

Sr. Big Data/Data Engineer Sr. Big Data/Data Engineer Sr. Big Data/Data Engineer - AT & T Jersey City, NJ ? Having 10+ years of overall IT experience in a variety of industries, which includes hands on experience of 7+ years in Big Data Analytics/Data Engineer and development. ? Expertise with the tools in Hadoop Ecosystem including Pig, Hive, HDFS, MapReduce, Sqoop, Storm, Spark, Kafka, Yarn, Oozie, and Zookeeper. ? Excellent knowledge on Hadoop ecosystems such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node and Map Reduce programming paradigm. ? Experience in manipulating/analysing large datasets and finding patterns and insights within structured and unstructured data. ? Strong experience on Hadoop distributions like Cloudera, MapR and Horton works. ? Good understanding of NoSQL databases and hands on work experience in writing applications on NoSQL databases like HBase, Cassandra and MongoDB. ? Experience in search engine Technologies SOLR, Elastic Search etc ? Worked with various HDFS file formats like Avro, Sequence File and various compression formats like Snappy, bzip2. ? Developed Simple to complex MapReduce streaming jobs using Python language that are implemented using Hive and Pig. ? Skilled in developing applications in Python language for multiple platforms. ? Hands on experience in application development using Java, Linux Shell Scripting. ? Experience in Oozie and workflow scheduler to manage Hadoop jobs by Direct Acyclic Graph (DAG) of actions with control flows. ? Experience in migrating the data using Sqoop from HDFS to Relational Database System and vice-versa according to client's requirement. ? Extensive Experience on importing and exporting data using stream processing platforms like Flume and Kafka. ? Strong Knowledge on Apache Spark with Scala Environment. ? Developed Spark scripts by using Scala shell commands as per the requirement. ? Good hands on experience in creating the RDD's, Data frames for the required input data and performed the data transformations using Spark Scala. ? Good knowledge on real time data streaming solutions using Apache Spark Streaming, Kafka and Flume. ? Experience in designing and developing applications in Spark using Scala to compare the performance of Spark with Hive and SQL/Oracle. ? Very good experience in complete project life cycle (design, development, testing and implementation) of Client Server and Web applications. ? Excellent Java development skills using J2EE, J2SE, Servlets, JSP, EJB, JDBC, SOAP and Restful

web services. ? Experience in MicroServices Architecture with Spring Boot and Docker. ? Strong Experience of Data Warehousing ETL concepts using Informatica Power Centre, OLAP, OLTP and Autosys. ? Experienced in working with Amazon Web Services (AWS) using EC2 for computing and S3 as storage mechanism. ? Strong experience in Object-Oriented Design, Analysis, Development, Testing and Maintenance. ? Excellent implementation knowledge of Enterprise/Web/Client Server using Java, J2EE. ? Experienced in using agile approaches, including Extreme Programming, Test-Driven Development and Agile Scrum. ? Worked in large and small teams for systems requirement, design & development. ? Key participant in all phases of software development life cycle with Analysis, Design, Development, Integration, Implementation, Debugging, and Testing of Software Applications in client server environment, Object Oriented Technology and Web based applications. ? Experience in using various IDEs Eclipse, IntelliJ and repositories SVN and Git. ? Experience of using build tools Ant, Maven. ? Preparation of Standard Code guidelines, analysis and testing documentations. ? Good interpersonal skills, committed, result oriented, hard working with a quest and deal to learn new technologies. Work Experience Sr. Big Data/Data Engineer AT & T - Middletown, NJ April 2018 to Present AT&T Inc. is an American multinational conglomerate holding company headquartered at Whit acre Tower in Downtown Dallas, Texas. It is the world's largest telecommunications company, the second largest provider of mobile telephone services, and the largest provider of fixed telephone services in the United States through AT&T Communications.

