Abhinav Reddy

EMAIL- *abhihad12@gmail.com*
Phone- *484 341 0318*

PROFESSIONAL SUMMARY

- Experience around 5+years in IT industry with complete software development of life cycle (SDLC) which includes business requirements gathering, system analysis & design, data modeling, development, testing and implementation of the projects.
- Experienced in configuration, deployments and managing of different Hadoop distributions like Cloudera (CDH4 & CDH5) and Hortonworks (HDP).
- Experience of import/export data using Sqoop from Hadoop distributed file systems to relational database systems and vice versa.
- Experience in handling various file formats like AVRO, Sequential, text, xml, JSON and Parquet with different compression techniques such as gzip, LZO, Snappy etc.
- Experienced on Spark Architecture including Spark Core, Spark SQL, Data Frames, Spark Streaming and Spark MLlib.
- Imported the data from source HDFS into Spark Data Frame for in-memory data computation to generate the optimized output response and better visualizations.
- Good knowledge on Spark (spark streaming, spark SQL), Scala and Kafka.
- Experience in implementing Spark integration with Hadoop ecosystem.
- Experience in designing and developing Spark applications using Scala.
- Expertise in writing Spark RDD transformations, actions, Data Frame's, case classes for the required input data and performed the data transformations using Spark-Core also convert RRD to Data Frame.
- Experienced on collection the real time streaming data and creating the pipeline for row data from different source using Kafka and store data into HDFS and NoSQL using Spark.
- Extending HIVE core functionality by using custom User Defined Function's (UDF) and User Defined Aggregating Functions (UDAF).
- Implemented POC for using Impala for data processing on top of HIVE for better utilization of C++ executions engines.
- Experience in NoSQL Databases HBase, Cassandra and it's integrated with Hadoop cluster.
- Implemented Cluster for NoSQL tools HBase as a part of POC to address HBase limitations.
- Exploring with Spark Beta version API to improve the performance, and optimization of the existing algorithms with different modes such as YARN, Mesos and standalone for POC.
- Expertise in using ETL Tool Informatica Power Center designer, workflow manager, repository manager, data quality and ETL concepts.
- Experienced with NiFi to automate the data movement between different Hadoop systems.
- Worked with different Hadoop Security such as Knox and Ranger integrated LDAP store with Kerberos KDC.
- Good understanding on security requirements for Hadoop and integrate with Kerberos authentication and authorization infrastructure.
- Experienced on cloud integration with AWS using Elastic Map Reduce (EMR), Simple Storage Service (S3), EC2, Redshift and Microsoft Azure.
- Experienced on different Relational Data Base Management Systems like Teradata, PostgresDB, DB2, Oracle and SQL Server.

- Experienced in scheduling and monitoring the production jobs using Oozie and Azkaban.

TECHNICAL SKILLS

Hadoop Ecosystems	Hadoop, HDFS, MapReduce, Hive, Spark Core, Spark SQL, Spark Streaming, Spark MLlib Impala, Kafka, YARN, Oozie, Zookeeper, Solar, Sqoop, NiFi, Knox, Ranger, and Kerberos.
Cloud Services	EMR, EC2, S3, Cloud Watch, RedShift, BigQuery and MS Azure
Languages	Java, Scala, Python, Pandas, R, PL/SQL, Unix Shell Scripting
UI Technologies	HTML5, JavaScript, CSS3, Angular, XML, JSP, JSON AJAX
Development Tools	IntelliJ, Postman's, Scala IDE, Jupyter, Zeppelin, Condo
Frameworks/Web Server	Spring, JSP, Hibernate, Web Logic, Web Sphere, Tomcat
SQL/ NoSQL Databases	Teradata, PostgreSQL, Oracle, HBase, MongoDB, Cassandra, MySQL and DB2
Other tools	GitHub, BitBucket, SVN, JIRA, Vagrant, Dockers, Maven

PROFESSIONAL EXPERIENCE:

Client: Santander Bank - New York, NY   **Dec 2017 – Present**
Role: Hadoop Developer

Roles & Responsibilities:
- Involved in installation, configuration and Design of Hadoop Distributed using Cloudera and Hortonworks of Hadoop.
- Involved in complete Big Data flow of the application data ingestion from upstream to HDFS, processing the data in HDFS and analyzing the data using several tools.
- Implemented Spark 2.0 using Scala 2.11.8 and Spark SQL for faster processing of data.
- Implemented algorithms for real time analysis in Spark.
- Used Spark for interactive queries, processing of streaming data and integration
- Installed and Configured Apache Hadoop clusters for application development and Hadoop tools like Hive, Pig, Oozie, Zookeeper, HBase, Flume and Sqoop.
- Implemented multiple Map Reduce Jobs in java for data cleaning and pre-processing.
- Worked in a team with 150 node cluster and increased cluster by adding Nodes, the configuration for additional data nodes were done by Commissioning process in Hadoop
- Imported the data from various formats like JSON, Sequential, Text, CSV, AVRO and Parquet to HDFS cluster with compressed for optimization.
- Experienced on ingesting data from RDBMS sources like - Oracle, SQL Server and Teradata into HDFS using Sqoop.

