Well versed working with Relational Database Management Systems as Oracle 9i/12c, MS SQL, MySQL Server   Hands on

experience in working on XML suite of technologies like XML, XSL, XSLT, DTD, XML Schema, SAX, DOM, JAXB.   Hands on experience in advanced Big-Data technologies like Spark Ecosystem (Spark SQL, MLlib, SparkR and Spark Streaming), Kafka and Predictive analytics   Knowledge of the software Development Life Cycle (SDLC), Agile and Waterfall Methodologies.   Experienced on applications using Java, python and UNIX shell scripting   Experience in consuming Web services with Apache Axis using JAX-RS(REST) API's.   Experienced in building tool Maven, ANT and logging tool Log4J.   Experience in working with Web Servers like Apache Tomcat and Application Servers like IBM Web Sphere and JBOSS.   Good knowledge of NoSQL databases such as HBase, MongoDB and Cassandra.   Experience in working with Eclipse IDE, Net Beans, and Rational Application Developer.   Experienced in Apache Flume for collecting, aggregating and moving huge chunks of data from various sources such as web server, telnet sources etc.   Extensively designed and executed SQL queries in order to ensure data integrity and consistency at the backend. Experienced in working with different scripting technologies like Python, UNIX shell scripts.   Strong experienced in working with UNIX/LINUX environments, writing shell scripts.   Expertise with frameworks like Angular JS, jQuery in web presentation layer with servlets, JSP, Spring MVC at the web controller layer.   Experienced in deploying J2EE applications on Apache Tomcat web server and WebLogic, WebSphere, JBoss application server.   Technical Tools/Technology   Hadoop Ecosystem: Hadoop2.7/2.5, MapReduce, Sqoop, Hive, Oozie, Pig, HDFS1.2.4, Zookeeper, Flume, Impala, Spark2.0/2.0.2, Storm, Hadoop (Cloudera), Horton works and Pivotal) Big Data Platforms: Horton works, Cloudera, Amazon AWS, and Apache.   Databases & NOSQL Databases: Oracle12c/11g, MYSQL, Microsoft SQL Server2016/2014, MongoDB, HBase and Cassandra. Work Experience Sr. Big Data/Hadoop Developer VF Corporation, Greensboro, NV May 2017 to Present Responsibilities: Worked as a Sr. Big/Hadoop Developer for providing solutions for big data problem.   Involved in full life cycle of the project from Design, Analysis, logical and physical architecture modelling, development, Implementation, testing.   Utilized SDLC Methodology to help manage and organize a team of developers with regular code review sessions.   Involved in Agile methodologies, daily scrum meetings, spring planning.   Developed Pig Latin scripts for replacing

the existing legacy process to the Hadoop and the data is fed to AWS S3.   Worked on MongoDB by using CRUD (Create, Read, Update and Delete), Indexing, Replication and Sharding features.   Created Talend jobs to read messages from Amazon AWS SQS queues & download files from AWS S3 buckets.   Developed MapReduce (YARN) jobs for cleaning, accessing and validating the data.   Worked on analyzing Hadoop cluster and different Big Data Components including Pig, Hive, Spark, HBase, Kafka, Elastic Search, database and SQOOP.   Written Hive jobs to parse the logs and structure them in tabular format to facilitate effective querying on the log data.   Wrote MapReduce jobs to filter and parse inventory data which was stored in the HDFS.   Configured Flume, Hive, Pig, Sqoop and Oozie on the Hadoop cluster for data pipelining.   Imported and exported data into the HDFS from the Oracle database using Sqoop.   Integrated MapReduce with Cassandra to import bulk amount of logged data.   Converted ETL operations to the Hadoop system using Hive transformations and functions.   Conducted streaming jobs with basic Python to process terabytes of formatted data for machine learning purposes.   Used Flume to collect, aggregate and store the web log data and loaded it into the HDFS.   Implemented Partitioning, Dynamic Partitions, Buckets in Hive.   Developed custom and Pig UDFs for product specific needs.   Implemented and configured workflows using Oozie to automate jobs.   Performed Hadoop cluster management and configuration of multiple nodes on AWS.   Involved in creating buckets to store the data in AWS and stored the data repository for future needs and reusability.   Worked along with Tableau developers to help performance tune the visualizations graphs/analytics.   Involved in the cluster coordination services through Zookeeper.   Participated in the managing and reviewing of the Hadoop log files.   Used Elastic Search & MongoDB for storing and querying the offers and non-offers data.   Created custom UDF'sfor Spark and Kafkaprocedure for some of non-working functionalities in custom UDF into Scala in production environment.   Import the data from different sources like HDFS/HBase into Spark RDD and developed a data pipeline using Kafka and Storm to store data into HDFS.   Worked independently with Cloudera support and Hortonworks support for any issue/concerns with Hadoop cluster.   Responsible for building scalable distributed data solutions using Hadoop Cloudera.   Used Spark streaming to receive real time data from the Kafka

