Position: **Principal Analyst – GCP**

📍 Bangalore, Karnataka

Factspan Overview:

Factspan is a pure play data and analytics services organization. We partner with fortune 500 enterprises to build an analytics center of excellence, generating insights and solutions from raw data to solve business challenges, make strategic recommendations and implement new processes that help them succeed. With offices in Seattle, Washington and Bangalore, India; we use a global delivery model to service our customers. Our customers include industry leaders from Retail, Financial Services, Hospitality, and technology sectors.

Job Description

As Principal Analyst,

➢ Knowledge of data engineering technologies, architecture, and processes. Specifically, GCP, Hadoop ecosystem, Kafka, and common third-party integration and orchestration tools.
➢ Good knowledge of multi-cloud data ecosystem and build scalable solutions on cloud (GCP)
➢ Good knowledge of Big Data Ecosystem-Spark, Hadoop, Databricks
➢ Work across 3-4 teams to develop practices which lead to the highest quality products and contribute transformation change within the cloud
➢ Experience building large scale data processing ecosystems with real time and batch style data as input using big data technologies
➢ Experience in any programming language like Scala or Python.

Responsibilities

➢ The Principal Analyst will be responsible for driving large multi-environment projects end to end and will act more of individual contributor
➢ He / She will work on designing the architecture, setting up the HDP/Cloudera cluster infrastructure, building data marts, data migration, and developing the scripts on Hadoop ecosystem
➢ Design and develop reusable classes for ETL code pipelines and responsible for optimistic ETL framework design.
➢ The candidate should be able to plan and execute the projects and be able to guide the junior folks in the team
➢ The person should be comfortable to engage with communication with internal and external stakeholders

Qualifications & Experience:

➢ Bachelor's or Master's degree in a technology related field (e.g. Engineering, Computer Science, etc.) required

➢ 5+ years of experience in developing Big data applicationsin Cloud, preferably GCP.

➢ Design and develop new solutions on the Google cloud Platform specifically for building Data ignition pipelines, Transformation, Data Validation and Deployments.

➢ Automate GCP data pipelines and work on Airflow.

➢ Create Complex Data Pipelines in GCP.

➢ Hands on experience with ETL pipeline development and functional programming

➢ Must be good in developing ETL layer for high data volume transaction processing

➢ Experience with any ETL tool (Informatica/DataStage/SSIS/Talend) with Data modelling, and Data warehousing concepts

➢ Good to have jobs execution/debugging experience with pyspark, pykafka classes, with combination of Docker containerization

➢ Agile/Scrum methodology experience is required.

➢ Excellent presentation and communication skills

Why Should You Apply?

Grow with Us: Be part of a hyper- growth startup with ample number of opportunities to Learn & Innovate.

People: Join hands with the talented, warm, collaborative team and highly accomplished leadership.

Buoyant Culture: Embark on an exciting journey with a team that innovates solutions everyday, tackles challenges head-on and crafts a vibrant work environment