**KARAN SAWLANI**

**Email id:** karan.sawlani3@gmail.com

**Contact:** 7065151049

**LinkedIn Profile Link:** <https://www.linkedin.com/in/karan-sawlani-138030136/>

|  |
| --- |
| **PROFESSIONAL SUMMARY:** |

* Design, plan, and develop programs to perform automated extract, transform and load data between datasources when working with large data sets
* Responsible for building and supporting a Hadoop-based ecosystem designed for enterprise-wide analysis of structured, semi-structured, and unstructured data
* Knowledge on big data frameworks and components : Hadoop, Hive, Spark, Sqoop, Flume, Oozie, Pig.
* Knowledge of performance tuning and optimization of big jobs in hadoop framework.
* Worked on Apache Nifi and Apache Airflow data ingestion and data transformation pipeline.
* Worked on AWS Services and created end to end generic pipelines for rest apis,ftp,rdbms sources using nifi and spark respectively along with integration to AWS Services like S3,Glue,Emr,Iam,Ec2,Rds,Lambda,Glue etc.
* Worked on Spark Streaming, built the reliable pipeline and done analysis on streaming data feeding it to a pre build ML model.

|  |
| --- |
| **WORK EXPERIENCE:** |

**TO THE NEW July (2019)-Present**

Data Engineer

**TO THE NEW Feb (2019)-June(2019)**

Big Data Trainee

|  |
| --- |
| **TECHNICAL SKILLS:** |

|  |  |
| --- | --- |
| Components | Skills |
| Language | Python(Primary),Scala(Secondary) |
| Operating System | Windows, Linux |
| Technologies | Hadoop Ecosystem, Apache Hive, Apache Sqoop, Apache Nifi  Apache Airflow, Apache Spark,Apache Oozie, Apache Pig, Apache Flume,Apache Kafka,Aws (S3,Emr,Ec2,Vpc,Ebs,Rds,Lambda,Glue Crawler) |
| Database | Mysql |
| Versioning | Git |

|  |
| --- |
| **PROJECTS HIGHLIGHTS:** |

|  |  |
| --- | --- |
| To The New (July-2019 -> March-2020) | |
| About Client | Seera Group is one of the biggest chains for online & offline Flight and Hotel Booking Systems. Here in this project EDL, we are setting up Data Lake from scratch and building Analytics Jobs to achieve KPIs & Reporting etc. Fetching data from n no of sources using Apache NiFi, Sqoop, and using Spark and Hive for processing that data to generate meaningful business reports. |
| Overview | Data lake need to be build which consumes data for all organization and  put that in HDFS incrementally and use Hadoop ecosystem and other tools to analyze this huge amount of data  visually using D3JS and click-sense after processing |
| Technology / Tools | Hadoop ecosystem, Hive, Sqoop, NIFI, Airflow, Oozie |
| Role | Data Engineer |
| Responsibilities | * Involved in end to end architecture level discussions and implementations. * Created many NIFI flows for data ingestion from different sources and other existing flows optimisation. * Created large Hive queries for data processing. * Created many Sqoop jobs and schedules using crontab. * Optimise many large data set tables for better efficiency using partitioning. * Hive query optimizations. |

|  |  |
| --- | --- |
| To The New ( March-2020 -> Present) | |
| About Client | TTN Internal Project (Accelerator) |
| Overview | Building generic pipelines for different sources like Rest Api,Ftp,Rdbms using different tools and integrating them end to end with front end UI these pipelines handles incremental, cdc capturing,full dumps along with maintaining states and storing in different destinations. |
| Technology / Tools | Spark(Scala),NIFI, Airflow, AWS(S3,Glue,Emr,Rds,Ec2,Lambda etc) |
| Role | Data Engineer |
| Responsibilities | * Involved in end to end architecture level discussions and implementations. * Created generic Nifi flow for data ingestion from different Rest Api sources and other existing flows optimisation. * Created Spark-Scala independent generic jar for On-premise and Aws (Emr) jar for generic Rest Apis and Ftp sources that places data in different destinations like Rdbms,Ftp.Hdfs,S3,Local while maintaining states in Aws (Rds)/Local Mysql along with cdc and data duplicacy checks. * Created end to end Airflow scheduler for triggering the generic Nifi,Spark and Sqoop flows from here along with proper notifications this orchestrates data quality checks and data validation checks frameworks. |

|  |
| --- |
| **PERSONAL PROJECTS HIGHLIGHTS:** |

|  |  |
| --- | --- |
|  | |
| Project | Streaming Data Pipeline Construction With Visualization |
| Overview | Build a streaming data pipeline that takes data from various datasets (edureka datasets,twitter api data) and then do data prediction using spark-ml in streaming fashion. |
| Technology / Tools | PySpark,Spark MLlib, Flume, Kafka,Spark Streaming,Hdfs,Python(regex,dash) |
| Role | Data Engineer |
| Responsibilities | * Created an end to end data pipeline using Flume(data ingestion)->Kafka(data ingestion and loss prevention)->Spark Streaming(data in streaming fashion)->Spark MLlib(data analysis) on Edureka provided datasets, and concurrently it is working to save the golden copy of data as well with Flume->Kafka->Hdfs**.** |

|  |  |
| --- | --- |
|  | |
| Project | Data Analysis and Data Prediction Using Python |
| Overview | Build python code and moulded it to scale up in pyspark analysing stocks data and covid data. |
| Technology / Tools | Python(Scikit-learn,Dash,Pandas,BeautifulSoup),pyspark,web scraping |
| Role | Data Engineer + Data Analyst + Data Visualization |
| Responsibilities | * Done data collection using python web scraping. Here collected and transformed data of yahoo finance using BeautifulSoup and pandas. * Done data analysis and prediction of stock data using python here used **S**cikit learn for ML, Panda**s** for data collection and analysis and worked on matplotlib,dash for basic data visualization (graphing purpose). * Analysed twitter data (Avengers Movie data) and kaggle (Covid-19 Dataset) twitter\_api,python(dash) and moulding this python code to scale up in pyspark. |

|  |
| --- |
| **CERTIFICATIONS:** |

|  |  |
| --- | --- |
|  | |
| Edureka | Apache Spark Developer |
| Edureka | Big Data Hadoop Expert |
| Cloudera | CCA-175 Spark and Hadoop Developer |

|  |
| --- |
| **ACADEMIC QUALIFICATION:** |

**Bachelor of Technology (CSE ) [Hons.] 2015-2019**

AKTU

**Higher Secondary 2014-2015**

CBSE

**Secondary 2012-2013**

CBSE