Sumeet Agrawal

Mobile : 09891487397
E-mail: sumeetagrawal.srm@gmail.com
Bangalore

Objective:

A Hadoop Developer with 5 years of experience with very good technical and team working skills for application development in BigData and its Ecosystem projects, Analytical Processing and SDLC. Looking for challenging roles in Spark/scala and Hadoop related activites.

Professional Summary:

- 2.5+ years of experience in BigData, Hadoop and Hadoop Ecosystem (Pig, Hive, Sqoop, Flume,Hbase,Oozie, Spark with Scala etc) and 2+ years of work experience in Quality Assurance in retail banking.
- Good Knowledge of offloading RDBMS like Teradata and Oracle Databases and effectively migrating into HDFS.
- Successful implementation of ingesting real time streaming data and processing.
- Implementing Business Solutions and migration to HDFS and process data using Pig and Hive.
- Skilled in analyzing business plans and providing best optimized performance solutions using Hadoop related tools.
- Good interpersonal skills, committed, Result oriented, Innovative, Hardworking with quest and zeal to learn new technologies.

Technical Skills:

Working knowledge on Hive, PIG, Sqoop, Flume, Spark(SQL, Streaming and core) with Scala, Oozie, MapReduce in Core Java and Cloudera CDH 5.x.
Exposure to SQL Programming, Unix and shell scripting, Data Modelling & Dataware Housing Concepts.
Advanced Conversant with Linux and Windows.
Hands on Java object oriented programming and scala functional programming
Build tools- sbt and maven, IDE – Eclipse IDE for Scala and java

Projects:

1) Title   : **Jenssen**
   Organization/Client   : Johnson & Johnson
                                  (Jan 2017-Present)
   Tools   : EMR, Spark 2.x, Scala 2.11, S3, Redshift, maven, sbt, git

Description: The health care project is about creating multiple reports for Jenssen to analyse patient solutions. Complex relations are to be carried out from multiple files captured from different sources.

Responsibilities:

- Understanding the business requirements and make plans and design for spark coding.
- Data migration from Amazon S3 to Redshift database using Spark code.
- Loading the data into spark data frames to perform the operations.
- Filter and transform columns and rows as per the requirements

- ⏱ Performed complex operation using scala on data frames to get the master tables which was used for tableau reporting.
- ⏱ Scheduling and writing scalatest testing scripts for the modules.

2) Title   : **Lead management system**
 Organization/Client : Cox Communications(USA)
         (June 2016-Present)
 Tools   : HDP2.3, Hive, Oozie, Sqoop, Oracle, Shell,
         Amazon Red shift and S3 bucket.

Description: The project is about tracking and monitoring the customers who tried to make interaction with COX in terms of complaints or looking for products. The system captures the various events or transactions and analyses the events and transform it to send to the analytics division.
Responsibilities:

- ⏱ Worked on real time data ingestion from Amazon s3 to Hadoop.
- ⏱ Parsing and designing of unstructured data into hive
- ⏱ creating schema for pre-staging, staging and Base tables
- ⏱ deriving complex variables records from the unstructured data to be consumes by analytical tools.
- ⏱ Automation of the whole process using shell scripts and exporting data into oracle databases in regular time intervals.

3) Title   : DAST – Data Analytics and Support Tool in
         Risk and Finance
 Organization/Client : Accenture/Swiss based Banking Client
         (Aug 2015 – May 2016)
 Tools   : Cloudera CDH 5.x, Spark 1.5-1.6, PIG,
         Hive, Sqoop, Eclipse , SBT

Description: The Risk Division acts as guardian of the bank's risk appetite and provides independent risk oversight. The Division supports the business in shaping its risk profile, and aligning its strategy execution with shareholders' and regulatory requirements.
Responsibilities:
- ⏱ Worked on Hive/Flume – Spark integration, Spark SQL, and spark core.
- ⏱ Writing scala codes using Spark API, Building and submitting on the yarn cluster, Performance benchmarking estimation.
- ⏱ Importing data from Finesis clusters to HDFS by sqoop and scheduling work flows.
- ⏱ Analyzed Splunk Enterprise and setting up Hadoop development cluster using CDH parcels..

