Sharath Kumar

P: +1 848 437 0005;
E: sharathchitulla@gmail.com

SENIOR DATA ENGINEER

Snapshot

- Career-focused and motivated professional with an exceptional record of delivering cost-effective, high-performance technology solutions to meet challenging business essentials.
- Demonstrated ability in planning, supervising, and managing various technical operations under extreme milieus with the determination to accomplish day-to-day tasks affirmatively, profound knowledge of technical and business experience in energy, compliance, and regulatory projects.
- Broad experience as a Data engineer with exceptional knowledge of technology software, languages, frameworks, and processes. Experienced in managing requirements from project inception through release.
- Initiative-taking collaborator with the capability to grasp technical concepts quickly and easily; successfully identify the requirement of the project, and handle project development activities.
- Preserve excellent organizational, communication, and people skills; an analytical person and decision-maker who balances the needs of employees with the organizational directives.

Core Competencies

Python | Pyspark | Databricks | Terraform | GCP | SQL Proficiency | ETL Development - Informatica | Snowflake | Airflow | AWS-S3 | Unix Scripting | Analytical Development | Data Warehousing | Data Validation | Data Integration | Data pipeline implementation | RDBMS | DevOps- CI/CD | Requirement Gathering | Production Support | Process Implementation | Process Management.

Technical Skills

▪ Python ▪ Pyspark ▪ Big Data ▪ Databricks ▪ Azure (ADF, DB, ADLS, Synapse, Log Analytics, Logic Apps, Key Vault, Monitoring) ▪ ETL - Informatica Power Center ▪ Amazon Web Services ▪ Snowflake ▪ Airflow ▪ GCP (Big Query, Dataproc, Cloud functions, cloud run, workflows) ▪ AWS-S3, Looker ▪ Robot Framework ▪ Microsoft SQL Server ▪ UNIX scripting ▪ C# ▪ ASP.Net ▪ Microsoft BizTalk Server ▪ Jinja2 ▪ Retool UI ▪ Rest API (Python Fast API)

Key Deliverables

- Developed data pipelines to process the semi-structured data by integrating raw data from different sources and send to the target.
- Demonstrate expertise in information systems which includes the development and execution of action plans and supporting and aligning efforts to meet customer and business needs.
- Establish business requirements and solutions required to solve customer issues by working closely with customers and appropriate teams.
- Execute projects by completing and updating project documentation, managing project scope, adjusting schedules, determining daily priorities, balancing project activities, and arranging the time for self-development activities.
- Provide technical support to the software development team by providing them with a detailed review of software specifications; deliver feedback to the development team on the required changes in software programs.
- Validate results by evaluating the developed programs and establishing specifications and control solutions.

Projects Undertaken

Data Core Engineering | Index Provider | as Senior Azure Data Engineer
Company: McKesson Corporation   **Nov 2022- Present**

- Data core Engineering team acts as a single point of truth for the source where we produce data, transform data and store data.
- Data is ingested from multiple Data vendors and transformed as per business requirements using python language and made available to stakeholders using API or Azure cloud which they use for analytical purposes.
- Expert in using Databricks with Azure Data Factory (ADF) to compute large volumes of data.
- Performed ETL operations in Azure Databricks by connecting to different relational database source systems using JDBC connectors.
- Developed Python scripts to do file validations in Databricks and automated the process using ADF.
- Worked on Python Fast API modules to create Rest API's and deployed them into Kubernetes cluster using terraform which are hosted on Azure.
- Developed Databricks ETL pipelines using notebooks, Spark Dataframes, SPARK SQL and python scripting.
- Data is stored in Microsoft Azure at every stage like raw, scrubbing, and standard zone.

- Also, worked on creation of front-end UI using Retool which connects to Azure, Rest API's, etc. and does all DML actions in the backend.
- Performed Data Analysis, Data Migration, Data Cleansing, Transformation, Integration, Data Import, and Data Export through Python.
- Lead the efforts for source system analysis, data discovery, enterprise data modeling, custody systems integration, data quality and automation.
- Primarily involved in Data Migration using SQL, SQL Azure, Azure Storage, Azure Data Factory, SSIS, Powershell.
- Created a process for migrated data to Azure Data Lake and Azure SQL Data Warehouse.
- Basic knowledge on JavaScript which is being used in Retool.
- Airflow is used to schedule the jobs, and which are being migrated to run deck.
- Grafana is used for monitoring purposes.

