

# Gautham Kolluru

GitHub | LinkedIn | Stack overflow | Hacker Rank

Mob: +1 240-639-9934 | Email: gautham.kolluru@gmail.com | Website: thegauthams.com

## Summary

I Love **solving problems** *in* consuming data (*Data Engineering*) and **by** consuming data (*AI/Data Science*) using **Programming** and **Mathematics**, along with learning new technologies and implementing them by building & optimizing systems & solutions.

## Skills

Programming Languages: **Python, Java, JavaScript, Rust, SQL, Bash, C, C++, BASIC**

Databases: **SQL Server, Mysql, SQLite3, Postgresql, MongoDB**

Cloud, Bigdata and Orchestration: **Google Cloud Platform (GCP), Hadoop, Spark, Dask, Snowflake, Airflow**

AI/ML (& related): **Scikit-Learn, TensorFlow, DialogFlow, RASA, Matplotlib, Seaborn**

Operating Systems: **Unix (macOS), Linux & Windows**

## Experience

### Data Engineer, Cisco; RTP, NC, USA — 2021-Present

Building a log/file aggregator with the capability of dumping the aggregated logs to a single file basing on different log levels/keywords or forwarding them as an email to the concerned with simple mail setup configuration along with a scheduler to automate the same.

*Leveraging **LLMs** for summary generation*

Built a network security *threat analytics* system involving scheduled multiple data transfers between different systems along with statistics based analytics system resulting in threat and compromise identification and subscriber segmentation.

- Scheduled and automated transfer of data between varied systems including on-premises and cloud
- Built the Statistical Analytics module for identifying threats and indicating compromises
- Leveraged upon MLFlow open-source python package for maintaining and visualising analytics process logs
- Integrated the system with Cisco's SecureX platform

*Built a customised (loosely coupled) **K-means** module along with the data transformations around it*

### Data Engineer, 84.51o; Cincinnati, OH, USA — 2020-2021

Built an automated system around a *data science* model that fetches data from multiple sources, aggregates, transforms, fed to the model for prediction and the resulting data is written on to the target location.

- Developed the workflows using Python and Airflow
- Developed the data jobs using Pyspark

*Executes **20 billion rows** with **20 columns** and **40 steps** of transformation in about **23 mins***

### Data Engineer, Cisco; Bengaluru, KA, IND — 2019-2020

Automated data pipelines involving data transformations and migrations between various storage services

- Smartsheet to Hadoop (on GCP)
- Hadoop (on GCP) to Snowflake
- Hadoop (on GCP) to Hadoop (on GCP)

*Each script executes **twice weekly** handling incremental updates*

### NLP Engineer, Cisco; Bengaluru, KA, IND — 2019-2020

Developed a chatbot that interacts with the user as first point of contact and transfers the conversation to a specific bot (with the context) for the specific resolution.

*Implemented **context transfers** between the bots*

### **Consultant, Talent Sprint; Hyderabad, TG, IND — 2018-2019**

Developed and deployed a *data science model* that suggests the optimized arguments for the execution parameters of the task at hand.

### **Software Engineer, Global Data; Hyderabad, TG, IND — 2018-2018**

Automated 2 data pipelines involving extraction of data from multiple sources including **HTML and XML pages & HTTP and FTP** servers and loading onto **on-prem SQL Servers** after required cleanings and transformations.

*Implemented **changing proxies** and leveraged on **multi-processing** execution and runs **thrice weekly** on multiple machines*

### **Software Developer, Invincible Software Solutions Integrated; Visakhapatnam, AP, IND — 2017-2018**

Developed **Backend and Database** for **12** web applications. Built descriptively analyzed data visualization reports using **Tableau** (& **SSRS**) for *USA* based *federal* clients.

### **Project Engineer, Royal Oman Police; Muscat, OM — 2013-2015**

Designed and developed database for Speed Cameras and developed *automated transfer of data* from the *primary (OLTP)* servers to *secondary (OLAP)* servers.

### **Automation Engineer, Mott MacDonald; Abu Dhabi, UAE — 2009-2012**

- Implemented automation multi-threaded document pooling using string search across systems from different geographies
- Automated emails sends to a list of email address with the above found documents as attachments

## **Education**

Andhra University, Visakhapatnam, AP, IND — B. Tech (**Computer Science Engineering**), 2009