Pradeep Kumar Chokkahalli Prabhakar

Richardson, TX, 75080; +1(214)-499-3315; GitHub; pradeep.chokkahalli@gmail.com; LinkedIn

EDUCATION

The University of Texas at Dallas May 2023
Master of Science, Business Analytics Major in Data Science |Dean's Excellence Scholar |Mentor **GPA 3.83/4.0**
Coursework: Google Cloud & Big Data Analytics, Applied Machine Learning, NLP, Data Visualization, Data Warehousing
Dayananda Sagar College of Engineering, Bengaluru, India August 2017
Bachelor of Engineering, Electronics and Communication **GPA 3.45/4.0**

CERTIFICATIONS AND TECHNICAL SKILL EXPERIENCE

Certifications: Graduate Certificate in Applied Machine Learning, Python-Data Engineer, Google Data Analytics, Intermediate Python & R, DP-900 Data Engineering on Microsoft Azure

Tools: Microsoft Azure Cloud Tech Stack, GCP, ETL/ELT Process, MySQL, Hadoop, Hive, Azure services, Docker, Tableau, PowerBI, Kubernetes, NoSQL MongoDB, Pig, PostgreSQL, MS Excel, Spark, SAP ERP, Jira, MATLAB, ArcGIS Pro, Oozie, Git Bash, GitHub, Kanban

Programming: Python (Pandas, NumPy, Matplotlib, Keras, Seaborn, PySpark, Scikit-learn, NLTK, spaCy), Spark Scala, SQL, R programming, PySpark, SAP ABAP, SAS, Stata, UNIX / Linus Scripting, Java

Business Logic: ETL/ELT Process**,** API architecture, HDFS, Data Structures, Virtual machines, Database Management, Azure DevOps, AI

Soft Skills: Collaboration, Critical Thinking, Strong Communication Skills, Result Oriented, Decision Making, Problem Solving, Innovative

PROFESSIONAL EXPERIENCE

Data Engineer, iTrending Solutions **11th August 2022 - Current**

- Designed and built a scalable data engineering solution on **GCP**, leveraging **Python**, **PySpark**, and **SQL** to develop **ETL** pipelines, **optimizing queries** and managing data dependencies, resulting in a 40% improvement in data processing efficiency.
- Implemented large-scale data structures and pipelines for a retail client using **Python**, **Hadoop**, and **HDFS**, organizing and standardizing data to improve reporting efficiency by 30% and enhance decision-making capabilities.
- Improved **data warehouse** architecture for an services company, enhancing infrastructure components and implementing reporting tools, reducing data latency by 30% and increasing reporting accuracy.
- Scheduled and orchestrated all the pipelines using **ADF** through **Kafka** improving data ingestion rate by 20%.
- Developed code for Data Transformations & Aggregations using Spark/Scala a in Azure Databricks Notebook & Scheduled Data Factory pipelines to execute over 100 spark jobs on **Azure Databricks**.
- Developed ADF Dataflows to merge the Incremental data from ADLS to Synapse tables using SCD1 approach.
- Designed and optimized queries against **HDFS** environments for a marketing firm, demonstrating strong knowledge of **Hadoop architecture** and enhancing data pipeline performance.

Data Analyst, Accenture, ABInBev **23rd March 2018 – 22nd Feb 2021**

- Reconciliated 15 days backlog data from **Hybris** through **Snaplogic** to EAP using a **Oozie** Workflow and Coordinator for Qlik reports.
- Utilized strong articulation skills to effectively convey complex data engineering concepts and project updates to non-technical stakeholders, resulting in a 25% increase in project approval rates
- Scrutinized and visualized large datasets using **Tableau**, integrating powerful models and algorithms into operational solutions for pattern recognition which succor reduced incidents by 30%.
- Optimizing data flow using **SQL** scripting and enhancing Business performance by finding RCA of **SOLMAN** (Solution Manager) errors and effectively reducing Problem Records from 10 to 4 a month.
- Lead as Agile Scrum Lead for change management, deploying **Business Intelligence** Applications 100% on time.

PROJECTS

Bike Rent Prediction, UTD – Python, Machine Learning Models **May 2022**

- Engineered a Model to predict Rental Bike Availability. Implemented a Gradient Descent Algorithm for Linear Regression. Incorporated and Documented SVM and Decision Tree for classifications with 92% accuracy.

Data Ingestion for Truck Risk Factor Data, UT Dallas - Big Data, SQOOP, Hadoop, Hive, Pig, Impala, Tableau, MS Excel **Nov 2022**

- Integrated all Big Data Platforms for data ingestion. Analyzing Hadoop cluster and different Big Data analytic tools.

UT DALLAS LEADERSHIP, AFFILIATIONS AND ORGANIZATION EXPERIENCE

MSBA Mentor, Business Analytics

- Mentored graduates in the field of Data visualization, AWS Cloud analytics and Advanced Big Data.

MAC Club, *Creative Officer*

- Designed events for club members for skillset enhancement in the field of Data science.

ADDITIONAL INFORMATION

Languages: Advanced English, Advanced Kannada, Basic Spanish, Proficient Hindi, Fluent Telugu
Eligibility: Eligible to work in the U.S. under OPT full-time employment for 36 months (STEM only) without sponsorship.