Sr. Hadoop Developer Sr. Hadoop Developer Sr. Hadoop Developer - Humana Health Insurance Louisville, KY   7+ years of practical experience in building industry specific Java applications and implementing Big-data technologies in core and enterprise software development.   4+ years of experience in developing applications that perform large scale Distributed Data Processing using Big Data ecosystem tools Hadoop, Hive, Pig, Sqoop, Hbase, Cassandra, Spark, Spark Streaming MLLib, Mahout, Oozie, Zoo Keeper, Flume, Yarn and Avro.   Working knowledge in installing and maintaining Cassandra by configuring the cassandra.yaml file as per the business requirement and performed reads/writes using Java JDBC connectivity.   Experience with Cloudera Manager for management of Hadoop cluster.   Good conceptual understanding and experience in cloud computing applications using Amazon EC2, S3, EMR.   Experience in analyzing data using HiveQL, Pig Latin, HBase and custom Map Reduce programs in Java.   Extensive Experience on Teradata database, analyzing business needs of clients, developing effective and efficient solutions and ensuring client deliverables with in committed timelines. Expertise in maintaining data quality, data organization and metadata.   Passionate about Big Data Analytics and skilled in exploring data, content and Expert in distributed computing, algorithms, and data analytics.   Hands on experience in using various Hadoop distributions (Apache, Cloudera, Hortonworks, MapR).   In-Depth knowledge and experience in design, development and deployments of Big Data projects using Hadoop / Data Analytics / NoSQL / Distributed Machine Learning frameworks.   Solid understanding of SQL & NOSQL databases such as Oracle, PostgreSQL, MySQL, MongoDB, HBase & Cassandra.   Knowledge on MongoDB NoSQL data modeling, tuning, disaster recovery backup used it for distributed storage and processing using CRUD.   Worked on MongoDB database concepts such as locking, transactions, indexes, Sharing, replication, schema design.   Very good understanding Cassandra cluster mechanism that includes replication strategies, snitch, gossip, consistent hashing and consistency levels.   Experience in Coding Teradata SQL, Teradata Stored Procedures, Macros and Triggers   Strong experience in Creating Database Objects such as Tables, Views, Functions, Stored Procedures, Indexes, Triggers, Cursors in Teradata   Experience data processing like collecting, aggregating, moving from various sources using Apache Flume and

Kafka.   Developed a data pipeline using Kafka and Spark Streaming to store data into HDFS and performed the real-time analytics on the incoming data.   Experience in importing the real-time data to Hadoop using Kafka and implemented the Oozie job for daily imports.   Experienced in working with in-memory processing framework like Spark Transformations, SparkQL, MLib and Spark Streaming.   Expertise in creating Custom Serdes in Hive.   Good working experience on using Sqoop to import data into HDFS from RDBMS and vice-versa.   Expertise in job scheduling and monitoring tools like Oozie and Zoo Keeper.   Experience in design and development of Map Reduce Programs using Apache Hadoop for analyzing the big data as per the requirement.   Experience in performing ad-hoc queries on structured data using Hive QL and used Partition and Bucketing techniques and joins with HIVE for faster data access.   Experience in performing ETL operations using Pig Latin scripts.   Implemented Java APIs and created custom Java programs for full-fledged utilization of Hadoop and its related tools.   Implemented work flows that involve Hadoop actions using Oozie co-ordinators.   Experienced in implementing POC using Spark Sql and Mlib libraries.   Improving the performance and optimization of the existing algorithms in Hadoop using Spark Context, Spark-SQL, Data Frame, Pair RDD's, YARN.   Hands on experience in handling Hive tables using Spark SQL.   Hands on experience on writing Queries, Stored procedures, Functions and Triggers by using SQL.   Used EMR (Elastic Map Reducing) to perform bigdata operations in AWS.   Proficient using version control tools like GIT, VSS, SVN and PVCS. Involvement in all stages of software development life cycle (SDLC) and follow agile methodologies and continuous delivery.   Strong skills in Object Oriented Analysis and Design (OOAD).   Well versed in enterprise software development methodologies and practices including TDD, BDD design patterns and performance testing.   Created User Defined Functions (UDFs), User Defined Aggregated Functions (UDAFs) in PIG and Hive.   Experience in dealing with log files to extract data and to copy into HDFS using Flume.   Experience with Testing MapReduce programs using MRUnit, JUnit and Easy Mock.   Implemented distributed searching capabilities using Solr to empower the geospatial search and navigation feature.   Experienced in using Solr to create search indexes to perform search operations faster.   Strong hands-on experience in Java and J2EE

