**Anil Ambati** **Phone :+91 9052191053**

**Software Developer E-mail:aakr739753@gmail.com**

Professional Summary



* 3 years of experience in development and maintenance of ETL and Big Data applications.
* My experience includes in open source Big data components like Hadoop, Spark technology frameworks.
* Worked on Spark code & sparksql with Scala to load the data into Hive warehouse tables.
* Building the code with various Spark & Hive Optimizations techniques.
* Worked on NoSQL databases like Cassandra.
* Having knowledge in Spark streaming technologies like Apache Kafka & Flume.
* Having good exposure on Agile software development process.
* Worked on Sqoop to load the tables data from RDBMS databases to Hive.
* Worked with different file formats in Hive like Text, Parquet, ORC.
* Good exposure on working with Agile & Waterfall methodologies.
* Experience on build tools like SBT, usage of version controller like GIT, IDE like IntelliJ and Eclipse.
* Experience on RDBMs like Oracle, MySQL, SqlServer.
* Experience on ETL tools like Diyotta,Informatica.
* Knowledge on PLSQL and TSQL.
* Knowledge on UNIX and UNIX SHELL SCRIPTING.

Work experience



**COMPANY NAME: DIYOTTA ,Hyderabad**

**Project #1:**

Name : Scotia AML

Duration : Sep 2018 — Present

Role : ETL,Spark and Hadoop Developer

Environment: Linux, Spark, Java, MySQL, Scala, Apache Hadoop, Hive, Map-reduce, Apache Ambari.

**Description:**

The purpose of this project is to deal with the huge gigabytes from the world wide scotia Customers. The data getting from the different RDBMS systems as well as from different file formats. So, to deal with the RDBMS for Ingestion using Sqoop and loading into hive tables and to deal with the files using spark and processing and loading into hive tables. After that based on the client requirement business extracts are provided by joining some tables and applying transformations, for further data analytics.

**Responsibilities:**

* Importing the data from relational data bases into the HDFS using Sqoop to load & analyze data.
* Transferring source files from edge node to HDFS for further processing the data.
* Creating Scala/Spark jobs for data transformation and aggregation.
* Creating External/Managed Hive tables and loading the data into Hive from HDFS.
* For performance gain we are using Partitioning and Bucketing Concept.
* Regularly tune performance of Hive queries using various hive joins to improve data processing and retrieving.
* Responsible for writing Hive Queries using SparkSQL for analyzing data in Hive.
* Importing the varies files data into the HDFS using Spark for analyzing data.
* Written Spark code to generate business extracts using Spark api’s.

**COMPANY NAME: ReferPad Technologies,Banglore.**

**Project #2:**

Name : Explorian

Duration : Nov 2017 — Aug 2018

Role : ETL,Spark and Hadoop Developer

Environment: Linux, Java, Scala, Python, MySQL, Apache Hadoop, Hive, Spark, Map Reduce, Sqoop.

**Description:**

Data from a source application is received to Hive using Sqoop, stored in HDFS. This solution is based on the open source technology Big Data Hadoop. The data which is stored in Hive warehouse is analyzed using Spark by applying transformations and actions. The data is then used by down streams for further processing.

**Responsibilities:**

* Importing data from various data sources like Teradata, MySQL using Sqoop, performed transformations using Hive and loaded data into HDFS.
* Loading and transform large sets of structured, semi structured and unstructured data.
* Responsible for managing data coming from different sources.
* Involved in creating Hive tables, loading with data and writing hive queries, which will run internally in spark engine instead of map, reduce way.
* Creation of spark core RDD’s and applying related transformations and actions to improve the efficiency of the program using Spark.
* Spark & Hive Performance tuning.
* Responsible for writing Hive Queries using SparkSQL for analyzing data in Hive.
* Created non-merged tables with partitioning, bucketing for optimizing the hive data.
* Good understanding and related experience with Hadoop stack and spark Internals.

Education



MCA Jun 2014 — Apr 2017

JNTU, Anantapur

Aggregate of 78% .



My Skills



**Languages** : Java, Scala, Python.

**Big Data Technologies** : Apache Hadoop, Map-Reduce, HDFS, Hive, Sqoop , Spark2 ,Pyspark.

**Tools :** IntelliJ , WinSCP, Putty, Sbt, Version Control Tool Git,Jira.

**Databases** : Oracle 9i/10g/11.2.1, MySQL 2000/2005/2008,SqlServer,Db2.

**Development Methodologies:** Agile

**Operating Systems** : Linux, Windows NT/2000/XP/2003

**Signature**

Anil Ambati