

Sr. Hadoop Developer Sr. Hadoop Developer Sr. Hadoop Developer - United Technologies Corporation Farmington, CT 7 Years of IT industry experience with 5 years of experience in dealing with Apache Hadoop components like HDFS, MapReduce, Spark, Hive, Pig, Sqoop, Oozie, Zookeeper, HBase, Cassandra, MongoDB and Amazon Web Services. 2 years of experience in the Application Development and Maintenance of SDLC projects using Java technologies. Developed applications for Distributed Environment using Hadoop, MapReduce and Python. Developed MapReduce jobs to automate transfer of data from HBase. Developing and Maintenance the Web Applications using the Web Server Tomcat. Experience in integrating Hadoop with Ganglia and have good understanding of Hadoop metrics and visualization using Ganglia. Good experience working with Hortonworks Distribution, Cloudera Distribution and MapR Distribution. Very good understanding/knowledge of Hadoop Architecture and various components such as HDFS, JobTracker, TaskTracker, NameNode, DataNode, Secondary Namenode, and MapReduce concepts. Experience in data extraction and transformation using MapReduce jobs. Proficient in working with Hadoop, HDFS, writing PIG scripts and Sqoop scripts. Performed data analysis using Hive and Pig. Expert in creating Pig and Hive UDFs using Java in order to analyze the data efficiently. Experience in importing and exporting data using Sqoop from Relational Database Systems to HDFS and vice-versa. Strong understanding of NoSql databases like HBase, MongoDB & Cassandra. Strong understanding of Spark real time streaming and SparkSQL and experience in loading data from external data sources like MySQL and Cassandra for Spark applications. Experience in performing in-memory data processing and real time streaming analytics using Apache Spark with Scala, Java and Python. Created HBase tables to load large sets of structured, semi-structured and unstructured data coming from UNIX, NoSQL and a variety of portfolios. Intensive working experience with Amazon Web Services(AWS) using S3 for storage, EC2 for computing and RDS, EBS. Well versed with job workflow scheduling and monitoring tools like Oozie. Loaded streaming log data from various web servers into HDFS using Flume. Experience in using Sqoop, Oozie and Cloudera Manager. Experience on Source control repositories like SVN, CVS and GIT. Experience in improving the search focus and quality in

ElasticSearch by using aggregations and Python scripts. Hands on experience in application development using RDBMS, and Linux shell scripting. Have experience with working on Amazon EMR and EC2 Spot instances. Solid understanding of relational database concepts. Extensively worked with Unified Modeling Tools (UML) in designing Use Cases, Activity flow diagram, Class diagrams, Sequence and Object Diagrams using Rational Rose, MS-Visio. Very good experience in complete project life cycle (design, development, testing and implementation) of Client Server and Web applications. Good Knowledge on Hadoop Cluster architecture and monitoring the cluster. Adequate knowledge and working experience in Agile & Waterfall methodologies. Support development, testing, and operations teams during new system deployments. Practical knowledge on implementing Kafka with third-party systems, such as Spark and Hadoop. Good team player and can work efficiently in multiple team environments and multiple products. Easily adaptable to the new systems and environments. Possess excellent communication and analytical skills along with a can-do attitude.

Work Experience

Sr. Hadoop Developer United Technologies Corporation - Farmington, CT February 2018 to Present

United Technologies Corporation is a financial services and bank holding company in United States. The objective should be analyzing the huge amount of data and producing large-scale programs that integrate with technology for the clients to achieve customer satisfaction. Implement and deploy custom applications on Hadoop clusters. Providing information security, customer satisfaction and fraud detection solutions using big data analytics. The technical environment includes Spark, Kafka, Scala, Hive and Oozie.