frameworks.   Experience working with JAVA J2EE, JDBC, ODBC, JSP, Java Eclipse, Java Beans, EJB, Servlets   Expert in developing web page interfaces using JSP, Java Swings, and HTML scripting languages.   Excellent understanding on Java beans and Hibernate framework to implement model logic to interact with RDBMS databases.   Experience in using IDEs like Eclipse, NetBeans and Maven.   Hands on experienced working with source control tools such as Rational Clear Case and Clear Quest. Authorized to work in the US for any employer Work Experience Sr. Hadoop Developer Humana Health Insurance - Louisville, KY September 2018 to Present Responsibilities: ? Analyzed large and critical datasets using Cloudera, HDFS, HBase, MapReduce, Hive, Hive UDF, Pig, Sqoop, Zookeeper and Spark. ? Gathered the business requirements from the Business Partners and Subject Matter Experts ? Developed environmental search engine using JAVA, Apache SOLR and MYSQL. ? Managed works including indexing data, tuning relevance, developing custom tokenizers and filters, adding functionality includes playlist, custom sorting and regionalization with SOLR Search Engine. ? Ingested data from RDBMS and performed data transformations, and then export the transformed data to Cassandra as per the business requirement. ? Written multiple MapReduce programs for data extraction, transformation and aggregation from multiple file formats including XML, JSON, CSV & other compressed file formats. ? Developed automated processes for flattening the upstream data from Cassandra which in JSON format. Used Hive UDFs to flatten the JSON Data. ? Optimized MapReduce Jobs to use HDFS efficiently by using various compression mechanisms ? Developed PIG UDFs to provide Pig capabilities for manipulating the data according to Business Requirements and worked on developing custom PIG Loaders and Implemented various requirements using Pig scripts. ? Experienced on loading and transforming of large sets of structured, semi structured and unstructured data ? Created POC using Spark Sql and Mlib libraries. ? Developed a Spark Streaming module for consumption of Avro messages from Kafka.   Implementing different machine learning techniques in Scala using Scala machine learning library, and created POC using SparkSql and Mlib libraries.   Experienced in Querying data using SparkSQL on top of Spark Engine, implementing Spark RDD's in Scala.   Expertise in writing Scala code using Higher order functions

for iterative algorithms in Spark for Performance considerations. ? Experienced in managing and reviewing Hadoop log files ? Worked with different File Formats like TEXTFILE, AVROFILE, ORC, and PARQUET for HIVE querying and processing. ? Loading data by using the Teradata loader connection, writing Teradata utilities scripts (Fast Load, Multiload) and working with loader logs. ? To monitor query run times using Teradata Performance Monitor ? Involved in loading of data into Teradata from legacy systems and flat files using complex MLOAD scripts and FASTLOAD scripts. ? Create and Maintain Teradata Tables, Views, Macros, Triggers and Stored Procedures ? Monitored workload, job performance and capacity planning using Cloudera Distribution. ? Worked on Data loading into Hive for Data Ingestion history and Data content summary. ? Involved in developing Impala scripts for extraction, transformation, loading of data into data warehouse. ? Used Hive and Impala to query the data in HBase. ? Created Impala tables and SFTP scripts and Shell scripts to import data into Hadoop. ? Developed Hbase java client API for CRUD Operations. ? Created Hive tables and involved in data loading and writing Hive UDFs. Developed Hive UDFs for rating aggregation ? Generated Java APIs for retrieval and analysis on No-SQL database such as HBase and Cassandra ? Provided ad-hoc queries and data metrics to the Business Users using Hive, Pig ? Did various performance optimizations like using distributed cache for small datasets, partition and bucketing in hive, doing map side joins etc ? Worked on importing and exporting data from Oracle and DB2 into HDFS and HIVE using Sqoop for analysis, visualization and to generate reports. ? Involved in file movements between HDFS and AWS S3 and extensively worked with S3 bucket in AWS ? Experienced with AWS services to smoothly manage application in the cloud and creating or modifying the instances. ? Created data pipeline for different events of ingestion, aggregation and load consumer response data in AWS S3 bucket into Hive external tables in HDFS location to serve as feed for tableau dashboards. ? Used EMR (Elastic Map Reducing) to perform big data operations in AWS. ? Worked on Apache spark writing python applications to convert txt, xls files and parse. ? Developed Python scripts, UDF's using both Data frames/SQL and RDD/MapReduce in Spark for Data Aggregation, queries and writing data back into RDBMS through Sqoop. ? Installed the application on AWS EC2 instances and configured the storage on S3

