

Sr. Big Data Developer Sr. Big Data Developer Sr. Big Data Developer - Anthem Inc Atlanta, GA

Overall 10+ 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.

Extensive knowledge in Software Development Lifecycle (SDLC) using Waterfall, Agile methodologies. Having Project experience on Microsoft Azure Cloud cluster environment and

Azure Data lake store Experience in importing and exporting data using Sqoop from HDFS to

Relational Database Systems and vice-versa. Experience in converting Hive queries into Spark

transformations using Spark RDDs and Scala. Good Knowledge in Amazon Web Service (AWS)

concepts like EMR and EC2 web services which provides fast. Good experience in Spark and its

related technologies like Spark SQL and Spark Streaming. Good knowledge of NOSQL databases

like Mongo DB, Cassandra and HBase. Experience on building the applications using Spark Core,

Spark SQL, Data Frames, Spark Streaming Strong experience in developing the workflows using

Apache Oozie framework to automate tasks. Experience in deploying J2EE applications on

Apache Tomcat web server and Web Logic, WebSphere, JBoss application server. Experience in

working with Eclipse IDE, Net Beans, and Rational Application Developer. Expertise on usage of

SQL queries to extract data from RDBMS databases - MySQL, DB2 and Oracle. Experience with

installing, backup, recovery, configuration and development on multiple Hadoop distribution

platforms Cloudera and Hortonworks including cloud platforms Amazon AWS and Google Cloud.

Experience on data ingestion tool Apache NiFi, used to extract data from various data sources into

Hadoop data lake Experience in working with NoSQL database Apache HBase and have

implemented performance improvement as per project requirement. Proficient in using and

deploying applications to Web Servers/Application servers like Tomcat, WebSphere, Micro services.

Experience in generating logging by Log4j to identify the errors in production test environment.

Extensive knowledge in programming with Resilient Distributed Datasets (RDDs). Experience in

working different version control tools like GIT, SVN and Continuous Integration tool Jenkins.

Hands on Experience in developing and implementing the Spring Rest and Restful Web Services.

Expertise in writing Hadoop Jobs to analyze data using MapReduce, Apache Crunch, Hive, Pig and

SOLR, Splunk. Experience in working with Web Servers like Apache Tomcat and Application Servers like IBM WebSphere and JBOSS. Experience in Apache Flume for collecting, aggregating and moving huge chunks of data from various sources such as web server, telnet sources etc. Experience in working with different scripting technologies like Python, UNIX shell scripts. Strong team player, ability to work independently and in a team as well, ability to adapt to rapidly changing environment, commitment towards learning, Possess excellent Work Experience Sr. Big Data Developer Anthem Inc - Atlanta, GA March 2018 to Present Responsibilities: As a Sr. Big Data Developer, I worked on Hadoop eco-systems including Hive, HBase, Oozie, Pig, Zookeeper, Spark Streaming MCS (MapR Control System) and so on with MapR distribution. Developed Java and J2EE applications using Rapid Application Development (RAD), Eclipse. Installed and configured Hadoop MapReduce, HDFS, Developed multiple MapReduce jobs in Java for data cleaning and Pre-processing. Create data integration and technical solutions for Azure Data Lake Analytics, Azure Data Lake Storage, Azure Data Factory, Azure SQL databases and Azure SQL Data Warehouse for providing analytics. Built code for real time data ingestion using Java, MapR-Streams (Kafka) and STORM. Involved in various phases of development analysed and developed the system going through Agile Scrum methodology. Involved in Design, development and testing of web application and integration projects using Object Oriented technologies such as Core Java, J2EE, Struts, JSP, JDBC, Spring Framework, Hibernate, Java Beans, Web Services (REST/SOAP), XML, XSLT, XSL and Ant. Worked on Apache Solr which is used as indexing and search engine. Primarily involved in Data Migration process using Azure by integrating with GitHub repository and Jenkins. Used Spark API over Hortonworks, Hadoop YARN to perform analytics on data in Hive. Developed various Servlets and Java Interfaces as part of the integration and process flow required for the system. Automated workflows using shell scripts and Control-M jobs to pull data from various databases into Hadoop Data Lake. Developed Spring Framework Controllers and worked on spring application framework features. Prepared Linux shell scripts to configure, deploy and manage Oozie workflows of Big Data applications. Worked with different data sources like Avro data files, XML files, JSON files, SQL server and Oracle to load data into

