Shubham Jain

Contact: +917828733244
Email : shubhamakachamp@gmail.com
Github : https://jnshubham.github.io

OBJECTIVE

A proactive and fast learning individual seeking an opportunity to work as a dynamic data analyst utilizing analytical & methodical skills and relevant expertise to help the company achieve business goals while sticking to vision, mission and values.

Summary

- 2+ years of IT experience in data analytics with apache hadoop ecosystem like Spark, Python, HDFS, Hive, Impala, Flume, Sqoop, Kafka, Oozie and Big Data Analytics
- Excellent understanding of Hadoop Architecture and underlying Hadoop framework including Storage Management
- Analyzed and processed complex data sets(structured and semi-structured) using advanced querying and analytics tools
- Experience in developing framework for importing data from various databases (Teradata, Oracle, SQL Server, DB2, Informix) to hadoop using python
- Proficient in developing data pipelines using Python
- Excellent analytical, Interpersonal and Communication skills, fast learner, hardworking and good team player

Achievements

- Received technical excellence award in 2018 for developing Pyspark framework for creating a data pipeline from RDBMS to Hadoop
- Received on the spot award in 2018 for proactively providing solutions across multiple teams

Work Experience

Tata Consultancy Services
System Engineer
Pune (IN)
December 2016 – Present

Assignment :   Data Ingestion and Transformation using BigData
Client Domain: Banking and Financial Services
Position: Hadoop Big Data Developer

Responsibilities:
- Handling of complex xml data and nested json data using Spark
- Developed Spark based reconciliation processes to maintain the data integrity in Hadoop
- Developed Python framework to import data from various RDBMS to Hadoop using Sqoop, Hive and Oozie
- Developed Spark Jdbc based data importing framework from Informix database
- Setup of data flow pipeline from Dynatrace to HDFS
- Handling of protobuf message format
- End to end setup of flume agent for fetching live transactions
- Development of the protobuf message handler code in java for storing it on HDFS using Flume sink
- Development of REST python setup for raising incident in case of any failure using BMC truesight
- Worked in unix environment and created various unix scripts as per the requirements
- Experience with publish subscribe architecture using kafka

Environment:
Hadoop 2, Spark 2.1.0, Sqoop, Hive 1.1.x, Impala, Oozie, Cloudera CDH 5.13.5, Flume 1.6.0, Python 2.7, Java 1.7, Maven, Git, Jenkins

Technical Skills

- Programming languages: Python, Java, Pyspark
- Hadoop ecosystem: Spark, Sqoop, Hive, Impala, Oozie, Flume
- RDMS: Teradata, DB2, Oracle, sql server
- Devops Tools: Git, Maven, Jenkins
- Operating System: Linux, Windows

QUALIFICATIONS

Bachelor of Engineering   2012-2016
Maharana Pratap College of Technology
Branch: CSE
CGPA: 8.11

Higher Secondary   2011-2012
Nehru Higher Secondary School
Scored: 84.2%

High School   2009-2010
Nehru Higher Secondary School
Scored: 82.3%

Interests

- Playing badminton
- Reading
- Playing multiplayer online games

STRENGTH

- Optimistic
- Innovative and creative thinker
- Quick learner

PERSONAL DETAILS

- DOB : 03/07/1994
- Languages known : Hindi and English
- Permanent address : Jain Mechanical Works, pali road, Sheopur, M.P. Pin: 476337
- Corresponding Address: A-1201, Lorelle Society, Wakad, Pune