RESUME

Prakash Aravind Raj

Data Engineer.

Email:m.prakasharavindraj@gmail.com

Mobile: +91-9600084407.

Objective:

      To seek the challenging position in Software industry that needs innovation, creativity, dedication and enable me to continue to work in a challenging and fast paced environment, leveraging my current knowledge and fostering creativity with many learning opportunities.

Experience Summary:

➢ 6.5 Years of experience in DWH [ETL and BI Basic] and DE pipelines using Big Data Ecosystem, Teradata, Databricks and Snowflake.

➢ Good Programmer in SQL, Python and Pyspark.

➢ Strong understanding of Object-Oriented concepts (OOPS).

➢ Good experience on IDE tools like Visual Studio and PyCharm.

➢ Designed and Developed Big data Applications in Apache Spark, Azure cloud, Airflow, Databricks, Synapse, Snowflake, Cosmos, data factory and GCP.

➢ Good Knowledge and exposure in data migration process to delta lake and data lake.

➢ Experience in utilizing Teradata, ORACLE, ab initio, Informatica, Unix commands and scripts and BI Tools [Power BI, tableau].

➢ Proven team player.

⭘ Experience in developing automation and bots in Python.

⭘ Experience in deploying and maintaining Data Engineering and Data Science pipeline in cloud using Jenkins and azure devops.

Education Qualification:

- BE from SRM Valliammai Engineering college in 2017.

Professional Experience:

- Currently working as a Data Engineer in **Delcosys IT Solution**.

Technical Skills:

Languages	Python, SQL and Pyspark[Known Core Java but currently working in mentioned language]
Database	Teradata, Snowflake, Big Query and ORACLE

CICD	Git, Jenkins and Gitlab.
Big Data	Azure Data Lake Gen 2, Apache Spark, Databricks, Snowflake, Airflow, Splunk and Basic hive commands
Bi Tools	Power BI [Basic]and Tableau [basic]

Project Experience:

Project 1#

Project Title	:	Staples
Team Size	:	10
Role	:	Data Engineer.
Environment	:	Azure Cloud, Azure Data Lake, Pyspark, Python, Databricks and Synapse.
Duration	:	April 2022 - Present

Description:

Staples, Inc. is an American retail corporation. Staples has been the industry leader in workspace products like furniture, technology, cleaning products, and traditional office supplies. We have thousands of experts ready to help our customers navigate, thrive and succeed in the ever-changing world of work.

Roles and Responsibilities:

- Gather Requirement, analysis and prepare documentation to walkthrough business stack holders.
- Manage, Plan and Resource allocation to complete project in time.
- Identify data volume in source and write pipeline to migrate data.
- Work with Infrastructure team and create required setup for project.
- Develop pipeline with business logic and load the curated data in synapse.
- Performance improvement in pipeline to improve pipeline runtime.
- Develop incremental logic in fact tables to capture delta tables.
- Help BI developers to create data model.
- Provide Hyper support to support team.
- Prepare documentation post completion of project.
- Project management, create Jira task and manage a team as scrum master.
- Exposure in working end to end project.
- Weekly review with product owner about project progress.
- L3 support for all developed Project.
- Data validation and verification after development completion.

Project 2#

Project Title	: CVS Pharmacy	
Team Size	: 1	
Role	: Data Engineer.	
Environment	: Azure Cloud, Azure Data Lake, Pyspark, Python, Databricks and Snowflake	
Duration	: MAY 2019 - April 2022	

Description:

CVS Pharmacy, Inc. is an American retail corporation. A subsidiary of CVS Health, it is headquartered in Woonsocket, Rhode Island United States. Client wants to develop Data Engineering pipeline for ML prediction and deploy the Data Science and Data Engineering pipeline in Azure cloud.

**Roles and
Responsibilities:**

- Developed Feature Engineering for ML algorithm in pyspark.
- Improve performance of Data Engineering and Data Science pipeline.
- Tunning performance of Data science pipeline.
- Implemented transformations and complex business logics using pyspark
- Deploy and maintain the Data science and Data Engineering pipeline in Azure Cloud.
- Scheduled the pipeline in airflow.

Project 3#

Project Title	: The Home Depot
Team Size	: 15
Role	: Developer
Environment	: GCP Cloud, Big query, Python, SQL, Teradata and Abinito.
Duration:	: October 2017 – May 2019

Description:

The Home Depot, Inc., commonly known as Home Depot, is the largest home improvement retailer in the United States. Client wants to extract data from different sources and store it in GCP container. This data can be further used for performing aggregations across different dimensions.

<u>Roles and Responsibilities:</u>

- Analyzed the requirements and source systems, designed the technical architecture for the migration.
- As part of analysis, utilized the Python Pandas and Numpy packages for data profiling.
- Developed programs that fetch data from different data sources.
   Automated the migration of 100TB data from on-premises using Sqoop jobs.
- Migrated the Teradata to Big query using hive jobs.
- Handled various complex scenarios like performance improvement and automation in python.
- Migrated Ab initio Jobs to big Query code in Google cloud. ● Created Data Pipelines using Big query and Jenkins for Data Flow Management.

<u>Declaration</u>

I hereby declare that the particulars furnished above are true to the best of my knowledge and belief.

Place: Chennai   **Prakash Aravind Raj**