Responsibilities: ? Configured Spark Streaming to receive real time data from the Kafka and store the stream data to Cassandra utilizing Scala. ? Developed Spark code to read data from Hdfs and write to Cassandra. ? Configured Spark Streaming to receive real time data from the Kafka and store the stream data to Cassandra utilizing Scala. ? Performed real-time analysis of the incoming data using Kafka consumer API, Kafka topics, Spark Streaming utilizing Scala. ? Integrated Apache Storm with Kafka to perform web analytics. Uploaded click stream data from Kafka to HDFS, HBase and Hive by integrating with Storm. ? Performed real-time analysis of the incoming data using Kafka consumer API, Kafka topics, Spark Streaming utilizing Scala. ? Developed Kafka producer and consumers, Cassandra clients and Spark along with components on HDFS. ? Used the Spark -

Cassandra Connector to load data to and from Cassandra. ? Experienced in designing and deployment of Hadoop cluster and various Big Data components including HDFS, Map Reduce, and Zookeeper in Cloudera distribution. ? Implemented advanced procedures like text analytics and processing using the in-memory computing capabilities like Apache Spark written in Scala. ? Implemented Kafka Custom partitions to send data to different categorized topics. ? Implemented messaging system for different data sources using apache Kafka and configuring High level consumers for online and offline processing. ? Used Pig as ETL tool to do transformations, event joins, filter and some pre-aggregations. Migrated ETL jobs to Pig scripts do Transformations, even joins and some pre-aggregations before storing the data onto HDFS. ? Loaded and transformed large sets of structured unstructured data. ? Was responsible to handle a Team of 4 members at Off-shore. ? Involved in daily SCRUM meetings to discuss the development/process. ? Worked with data delivery teams to setup new Hadoop users. This job included setting up Linux users, setting up Kerberos principals and testing HDFS, Hive. ? Worked with data Ingestion Team, have Good understanding of Apache Nifi and its transformations. ? Done Scaling Cassandra cluster based on lead patterns. ? Good understanding of Cassandra Data Modelling based on applications. ? Create Solution Architecture based upon Microsoft Azure PaaS Services ? Design solution for various system components using Microsoft Azure ? Configure & Setup Azure Hybrid Connection to pull data from SAP Systems ? Involved in code deployment to Production and providing support to App support team. ? Good understanding of Hadoop admin work - maintained Hadoop cluster using Ambari. Environment: Hadoop, Spark, Scala, Java, Map Reduce, HDFS, Cassandra, Ambari, Hive, Pig, Sqoop, Flume, Linux, Python, Kafka Storm, Shell Scripting, XML, ETL, Eclipse, Cloudera, DB2, SQL Server, MySQL, AWS, HBase Sr. Big Data Hadoop Developer Scripps Health - San Diego, CA February 2015 to March 2018 Scripps Health is a non-profit health care system based in San Diego, California. The system includes five hospitals and 19 outpatient facilities and treats a half-million patients annually through 2,600 affiliated physicians. The system also includes clinical research and medical education programs. The organization has a number of projects planned, including the Scripps Prebys Cardiovascular Institute, which will serve as a centre for heart disease

treatment, research and graduate medical education. Responsibilities: ? Responsible for building scalable distributed data solutions using Hadoop. ? Developed Pyspark code to read data from Hive, group the fields and generate XML files. Enhanced the Pyspark code to write the generated XML files to a directory to zip them to CDAs. ? Implemented REST call to submit the generated CDAs to vendor website Implemented Impala to support JDBC/ODBC connections for Hiveserver2. ? Configured Spark Streaming to receive real time data from the Kafka and store the stream data to HDFS. ? Integrated Apache Storm with Kafka to perform web analytics. Uploaded click stream data from Kafka to HDFS, HBase and Hive by integrating with Storm. ? Performed real-time analysis of the incoming data using Kafka consumer API, Kafka topics, Spark Streaming utilizing Scala. ? Developed Kafka producer and consumers, Cassandra clients and Spark along with components on HDFS, Hive. ? Good Understanding of DAG cycle for entire Spark application flow on Spark application WebUI. ? Used the Spark - Cassandra Connector to load data to and from Cassandra. ? Written Storm topology to emit data into Cassandra DB. ? Experience with developing and maintaining Applications written for Amazon Simple Storage, AWS Elastic Beanstalk, and AWS Cloud Formation. ? Implemented Python script to call the Cassandra Rest API, performed transformations and loaded the data into Hive. ? Experienced in designing and deployment of Hadoop cluster and various Big Data components including HDFS, MapReduce, Hive, Sqoop, Pig, Oozie, and Zookeeper in Cloudera distribution. ? Implemented advanced procedures like text analytics and processing using the in-memory computing capabilities like Apache Spark written in Scala. ? Implemented Kafka Custom partitioning to send data to different categorized topics. ? Developed end to end data processing pipelines that begin with receiving data using distributed messaging systems Kafka through persistence of data into HBase. ? Implemented messaging system for different data sources using apache Kafka and configuring High level consumers for online and offline processing. ? Written Shell scripts that run multiple Hive jobs which helps to automate different hive tables incrementally which are used to generate different reports using Tableau for the Business use. ? Worked with data delivery teams to setup new Hadoop users. This job included setting up Linux users, setting up Kerberos principals and testing HDFS, Hive. ? Done