Hive tables. Developed Use Case Diagrams, Object Diagrams and Class Diagrams in UML using Rational Rose. Exported event weblogs to HDFS by creating a HDFS sink which directly deposits the weblogs in HDFS. Implemented a distributed messaging queue to integrate with Cassandra using Apache Kafka and Zookeeper. Imported weblogs & unstructured data using the Apache Flume and stores the data in Flume channel. Developed scripts to run scheduled batch cycles using Oozie and present data for reports. Managed real time data processing and real time Data Ingestion in MongoDB using Storm. Developed using the framework builds the graphical components and define actions, popup menus in XML. Primarily involved in Data Migration process using Azure by integrating with Git-hub repository and Jenkins. Used Jenkins for build and continuous integration for software development. Provided connections using JDBC to the database and developed SQL queries to manipulate the data. Monitored Hadoop cluster using tools like Cloudera manager, managing and scheduling the jobs on Hadoop cluster. Used Spark to create the structured data from large amount of unstructured data from various sources. Designed and implemented scalable Cloud Data and Analytical architecture solutions for various public and private cloud platforms using Azure. Extracted Real time feed using Spark streaming and convert it to RDD and process data into Data Frame and load the data into Cassandra. Performed importing data from various sources to the Cassandra cluster using Sqoop. Environment: Hadoop 3.0, Hive 2.3, HBase 2.2, Oozie 5.1, Pig 0.17, Zookeeper 3.4, Spark 2.4, MapR, MapReduce, HDFS, Java, MS Azure, Kafka 2.2, Agile, Jenkins 2.16, Hortonworks, Cassandra 3.11, XML, MongoDB Sr. Spark/Hadoop Developer Pacific Life - Newport Beach, CA October 2015 to February 2018 Responsibilities: Responsible for building scalable distributed data solutions using Hadoop. Participate in all stages of Software Development Life Cycle (SDLC) including requirements gathering, system Analysis, system development, unit testing and performance testing. Installed Oozie workflow engine to run multiple Hive, Shell Script, Sqoop, pig and Java jobs. In depth understanding of Hadoop Architecture including YARN and various components such as HDFS, Resource Manager, Node Manager, Name Node, Data Node and MR v1 & v2 concepts. Data Ingestion: Experience in importing and exporting data from different RDBMS Servers like MySQL,

Oracle and Teradata into HDFS and Hive using Sqoop. Experience in ingesting data from FTP/SFTP servers using Flume. Experience in developing Kafka Consumer API using Spark Scala applications. Data Processing: Developed MapReduce programs in Java for data cleansing, data filtering, and data aggregation. Experienced in analyzing the data using PIG Latin scripts. Experience in designing table partitioning, bucketing and optimized hive scripts using different performance utilities and techniques. Experience in developing Hive UDF's and running hive scripts using different execution engines like Tez and Spark (Hive on Spark). Experience in designing tables and views for reporting using Impala. Experienced in Developing Spark application using Spark Core, Spark SQL and Spark Streaming API's. Developed Java Beans for business logic. Design of REST APIs that allow sophisticated, effective and low cost application integrations. Implemented Spark using Scala and Spark SQL for faster testing and processing of data. Involved in story-driven Agile development methodology and actively participated in daily scrum meetings. Developed Apache spark-based programs to implement complex business transformations. Used Hive to analyze partitioned and bucketed data and compute various metrics for reporting. Developed unloading micro services using Scala API in Spark Data frame API for the semantic layer. Used SparkSQL and Spark Data frame extensively to cleanse and integrate imported data into more meaningful insights. Migrated associated business logic (PL/SQL procedures/functions) to Apache Spark. Involved in Unit testing and delivered Unit test plans and results documents using JUnit and MrUnit. Used Spark API over ClouderaHadoopYARN to perform analytics on data in Hive. Created scripts to sync data between local MongoDB and Postgres databases with those on AWS Cloud. Used Oozie engine for creating workflow and coordinator jobs that schedule and execute various. Created Kafka producer API to send live-stream json data into various Kafka topics. Created concurrent access for Hive tables with shared and exclusive locking that can be enabled in Hive with the help of Zookeeper implementation in the cluster. Created web-based User interface for creating, monitoring and controlling data flows using Apache Nifi. Worked on a POC to compare processing time of Impala with Apache Hive for batch applications to implement the former in project. Extensively used Spark - Cassandra