ABN Amro | Finance Company | as Azure Data Engineer   Jan 2020- July 2021

- Developed frameworks to read the data from Data Lake using Azure Databricks and applied business transformation rules using Pyspark and orchestrate the jobs through Airflow.
- Involved in the project life cycle including the design, development, and implementation of verifying data received in the data lake.
- Translated data access, transformation, and movement requirements into functional requirements and mapping designs.
- Worked on Dimensional Data modelling in Snowflake schemas.
- Worked on developing Data mart with snowflake schema which has dimension and fact tables concept.
- Automated the transformation logic of pyspark for Data mart.
- Implemented simple and complex spark jobs in python for Data Analysis across different data formats.
- Developed ETL pipelines using Pyspark to process and transform large volumes of data, resulting in a 20% reduction in processing time.
- Optimized Pyspark code for performance, using techniques such as partitioning and caching to reduce processing times by up to 50%.
- Applied complex pyspark transformations on dataframes and applied rules on the dataframes which includes joining of multiple tables, transforming columns, selecting columns based on different cases.
- Ingested data from a data lake to Blob storage in the form of parquet which are to be curated further.
- Created Integrated Datasets from the curated datasets which are source for Data mart.
- Worked on sending data across different teams using ADF pipeline. Python being the most used language.
- Reading the data from various sources and triggering the ADF pipelines where the Databricks notebook gets executed.
- Transforming the data using Pyspark and running them through Airflow Jobs in deployable environment.
- Also automating the entire process using Python and jinja_template formatting right from reading the data and transforming the data based on business rule file and deploying into airflow jobs.
- Created and modifying the CI/CD (azure_pipelines.yml) pipeline to execute scripts present in other repos and developed a proper flow within the CICD template.
- Designing the pipeline and better understanding the technologies like Delta Lake to view the reports and Data check quality framework.
- Knowledge in automated deployments leveraging Azure DevOps and repository for Automation and usage of Continuous Integration (CI/CD).
- Implemented Azure Storage - Storage accounts, blob storage, and Azure SQL Server. Explored on the Azure storage accounts like Blob storage.
- Built workflows on Databricks which run on a scheduled basis.
- Implemented and analyzed pyspark performance issues in databricks to load TBs of data into ADLS.
- Worked on multiple file formats like parquet, delta, csv, etc.
- Have written a script to send logs to Log Analytics workspace.
- Masked the data using function md5/sha2 from higher environment to lower environments.
- Responsible for design development of Spark SQL Scripts based on Functional Requirements and Specifications.

Nerdwallet | Finance Company | as Data Engineer   Aug 2017 - Dec 2019

- Developed ETL process Flow using Python script and run it through Airflow.
- Migrated the Data flow process from Legacy to Delphi i.e., Redshift to Snowflake.
- Source/Target of the data is stored in file formats in Amazon Web Services- S3 bucket.
- Worked on Data analysis part of the Production failures or any mismatch of Data from Legacy to Delphi.
- Data engineering work on developing solutions to generate data insights.
- Developed frameworks, metrics such that end data is used for reporting using looker etc.
- Design and develop the data pipeline architecture to enable extraction of data, transformation of data as per requirement and load into target.

Capital Group | Investment Banking as Intern Application Development Analyst   **March 2017- Aug 2017**

- Performed ETL operations over business data including data extraction from source Files/DB in different formats. Its transformation and loading it into DB.
- Automated ETL processes, making it easier to wrangle data and reducing time by as much as 40% using Robot Framework and Python.
- Wrote python scripts to read, validate and load data into DB, reading logs from UNIX server etc. Used Pandas and NumPy for data manipulation and retrieval.
- Implemented change requests, production support, bugs, enhancement, and development of the existing applications using SQL, Informatica Power Center, UNIX and Shell Scripting.
- Hands-on knowledge of working on Atlassian Tools like JIRA, Confluence, Bitbucket, Bamboo, and XL Deploy.
- Experienced with stories development using SQL DB changes, ETL tool, Autosys Jobs, batch processing and testing the developed stories as part of Quality Analysis.
- Practically worked on Robot Framework and Python involves text processing services, file and directory access, file formats, data libraries, importing modules, Pandas, NumPy, Selenium Libraries etc.
- Proactively involved in production deployments and delivered 24/7 on-call support concerning the occurrence of any major production failures.
- Have a basic knowledge of RabbitMQ, ActiveMQ queues and development of API's using MuleSoft.

Professional Achievements

- Az-104 Microsoft Azure Administrator certified.
- AZ-400 Microsoft Azure Devops certified
- Kubernetes Administrator certified – The Linux Foundation

Education

2017   **B.Tech. in ECE** from Aurora Scientific and Technological Institute, Hyderabad, India
2022   **Masters in Business Analytics** from Saint Peters University, Jersey City, New Jersey, USA