

Sr. Big Data Developer Sr. Big Data Developer Sr. Big Data Developer - JPMorgan Chase Texas City, TX 6+ years of IT experience including 4 years of working with Big Data and Cloudera. Involved in various SDLC methods from analysis, design, development, testing, implementation and maintenance with timely delivery against aggressive deadlines in both Agile/Scrum and Waterfall methodology. Good experience in installing, configuring, and leveraging the Hadoop ecosystem to glean meaningful insights from semi-structured and unstructured data. Excellent Communication, Management and Presentation skills. Strong working experience with ingestion, storage, processing and analysis of big data. Successfully loaded files to HDFS from Oracle, Sql Server and Teradata using Sqoop. Extensive experience with ETL and Query tools for Big Data like Pig Latin and HiveQL. Experience in developing data pipeline using Kafka, Spark and Hive to ingest, transform and analyzing data. Experience in Data modeling and connecting Cassandra from Spark and saving summarized data frame to Cassandra. Exploring with Spark for improving the performance and optimization of the existing algorithms in Hadoop using Spark Context, Spark-SQL, Data Frame, Pair RDD's, Spark Yarn. Developing applications using Scala, Spark SQL and MLlib along with Kafka and other tools as per requirement then deployed on the Yarn cluster. Experience in building high performance and scalable solutions using various Hadoop ecosystem tools like Pig, Hive, Sqoop, Spark, Solr and Kafka. Good knowledge in distributed coordination system ZooKeeper and experience with Data Warehousing and ETL Good knowledge in job workflow scheduling and monitoring tools like Oozie and Zookeeper. Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs, Spark SQL using Scala. Very good understanding of Partitions, bucketing concepts in Hive and designed both Managed and External tables in Hive to optimize performance Experience in developing ETL scripts for data acquisition and transformation using Informatica and Talend. Experience in working with databases, such as Oracle, SQL Server, My SQL. Good understanding of Software Development Life Cycle (SDLC) and sound knowledge of project implementation methodologies including Waterfall and Agile. Proficient in Java, J2EE, Servlets, JSP, spring, Hibernate. Strong Experience on Application/Platform Consolidation and Re-hosting, Legacy Conversion/Retirement,

ETL & ELT data pipeline development for operational data stores and analytical warehouses. Well verse and hands on experience in Version control tools like GIT and SVN. Experience on working structured, unstructured data with various file formats such as Avro data files, xml files, JSON files, sequence files, ORC and Parquet. Experience in job workflow scheduling and monitoring tools like Oozie. Experience in Amazon AWS to spin up the EMR cluster to process the data which is stored in Amazon S3

Work Experience

Sr. Big Data Developer JPMorgan Chase - Plano, TX February 2018 to Present

Responsibilities: Involved in project Life Cycle - from analysis to production implementation, with emphasis on identifying the source and source data validation, developing logic and transformation as per the requirement and creating mappings and loading the data into different targets

- Loaded periodic incremental imports of structured batch data from various RDBMS to HDFS using Sqoop
- Implemented Kafka consumers for HDFS and Spark Streaming
- Used Spark Streaming to preprocess the data for real-time data analysis
- Worked on basic Shell Scripting to put data from sources to HDFS and S3. Scheduled the scripts using cron tab
- Involved in writing query using Impala for better and faster processing of data. Implemented Partitioning in Impala for faster and efficient data access
- Worked on reading multiple data formats such as Avro, Parquet, ORC, JSON including Text
- Spark transformation scripts using API's like Spark Core and Spark SQL in Scala
- Worked on writing custom Spark Streaming API's to ingest the data to Elastic Search post the data enrichment in Spark.
- Worked on Apache NiFi in implementing basic workflows using prebuilt processors
- Worked with the team in visualizing data using Tableau

Experienced as Senior ETL Developer (Hadoop ETL/ Teradata /Vertica / Informatica / Datastage/ Mainframe), Subject Matter Expertise (SME), Production Support Analyst, QA Tester

Environment: Spark Streaming, Spark SQL, Spark Core, HDFS, S3, EMR, Impala, Kafka, Sqoop, Oozie, Cloudera Manager, Apache NiFi, Zoo Keeper.