and store the stream data to HDFS using Scala and NoSQL databases such as HBase and Cassandra.   Implementing Hadoop Security on Hortonworks Cluster using Kerberos and Two-way SSL   Worked with teams in setting up AWS EC2 instances by using different AWS services like S3, EBS, Elastic Load Balancer, and Auto scaling groups, VPC subnets and CloudWatch. Implemented Hortonworks NiFi and recommended solution to inject data from multiple data sources to HDFS and Hiveusing NiFi.   Developed Restful web services using JAX-RS and used DELETE, PUT, POST, GET HTTP methods   Created scalable and high-performance web services for data tracking and done High-speed querying.   Used Java Messaging Services (JMS) for reliable and asynchronous exchange of important information such as payment status report on IBM WebSphere MQ messaging system.   Executed Hive queries on Parquet tables stored in Hive to perform data analysis to meet the business requirements.   Worked on migrating MapReduce programs into Spark transformations using Spark and Scala.   Developed NIFI flow to move data from different sources to HDFS and from HDFS to S3 buckets-In current project   Created and maintained various Shell and Python scripts for automating various processes and optimized MapReduce code, pig scripts and performance tuning and analysis.   Worked on Oozie workflow engine for job scheduling. Involved in Unit testing and delivered Unit test plans and results documents.   Involved with ingesting data received from various providers, on HDFS for big data operations.   Wrote MapReduce jobs to perform big data analytics on ingested data using Java API.   Wrote MapReduce in Ruby using Hadoop Streaming to implement various functionalities.   Performed transformations, cleaning and filtering on imported data using Hive, Map Reduce, and loaded final data into HDFS.   Conducted meetings with data analysts with basic Python and wrangled data for data repositories.   Environment: Hadoop 3.0, Java, MapReduce, AWS, HDFS, Scala 2.12, Python 3.7, MongoDB 4.0, Spark 2.3, Hive 2.3, Pig 0.17, Linux, XML, Cloudera, CDH4/5 Distribution, Oracle 12c, PL/SQL, EC2, Apache Flume 1.8, Zookeeper 3.4, Cassandra 3.11, Hortonworks, Elastic search, IBM WebSphere Big Data Engineer MFS Investments - Boston, MA January 2016 to April 2017 Responsibilities: Architected, Designed and Developed Business applications and Data marts to facilitate the reporting.   Involved in all phases of Software Development Life Cycle (SDLC)

and Worked on all activities related to the development, implementation and support for Hadoop. Performed Requirements gathering, Analysis, Design, Code development, Testing using Agile methodologies.   Primary responsibilities include building scalable distributed data solutions using Hadoop ecosystem.   Worked on Hortonworks Data Platform Hadoop distribution for data querying using Hive to store and retrieve data.   Implemented Hive optimized joins to gather data from different sources and run ad-hoc queries on them.   Performed custom aggregate functions using Spark SQL and performed interactive querying.   Co-ordination with Hortonworks, development and the operations team on the platform level issues.   Extensively worked on creating combiners, partitioning and distributed cache to improve performance of MapReduce jobs.   Worked on Spark SQL and Data frames for faster execution of Hive queries using Spark Sql Context.   Used Sqoop transfer data between databases and HDFS and used Kafka to stream the log data from servers. Used Pig to perform data transformations, event joins, filter and some pre-aggregations before storing the data onto HDFS.   Implemented different analytical algorithms using MapReduce programs to apply on top of HDFS data.   Worked on MongoDB database concepts such as locking, transactions, indexes, sharding, replication and schema design.   Implemented read references in MongoDB replica set.   Used Apache Tez for processing data and storing it in MongoDB.   Write concern in MongoDB to avoid loss of data during system failures.   Created HBase tables to load large sets of structured, semi-structured and unstructured data coming from Unix, NoSQL and a variety of portfolios.   Extensively performed CRUD operations like put, get, scan, delete, update etc. on HBase database.   Wrote Hive Generic UDF's to perform business logic operations at table level.   Developed workflow in Oozie to automate the tasks of loading the data into HDFS and preprocessing with Pig, Hive and Sqoop.   Analyzed large amounts of data sets to determine optimal way to aggregate and report on it.   Used Hive join queries to join multiple tables of a source system and load them into Elastic Search Tables.   Used Apache Kafka as messaging system to load log data, data from applications into HDFS system.   Developed POC using Scala and deployed on the Yarn cluster, compared the performance of Spark, with Hive and SQL.   Involved in converting Hive queries into Spark transformations using Spark RDDs, Python