connector to load data to and from Cassandra. Involved in the Design Phase for getting live event data from the database to the front-end application using Spark Ecosystem. Worked on various compression and file formats like Avro, Parquet and Text formats. Created a complete processing engine, based on Cloudera distribution, enhanced performance. Worked with NoSQL databases like HBase in creating HBase tables to load large sets of semi structured data coming from various sources. Written event-driven, link tracking system to capture user events and feed to Kafka to push it to HBase. Implemented Spark using Scala and performed cleansing of data by applying Transformations and Actions. Developed MapReduce programs in Java to search production logs and web analytics logs for application issues. Involved in build/deploy applications using Maven and integrated with CI/CD server Jenkins. Environment: Hadoop 3.0, Oozie 5.0, Hive 2.3, Sqoop 1.4, Pig 0.17, Java, Spark 2.4, Scala 2.13, Agile, PL/SQL, JUnit 5.4, AWS, Kafka 2.2, Zookeeper 3.4, Cassandra 3.11, HCatalog, NoSQL, Maven 3.6, Jenkins 2.16

Java Hadoop Developer Comcast - Philadelphia, PA January 2013 to September 2015 Responsibilities:

Worked on analyzing Hadoop cluster and different big data analytic tools including MapReduce, Hive and Spark. Implemented Spark RDD transformations to map business analysis and apply actions on top of transformations. Developed the server side scripts using JMS, JSP and Java Beans. Adding and modifying Hibernate configuration code and Java/SQL statements depending upon the specific database access requirements. Implemented POC to migrate MapReduce programs into Spark transformations using Spark and Scala. Implemented Storm topologies to pre-process data before moving into HDFS system. Worked with various Hadoop file formats, including Text, Sequence File, RCFILE and ORC File. Written a Pig Scripts to read data from HDFS and write into Hive Table. Managed servers on the Amazon Web Services (AWS) platform instances using Puppet, Chef Configuration management. Interacted with Cloudera support and log the issues in Cloudera portal and fixing them as per the recommendations. Migrated complex MapReduce programs into Spark RDD transformations, actions. Involved in installing, configuring and managing Hadoop Ecosystem components like Pig, Sqoop, Kafka and Flume. Involved in the process of data acquisition, data pre-processing and data exploration of telecommunication project in Scala.

Developed and Deployed Oozie Workflows for recurring operations on Clusters. Created POC on Hortonworks and suggested the best practice in terms HDP, HDF platform. Developed and implemented Apache NIFI across various environments, written QA scripts in Python for tracking files. Worked in AWS environment for development and deployment of Custom Hadoop Applications. Converted text files into Avro then to parquet format for the file to be used with other Hadoop Eco-system tools. Implemented Apache Pig scripts to load data from and to store data into Hive using HCatalog. Loaded data into the cluster from dynamically generated files using Flume and from relational database management systems using Sqoop. Created HBase tables to store various data formats of incoming data from different portfolios. Designed and implemented Cassandra and associated Restful webservice. Expertise in writing Hadoop Jobs for analyzing data using Hive QL (Queries), Pig Latin (Data flow language), and custom MapReduce programs in Java. Implemented AJAX, JSON, and JavaScript to create interactive web screens.

