AGASTHYA NAIB GURUPRASAD

Lane Cove North, Sydney

https://www.linkedin.com/in/agasthya-naib-guruprasad-48698918b

PROFESSIONAL EXPERIENCE

Passionate and results-driven Data Engineer specialising in infrastructure as code and Terraform, with Python as the primary language for developing scalable and efficient data pipelines.

Experienced in leveraging cloud platforms such as Google Cloud Platform and AWS to architect robust and automated solutions for real-time web analytics data processing, including expertise in Snowplow web analytics and NPAW (Nice People at Work) solutions. Additionally, adept in DevOps practices, ensuring seamless integration and deployment of data-driven applications. Proficient in Go language development, offering versatility and expertise in optimising performance and scalability. Skilled in creating interactive dashboards and analytical tools to drive data-driven decision-making. Actively seeking opportunities to apply my Python proficiency, complemented by Go capabilities and DevOps expertise, to innovate and drive business success in dynamic environments.

Nine Entertainment Co, Sydney, Australia DEC 2021 – Present **Data Engineer**

Experienced Cloud Engineer with a strong proficiency in Google Cloud Platform (GCP) resources, including Composer, Cloud Run, Cloud Build, Cloud Storage, Dataflow, Compute Engine, Pub/Sub, GCR/Artifact Registry, and monitoring/logging tools. Adept at creating pipelines on Airflow in Composer for seamless data transfer between AWS and GCP environments.

Key Accomplishments:

- Designed and implemented ETL pipelines, migrating data from Adobe Analytics and Adobe Campaign to GCP for advanced analytics. Automated processes with Airflow, enhancing data availability and accuracy.
- Architected the migration of Snowplow infrastructure from AWS to GCP, deploying scalable, real-time data processing pipelines using Terraform and Cloud Build. Achieved 99.99% uptime during high-traffic events like the Olympics.
- Streamed data into the Customer Data Platform (CDP) using Dataflow, enabling real-time analytics. Integrated video analytics data into CDP using Kafka and Dataflow.
- Utilised Terraform for IaC, automating infrastructure deployment for multiple projects, including server-side tagging with load balancing and SSL certificates.
- Developed monitoring dashboards and alerts for Pub/Sub and Dataflow, identifying bottlenecks, ensuring high performance, and implementing cost-cutting strategies.
- Created CI/CD pipelines for continuous delivery, deploying Docker images, and integrating security scanning tools like Snyk to maintain service integrity.
- Built a Redis-backed caching API for ad-serving, improving performance through gRPC communication between frontend and backend components.
- Enhanced GDPR compliance by introducing Google Tag Manager (GTM) tags for user consent management and data privacy.

- Contributed to agile process improvements with Jira data models and dashboards for sprint-level analysis, supporting effective decision-making.
- Maintained Snowplow as the world's most cost-efficient self-hosted service, optimising GCP resources through log sinking and scheduling.
- Created a flattened, optimised table for multiple schemas used by Snowplow, utilising dbt to help analysts and other data engineers easily use the table, rather than coalescing between multiple schemas.

Additional Experience:

- Decommissioned Go lang repositories and updated team-owned repositories to newer environments in collaboration with Platform Engineering.
- Contributed to an ad-serving API using Go on AWS, integrating Redis for caching and querying data from BigQuery (GCP) to enhance performance.
- Streamed NPAW data into the CDP using dataflow pipelines, enabling real-time processing and profiling to improve data quality and insights.
- Consumed video analytical data from a broker set up by NPAW using Kafka and SASL security.
- Set up infrastructure for server-side tagging on GCP using Cloud Run with Terraform. - Configured required load balancing and SSL certificates for server-side tagging on an existing network.

Ericsson, Sydney, Australia Aug 2021 – NOV 2021 **Data Scientist student intern**
- Created Deep learning models to help the company identify if procedures are followed or not.
- Pytorch was used to develop an image classification model.
- The object detection model was created to help the company detect the required object in images.
- The OCR model was created to extract useful textual information from the images. Helped the company accept or reject an installation process based on images provided and saved time and effort required to do the work manually.

Investa Property Groups, Sydney, Australia AUG 2020 –FEB 2021 **Data Engineer intern**
- Built production-ready data pipelines with Airflow, processed and cleaned data and created a data warehouse and data mart for analytical purposes.
- Collected data from websites using python's beautiful soup library.
- Performed EDA process on acquired datasets using python's pandas.
- Performed Sentimental analysis on tenant's briefings using Spacy.
- Created a Time Series model using python that shows the tenant's engagement. Created Dashboard using Power BI to show results.
- The project helped Investa in identifying events that show traction and events that don't.

Trainingworld360 Solutions Pvt Ltd, Bangalore, India SEP 2019 – JAN 2020 **Software**

Developer and Tester
- Collaborated with client-side developers by performing Web-based testing using selenium. Researched to find a seamless way of identifying trucks working in a dusty environment using Arduino with IR sensors.

EDUCATION

University of Technology, Sydney, Sydney, Australia 2022 **Master of Data Science and innovation**

Dr Ambedkar Institute of Technology, Bangalore, India 2020 **Bachelor of Computer Science and Engineering**

SKILLS & OTHER

- Programming Languages: Python (ML/DL and web scraping), JAVA, R, MongoDB, SQL, Go, Javascript
- BI tools: Power BI, Tableau, Git
- Data Engineering: Kafka, Airflow, Spark, Beam, DBT
- Cloud: Azure, GCP
- Docker, YAML (Cloud Build)
- CI/CD terraform, cloud build, Docker and Kubernetes
- GCP resources: Cloud Run and Cloud run jobs, Pub/Sub, Cloud Functions, Compute Engine, Monitoring, maintain IAM and service accounts, Cloud build, Big Query, Composer, Load Balancer
- Load testing using Locust
- Google tag manager
- Setting up server side tagging

Certifications and Courses
- Azure AI-900 Fundamentals-04/2021- Microsoft Certified