Scaling Cassandra cluster based on load patterns. ? Worked with Kafka for the proof of concept for carrying out log processing on a distributed system. ? Performed real-time analysis of the incoming data using Kafka consumer API, Kafka topics, Spark Streaming utilizing Scala. ? Good understanding of Cassandra Data Modelling based on applications. ? Created PIG script jobs in maintaining minimal query optimization. Environment: Hadoop, Java, MapReduce, HDFS, Hive, Pig, Sqoop, Flume, Linux, Python, Spark, Impala, Scala, Kafka Storm, Shell Scripting, XML, Eclipse, Cloudera, DB2, SQL Server, MySQL, Autosys, Talend, AWS, HBase. Hadoop Developer Merck Pharmaceuticals Ltd - Kenilworth, NJ October 2013 to January 2015 Merck is adopting Hadoop to overcome three data challenges to its goal of improving yields in its manufacturing process. First, it needed to combine years of data from multiple data silos within its organization. Secondly, Merck needs to extend both the amount of data it can capture and its ability to retain that data for longer. Thirdly, the Merck team wants to test new hypotheses virtually at a far lower cost than testing those ideas with real-world material and equipment. With Hadoop, Merck plans to overcome these challenges to combine 10 years of vaccine manufacturing data and conduct 5.5 million cross-batch comparisons over 10 billion records. The resulting yield improvement could grow profits by \$10 million dollars for just one vaccine. Responsibilities: ? Installed and configured Hadoop MapReduce, HDFS, developed multiple MapReduce jobs in Java for data cleaning and pre-processing. ? Importing and exporting data into HDFS from Oracle 10.2 database and vice versa using SQOOP. ? Experienced in defining and coordination of job flows. ? Gained experience in reviewing and managing Hadoop log files. ? Extracted files from NoSQL database (MongoDB), HBase through Sqoop and placed in HDFS for processing. ? Involved in Writing Data Refinement Pig Scripts and Hive Queries. ? Good knowledge in running Hadoop streaming jobs to process terabytes of xml format data. ? Load and transform large sets of structured, semi structured and unstructured data. ? Coordinated cluster services using Zookeeper. ? Used XML Technologies like DOM for transferring data. ? Object relational mapping and Persistence mechanism is executed using Hibernate ORM. ? Developed custom validator in Struts and implemented server side validations using annotations. ? Used Oracle for the database and Web Logic as the application