buckets.   Use different components of Talend (tOracleInput, tOracleOutput, tHiveInput, tHiveOutput, tHiveInputRow, tUniqeRow, tAggregateRow, tRunJob, tPreJob, tPostJob, tMap, tJavaRow, tJavaFlex, tFilterRow etc) to develop standard jobs.   Loading data from different source (database & files) into Hive using Talend tool.   Load and transform data into HDFS from large set of structured data /Oracle/Sql server using Talend Big data studio.   Implemented Spark using Python/Scala and utilizingSpark Core, Spark Streaming and Spark SQL for faster processing of data instead of MapReduce in Java   Experience in integrating Apache Kafka with Apache Spark for real time processing.   Exposure on usage of Apache Kafka develop data pipeline of logs as a stream of messages using producers and consumers. ? Scheduled Oozie workflow engine to run multiple Hive and Pig jobs, which independently run with time and data availability ? Worked on custom Pig Loaders and Storage classes to work with a variety of data formats such as JSON, Compressed CSV etc ? Involved in running Hadoop Streaming jobs to process Terabytes of data   Used JIRA for bug tracking and CVS for version control.   Environment: Hadoop, Map Reduce, Hive, HDFS, PIG, Sqoop, Oozie, Cloudera, Flume, HBase, SOLR, CDH3, Cassandra, Oracle, Unix/Linux, Hadoop, Hive, PIG, SQOOP, Flume, HDFS, J2EE, Oracle/SQL & DB2 , Unix/Linux, JavaScript, Ajax, Eclipse IDE, CVS, JIRA. Sr. Hadoop Developer Microchip Computer Solutions - Springfield, IL April 2015 to August 2018 Responsibilities: ? Installed and configured Hadoop MapReduce, HDFS and developed multiple MapReduce jobs in Java for data cleansing and preprocessing.   ? Responsible to manage data coming from different sources. ? Involved in gathering the business requirements from the Business Partners and Subject Matter Experts. ? Proactively monitored systems and services, architecture design and implementation of Hadoop deployment, configuration management, backup and disaster recovery systems and procedures. ? Involved in works including indexing data, tuning relevance, developing custom tokenizers and filters, adding functionality includes playlist, custom sorting. ? Supported MapReduce Programs those are running on the cluster. ? Involved in HDFS maintenance and loading of structured and unstructured data. ? Involved in file movements between HDFS and AWS S3 and extensively worked with S3 bucket in AWS ? Worked extensively with importing metadata into Hive using Python and migrated existing

