

Hadoop Developer Hadoop Developer Hadoop Developer Overall, 5+ years of IT experience & result oriented Hadoop Developer possessing a proven track record of effectively administering Hadoop ecosystem components & architecture and managing file distribution systems in the Big Data arena. Proficient in collaborating with key stakeholders to conceptualize & execute solutions for resolving systems architecture-based technical issues. Highly skilled in processing complex data designing Machine Learning modules for effective data mining & modeling. Adept at Hadoop cluster management & capacity planning for end-to-end data management & performance optimization.

Hadoop Developer with work experience in using HDFS, MapReduce, Hive, Pig, Spark, Sqoop, Oozie, Kafka, zookeeper, and HBase. Experienced working with various Hadoop Distributions (Cloudera, Hortonworks, Map R, Amazon EMR, Microsoft Azure HDInsight) to fully implement and leverage new Hadoop features. Proficient in using Unix based Command Line Interface.

Experience in moving data into and out of the HDFS and Relational Database Systems (RDBMS) using Apache Sqoop. Expertise in working with Hive data warehouse infrastructure-creating tables, data distribution by implementing Partitioning and Bucketing, developing and tuning the HQL queries. Involved in creating Hive tables, loading with data and writing Hive Ad-hoc queries that will run internally in MapReduce and Spark. Significant experience writing custom UDF's in Hive and custom Input Formats in MapReduce. Experience in managing and reviewing Hadoop log files. Knowledge of job work flow management and monitoring tools like Oozie, and zookeeper.

Experience working with NoSQL database technologies, including MongoDB, Cassandra, and HBase. Strong experience building end to end data pipelines on Hadoop platform. Experience in developing Spark Applications using Spark RDD, Spark SQL and Data frame APIs. Replaced existing MR jobs and Hive scripts with Spark SQL & Spark data transformations for efficient data processing. Deep knowledge of troubleshooting and tuning Spark applications and Hive scripts to achieve optimal performance. Strong understanding of real time streaming technologies Spark and Kafka. Worked with real-time data processing and streaming techniques using Spark streaming and Kafka. Good understanding on Machine Learning Methodologies to uncover the hidden patterns, user behaviour, and modeling using Spark MLlib. Java Developer with a year of

diversified experience in analysis, design, implementation, integration, testing and maintenance of applications using Java/J2EE and Object-Oriented Client-Server technologies. Strong understanding of Java Virtual Machines and multi-threading process. Expert in Core Java with strong understanding of Collections, Multithreading, Event handling, Exception handling and Generics. Strong experience with Software methodologies like Agile (Scrum), Waterfall and Test Drive Development. Significant experience in UI frame works such as JSP, HTML5, XML, CSS3, JavaScript, Angular JS, JQuery. Extensive experience in developing web applications using Java, JEE, Spring, Servlets, Hibernate, JDBC, Eclipse. Hands-on experience working with Continuous Integration (CI) build-automation tools such as Maven, SVN, CVS, Jenkins and Apache Ant. Experienced in generating logging by Log4j to identify the errors in production test environment and experienced in Ant and Maven tools. Database design, modeling, migration and development experience in using stored procedures, triggers, cursor, constraints and functions. Used My SQL, MS SQL Server, DB2, and Oracle. Effective communication and interpersonal Skills, an excellent team player work towards the growth of an organization.

**Work Experience Hadoop Developer**

**Express Scripts - Franklin Lakes, NJ August 2015 to May 2017**

**Description:** Express Scripts is one of the most significant pharmacy benefit management (PBM) organization in the United States produces an incredible amount of data. Using big data to generate actionable insights helps to improve healthcare by making it more accessible, affordable, and effective. The scope of the project was to address the opioid epidemic through proactive intervention by advanced analytics using big data, avoiding downstream costs and adverse health events using machine learning algorithms, and predicting personalized risk to influence better health using advanced analytics, and artificial intelligence provides the cutting-edge capability to determine the health risk of individuals.

**Responsibilities:** Utilized Sqoop, Kafka, Flume and Hadoop File System API's for implementing data ingestion pipelines. Worked on real time streaming, performed transformations on the data using Kafka and Spark Streaming. Created storage with Amazon S3 for storing data. Worked on transferring data from Kafka topic into AWS S3 storage. Created Hive tables, loaded with data, and wrote Hive queries to process the data. Created Partitions and used Bucketing on Hive tables

and used required parameters to improve performance and developed Hive UDFs as per business use-cases. Developed Hive scripts for source data validation and transformation. Automated data loading into HDFS and Hive for pre-processing the data using Oozie. Designed and implemented an ETL framework using Java and Pig to load data from multiple sources into Hive and from Hive into Vertica. Collaborated in data modeling, data mining, Machine Learning methodologies, advanced data processing, ETL optimization. Worked on various data formats like Avro, Sequence File, JSON, Map File, Parquet, and XML. Worked extensively on AWS Components such as Airflow, Elastic Map Reduce (EMR), Athena, Snowflake. Used Apache NiFi to automate data movement between different Hadoop components. Used NiFi to perform conversion of raw XML data into JSON, Avro. Experienced in working with Hadoop from Cloudera Data Platform and running services through Cloudera manager. Assisted in Hadoop administration and support activities for installations and configuring Apache Big Data Tools and Hadoop clusters using Cloudera Manager. Experienced in Hadoop Production support tasks by analysing the Application and cluster logs. Used Agile Scrum methodology/ Scrum Alliance for development.

