Rakesh Chaudhary



arcsle09plus@gmail.com

9998006038/7016731487

Detailed CV

Profile

An accomplished Technology & Research Professional having 6 + Years of experience in all phases of agile software development lifecycle. Passionate about Analytics Engineering Technologies and delivery of effective solutions through creative problem-solving.

Education

Master of Science - Financial Mathematics, Maharaja Sayajirao University of Baroda ☑ Jun 2016 - Jun 2018

Certification

- AWS Certified Data Engineer □
- Certified SAFe® 6 Scrum Master □
- Databricks Lakehouse Fundamentals
- HackerRank ☑

Professional Experience

Manager - TB Cash Analytics, Standard Chartered Bank ☑

May 2024 - present | Bangalore, India

- Contributing as Analytics Engineer to build Cash Data Lake using Bank's various tranctional processing systems.
- Leading the Reporting Automations and Cash Data Assets Initiatives aiming to generate Bank's Cash Revenue Opportunities.
- Conducting various POCs to enhance and optimize the existing cash data models.

Technical Skills

Programming Languages Python SQL	• •	•	• •
Databases/DWH MySQL Oracle Redshift	• •	•	• •
Data Transformation Tools Pyspark Pandas Polars DB		•	• •
DevOps Tools Git Bitbucket Docker	• •	•	• •
Cloud Services EC2 S3 Lambda Glue Atho Databricks	• • ena	•	• •
Project Tracking Tools JIRA and Confluence	• •	•	• •
Front End Tools Tableau Streamlit	• •	•	• •

Projects Summary

Transaction Info Data Model (StanChart) Tech: Pyspark,Polars,S3,Airflow,Git

- Implemented an end-to-end ETL pipeline to integrate critical billing & payments transaction data into the HDFS Cash Data Lake.
- Collaborating with stakeholders to map relational sources.
- Leveraged PySpark for large-scale transformation and ingestion.
- Built cost effective data validation modules using python-polars to implement business checks.
- Automated and orchestrated workflows using Apache Airflow for reliable, scheduled data operations.
- Enabled downstream teams to access high-quality, up-to-date transaction data, reducing manual data correction efforts by 60%.

Cash Performance View (StanChart) Tech: Pyspark, Docker, GIT, JFROG, Shell

- Architected, Developed and Deployed Reporting Application to monitor the Bank's Cash Revenue (NFI and NII) and Liabilities on Monthly basis.
- Built Analytical Data Models using Bank's Internal OLTP system(Oracle) and data lake(HDFS)
- Automated data extraction using SQL, Pyspark and Shell Scripting.
- Transformed the Data using Python-Polars library to build cost effective data processing modules.
- Discovered the excel workbook level Encryption Method in Python to meet bank compliance.
- 2000 Analytical Sales Level Reports + 42 Country Level Reports to Country Sales Leadership in Sharepoint Site.

Sr. Analyst - Innovation & Automation, NielsenIQ ☑

Nov 2018 - May 2024 | Remote

- An Integral Part of Global Innovation and Research Team.
- Developed the technical solutions for mid and large scale data collection, Ingestion and Validation activities.

Global Item Xcode App (NielsenIQ) Tech: Databricks, Pyspark, Azure-Blob, python-polars

- Developed and schedule the data extraction and transformation pipelines to build country level product info vocab using pyspark in databricks.
- The data loaded/written to azure blob as parquet files.
- Developed the string match module using polars and jaro-winkler algorithm to produce top 4 matches for respective user input product description parsed from end users via Django Web App
- Wrote unit test cases for string match module.
- The country level data gets autorefreshed every week.
- Reduce the manual item coding efforts to 30%

Ecom Data Acquisition (NielsenIQ) Tech: Scrapy, Selenium, BeautifulSoup, Requests,RapidFuzz, EasyOCR, Polars, Pandas, Multiprocessing,EC2,S3,Docker

- Acted as Lead Developer for various mid to large scale webscraping Initiatives.
- Explored and Implemented various robust webscraping techniques like Authentication Tokenization, chaining APIs, Sitemaps, Bypass Antibots, Optimal Usage of external proxy services to extract dynamic and static data
- Used Various Python libraries(Scrapy,Selenium,BS,Requests, Selenium Wire) for webscraping data.
- Designed and Developed various Web Scrapers to scrape product images, price, sales, units and categories data (Retailers and Google sites) to extract ~2 million products data and stored in AWS S3 every week.
- Developed and deployed 105 webscrapers for Italy based FMCG retailer websites in order to extract the weekly promotional flyers(Avg 30 GB Data Extraction) and stored in AWS S3.
- Built various shell script utilities to schedule the webscraper jobs.

ECOM Data Validation Tool (NielsenIQ) Tech: Python, Polars, Streamlit, S3, EC2 & Docker

 Developed and Deployed the Web App that performs various validation methods on the large scale web scraped datasets.