Environment: Hadoop 2.3, MapReduce, Hive 1.8, Spark 2.0, Scala 1.7, HDFS, AWS, Puppet, Oozie 3.0, Hortonworks, AJAX, JSON, JavaScript, Java

Java/J2EE Developer Amica Mutual Insurance Company - Lincoln, RI August 2011 to December 2012

Responsibilities: Developed and utilized J2EE Services and JMS components for messaging communication in WebSphere Application Server. Developed the code based on the design using Tiles MVC (Struts framework) and using J2EE patterns. Developed the MVC applicationmodel using Spring framework, Spring Boot, Micro services, and used Hibernate framework to interact with the database. Developed the custom Logging framework used to log transactions executed across the various applications using Log4j. Used Maven in building the application and auto deploying it to the environment. Extensively used JQuery to provide dynamic User Interface and for the client-side validations. Developed dynamic proxies to consume the web services developed in JAX-WS standards for CRM module. Written and executed test cases for unit testing using Mockito, JUNIT framework. Designed and developed business components using Session and Entity Beans in EJB. Extensively used Eclipse IDE for developing, debugging, integrating and deploying the application. Used Bit bucket for the repository and version management through SourceTree for GIT. Developed user interface

using JSTL, HTML, JavaScript, JQuery and CSS. Created clean and readable API documentation using Slate to assist our current and future developers. Involved in writing test cases using JUnit and integrated these tests with Jenkins Continuous integration tool. Extensively used Eclipse IDE for developing, debugging, integrating and deploying the application. Involved in exporting and importing integrations and jar files from development, staging and production environments using WebLogic. Developed real time tracking of class schedules using Node JS (socket.io based on socket technology, Express JS framework) Involved in Unit integration, bug fixing, acceptance testing with test cases, Code reviews Extensively used Java Multi-Threading concept for downloading files from a URL. Implemented Service Oriented Architecture by developing Java web services using WSDL and SOAP. Used frameworks such as Angular, backbone.js to a handful of web applications. Involved in fixing defects in application worked in JSF managed beans, converters, validator and configuration files. Mapped business objects to database using Hibernate and used JPA annotations for mapping DB to objects. Environment: JQuery, JavaScript, HTML5, CSS3, JUNIT, Maven, MVC, Struts, Eclipse, JUnit Java Developer Glansa Solutions - Hyderabad, Telangana April 2009 to July 2011 Responsibilities: Worked as a Java Developer and involved in application and database design. Involved in requirement analysis and participated in the design of the application using UML and OO Analysis Design and Development. Around 6 years of IT Industry experience as java developer in software development life cycle core area of object oriented and web-based enterprise applications using java/j2ee technology. Experience in application/web servers like IBM Web Sphere, Web Logic Application Servers, JBoss and Tomcat Web Servers. Proficient in n-tier application design and development using Java, J2EE, JSP, Servlets, Struts 2.0, Spring and Oracle. Expertise in developing applications using Core Java, Multithreading, Collections, Swing development, Memory Management, Application utilities, J2EE, JSP, Servlets, Java Beans, Web Services (SOAP, WSDL, UDDI) AngularJS, JMS, JDBC, JSON, Ajax. Expertise in HTML, XHTML, HTML5, ML5, CSS, AJAX, jQuery, JSTL specification, XML (SAX, DOM), XSL, XSLT, JAXP, JAXB, DTD, Resource properties. Expert in Various Agile methodologies like SCRUM, Test Driven Development (TTD), Incremental and Iteration