tables and applications to work on AWS cloud(S3). ? Installed and configured Pig and written Pig Latin scripts. ? Imported data using Sqoop to load data from MySQL to HDFS on regular basis. ? Expert knowledge on MongoDB NoSQL data modeling, tuning, disaster recovery backup used it for distributed storage and processing using CRUD. ? Extracted and restructured the data into MongoDB using import and export command line utility tool. ? Designed and Maintained Tez workflows to manage the flow of jobs in the cluster. ? Developed Scripts and Batch Job to schedule various Hadoop Program. ? Installation of Oozie workflow engine to run multiple Hive and pig jobs. ? Writing Hive queries for data analysis to meet the business requirements. ? Loading log data into HDFS using Flume and performing ETL Integration. ? Collecting and aggregating large amounts of log data using Flume and staging data in HDFS for further analysis. ? Developed Flume ETL job for handling data from HTTP Source and Sink as HDFS. ? Created Hive tables and working on them using Hive QL. ? Good Understanding of DAG cycle for entire Spark application flow on Spark application WebUI. ? Developed Spark SQL scripts and involved in converting Hive UDF's to Spark SQL UDF's. ? Implemented procedures like text analytics and processing using the in-memory computing capabilities like Apache Spark written in Scala. ? Developed multiple Spark jobs in Scala/Python for Data cleaning, pre-processing and Aggregating. ? Developed Spark programs using Scala, Involved in Creating SparkSQL Queries and Developed Oozie workflow for Spark jobs. ? Push data as delimited files into HDFS using Talend Big data studio. ? Analyzed and performed data integration using Talend open integration suite. ? Wrote complex SQL queries to take data from various sources and integrated it with Talend. ? Used Storm for an automatic mechanism for repeating attempts to download and manipulate the data when there is a hiccup. ? Designing and development of technical architecture, requirements and statistical models using R. ? Used Storm to analyze large amounts of non-unique data points with low latency and high throughput. ? Developed UI application using AngularJS, integrated with Elastic Search to consume REST. ? Writing the shell scripts to monitor the health check of Hadoop daemon services and respond accordingly to any warning or failure conditions. Utilized Agile and Scrum Methodology to help manage and organize a team of developers with regular code review sessions. Environments:-

Hadoop, Hive, Linux, Map Reduce, HDFS, Hive, Pig, HBase, Sqoop, Kafka, Flume, Shell Scripting, Storm, Java (JDK 1.6), Java 6, Eclipse, Oracle 10g, PL/SQL, SQL*PLUS, Toad 9.6, Linux, JIRA 5.1, Storm, CVS, JIRA 5.2. Hadoop Developer Dell - Austin, TX May 2012 to March 2015 Responsibilities: ? Launching and Setup of HADOOP Cluster which includes configuring different components of HADOOP. ? Hands on experience in loading data from UNIX file system to HDFS. ? Wrote the Map Reduce jobs to parse the web logs which are stored in HDFS. ? Developed Simple to complex MapReduce Jobs using Hive and Pig. ? Developed multiple Map Reduce jobs in PIG and Hive for data cleaning and pre-processing. ? Cluster coordination services through Zookeeper. ? Designed and implemented Hive queries and functions for evaluation, filtering, loading and storing of data. ? Installed Oozie workflow engine to run multiple Hive and Pig jobs which run independently with time and data availability. ? Expertise in Partitions, Bucketing concepts in Hive and analyzed the data using the HiveQL ? Installed and configured Flume, Hive, PIG, Sqoop and Oozie on the Hadoop cluster. ? Involved in creating Hive tables, loading data and running Hive queries in those data. ? Extensive Working knowledge of partitioned table, UDFs, performance tuning, compression-related properties, thrift server in Hive. ? Used Hive to analyze data ingested into HBase by using Hive-HBase integration and compute various metrics for reporting on the dashboard. ? Loading the data to HBase Using Pig, Hive and Java API's. ? Worked with NoSQL databases like HBase in creating HBase tables to load large sets of semi-structured. ? Experienced with performing CURD operations in HBase. ? Collected and aggregated large amounts of web log data from different sources such as web servers, mobile using Apache Flume and stored the data into HDFS/HBase for analysis. ? Developed Flume ETL job for handling data from HTTP Source and Sink as HDFS. ? Involved in writing optimized PIG Script along with involved in developing and testing PIG Latin Scripts. ? Created Map Reduce programs for some refined queries on big data. ? Working knowledge in writing PIG's Load and Store functions.   Environment: Apache Hadoop 1.0.1, MapReduce, HDFS, CentOS, Zookeeper, Sqoop, Cassandra, Hive, PIG, Oozie, Java, Eclipse, Amazon EC2, JSP, Servlets. Java Developer Cubic Technologies - Irving, TX June 2010 to April 2012 Responsibilities: ? Implemented the project

