Naveen
Data Engineer

Data engineer passionate about creating data driven automated solutions. Demonstrated history of working in Data domain with more than 3.5 years of Big Data experience. Skilled in Spark , Python, SQL, Hadoop. Worked mainly in importing data from different sources, extracting, transforming, processing and loading them on Warehouse. Expertise in Big-Data/Hadoop technology stack using Apache Spark (Batch Processing), SQL, Python, Scala, Hive, AWS(EMR, S3, Redshift).

📞 8897573407   ✉ nknaveen379@gmail.com   🐙 github.com/naveen7496   📍 Navi Mumbai

WORK EXPERIENCE

Sr. Executive Programmer
R-Pac Internationals
09/2021 – Present   Mumbai, India

- Responsible for Building Scalable data pipelines Using **Bigdata/Hadoop** technology stack.
- Used **Scala** and **Spark** dataframes for data processing and transformation to load data to warehouse tables.
- Worked on various **Spark optimization** techniques such as code level and resource level, which gave a lot of performance gain.
- Worked on **Hive** optimization performance using **Partitioning**, **Bucketing**.
- Followed agile-sprints and code versioning through **GIT**.
- Interacted with Business Analysts and stake holders to understand business requirements.

Sr. Analyst
Capgemini
03/2019 – 09/2021   Mumbai, India

- Worked with AWS **EMR** cluster to load data from **S3** bucket to AWS **Redshift**.
- Hands on experience in developing **Spark** applications using spark objects like Dataframe, Datasets using tools like **Spark core** and **Spark SQL.**
- Used Spark-Redshift community to connect spark with Redshift.
- Used **Python**, **Pandas**, **SQL**, **FTP**, **SFTP**, **GIT** for processing data.
- Heavily used **Pandas Dataframes APIs** and **Bulk Copy Program (BCP)** to process and insert data into **MSSQL**.
- Involved in scrum ceremonies (stand-up, grooming, planning, demo/review) with the teams to ensure successful project forecasting.

SKILLS

Spark Python SQL Scala Hive Hadoop AWS Pandas EMR S3 MSSQL PostgreSQL Map Reduce Django Redshift Airflow Linux

PERSONAL PROJECTS

- Analysis of New York Cab sharing dataset with Spark Dataframe in Scala in Maven setup.
- Analysis of 1 Million Movie dataset using Spark, Scala with SBT setup. Used Dataset, broadcast joins, case class, joins etc.
- Used spark-submit command to execute project locally.
- Deployed a spark project on AWS Elastic Map Reduce, S3 and performed computation.
- Good knowledge on Hadoop, Spark cluster, Map-Reduce, Airflow.

EDUCATION

B.Tech
Godavari Institute of Engineering and Technology
08/2013 – 05/2017   Rajahmundry, Andhra Pradesh

Senior Secondary
Government Model Senior Secondary School
03/2011 – 05/2013   Aberdeen Bazaar, Port Blair

Secondary
Government Model Senior Secondary School
03/2010 – 05/2011   Aberdeen Bazaar, Port Blair

LANGUAGES

- English
- Hindi
- Telugu