Environment: Hadoop, HDFS, AWS, Vertica, Scala, Kafka, MapReduce, YARN, Drill, Spark, Pig, Hive, Scala, Java, NiFi, HBase, MySQL, Kerberos, Maven. Java Developer Hadoop - Hyderabad, Telangana June 2014 to June 2015 Indus Group - Hyderabad, India Description: Indus Group is a technology consulting firm founded by proficient performers in the technical solutions with a mission to provide on-time, on-budget and quality service to the clients across globe and consistently meet their expectations. As a consulting partner, Indus group also suggest best practices, appropriate frameworks, and optimal solutions to the clients. The project is responsible for design, development, management of java-based application, and had deep insight about using Big Data technologies.

Responsibilities: Standardized practices for data acquisition & analysis to deliver new products using Big Data technologies. Directed data streaming in Kafka & deployed Scrum methodologies for data management and analytics in Jira Involved in complete project Life Cycle, i.e. Design, Implementation, Unit Testing. Extensively used agile development methodology and involved in sprint planning. Designed and modified User Interfaces using JSP, JavaScript, HTML5, Angular

JS and jQuery with the help of several design patterns like Singleton, Factory and MVC. Involved in migrating legacy projects to latest versions of spring and hibernate. Used DAO to handle connection and to retrieve data from data storage elements. Written Microservices to export/import data and task scheduling using Spring Boot, Spring and Hibernate. Also Used Swagger API tools while developing the microservices. Implemented Hibernate to persist the data into Database and wrote HQL based queries to implement CRUD operations on the data. Maintained Sessions using Spring MVC session management tools. Annotated POJOs are created using Hibernate annotations. Familiarized with Named Queries and Parameterized Queries in Hibernate. Also Worked on SQL, PL/SQL using SQL Developer for Oracle database. Involved in deploying the application under Apache Tomcat and maintained application logs Using Log4j. Involved in unit testing using JUnit. Used MAVEN to define the dependencies / plug-in and build the application. Used SVN version Control tools. Used Jenkins for deploying the application to test and production environments. Designed and Developed Web services using SOAP to make submissions. Created and maintained various Message Queues and Message brokers that were a part of the application. JMS is used extensively in the application for sending budget related alerts through SMS, email etc. Environment: Hadoop, Kafka, Java 7, Spring 4, Spring MVC, Spring AOP, Spring Data, JPA, Hibernate 3, SQL, Microservices, Spring Boot, RESTful web services, JSON, JUnit 4, SVN, Java script, Log4J, Jenkins, Tomcat, Jira. Education Master's Skills APACHE KAFKA (2 years), JAVA (2 years), JSON (2 years), KAFKA (2 years), APACHE HADOOP HDFS (1 year)

**Additional Information TECHNICAL SKILLS:** Operating Systems Linux (Ubuntu, CentOS), Windows, Mac OS Hadoop Ecosystem Hadoop, MapReduce, Yarn, HDFS, Pig, Oozie, Zookeeper Big Data Ecosystem Spark, Spark SQL, Spark Streaming, Spark MLlib, Hive, Impala, Hue Data Ingestion Sqoop, Flume, NiFi, Kafka NOSQL Databases HBase, Cassandra, MongoDB, CouchDB Programming Languages C, C++, Scala, Core Java, J2EE Frameworks Model View Controller (MVC), Spring, Hibernate Web Technologies Spring 4, Spring MVC, Hibernate 3, JSP, JavaScript, AngularJS, HTML 5, CSS 3, AJAX, JQuery, XML, XSD, WSDL, JSON, Web services Scripting Languages Java Script, UNIX, Python, R Language Databases

Oracle 10g/11g, PostgreSQL 9.3, MySQL, SQL-Server, Teradata    IDE   IntelliJ, Eclipse, Visual Studio, IDLE    Web Services   Restful, SOAP    Tools    Ant, Tortoise SVN, Putty, Win SCP, Maven, log4j, JUnit, SOAPUI, Git, Jasper reports, Jenkins, Tableau, Mahout    Methodologies   SDLC, Agile, Scrum, Iterative Development, Waterfall Model    image1.emf

Name: Albert Barnes

Email: mariajones@example.com

Phone: 001-820-833-7075