

Ashutosh Reddy FNU
Phone: 510-901-8303
Email: asuthosh.hadoopdev@gmail.com

Summary:

- Over 8+ years of professional IT experience which includes experience in Big data ecosystem experience in complete project life cycle (design, development, unit testing and implementation) of which over 3 years of work experience in ingestion, storage, querying, processing and analysis of Big Data with hands on experience in Hadoop Ecosystem (YARN, HDFS) and its components Hive, HBase, Sqoop, Hue, Kafka, Oozie, Zookeeper, Spark, Spark SQL and Spark Streaming.
- Hands on experience in improving the performance and optimization of the existing algorithms in Hadoop using Spark context, Spark-SQL, Data Frame, pair RDD's, Spark YARN and experience in application in Apache Spark using python (PySpark).
- Exposure to Data Lake Implementation using Apache Spark and developed Data pipe lines and applied business logics using Spark and used Scala and Python to convert Hive/SQL queries into RDD transformations in Apache Spark.
- Experienced in Cluster maintenance and Commissioning /Decommissioning of Data Nodes and good understanding/knowledge of Hadoop Architecture and various components such as HDFS, Job Tracker, and Task Tracker, NameNode, DataNode and MapReduce concepts.
- Experienced in performance tuning of Yarn, Spark, and Hive and experienced in developing MapReduce Programs using Apache Hadoop for analysing the big data as per the requirement.
- Working experience on building spark applications using build tools like SBT, Maven and Gradle.
- Good experience in dealing with different file formats like text, Sequence, RCFILE, ORC, Parquet, Avro and JSON and different compression formats like GZip, LZO, BZip2 and snappy.
- Good knowledge on relational databases like MySQL, Oracle and NoSQL databases like HBase.
- Working experience in handling semi/un-structured data from different data sources.
- Working experience in developing Map side join, Reducer side join, Distributed Cache, Compression techniques, Multiple Input & output.
- Good working experience in performing ad-hoc analysis on structured data using HiveQL.
- Good knowledge in using IDE like Net Beans, Eclipse, IntelliJ.
- Experienced Schedule Recurring Hadoop Jobs with Apache Oozie and experience in Jumpstart, Kickstart, Infrastructure setup and Installation Methods for Linux.
- Major strengths are familiarity with multiple software systems, ability to learn quickly new technologies, adapt to new environments, focused adaptive and quick learner with excellent interpersonal, technical and communication skills.

Technical Expertise

Operating System:	Windows, Ubuntu, Unix
Hadoop Ecosystem:	MapReduce, HDFS, Spark, HBase, Hive, Kafka, Zookeeper, Sqoop, Oozie, PySpark
Programming Language:	C, C++Python, Java, Scala, SQL
Databases:	MySQL, Oracle, MongoDB
Scripting Languages:	Shell Scripting, HTML scripting
Web Technologies:	JavaScript, HTML, CSS, XML
Version Control:	Git
IDEs:	Eclipse for Java, Visual Studio, IntelliJ

PROFESSIONAL EXPERIENCE:

Client: East West Bank, Palo Alto, CA

Role: Hadoop/Spark Developer

September 2017 - August 2019

Responsibilities:

- Using Sqoop, ingested the Data from data warehouse to HDFS.
- Using Kafka, collected real-time streaming and log data from web applications and click stream data, analyzing a part of data using spark streaming and rest stored into HDFS for future use.
- Worked in creating External and Managed Hive tables.
- Worked in writing Hive Queries for analyzing data in Hive warehouse using Hive Query Language (HiveQL) and worked with Hive Tables, Hive queries, Partitioning, Bucketing.
- Performed Data Profiling, identify data quality and validating rules regarding data integrity and data quality as it relates to the impact on business requirements.
- Build spark applications using SBT builds.
- Involved in migration of ETL processes from Oracle to Hive to test the easy data manipulation and worked on importing and exporting data from Oracle and DB2 into HDFS and HIVE using Sqoop.
- Managed and reviewed Hadoop log files.
- Used Spark SQL to process the huge amount of structured data.
- Worked on Python plugin on MySQL workbench to upload CSV files
- Created Rich dashboards using Tableau Dashboard.
- Setup and benchmarked **Hadoop/HBase** clusters for internal use
- Connected Tableau server to publish dashboard to a central location for portal integration.