Sr. Hadoop Developer Allergan - Chicago, IL June 2016 to December 2017

Responsibilities:

- Extracted the data from Oracle, Teradata into HDFS using Sqoop
- Created and worked Sqoop (version 1.4.3) jobs with incremental load to populate Hive External tables
- Used Flume to collect, aggregate, and store the log data from web servers
- Developed UDFs in Java as and when necessary to use in Hive queries
- Used Spark Core and

Spark SQL for transformations in python based of the business requirements Developed Oozie workflow for scheduling and orchestrating the ETL process Used JIRA for incident creation, bug tracking and change management process Worked on developing automation scripts using python

Good Working knowledge of Tableau Environment: HDFS, PySpark, Core Spark, Spark SQL, Sqoop, Flume, Oozie, WinSCP, UNIX Shell Scripting, HIVE, Cloudera (Hadoop distribution), AWS, EC2, EMR, S3, JIRA, Oracle etc. Hadoop Developer Sabre - Dallas, TX May 2015 to May 2016

Responsibilities: Ingested Batch Files into HDFS using shell scripting. Used flume to ingest near-real-time data and perform necessary transformations and aggregations on the fly and persisted the data in Hive. Used Hadoop's Pig, Hive and Map Reduce for analyzing the data and to help by extract data sets for meaningful information. Building high resource intensive, low latency and SLA stringent processes by establishing Data Ingestion from Data lake in Hadoop, ETL/ELT jobs development using Scala API in Spark, Spark SQL, Hive QLs for few ETL activities, Hive as operational data store, Hbase/Teradata as historical/production data. Developed workflow in Oozie to orchestrate a series of Pig scripts to cleanse data, such as removing irrelevant information or merging many small files into a handful of very large, compressed files using pig pipelines in the data preparation stage. Extensively used PIG to communicate with Hive using HCatalog. Implemented exception tracking logic using Pig scripts. Saved the analyzed data to the Hive Tables for visualization and to generate reports for the BI team. Environment: Hadoop, Map Reduce, HDFS, Hive, Pig, Oozie, Core Java, Python, Eclipse, Flume, Cloudera, Oracle, UNIX Shell Scripting. Java/J2EE Developer Bajaj Allianz - Mumbai, Maharashtra January 2014 to March 2015 Responsibilities Worked in SDLC methodology followed Waterfall environment including Acceptance Test Driven Design and Continuous Integration/Delivery. Responsible for analyzing, designing, developing, coordinating and deploying web based application. Developed the application using Spring MVC Framework that uses Model View Controller (MVC) architecture with JSP as the view. Used Spring MVC for the management of application flow by developing configurable handler mappings, view resolution. Used Spring Framework to inject the DAO and Bean objects by auto wiring the components. Developed front end applications using the HTML,

CSS, JavaScript, and JQuery. Designed and developed XSLT transformation of components to convert data from XML to HTML. Implemented the project using JAX-WS based Web Services using WSDL, UDDI, and SOAP to communicate with other systems. Monitored the error logs using Log4J Maven is used as a build tool and continuous integration is done using Jenkins. Used complex queries like SQL statements and procedures to fetch the data from the database. Used version control repository GIT and Service now for issue tracking. Developed test cases and performed unit testing using Junit Test cases. Used ANT as build tool and developed build file for compiling the code of creating WAR files. Used Tortoise SVN for Source Control and Version Management. Environment: Java, J2EE, MVC, JUnit, JavaBeans, HTML, CSS, JavaScript, JQuery, Oracle, Hibernate, SQL, Soap, Eclipse, ANT, Maven Software Developer Infor - Hyderabad, Telangana May 2013 to December 2014 Responsibilities: Involved in daily scrums and weekly meeting with the project sponsors. Designed and developed abstract classes, interfaces classes to construct the business logic using Object Oriented Concepts. Developed the classes using Java, which incorporate N -tier architecture and database connectivity. Developed and tested user-friendly navigators by utilizing JavaScript and JQuery. Performed Manual Unit Testing for all units in developed pages. Designed and developed rich user interfaces with HTML/CSS/JavaScript/Bootstrap. Involved in writing Stored procedures and Views as per the requirement. Environment: Java, Spring, Hibernate, JSP, JavaScript, JQuery, Tomcat Apache, MySQL, HTML, CSS, Bootstrap, Eclipse IDE, Maven, MySQL workbench, Eclipse IDE, Apache Subversion(SVN), Junit, JIRA. Education Bachelor's in information Technology in information Technology Gurunanak Institute of Technology Skills Dynamodb, Cassandra, Ambari, Hdfs, Impala, Sqoop, Hbase, Kafka, Data visualization, Etl, Flume, Hadoop, Map reduce, MongoDB, Nosql, Avro, Git, Gradle, Hadoop, Hbase

Name: Michael Bishop

Email: ashley66@example.org

Phone: 333.205.0744x60773