Responsibilities:

Involved in installation, configuration and maintenance of Hadoop clusters for application development with Cloudera distribution. Developed Kafka consumer's API in Scala for consuming data from Kafka topics. Developed end-to-end scalable distributed data pipelines which receiving data using distributed messaging systems Kafka through persistence of data into HDFS with Apache Spark using Scala. Involved in performance tuning of Spark jobs using Cache and using complete advantage of cluster environment. In the framework we just need to mention the table names, schemas and location of source file/ Sqoop parameters etc. and the framework will generate the entire code which includes Workflow.xml. Performed advanced operations like text analytics and

processing, using in-memory computing capabilities of Spark using Scala. Experience in query data using Spark SQL on Spark to implement Spark RDD'S in Scala. Experienced in working with different scripting technologies like Python, UNIX shell scripts. Performed POC on writing the spark applications in Scala, Python and R programming language. Worked on Partitioning, Bucketing, Parallel execution, Map side Joins for optimization of necessary hive queries. Performed Hive QL to create Hive tables and to write Hive queries to perform the data analysis. Experience in collecting log data from web servers and pushed to HDFS using Flume and NoSql database Cassandra. Used Oozie workflow to Manage and scheduling Jobs on a Hadoop Cluster and used Zookeeper for cluster coordination services. Used NIFI for the transformation of data from different components of Big data ecosystem. Worked on different data sources like Oracle, Netezza, MySQL, Flat files etc. and experience with AWS components like Amazon Ec2 instances, S3 buckets and Cloud Formation templates. Used Qlik sense to build customized interactive reports, worksheets, and dashboards. Worked on different file formats like Sequence files, XML files and Map files using Map Reduce Programs. Strong expertise on MapReduce programming model with XML, JSON, CSV file formats. Involved in managing and organizing developers with regular code review sessions by utilizing Agile and Scrum Methodologies. Experience in implementing Spark RDD transformations, actions, data frames, case classes to required data by using Spark core. Implemented Jira for bug tracking and Bit-bucket to code and code review. Implemented apache airflow DAG to find popular items in Redshift and ingest in the main PostgreSQL via a web service call. Implemented Spark applications in data processing project to handle data from various sources and creating DStreams, Data frames on input data which we get from streaming service like Kafka. Environment: Map Reduce, HDFS, Hive, Spark, Spark-SQL, Sqoop, Apache Kafka, Java 7, Cassandra, Scala, Apache Pig 0.14.0, Apache Hive 1.0.0,Oozie, Linux, AWS EC2, Agile development, Oracle 11g/10g, UNIX Shell scripting, Ambari ,TezEclipse and Qlik sense, Cloudera. Hadoop Developer Transamerica - Dallas, TX July 2016 to January 2018 The Transamerica Corporation is an American private holding company for various life insurance companies and investment firms. Transamerica companies are recognized as leading providers of

life insurance, savings and retirement and investment solutions, serving millions of customers throughout the United States and Canada. Responsibilities: Used Cassandra Query Language to design Cassandra database and tables with various configuration options. Developed PIG UDF'S for manipulating the data according to Business Requirements and also worked on developing custom PIG Loaders. Experience in integrating Hadoop with Ganglia and have good understanding of Hadoop metrics and visualization using Ganglia. Involved in the review of functional and non-functional requirements. Practical experience in developing Spark applications in Eclipse with Maven. Strong understanding of Spark real time streaming and SparkSQL. Loading data from external data sources like MySQL and Cassandra for Spark applications. Developed Python and Shell scripts to automate the end-to-end implementation process of AI project. Experience in selecting and configuring the right Amazon EC2 instances and access key AWS services using client tools and AWS SDKs. Knowledge on using AWS identity and Access Management to secure access to EC2 instances and configure auto-scaling groups using CloudWatch. Firm understanding of optimizations and performance-tuning practices while working with Spark. Good knowledge on compression and serialization to improve performance in Spark applications. Performed interactive querying using SparkSQL. Worked on different file formats like Sequence files, XML files and Map files using Map Reduce Programs. Strong expertise on MapReduce programming model with XML, JSON, CSV file formats. Designed and Implemented Partitioning (static, Dynamic) and Bucketing in HIVE.AWS Practical knowledge on Apache Sqoop to import datasets from MySQL to HDFS and vice-versa. Good knowledge on building predictive models focusing on customer service using R programming. Experience in reviewing and managing Hadoop log files. Experience in building batch and streaming applications with Apache Spark and Python. Used the libraries built on Mlib to perform data cleaning and used R programming for dataset reorganizing. Debug CQL queries and implement performance enhancement practices. Strong knowledge on Apache Oozie for scheduling the tasks. Practical knowledge on implementing Kafka with third-party systems, such as Spark and Hadoop. Experience in configuring Kafka brokers, consumers and producers for optimal performance.