methodology, Agile Development & Testing using Software Development Life Cycle. Experienced in Development, Testing and Deployment of enterprise applications on Windows, Linux and UNIX platforms using IDEs such as Eclipse, STS, Rational Application Developer (RAD), NetBeans, IntelliJ14.x/15. Experienced in developing complex PL/SQL queries, Procedures, Triggers, Stored Procedures, Packages and Views in various databases such as Oracle, DB2 and MySQL, SQL Server. Working Knowledge in JSON and XML technologies such as XSL, XSLT, XPath, parsers like SAX, DOM and JAXB. Developed the Application using Spring MVC Framework by implementing Controller, Service classes. Performed JUnit testing to test the implemented services.

Used ANT scripts to build the application and deployed on Web Sphere Application Server. Implemented spring java-based SOAP Web Services for authorization and JUnit tests for part of my code. Managed Spring Core for dependency injection of control (IOC), and integrated with Hibernate. Worked on core Java concepts like Collections and Exception Handling for writing the backend API's. Developed DAO and service layers using the Spring Dao support and Hibernate ORM mapping. Created test cases for DAO Layer and service layer using JUnit and bug tracking using JIRA. Decomposed existing monolithic code base into Spring Boot Microservices. Extensively used Eclipse IDE for developing, debugging, integrating and deploying the application. Involved in implementation of MVC pattern using Angular JS, and Spring Controller. Used JavaScript user input validated using regular expressions and also in the server side. Developed business logic using Java, Struts Action classes and deployed using Tomcat. Used SVN for version control across common source code used by developers. Implemented MVC web frameworks for the web applications using JSP/Servlet/Tag libraries that were designed using JSP. Integrated Subversion (SVN) into Jenkins to automate the code check-out process. Implemented modules using Core Java APIs, Java collection, Threads, XML and integrating the modules. Environment: Java, MVC, JUnit, ANT, Hibernate, JUnit, Eclipse, Angular JS, JavaScript, Struts, MVC Education Bachelor's Skills CASSANDRA, HDFS, IMPALA, MAPREDUCE, OOZIE, SQOOP, HBASE, KAFKA, FLUME, HADOOP, JBOSS, JMS, MONGODB, NOSQL, ANALYSIS SERVICES, APPLICATION SERVER, Git, Hadoop, HBase, Hive Additional Information Technical Skills



Hadoop/Big Data Technologies: Hadoop 3.0, HDFS, MapReduce, HBase 1.4, Apache Pig, Hive 2.3, Sqoop 1.4, Apache Impala 2.1, Oozie 4.3, Yarn, Apache Flume 1.8, Kafka 1.1, Zookeeper    Cloud Platform: Amazon AWS, EC2, EC3, MS Azure, Azure SQL Database, Azure SQL Data Warehouse, Azure Analysis Services, HDInsight, Azure Data Lake, Data Factory    Hadoop Distributions: Cloudera , Hortonworks, MapR    Programming Language: Java, Scala, Python 3.6, SQL, PL/SQL, Shell Scripting, Storm 1.0, JSP, Servlets    Frameworks: Spring 5.0.5, Hibernate 5.2, Struts 1.3, JSF, EJB, JMS    Web Technologies: HTML, CSS, JavaScript, JQuery 3.3, Bootstrap 4.1, XML, JSON, AJAX    Databases: Oracle 12c/11g, SQL    Operating Systems: Linux, Unix, Windows 10/8/7    IDE and Tools: Eclipse 4.7, NetBeans 8.2, IntelliJ, Maven    NoSQL Databases: HBase 1.4, Cassandra 3.11, MongoDB, Accumulo    Web/Application Server: Apache Tomcat 9.0.7, JBoss, Web Logic, Web Sphere    SDLC Methodologies: Agile, Waterfall    Version Control: GIT, SVN, CVS

Name: Audrey Lopez

Email: nstevenson@example.net

Phone: 001-299-228-6637x3876