according to the Software Development Life Cycle (SDLC) ? Implemented JDBC for mapping an object-oriented domain model to a traditional relational database ? Created Stored Procedures to manipulate the database and to apply the business logic according to the user's specifications ? Developed the Generic Classes, which includes the frequently used functionality, so that it can be reusable ? Exception Management mechanism using Exception Handling Application Blocks to handle the exceptions ? Designed and developed user interfaces using JSP, JavaScript and HTML ? Involved in Database design and developing SQL Queries, stored procedures on MySQL ? Used CVS for maintaining the Source Code ? Logging was done through log4j   Environment:- JAVA, Java Script, HTML, log4j, JDBC Drivers, Soap Web Services, UNIX, Shell scripting, SQL Server

Java/J2EE Developer FINRA - Rockville, MD January 2009 to May 2010 Responsibilities: ? Responsible for programming, and troubleshooting web applications using Java, JSP, HTML, JavaScript, CSS, and SQLServer. ? Responsible for Cross-browser testing and debugging, and creating graphics. ? Involved in creating CSS for a unified look of the Front End User Interface. ? Created optimizing web graphics, including designing and incorporating graphic user interface (GUI) features. ? Worked with the business stakeholders to determine navigational schemes, site flow and general web page functionality. ? Prepared Technical Design Documentation for the modules designed. ? Involved in all facets of software development life cycle, from requirements analysis, architecture, design, coding, testing and implementation.   ? Developed and maintained the application UI based on Eclipse. ? Actively participated in requirements gathering, analysis, design, and testing phases. ? Developed and implemented the MVC architectural pattern, JSPs as the view, Struts as Controller and as model. ? Created graphical user interfaces (GUIs) front-end using JSP, JavaScript and JSON. ? Struts Action Servlet is used as Front Controller for redirecting the control to the specific J2EE component as per the requirement. ? Developed JSP with Custom Tag Libraries for control of the business processes in the middle-tier and was involved in their integration. ? Responsible for developing the client side validations using JavaScript and JQuery. ? Developed the XML Schema for the data maintenance and structures. ? Prepared documentation and participated in preparing user's manual for the application. ? Involved in unit testing, integration

testing, user-acceptance testing and bug fixing.   Environment:- JAVA EE5, J2EE, XML, HTML, Struts2, Servlets, Java Script, JSP, CSS, JDBC, SQL Server,WebSphere8,Windows. Education Bachelor's Additional Information Skill set: Hadoop Core Services HDFS, Map Reduce, Spark, YARN   Hadoop Distribution Horton works, Cloudera, Apache   NO SQL Databases HBase, Cassandra Hadoop Data Services Hive, Pig, Sqoop, Flume, Sqoop Hadoop Operational Services Zookeeper, Oozie Monitoring Tools Ganglia, Cloudera Manager Cloud Computing Tools Amazon AWS   Languages C, Java/J2EE, Python, SQL, PL/SQL, Pig Latin, HiveQL, Unix Shell Scripting Java & J2EE Technologies Core Java, Servlets, Hibernate, Spring, Struts, JMS, EJB Application Servers Web Logic, Web Sphere, JBoss, Tomcat. Databases Oracle, MySQL, Postgress, Teradata Operating Systems UNIX, Windows, LINUX Build Tools Jenkins, Maven, ANT Development Tools Microsoft SQL Studio, Toad, Eclipse, NetBeans Development methodologies Agile/Scrum, Waterfall

Name: Amanda Smith

Email: lisafoster@example.org

Phone: +1-238-489-8749x178