TCS Big data R&D

4) Title   : Project Babar
 Organization/Client : Tata consultancy services/ **HSBC Bank** (Mar
         2015 – July 2015)
 Technologies   : HDP2.2, global ID's, hive, pig, sqoop,
         Apache Kafka, java, Python

Description: A central data lake needs to be formed for analytics factory from the golden data sources. Data migration from Oracle and Teradata to form a data lake which will be a source for many big data initiatives. The data is then passed to analytics teams a feed for machine learning algorithms.

Responsibilities:

- Ingesting historic trade data and streaming data upto 300messages/sec
- Providing Technical Architecture [Selections of tools and technologies]
- Pre-preparation, transformation, cleansing of data and setting up rule engines on global ID's and python
- Writing ETL jobs using Pig/Hive and UDFs in Python.
- Writing Hive scripts to generate the reports.

5) Title   : R&D Poc for Barclays
   Organization/Client   : Tata consultancy services/Barclays (Jan'
                                    2015- Feb'2015)
   Technologies   : Pig, Hive, R, Java

Description: Worked in TCS R&D team for Barclay proof of concept. Master Files is a central repository for non-transaction specific information such as Client or Partner information, Fixed Income, Instruments information and Reference information.
In this Poc, we used to collect the data from OLTP systems and stored in the HDFS.
Implementation of data de-duplication and skewed data storage.

Responsibilities:

- Understanding the data, customer requirements and business use case.
- Implementation of data de-duplication algorithm from Map Reduce and python.
- Setting up a few Hadoop cluster and installed Ecosystem components.
- Configure end to end of Rhadoop packages(rjava,rhive.rmr2,rhbase) on Hadoop environment.

6) Title   : R&D Poc for Amtech
   Organization   : Tata consultancy services/Amtech (Nov'
                                    2014- Dec'2014)
   Tools   : Python, Apache Tika, Beautiful soup,
                                    NLTK

Description: The proof of concept was to develop an analytical model to find out the sentiment analysis of news articles related to AMTECH group and find out the correlation of news articles with the stocks fluctuation.

Responsibilities:

- Setting up and running Hadoop cluster.
- Python script to run sentiment analysis algorithm on news articles.
- Fetching data from yahoo finance,bloomberg and reuters using beautiful soup.
- Statistical analysis and visualization of co-relation between news feeds and Stocks fluctuation

6) Title   : Nielsen BI topological Analysis
   Organization   : Tata consultancy services/Nielsen (Jan'
                                    2014- Oct' 2014)
   Tools   : Flume, Hive, Oozie

Description: Nielsen is a leading global information and measurement company that enables companies to understand consumers and consumer behaviour. Wants to develop an analytical model where unstructured data can be stored. The data has to be ingested on regular intervals and stored into tables to generate reports for BI.

Responsibilities:

- Being a Team member at Nielsen was to develop test cases and generate reports

using media data.
- Ingest data from the servers using flume at regular intervals and then pass it to the BI team for customer predictions.
- Interact with stakeholders to provide complete understanding of requirements by actively participating in requirements gathering.
- Automation of scripts using oozie.

7) Title   :   Core Banking system LLOYDS Bank

Organization   : Tata consultancy services/Lloyds bank
                        (Oct' 2011- Dec' 2013)
Tools : HP Quality center, DB2, SM7, JCL Description:

Lloyds Bank is one of the largest retail bank of UK. It wants to manage its core banking system dev environments, deploy reference data as per the release implementation, code check in on 190 + Application groups and manage consistency. New API implementation, job scheduling, dependencies and implementation of group releases on test environments

Responsibilities:
- Part of Core banking system team. Running batches for group releases, trouble shooting and root cause analysis during batch failures.
- Raise defects for job failures while implementing group release codes.
- Part of the successful major segregation of Lloyds with TSB bank as most complex migration in CBS.
- Implementation of reference data of new branch addition, products introduction, Process record control in CBS database.

Professional Experience:

- ❖ Working as Senior Software Engineer in Accenture,
    Bangalore. o Duration: Aug 2015 – Present
- ❖ Worked as Systems Engineer in TCS pune/Gurgaon.

    o Duration: Oct 2011 – July 2015

Educational Profile

2007-2011 -B.Tech (Information technology) – SRM University(78%)