Knowledge of creating Apache Kafka consumers and producers in Java. Involved in creating Hive tables, loading with data and writing hive queries which will run internally in map reduce way.

Practical knowledge of monitoring a Hadoop cluster using Nagios and Ganglia. Experience with GIT for version control system. Involved in loading data from UNIX file system to HDFS and developed UNIX scripts for job scheduling, process management and for handling logs from Hadoop. Understanding technical specifications and documenting technical design documents.

Strong skills in agile development and Test-Driven development. Have practical knowledge on implementing Internet of Things (IoT) Environment: Hadoop Cloudera Distribution (CDH4), Java 7, Hadoop 2.5.2, Spark, SparkSQL, Mlib, R programming, Scala, Cassandra, IoT, MapReduce, Apache Pig 0.14.0, Apache Hive 1.0.0, HDFS, Sqoop, Oozie, Kafka, Maven, Eclipse, Nagios, Ganglia, Zookeeper, AWS EC2, GIT, Ambari ,TezEclipse ,UNIX Shell scripting, Oracle 11g/10g, Linux, Agile development.

Hadoop Developer Pacific Life - Newport Beach, CA August 2015 to June 2016 Pacific Life, a leading diversified international group of companies and is one of the top providers of Life insurance in United States. E-Commerce is a web application that facilitates customers to get a fast quote online, originate a policy and service an account. The company has many regional offices in the business of home and insurance. Each regional office sends the details of monthly transactions to the central system. Most of the feed is in the raw text format and the data is in fixed line format.

Responsibilities: Analyzed large amounts of data sets to determine optimal way to aggregate and report on it. Designed and implemented Incremental Imports into Hive tables. Developed and written Apache PIG scripts and HIVE scripts to process the HDFS data.

Involved in defining job flows, managing and reviewing log files. Involved in Unit testing and delivered Unit test plans and results documents using Junit and MR unit. Supported Map Reduce Programs those are running on the cluster.

As a Big Data Developer, implemented solutions for ingesting data from various sources and processing the Data-at-Rest utilizing Big Data technologies such as Hadoop, MapReduce Frameworks, HBase, Hive, Oozie, Flume, Sqoop etc. Imported Bulk Data into HBase Using Map Reduce programs. Perform analytics on Time Series Data exists in HBase using HBase API. Involved in collecting, aggregating and moving data from servers to

HDFS using Apache Flume. Written Hive jobs to parse the logs and structure them in tabular format to facilitate effective querying on the log data. Responsible for continuous monitoring and managing Elastic MapReduce cluster through AWS console. Wrote multiple java programs to pull data from HBase. Involved with File Processing using Pig Latin. Involved in creating Hive tables, loading with data and writing hive queries that will run internally in map reduce way. Experience in optimization of Map reduce algorithm using combiners and partitions to deliver the best results and worked on Application performance optimization for a HDFS cluster. Worked on debugging, performance tuning of Hive & Pig Jobs. Used Hive to find correlations between customer's browser logs in different sites and analyzed them to build risk profile for such sites. Implemented business logic by writing UDFs in Java and used various UDFs from Piggybanks and other sources. Created and maintained Technical documentation for launching HADOOP Clusters and for executing Hive queries and Pig Scripts. Environment: Java, Hadoop 2.1.0, Map Reduce2, Pig 0.12.0, Hive 0.13.0, Linux, Sqoop 1.4.2, Flume 1.3.1, Eclipse, AWS EC2, and Cloudera CDH 4.