- Configured Hive and written Hive UDF's and UDAF's Also, created partitions such as Static and Dynamic with bucketing.
- Used Scala to convert Hive/SQL queries into RDD transformations in Apache Spark.
- Implemented advanced procedures like text analytics and processing using the in-memory computing capabilities like spark.
- Importing and exporting data into HDFS and hive using Sqoop and Kafka with batch and streaming.
- Collected data using Spark Streaming from AWSS3 bucket in near-real- time and performs necessary Transformations and Aggregations to build the data model and persists the data in HDFS.
- Experienced with Spark-Streaming APIs to perform transformations and actions on the fly for building the common learner data model which gets the data from Kafka in near real time and Persists into HBase.
- Performance analysis of Spark streaming and batch jobs by using Spark tuning parameters.
- Enhanced and optimized product Spark code to aggregate, group and run data mining tasks using the Spark framework.
- Developed Spark scripts by using Python shell commands as per the requirement.
- Using Hive join queries to join multiple tables of a source system and load them to Elastic search tables.
- Experience in managing and reviewing huge Hadoop log files.
- Implemented POC to introduce Spark Transformations.
- Worked with NoSQL database HBase, MongoDB and Cassandra to create tables and store data
- Handled importing data from various data sources, performed transformations using Hive and Map Reduce, streamed using Flume and loaded data into HDFS
- Worked in transforming data from map reduce into HBase as bulk operations.
- Implemented CRUD operations on HBase data using thrift API to get real time insights.
- Installed Oozie workflow engine to run multiple MapReduce, Hive, Impala, Zookeeper and Pig jobs which run independently with time and data availability
- Developed workflow in Oozie to manage and schedule jobs on Hadoop cluster for generating reports on nightly, weekly and monthly basis.
- Expertise in designing and creating various analytical reports and Automated Dashboards to help users to identify critical KPIs and facilitate strategic planning in the organization.
- Involved in Cluster maintenance, Cluster Monitoring and Troubleshooting
- Created Data Pipelines as per the business requirements and scheduled it using Oozie Coordinators.
- Maintaining technical documentation for each and every step of development environment and launching Hadoop clusters.
- Built the automated build and deployment framework using GitHub and Maven etc.
- Worked on IntelliJ IDEA to develop the code and dubbing.
- Worked on BI tools as Tableau to create dashboards like weekly, monthly, daily reports using tableau desktop and publish them to HDFS cluster.

Client: Walt Disney - Burbank, CA   **Aug 2016 – Nov 2017**
Role: Hadoop Admin/Developer

Roles & Responsibilities:

- Worked on Hadoop Hortonworks distribution which managed services viz. HDFS, MapReduce2, Hive, Pig, HBase, Sqoop, Spark, Ambari Metrics, Zookeeper, Falcon and Oozie etc. for 4 cluster ranges from LAB, DEV, QA to PROD.
- Monitor Hadoop cluster connectivity and security on Ambari monitoring system.
- Led the installation, configuration and deployment of product software's on new edge nodes that connect and contact Hadoop cluster for data acquisition.
- Developed custom Apache Spark programs for data validation to filter unwanted data and cleanse the data.
- Designed and Developed Sqoop scripts to extract data from a relational database into Hadoop.
- Developed Map Reduce programs to cleanse the data in HDFS obtained from heterogeneous data sources.
- Developed workflow in Oozie to manage and schedule jobs on Hadoop cluster to trigger daily, weekly and monthly batch cycles.
- Responsible for Cluster maintenance, Monitoring, commissioning and decommissioning Data nodes, troubleshooting review data backups, review log files.
- Day to day responsibilities includes solving developer issues, deployments moving code from one environment to other environment, providing access to new users and providing instant solutions to reduce the impact and documenting the same and preventing future issues.
- Collaborating with application teams to install operating system and Hadoop updates, patches, version upgrades.
- Involved in Analyzing system failures, identifying root causes, and recommended course of actions.
- Interacting with HDP support and log the issues in portal and fixing them as per the recommendations.
- Retrieved data from HDFS into relational databases with Sqoop. Parsed cleansed and mined useful and meaningful data in HDFS using Map Reduce for further analysis.
- Responsible for developing data pipeline using HDInsight, Sqoop and pig to extract the data from weblogs and store in HDFS.
- Worked with Sqoop to export analyzed data from HDFS environment into RDBMS for report generation and visualization purpose.
- Experience with cloud-based tools like Jira, Jenkins, Git, Artifactory, creation of CI/CD pipelines.
- Experience in developing a Continuous integration and Continuous deployment system that would not only
- deploy the code onto environment like Dev, Staging and Production but also help's in scaling the system along with continuous monitoring.

