# Pradeep Kumar Gontla

(585) 530-9572 | pg3328@rit.edu | https://www.linkedin.com/in/pradeep-gontla/ | portfolio | github.com/pg3328

## Education

**Master of Science: Rochester Institute of Technology**　　　　　　　　Rochester, NY, USA (Jan 2022 - May 2024)
- Master of Science in Computer Science, CGPA - 3.72/4.0

**Bachelor of Technology　Vellore Institute of Technology**　　　　　　　Vellore, India (June 2014 - June 2018)
- Bachelor of Technology in Mechanical Engineering, CGPA – 8.58/10

## Technical Skills

- **Proficiency**: Python, Java, Lang chain, SQL, Spark, Hive, Hadoop, AWS Cloud, Azure Cloud, Snowflake platform, GitHub, Pandas, Shell scripts, ETL, Tableau, Databricks, Autosys, Supervised and Unsupervised Machine Learning Algorithms.
- **Exposure:** Scala, Kafka, Flink, Linux, Jenkins.

## Professional Experience

- **Auxiliary Services Tech – RIT,** *Data Scientist Intern*　　　　　　　　　　(Feb 2020 – May 2022)
  **Tech Stack**: Microsoft Logic Apps, Azure Data Factory, Azure Databricks, Azure Data Lake Storage, PySpark.
  - Automated the process of meal plan cancellations by integrating vendor portals using Restful APIs and Microsoft Logic Apps, which reduced the time spent on managing the cancellations from 2 hours to 15 minutes.
  - Built and scheduled reports in Atrium campus (MySQL Server) using complex SQL queries for RIT Leadership, which reduced transaction clarification tickets by 30% per month.
  - Contributed to the RIT datahub project, by developing an ETL pipeline to load data from resident and fund management portals to ADLS gen2 using Azure Data Factory, Azure Databricks, and Pyspark.
  - Performed data profiling, cleansing, and transformations using PySpark, Spark SQL, and Azure Databricks.
- **e4e Labs Private Limited, Bengaluru,** *Data Engineer*　　　　　　　　　　(Sep 2020 - Nov 2021)
  **Tech Stack**: Hadoop, Apache HDFS, Apache Sqoop, Apache Hive, RDBMS, Project management, Data Architecture
  - Developed a single source of truth within HDFS data lake by creating and maintaining the ETL pipeline with the help of Sqoop, Hive, Shell scripting, and Autosys.
  - Finetuned HQL queries that transform data, by implementing optimized bucketing and partitioning strategies, slashing the query execution time by 44%.
  - Conducted proof-of-concept on utilizing Pyspark for performance benefits over using Hive with spark core.
- **Alfa Laval India Private Limited,** *Business Intelligence Developer*　　　　　(Aug 2018 – Jun 2020)
  **Tech Stack**: MSBI, SSIS, SSAS, SSRS, Tableau, ad-hoc reporting, Python
  - Performed ad-hoc data analysis utilizing Python for transformations and aggregations and reported the analysis using Excel to generate insights into the equipment life cycle, leading to a 33% enhancement in number of contracts.
  - Automated the data analysis process using Python scripts and cron scheduler from databases, data marts, and data warehouses to perform data analysis and reporting.
  - Designed and developed ETL processes using SSIS for data ingestion from various RDBMS and flat file sources which include installed base, spare purchases, and billing information.
  - Collaborated with business stakeholders to understand data requirements and design data models, contributing to 5+ data modeling sessions.

## Academic Projects

- Business Reporting Chatbot:
  - Created a React Agent, backed and orchestrated by langchain with OpenAI LLM to interact with the relational databases and answer business questions with data-backed AI-powered insights on streaming data.
  - Developed prompts along with examples to ensure optimal performance and utilized performance testing mechanisms (using Apache Jmeter) to finetune the prompts and achieved an average response time of 2.3 s.
- Dynamic Query Scheduling in streaming data pipelines:
  - Developed a fault-tolerant streaming data pipeline using Apache Kafka, and Apache Flink as a part of an experimental setup on 3 node cluster hosted on a research computing cluster at RIT.
  - Reduced latency in stream processing pipelines by scheduling the queries based on predicted deadlines improving real-time responsiveness by 22% over default scheduling on linear road benchmark.
- Docker Chef:
  - Developed a containerized full-stack application utilizing Python, Flask, Docker, HTML, CSS, and JavaScript to enable recipe searches based on input keywords.
  - Implemented a vectorized comparison of input keywords against a recipe dataset using Python's Scikit-learn TF-IDF vectorizer. Applied cosine similarity to rank and present the top 3 matching results.

## Certifications

- Databricks Certified Associate Developer for Apache Spark - June 2024.
- Prompt Engineering for Developers - April 2024
- Langchain with python boot camp - February 2024
- AWS Cloud Practitioner - March 2023
- Certified Scrum Master - December 2022