and Scala.   Worked on various file formats and compression Text, Avro, Parquet file formats, snappy, bz2, gzip compression.   Implemented test scripts to support test driven development and continuous integration.   Scheduling Cron jobs for file system check using fsck and wrote shell scripts to generate alerts.   Data scrubbing and processing with Oozie.   Loading the analyzed Hive data into NOSQL databases like HBase, MongoDB.   Provide Technical support for the Research in Information Technology program   Manage and upgrade Linux and OS X server systems.   Responsible for installation, configurations and management for Linux Systems   Environment: Hadoop 3.0, Java, MapReduce, HDFS, Hive 2.3, Pig 0.17, Sqoop 1.4, Flume 1.8, Python 3.7, Spark 2.3, Impala, Scala, Kafka, Shell Scripting, Eclipse, Cloudera, MySQL, Talend, Cassandra 3.11 Sr. Java/Hadoop Developer Optum - Eden Prairie, MN February 2014 to December 2015 Responsibilities   Involved in analysis, design, testing phases and responsible for documenting technical specifications.   Worked as part of the Agile Application Architecture (A3) development team responsible for setting up the architectural components for different layers of the application.   Involved in end to end data processing like ingestion, processing, and quality checks and splitting.   Real time streaming the data using Spark Streaming with Kafka   Developed Spark scripts by using Scala as per the requirement.   Load the data into Spark RDD and performed in-memory data computation to generate the output response.   Performed different types of transformations and actions on the RDD to meet the business requirements.   Extremely used plain JavaScript and JQuery, JavaScript Library to do the client side validations.   Developed a data pipeline using Kafka, Spark and Hive to ingest, transform and analyzing data.   Developed Pig Scripts, Pig UDFs and Hive Scripts, Hive UDFs to analyze HDFS data.   Used Sqoop to export data from HDFS to RDBMS.   Performed ETL using Talend.   Worked on Hadoop eco system components HDFS, MapReduce, Hive, Pig, Sqoop and HBase.   Designed & developed web based GUI architecture using HTML, CSS, AJAX, JQuery, AngularJS, and JavaScript.   Developed Map Reduce programs for some refined queries on big data.   Involved in loading data from UNIX file system to HDFS.   Used Pig as ETL tool to do transformations, event joins and some pre-aggregations before storing the data onto HDFS.   Worked in the cluster disaster recovery plan for the Hadoop cluster by

implementing the cluster data backup in AmazonS3 buckets. Implemented SparkSQL to access Hive tables into Spark for faster processing of data. Extracted the data from Databases into HDFS using Sqoop. Handled importing of data from various data sources, performed transformations using Hive, PIG and loaded data into HDFS. Used PIG predefined functions to convert the fixed width file to delimited file. Used HIVE join queries to join multiple tables of a source system and load them into Elastic Search Tables. Manage and review Hadoop log files. Implemented lambda architecture as a solution. Adept at understanding Partitions, bucketing concepts managed and created external tables in Hive to optimize performance. Written Hadoop Jobs for analyzing data using HiveQL (Queries), Pig Latin (Data flow language), and custom MapReduce programs in Java. Used Hadoop streaming jobs to process terabytes data in Hive. Developed workflow in Oozie to automate the tasks of loading the data into HDFS and pre-processing with Pig. Created reports for the BI team using Sqoop to import data into HDFS and Hive. Responsible for the designing, coding and developed the application in J2EE using MVC architecture. Configured development environment using WebSphere application server for developer's integration testing Used Web services for sending and getting data from different applications using SOAP messages. Used ANT scripts to build the application and deployed on WebSphere Application Server. Used JUnit framework for Unit testing of application. Environment: CDH, Hadoop 2.8, HDFS, MapReduce, Yarn, Hive 2.1, PIG 0.16, Oozie, Sqoop 1.2, Linux, Shell scripting, Java, SBT, Amazon S3, JIRA, Git Stash, HDFS, Eclipse, SQL, Oracle 11g. Sr. Java/J2EE Developer Exchange Bank - Santa Rosa, CA October 2011 to January 2014 Responsibilities: Responsible for system analysis, design and development using J2EE architecture. Actively participated in requirements gathering, analysis, design and testing phases. Developed the application using Spring Framework that leverages classical Model View Controller (MVC) architecture. Involved in Software Development Life cycle starting from requirements gathering and performed OOA and OOD Used Spring JDBC to execute database queries. Created row mappers and query classes for DB operations. Created a Transaction History Web Service using SOAP that is used for internal communication in the workflow process. Designed and created components for company's object framework using best