- Develop and manage application code, User Interface, and third-party integration components.
- Strong experience in process improvement and re-engineering, business requirements gathering and process flowcharts.
- Experience in maintaining the necessary development documentation as needed (ex: technical design, developer notes etc.).
- Worked on Service Catalog, Incident Management, GRC, Configuration & Change Management and Release Management with far reaching learning on Content Management System.
- Created Workflow activities and approvals. Implemented new workflows that use a variety of activities to understand how records are generated from workflows.
- Developed client scripts, UI Policies, Script Includes, business Rules across the application as per the requirement.
- Implemented custom interceptors for Sqoop to filter data and defined channel selectors to multiplex the data into different sinks.
- Extracted data from SQL Server 2008 into data marts, views, and/or flat files for Tableau workbook
- consumption using T-SQL. Partitioned and queried the data in Hive for further analysis by the BI team.
- Managed Tableau extracts on Tableau Server and administered Tableau Server.
- Supported in setting up QA environment and updating configurations for implementing scripts with Pig and Sqoop.
- Environment: Hadoop Horton Works, HDFS, Map Reduce Hive, Elastic Search Pig, Tableau, HBase, Spark, Sqoop, Oozie, Cassandra, Zookeeper and Windows.

Role: Jr Hadoop Developer   **June 2015 – May 2016**
Creative Technosoft Systems

Roles & Responsibilities:
- Worked under the Cloudera distribution.
- Handled importing of data from various sources, performed transformations using Pig and loaded data into HDFS and Extracted the data from MySQL into HDFS using Sqoop.
- Experienced in implementing different kind of joins to integrate data from different data sets like Map and reduce side join.
- Involved in loading data from edge node to HDFS using shell scripting
- Involved in loading data from UNIX file system to HDFS.
- Managing and scheduling Jobs on a Hadoop Cloudera cluster using Oozie workflows and java schedulers.
- Integrated Oozie with the rest of the Hadoop stack supporting several types of Hadoop jobs out of the box (such as Map - Reduce, Pig, Hive, and Sqoop) as well as system specific jobs (such as Java programs and shell scripts.
- Used Impala to read, write and query the data in HDFS
- Involved in loading data from LINUX file system to HDFS.
- Monitored Hadoop cluster job performance and capacity planning.

- Created Hive tables to load the transformed Data.
- Importing the data from RDBMS to HDFS using Sqoop.
- Continuous monitoring and managing the Hadoop cluster through Cloudera Manager.
- Developed hive queries to analyze/transform the data.
- Developing predictive analytic using Apache Spark Scala API's. Knowledge in Spark Core, Streaming, Data Frames and SQL, MLib, GraphX.
- Used Spark stream processing get data into in-memory, implemented RDD transformations, actions to process as units.
- Created and worked Sqoop jobs with incremental load to populate Hive External tables.
- Automated the data processing with Oozie to automate data loading into the Hadoop Distributed File System.
- Involved in agile methodologies, daily scrum meetings, planning's.

EDUCATION:

Bachelor's in computer science
Master's in information system

Abhinav Reddy

- New York, NY, USA

Contact Information

- abhihad12@gmail.com (Preferred)

- 4843410318 (Preferred)

Work History

Total Work Experience: 4 years

- **Hadoop Developer | Santander Bank**
  Dec 01, 2017 - No End Date | New York NY United States

- **Hadoop Admin/Developer | Walt Disney**
  Aug 01, 2016 - Nov 01, 2017 | Burbank CA United States

- **Jr Hadoop Developer | Creative Technosoft Systems**
  Jun 01, 2015 - May 01, 2016

Skills

- **apache pig** | 4yrs | 2019
- **cluster** | 4yrs | 2019
- **mapreduce** | 4yrs | 2019
- **rdbms** | 4yrs | 2019
- **workflow** | 4yrs | 2019
- **apache hadoop** | 4yrs | 2019
- **apache hive** | 4yrs | 2019
- **apache oozie** | 4yrs | 2019
- **apache spark** | 4yrs | 2019
- **apache sqoop** | 4yrs | 2019
- **hdfs** | 4yrs | 2019
- **cloudera** | 3yrs | 2019
- **java** | 3yrs | 2019

- **scala** | 3yrs | 2019
- **apache cassandra** | 3yrs | 2019
- **apache hbase** | 3yrs | 2019
- **apache zookeeper** | 3yrs | 2019
- **cloudera impala** | 3yrs | 2019
- **hortonworks** | 3yrs | 2019
- **streaming** | 3yrs | 2019
- **apache**
- **cassandra**
- **hadoop**
- **hbase**
- **hive**

Work Preferences

- Likely to Switch: Most Likely
- Willing to Relocate: Yes
- Work Authorization:
    - o US
- Work Documents:
    - o Employment Auth Document
- Security Clearance: No
- Third Party: Yes
- Employment Type:
    - o Contract - Corp-to-Corp
    - o Contract to Hire - Corp-to-Corp

Profile Sources

- Dice:
  https://www.dice.com/employer/talent/profile/4d4f00e5ba2a3fbc1889b09c5019e739a53d
  b2c7