server. ? Used Flume to transport logs to HDFS. ? Experienced in moving data from Hive tables into Cassandra for real time analytics on hive tables. ? Organize documents in more useable clusters using Mahout. ? Configured connection between HDFS and Tableau using Impala for Tableau developer team. ? Responsible to manage data coming from different sources. ? Got good experience with various NoSQL databases. ? Experienced with handling administration activations using Cloudera manager. ? Supported MapReduce programs those are running on the cluster. ? Involved in loading data from UNIX file system to HDFS. ? Installed and configured Hive and also written Hive UDFs. ? Involved in creating Hive tables, loading with data and writing Hive queries which will run internally in MapReduce way. ? Worked on Talend ETL tool, developed and scheduled jobs in Talend integration suite. ? Modified reports and Talend ETL jobs based on the feedback from QA testers and Users in development and staging environments.      Environment: Apache Hadoop, Java, JDK1.6, J2EE, JDBC, Servlets, JSP, Linux, XML, Web Logic, SOAP, WSDL, HBase, Hive, Pig, Sqoop, ZooKeeper, NoSQL, HBase, R, MAHOUT Map-Reduce, Cloudera, HDFS, Flume, Impala, Tableau, Talend, MySQL, HTML5, CSS, MongoDB Java developer Dedicated Tech Services, Inc - Columbus, OH November 2011 to September 2013 The Fuse Box Electronic Cash Register shall provide basic set of services to allow the retail merchant to perform with great ease any function of a standard electronic cash register on the market today. This set of services shall provide a powerful web enabled cash management tool for any merchant.ss      Responsibilities: ? Developed JSP for UI and Java classes for business logic. ? Used XSLT for UI to display XML Data. ? Utilized JavaScript for client side validation. ? Utilized Oracle PL/SQL for database operations using JDBC API. ? Implemented DAO for Oracle 8i for DML Operations like Insert, Update, Delete the records. ? VSS is used for Software Configuration Management. ? Involved in the design, development and deployment of the Application using Java/J2EE Technologies. ? Used IDE tool WSAD for development of the application. ? Developed Application in Jakarta Struts Framework using MVC architecture. ? Customizing all the JSP pages with same look and feel using Tiles, CSS (Cascading Style Sheets). ? Involved in coding for the presentation layer using Apache Struts, XML and JavaScript. ? Created Action Forms and Action classes for the modules. ?

Developed JSP's to validate the information automatically using Ajax. ? Implemented J2EE design patterns viz. Factory pattern, Singleton Pattern. ? Created struts-config.xml and tiles-def.xml files. ? Developed Ant script to create war/ear file and deploy the application to application server. ? Extensively involved in database activities like designing tables, SQL queries in the application and maintained complex reusable methods which implements stored procedures to fetch data from the database. ? Used CVS for version control. ? Also involved in testing and deployment of the application on Web Logic Application Server during integration. Environment: Java/J2EE, JSP, Servlets, Struts 1.1, Spring, JUnit, Eclipse, Apache Ant, JSP, JavaBeans, JavaScript, Tomcat 4.1, Oracle 9i, XML, XSLT, HTML/DHTML/XHTML, CSS, Tiles, Ajax, DB2 UDB, PL/SQL, XML SPY.

Java Developer General Dynamics Information Technology May 2009 to October 2011 Apollo Munich is an insurance company intended to offer medical and health insurance plans that caters to individuals, families and corporate houses as well. Played the role of a Java developer for customer registration plans and claims plans. Responsibilities: ? Involved in projects utilizing Java, Java EE web applications to create fully-integrated client management systems ? Developed UI using HTML, Java Script, JSP and developed business Logic and interfacing components using Business Objects, JDBC and XML. ? Participated in user requirement sessions to analysis and gather Business requirements. ? Development of user visible site using Perl, back end admin sites using Python and big data using core java. ? Involved in development of the application using Spring Web MVC and other components of the ? Elaborated Use Cases based on business requirements and was responsible for creation of class Diagrams, Sequence Diagrams. ? Implemented Object-relation mapping in the persistence layer using Hibernate (ORM) framework. ? Implemented REST Web Services with Jersey API to deal with customer requests ? Experienced in developing Restful web services: consumed and also produced. ? Used Hibernate for the Database connection and Hibernate Query Language (HQL) to add and retrieve the information from the Database. ? Implemented Spring JDBC for connecting Oracle database. ? Designed the application using MVC framework for easy maintainability ? Provided bug fixing and testing for existing web applications. ? Involved in full system life cycle and responsible for Developing, Testing, Implementing. ? Involved

in Unit Testing, Integration Testing and System Testing. ? Implemented Form Beans and their Validations. ? Written Hibernate components. ? Developed client side validations with Java script.

Environment: Spring, JSP, Servlets, REST, Oracle, AJAX, Java Script, JQuery, Hibernate, Web Logic, Log4j, HTML, XML, CVS, Eclipse, SOAP Web Services, XSLT, XSD, UNIX, Maven, Jenkins, shell scripting, MVS, ISPF. Education Bachelor's Skills Amazon web services, Git, Hbase, Hdfs, Hive, Javascript, Jenkins, Json, Pig, Python, Scripting, Svn, Zookeeper, Cassandra, Impala, Oozie, Sqoop, Hbase, Kafka, Db2

Name: Karen Smith

Email: mary68@example.com

Phone: +1-903-715-8956