ANNU KUMAR

DATA ENGINEERING | TECHNOLOGY ANALYST

annukmr484@gmail.com | +91-7875187037 | www.linkedin.com/in/annu-kumar-058776a9

EDUCATION

BACHELOR OF ENGINEERING

Sinhgad Institute of Technology / Pune / 2016

BE – Computer Engineering

- With Distinction 66.45%

HIGH SCHOOL – HSC

Kendriya Vidyalaya No. 2 / Mumbai / 2012

- With 68%

SSC

Kendriya Vidyalaya No. 2 / Mumbai / 2010

- With 80%

CERTIFICATIONS

GOOGLE CLOUD CERTIFIED PROFESSIONAL DATA ENGINEER

Dec 2020 – Dec 2022

SKILLS

Programming Languages & Scripting | J2EE, Python, Shell scripting

Bigdata | Hive, PigLatin and Hadoop Ecosystem.

Platforms, Frameworks | Hibernate, Spring Boot Microservices, AngularJS, Postman, Junit

Scheduling & Dataflow | Apache Airflow, Apache NiFi, Apache Beam

IDE & Tools | Jenkins, Maven, Github, Eclipse, IntelliJ, Anaconda, Jupyter Notebook, Jira, Docker

Cloud | Google Cloud Platform

PROFESSIONAL SUMMARY

Result Oriented and problem-solving software engineering professional and a strong team player with a focused Cloud Data Migration, Big Data background. Over 4 years of experience, well versed with Agile methodologies. Has proven performance with writing dataflow jobs for ingesting data to various components of Google Cloud storages, both in batch as well as in near real time scenario along with developing REST API's for present firm's Google Cloud solution.

WORK HISTORY

INFOSYS LTD | DEC 2016 – PRESENT

BIG DATA ON-PREM MIGRATION TO GOOGLE CLOUD PLATFORM

Collaborator and developer of cloud data migration solution in collaboration with Google. The solution is a petabyte scale, multi-cloud data management platform that is uniquely going to be open source. No matter which cloud provider you're using, it manages a quick, seamless, secure and reliable transfer of mission-critical data. It further automates the collection and delivery of data from a variety of on-premise locations to the cloud. (or other on-prem sinks)

Development of Spring based microservices to help non-technical end users configure their data migration and transformation through a GUI.
Microservices deployment through CI/CD pipelines at target servers, leveraging Jenkins for builds, Nexus for staging the jars/scripts & Ansible tower for final deployment of updated services.

Packaging each of these microservices in a docker image, so that we can maintain a high availability of the services by hosting them in multiple containers across several regions within the organization/s.
Hands-on knowledge of software containerization platforms like Docker and container orchestration through Docker Swarm.

Masking the fields identified as PII by the user. This leverages pandas library for columnar processing.
Data cleansing and transformation though Apache NiFi.

Event based ingestion of files in cloud's data warehouse solution, on receipt of files at cloud storage (Cloud pub/sub, big query in case of google cloud)
Running analytics queries against the data finally stored/ingested in bigquery.

Training and consulting customer teams to solve their critical business use case and problems in the most efficient way on the cloud solution.

Some concrete outcomes of the Solution

- Over 50% cost reduction in data ingestion and overall ownership.
- On-prem Data and file availability at cloud reduced to 10-15 mins from previous 2 hours.
- Real time feed and job monitoring.
- 5 mins : Time for Element level data lineage report generation.
- Flexible job configuration for the feeds (masking, custom attribute generation).
- Time taken to deploy a new data feed reduced to only few days as compared to several weeks conventionally