practices and design Patterns such as Model-View-Controller (MVC).   Used DOM and DOM Functions using Firefox and IE Developer Tool bar for IE.   Debugged the application using Firebug to traverse the documents.   Involved in writing SQL Queries, Stored Procedures and used JDBC for database connectivity with MySQL Server.   Developed the presentation layer using CSS and HTML taken from Bootstrap to develop for browsers.   Did core Java coding using JDK 1.3, Eclipse Integrated Development Environment (IDE), clear case, and ANT.   Used Spring Core and Spring-web framework. Created a lot of classes for backend.   Involved in developing web pages using HTML and JSP.   Exposed business functionality to external systems (Interoperable clients) using Web Services (WSDL-SOAP) Apache Axis.   Developed POJO classes and writing Hibernate query language (HQL) queries.   Used PL/SQL for queries and stored procedures in SQL as the backend RDBMS.   Involved in the Analysis and Design of the front-end and middle tier using JSP, Servlets and Ajax.   Implemented Spring IOC or Inversion of Control by way of Dependency Injection where a Factory class was written for creating and assembling the objects.   Implemented modules using Core Java APIs, Java collection, Threads, XML, and integrating the modules and used SOAP for Web Services by exchanging XML data between applications over HTTP.   Created EJB, JPA and Hibernate component for the application.   Implemented XML parsers with SAX, DOM, and JAXB XML Parser Libraries to Modify User view of Products and Product information in Customized view with XML, XSD, XSTL in HTML, XML, PDF formats.   Established continuous integration with JIRA, Jenkins.   Developed data mapping to create a communication bridge between various application interfaces using XML, and XSL.   Used Hibernate to manage Transactions (update, delete) along with writing complex SQL and HQL queries.   Used Microsoft VISIO for developing Use Case Diagrams, Sequence Diagrams and Class Diagrams in the design phase.   Developed Restful Web services client to consume JSON messages using Spring JMS configuration. Developed the message listener code.   Providing production support which includes handling tickets & providing resolution. Used BMC Remedy Tool to add issues & update resolutions.

Create database objects like tables, sequences, views, triggers, stored procedures, functions packages.   Used Maven as the build tool and Tortoise SVN as the Source version controller.

Environment: Core Java, UNIX, J2EE, XML Schemas, XML, JavaScript 2014, JSON, CSS3, HTML5, spring, Hibernate, Design Patterns, Servlets, JUnit, JMS, MySQL, Restful Web Services, SOAP, Tortoise SVN 1.5, Web Services, Apache Tomcat 8.0, Windows XP Java/J2EE Developer ITC Infotech July 2009 to September 2011 Responsibilities: As a Java Developer involved in back-end and front-end developing team.   Designed and implemented the User Interface using JavaScript, HTML, XHTML, XML, CSS, JSP, and AJAX.   Wrote web service client for tracking operations for the orders which is accessing web services API and utilizing in our web application. Implemented data archiving and persistence of report generation meta-data using Hibernate by creating Mapping files, POJO classes and configuring hibernate to set up the data sources. Developed Spring framework DAO Layer with JPA and EJB3 in Imaging Data model and Doc Import.   The business logic is developed using J2EE framework and deployed components on Application server where Eclipse was used for component building.   Actively involved in deployment EJB service jars, Application war files in WebLogic Application server.   Developed GUI screens for login, registration, edit account, forgot password and change password using Struts. Used JUnit framework for unit testing of application and JUL logging to capture the log that includes runtime exceptions.   Wrote various SQL stored procedures and SQL commands to retrieve the data from the SQL server.   Used Spring MVC Framework to develop Action classes and Controllers along with validation framework and annotations.   Writing SQL queries for data access and manipulation using Oracle SQL Developer.   Developed Session Bean to encapsulate the business logic and Model and DAO classes using Hibernate   Designed and coded JAX-WS based Web Services used to access external financial information.   Implemented EJB Components using State less Session Bean and State full session beans.   Used spring framework with the help of Spring Configuration files to create the beans needed and injected dependency using Dependency Injection.   Utilized JPA for Object/Relational Mapping purposes for transparent persistence onto the Oracle database.   Involved in creation of Test Cases for JUnit Testing.   Used Oracle as Database and used Toad for queries execution and also involved in writing SQL scripts, PL/SQL code for procedures and functions.   Used SOAP as a XML-based protocol for web service operation

invocation. Packaged and deployed the application in IBM WebSphere Application server in different environments like Development, testing etc. Used Log4J to validate functionalities and JUnit for unit testing. Environment: Java, Servlets, JSP, Struts 1.0, Hibernate 3.1, spring core, Spring JDBC, HTML, JavaScript 2012, AJAX, XSL, XSLT, XSD schema, XML Beans, Web logic, Oracle9i Education Bachelor's

Name: Megan Villa

Email: sfrederick@example.net

Phone: (408)656-7884