DIVYANGANA PANDEY

Boston, MA 02115 | (+1) 617-412-9519 | divyangana75@gmail.com | LinkedIn

EDUCATION

Northeastern University   Boston, USA
Master of Science in Data Analytics Engineering   Expected May 2023
Relevant Coursework: Foundations of Data Analytics, Data Mining, Data Management, Operations Research

Banasthali Vidyapith   Rajasthan, India
Bachelor of Engineering in Biotechnology   July 2015
Relevant Coursework: Mathematics I-III, Probability and Statistics, Computer Fundamentals and Programming Object Oriented Programming, Data Structures

TECHNICAL SKILLS

Programming Languages: Python (NumPy,SciPy,Pandas), R(matplotlib,dplyr, ggplot,tidyr), Java, Scala
Big Data: Hadoop, Hive, Kafka, HBase, Oozie, Sqoop, Shell Scripting, Databricks, Spark
AI/NLP: Rasa, Google Dialogflow
Google Cloud: BigQuery, DataFusion, Dataflow, Dataprep, Cloud SQL, Cloud Composer, IAM, Apache Airflow, Cloud Functions
Databases: MySQL, MongoDB
IDE/Tools: IntelliJ, Eclipse, Jira, Git/Gerrit, Gitlab, Pycharm, Tableau
DevOps: Jenkins, Sonarqube

EXPERIENCE

ACCENTURE SOLUTIONS PRIVATE LIMITED   Gurgaon, India
Data Engineer for Market Intelligence Research McDonalds   December 2020 - July 2021
- Spearheaded an initiative to migrate, deduplicate & denormalize multidimensional data from AWS S3 to GCS buckets by designing automated ETL pipelines for data ingestion and warehousing for complex and disparate datasets
- Enabled sales teams to increase performance by 10% in its first year by developing & training various classification & prediction ML models by leveraging customer's transactional data, generating key features like churn probabilities & product propensity
- Deployed a Flask-based application using GCP platform to report, visualize & formulate effective marketing campaigns for 10+ countries spanning a market of 800k+ customers with KPIs relevant to stakeholders

Data Engineer for Accenture Internal Resource Monitoring Initiative   April 2020 - November 2020
- Designed an internal resource monitoring tool using Google Cloud Platform for Accenture's global leadership to help monitor and visualize resources across more than 10 countries in real-time
- Configured Google Compute Engine as data landing zone for data retrieval & developed multiple ingestion and transformation scripts to transfer data to secured GCS buckets while utilizing Dataprep for data wrangling and cleaning
- Orchestrated and scheduled CI/CD pipelines with Cloud Composer eliminating operational overheads
- Revamped the existing manual architecture by curating a Google Data Studio dashboard with illustrative visualizations using data stored in BigQuery datasets

Data Engineer for Helwett Packard Enterprises   April 2019 - March 2020
- Collaborated with multiple cross functional teams to enhance business opportunities by translating business requirements to design a technical delivery roadmap & presenting useful actionable insights like potential new customers via Qlik dashboards
- Created a MongoDB Datalake using an SBT Scala project to implement functional and logical transformations over large multi-relational tables. Optimized transformation logics using sophisticated partitioning & indexing techniques. Calibrated memory & cluster configurations accelerating the query performance for time expensive jobs up to 10%

TATA CONSULTANCY SERVICES   Gurgaon, India
System Engineer for Lloyds Bank   June 2015 - October 2018
- Configured HDFS platform for optimizing data storage to expand services portfolio for a major federal client. Developed shell-based sqoop scripts to ingest & migrate raw data from legacy storage systems to Hive & Hbase using data modeling techniques
- Reduced the number of existing Data Governance & Quality issues by 30% per week by building exhaustive data validation scripts identifying special characters and outliers

PROJECTS

Extract Time Series features for human activities using R
Performed time series modeling using acceleration level data of the activities of fifteen subjects. Applied visibility graphs, utilized ggplot to visualize patterns and calculated permutation entropy and complexity

Recommendation System using Amazon Reviews Data
Built a recommendation system using Amazon.com reviews to predict top 5 items in each category using multiple regression models