Java Developer Aegis Consulting Services - IN December 2013 to May 2015 Aegis Consulting Services keeps abreast with the state of art technologies. Knowledge of the technologies are carefully analyzed for suitability and appropriately deployed with the solution. The company has in-depth domain as well as functional knowledge. The company's vision is to create a cohesive group of technology creative talent, which holds the capacity to deliver effective results while deploying optional resources. Responsibilities: Involved in Analysis, Design, Implementation and Bug Fixing Activities. Designing the initial Web-WAP pages for a better UI as per the requirement. Involved in design of basic Class Diagrams, Sequence Diagrams and Event Diagrams as a part of Documentation. Developed SQL queries and Stored Procedures using PL/SQL to retrieve and insert into multiple database schemas. Involved in Functional & Technical Specification documents review and the code review. Undergone training on the Domain Knowledge. Discussions and meetings with the Business Analysts for understanding the functionality involved in Test Cases review. Prepared the Support Guide containing the complete functionality. Environment: Core Java, ApacheTomcat5.1, Oracle 9i, Java Script, HTML, PL/SQL, Rational Rose, Windows XP,

UNIX. Software Engineer Wind Stream Communications - IN June 2012 to November 2013 Wind Stream Communications, Inc. was established in 2008 with the goal of designing, prototyping and manufacturing affordable and scalable renewable energy technologies for a global marketplace.

Responsibilities: Designed, developed and executed Data Migration from Db2 Database to Oracle Database using Linux scripts, Java and SQL loader concepts. A key member of the team and playing a key role in articulating the Design requirements for the Development of Automated tools that perform error free Configuration. Developed UNIX and java utilities for Data migration from Db2 to Oracle. Sole developer and POC for the migration Activity. Developed JSP pages, Servlets and HTML pages as per requirement. Developed the necessary Java Beans, PL/SQL procedures for the implementation of business rules. Developed user interface using JAVA Server Pages (JSP), HTML and Java Script for the Presentation Tier. Developed JSP pages and client-side validation by java script tags. Developed an own realm for Apache Tomcat Server for authenticating the users. Developed front end controller in Servlet to handle all the requests. Developed the web interface using JSP and developed struts action classes. Responsible for both functional and non-functional requirements gathering, performing impact analysis and testing the solutions build on build basis. Coding using Java, Java Script and HTML. Used JDBC to provide database connectivity to database tables in Oracle. Used WebSphere Application Server for application deployment. Implemented Software Development Life Cycle (Requirements Analysis, Design, Development, Testing, Deployment and Support). Environment: J2EE, IBM DB2, IBM WebSphere Application Server, EJB, JSP, Servlets, HTML, CSS, JavaScript, Oracle database, Unix Scripting and Windows 2000. Skills CASSANDRA, MAPREDUCE, OOZIE, SQOOP, HBASE, KAFKA, ELASTICSEARCH, FLUME, HADOOP, APPLICATION SERVER, Git, Hadoop, HBase, Hive, HTML, JAVASCRIPT, MapReduce, Pig, PYTHON, PYSPARK Additional Information TECHNICAL SKILLS: Programming languages C, Java, Python, Scala, SQL HADOOP/BIG DATA MapReduce, Spark, SparkSQL, PySpark, SparkR, Pig, Hive, Sqoop, HBase, Flume, Kafka Cassandra, Yarn, Oozie, Zookeeper, ElasticSearch Databases MySQL, PL/SQL, Mongo DB, HBase, Cassandra. Operating Systems Windows, Unix, Linux, Ubuntu. Web Development HTML,

JSP, JavaScript, JQuery, CSS, XML, AJAX. Web/Application Servers Apache Tomcat, Sun Java
Application Server Tools IntelliJ, Eclipse, Net Beans, Nagios, Ganglia, Maven Scripting BASH,
JavaScript Version Controls GIT, SVN

Name: Troy Martin

Email: shelbydudley@example.net

Phone: 207.605.8999x286