Shubham Jain

Contact: +917828733244
Email   : shubhamakachamp@gmail.com

Summary

- 2 Years and 1 month of IT experience in Data Analysis with Apache Hadoop Ecosystem like Spark, Python, HDFS, Hive, Impala, Flume, Sqoop, Oozie and Big Data Analytics.
- Excellent understanding of Hadoop Architecture and underlying Hadoop framework including Storage Management.
- Effectively worked upon the pandas dataframe and performed data analysis using python.
- Experience in working with the spark Dataframes and spark sql.
- Experience in developing framework for importing data from various databases (Teradata, Oracle, SQL Server, DB2, Informix) to hadoop using python.
- Experience in handling the transactional data received as a protobuf message using flume.
- Good understanding of Object Oriented Programming (OOPS), J2EE 7, Web Services (REST) and JDBC.
- Excellent analytical, Interpersonal and Communication skills, fast learner, hardworking and good team player.

Achievements

- Received Technical Excellence Award in 2018.
- Received On the spot award in 2018.

Work Experience

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

Assignment 1:	Customer Experience Dynatrace Transactions in Big Data
Client Name:	PNC Financial Services Group
Period:	Oct 2018 – Till Date
Position:	Hadoop Big Data Developer
Description:	PNC offers a wide range of services for all our customers, from individuals and small businesses, to corporations and government entities. It provides deposit, lending, cash management and investment services to more than 5 million consumers and small business customers across 15 states and the District of Columbia.I am working in Customer Experience Project where we need to fetch the live transactions generated by the dynatrace as a protobuf messages in hadoop. For this the Dynatrace is configured to transmit the live transaction as a batch on UNIX port and we need to setup a flume data pipeline to handle the protobuf transactions and store them in hadoop in real time.

Responsibilities:
- 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.

Environment:
Hadoop 2, Spark 2.1.0, Sqoop, Hive 1.4.x, Impala, Oozie, Cloudera CDH 5.13.5, Flume 1.6.0, Python 2.7, Java 1.7, Maven, Git, Jenkins

Assignment 2:   Blackknight Reconciliation in Big Data
Client Name: PNC Financial Services Group
Period: May 2018 – Oct 2018
Position: Hadoop Big Data Developer
Description: PNC offers a wide range of services for all our customers, from individuals and small businesses, to corporations and government entities. It provides deposit, lending, cash management and investment services to more than 5 million consumers and small business customers across 15 states and the District of Columbia.I am working in Blackknight team where we need to reconcile the number of files generated by the blackknight vendor. In this, big data platform would be used to store the files received on blackknight vendor and then the files will be reconciled for integrity. This data will serve as input the consumer team for applying transformation and for creating reports.

Responsibilities:

- Developed spark based reconciliation processes to maintain the data integrity in hadoop.
- Developed spark based application to parse complex xml data received as file.
- Developed various spark transformation as per the requirement using python.
- Worked with spark Dataframes, python pandas Dataframes.
- Worked in unix environment and created various unix scripts as per the requirements.

Environment:
Hadoop 2, Spark 2.1.0, Sqoop, Hive 1.4.x, Impala, Oozie, Cloudera CDH 5.13.5, Flume 1.6.0, Python 2.7, Java 1.7, Maven, Git, Jenkins

Assignments 3: Anti – Money Laundering Data Ingestion in Big Data
Client Name: PNC Financial Services Group
Period: May 2017 – May 2018
Position: Hadoop Big Data Developer
Description: PNC offers a wide range of services for all our customers, from individuals and small businesses, to corporations and government entities. It provides deposit, lending, cash management and investment services to more than 5 million consumers and small business customers across 15 states and the District of Columbia.   I am working in AML (Anti Money Laundering) Enterprise Data Management Technology team for introducing AML STX required data into Enterprise Data Warehouse. In this, big data platform would be used to ingest data from different sources such as mainframe, Relational database like Oracle, Teradata. We will ingest whole data and will store them in HDFS as Hadoop distributed file system. This data will serve as input the consumer team for applying transformation and for creating reports.

Responsibilities:

- Developed python based framework to import data from various relational databases to hadoop using Sqoop, hive and Oozie.
- Developed spark jdbc based data importing framework from Informix database.
- Used spark 2.1.0 over cloudera CDH5.13 to perform analytics on data in hive.
- Developed various spark transformation as per the requirement using python.
- Developed Sqoop export framework for exporting the data from hadoop to Teradata.
- Effectively worked with python to develop the solutions as per the requirements.
- Worked with spark Dataframes, python pandas Dataframes.
- Worked in unix environment and created various unix scripts as per the requirements.

Environment:
Hadoop 2, Spark 2.1.0, Sqoop, Hive 1.4.x, Impala, Oozie, Cloudera CDH 5.13.5, Flume 1.6.0, Python 2.7, Java 1.7, Maven, Git, Jenkins

OBJECTIVE

To constantly work in a challenging environment on cutting edge technologies and to contribute meaningfully towards the organizational and personal growth.

Technical Skills

- Programming Languages: Python, Java, pyspark
- RDMS: Teradata, DB2, Oracle, Sql Server
- Devops Tools: Git, Maven, Jenkins, IBM uDeploy
- Hadoop Tools: Spark, Hive, Impala, Oozie, Flume
- 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
- Listening Poetry
- Playing Multiplayer Online games.

STRENGTH

- Optimistic
- Innovative and creative thinker
- Requires minimum input to learn

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