Environment: Cloudera Manager, Sqoop, Java (jdk1.8 Version), Hive, Spark, Spark-SQL, Spark Streaming, Scala, Python, PySpark, Oracle, Zookeeper.

Client: United Airlines, Chicago, IL

Role: Hadoop Developer

Dec 2015 – Aug 2017

Responsibilities:

- Involved in data extraction from distributed RDBMS like Teradata and Oracle.
- Implemented Partitioning, Dynamic Partitioning, and Bucketing in HIVE.
- Exported the result set from HIVE to MySQL using Shell scripts.
- Involved in loading data from UNIX file system to HDFS.
- Good experience in handling data manipulation using python Scripts.
- Developed data pipeline using Kafka to ingest customer behavioral data and purchase histories into HDFS for analysis.
- Troubleshooting the cluster by reviewing Hadoop LOG files. Involved in managing and reviewing Hadoop log files.
- Used Oozie to manage the Hadoop jobs.
- Involved in running Hadoop streaming jobs to process terabytes of text data.
- Load and transform large sets of structured, semi structured and unstructured data.
- Imported data using Sqoop from Teradata using Teradata connector.
- Used Zookeeper for various types of centralized configurations.
- Involved in maintaining various Unix Shell scripts.
- Automated all the jobs starting from pulling the Data from different Data Sources like MySQL to pushing the result set Data to Hadoop Distributed File System using Sqoop.
- Used GIT for version control.
- Maintain System integrity of all sub-components (primarily HDFS, and HBase).
- Monitor System health and logs and respond accordingly to any warning or failure conditions.

Environment: Hadoop, HDFS, Hive, Sqoop, Hue, NoSQL, Java 1.6.

Client: Amedha Info systems

Feb 2014 - Oct 2015

Role: Sr. Java Developer

Roles and Responsibilities:

- Involved in various phases of Software Development such as modeling, system analysis and design, code generation and testing using AGILE Methodology
- Participated in daily Stand up meetings with Scrum Master.
- Responsible for estimating the time for given task (Service Request)
- Designed, developed and deployed application using Eclipse and Tomcat application Server
- Developed SQL queries and used JDBC to interact with the Oracle database
- Configured and customized logs using Log4J
- Involved in installing and configuring Eclipse and Maven for development
- Used SVN as a version management tool
- Written Test Cases for Unit Level Testing using JUnit
- Developed database objects like Stored Procedures, Functions, Triggers, index, views to maintain referential integrity of the database
- Responsible to monitor all the tasks assigned to offshore and ensure they are delivered with in SLA
- Part of the Maintenance and Enhancement Group which provides on-going support for systems after implementation occurs

Environment: Spring Framework, Hibernate, PL/SQL, Tomcat 7.x, log4j, JUnit, Oracle 10g, Eclipse, AGILE Methodology, SCRUM, SOAP Webservices, Unix Shell Scripting, SVN, Oracle 9i, Maven 4

Client: Nishi Technologies, Hyderabad, India

June 2011- Jan 2014

Role: Java/Web Developer

Responsibilities:

- Involved in various phases of Software Development Life Cycle (SDLC) as design development and unit testing.
- Developed and deployed UI layer logics of sites using JSP, XML, JavaScript, HTML/DHTML and Ajax.
- CSS and JavaScript were used to build rich internet pages.
- Agile Scrum Methodology been followed for the development process.
- Designed different design specifications for application development that includes front-end, back-end using design patterns.
- Developed proto-type test screens in HTML and JavaScript.
- Involved in developing JSP for client data presentation and, data validation on the client side with in the forms.
- Developed the application by using the Spring MVC framework.
- Collection framework used to transfer objects between the different layers of the application.
- Developed data mapping to create a communication bridge between various application interfaces using XML, and XSL.
- Spring IOC being used to inject the parameter values for the Dynamic parameters.
- Developed JUnit testing framework for Unit level testing.
- Actively involved in code review and bug fixing for improving the performance.
- Documented application for its functionality and its enhanced features.
- Created connection through JDBC and used JDBC statements to call stored procedures.

Environment: Spring MVC, J2EE, Java, JDBC, Servlets, JSP, XML, Design Patterns, CSS, HTML, JavaScript, Junit, Apache Tomcat, My SQL Server 2008.

Education

Bachelor's in Computer Science

2011

Nitte Meenakshi Institute of Technology, Bangalore