

Akash Soni

Residency Status: nagasandra circle Road bangalore

Telephone (India): 9620746355

Email : akashsoni199218@gmail.com

GitHub:<https://github.com/akashbing>

CAREER OBJECTIVE

To be associated with a professionally managed organization where I can enhance my skills by contributing in the best way to the Organization and add value to everything that I deliver through integrity, innovation and excellence.

Real Time Experiance:

- Currently Working for DHL and overall 2.1 year of Experience
- Involving in Hive(1.2.1) Creating external tables and partitioning by different clause.
- Extracting data from **Postgresql by Sqoop and from PySpark JDBC Connector**
- Excellent knowledge on Hadoop ecosystems such as **HDFS,Resource Manager, Node Manager, Name Node, Data Node**
- Good Experience on **Spark (Pyspark)** in designing and developing applications in Spark using Python
- Experience in manipulating/analyzing large datasets and finding patterns and insights within structured
- Analytics on **PySpark(2.2.0) Dataframe**
- Good Knowledge of **DataMart Star schema ,data warehousing** creating fact and dimension tables
- Basic knowledge in Django framework
- Good knowledge in Pyspark **ML(pyspark-pipelines,Transform and Action Estimator,StringIndexer,VectorIndexer etc)**
- Visualise the Data in **Python Matplotlib**
- Good knowledge in Jupyter notebook,git
- Good knowledge in **Machine learning linear-regression,k-means clustering**

ACADEMIC CREDENTIALS

S.NO.	DEGREE	INSTITUTE	BOARD/ UNIVERSITY	RESULT (%)
1	B.tech	Greater Noida Institute Of Technology	AKTU	60
2	HSC	HANSRAJ MODERN PUBLIC SCHOOL	C.B.S.E	59.8
3	SSC	K.V	C.B.S.E	64

SOFTWARE EXPOSURES:

Hadoop,Hdfs,Pig,Hive,Sqoop,PySparkSQL,PySpark-Dataframe,PySpark-ML,Sql,python, Machine learning(Linear & multi regression ,K-means clustering with elbow method,Django)

Company:DigiKraft Technology

Project Details:

Project #1:

Client:DHL

Project Name: Logistics (EzyloadsIndia)Business Relation Between Carrier And Shipper

Environment : PySpark, PySpark Sql, PySpark DataFrame, MySQL,Python,HDFS,hive,sqoop

Data visualization:Python,Matplotlib

Description:

In this project I have to process data and import the data from Postgresql server to hdfs with Millions of records from different tables Reason of this datasets to find out the Business potential between Carrier and Shipper I joined different tables clean the data and make the table which is going to analytics and also create star schema and fact tables for OLAP find the number of trips,carrier performance,give the availability to the shipper .

Roles and Responsibilities:

- Importing the data from Postgresql Server into HDFS
- Analyse the Data by Joining different tables ,grouping and aggregation the dataframe
- Convert the tables into Data Frame and taking different column from different tables
- Aggregating and grouping the data for DataMart
- Visualize the Data in Python Matplotlib

Project Details:

Project #2:

Client:United Health group

Project Name: Health Care

Environment : Hadoop,Hive, SQOOP, MySQL,PySpark

Roles and Responsibilities:

- Analyzing the metadata as per the requirements
- Loaded the data in HDFS from local system.
- Applying transformation and action to aggregate data by Pyspark
- Created hive external tables and loaded the data using Sqoop.
- Writing queries in hive to map the data resided in HDFS.

HOBBIES

Cricket, Football, Music, Travelling

PERSONAL DETAILS

Father's Name : Mr R.k Soni
Date of Birth : OCT 18th 1992
Language Known : English, Hindi
Marital status : Single

I hereby affirm that the information in this document is accurate and true to the best of my knowledge.

Akash Soni