

# Hassan Amir

Lahore, Pakistan | (+92)-3474287703 | [hassanamir706@gmail.com](mailto:hassanamir706@gmail.com) | [LinkedIn](#)

## EDUCATION

**Forman Christian College University**  
Bachelor's of Computer Science

**Lahore, Pakistan**  
2015- 2019

## WORK EXPERIENCE

**Systems Limited - Visionet Systems**

**Lahore, Pakistan**

- 5 years' experience in analysis, design and development of Data engineering and warehousing solutions with good knowledge of data modelling and data architecture.
- Experience in designing, development and deployment of data pipelines in cloud platforms like Google Cloud Platform (Cloud Functions, Composer, Big Query), Microsoft Azure (Data Lake Gen 2, Data Factory, Databricks, Azure Synapse) and AWS (Glue, Redshift, S3)
- Utilized snowflake for data warehousing and analysis taking advantage of its scalability and performance to build dimension model incorporating slowly changing dimensions and normalization with change data capture mechanism integration to improve functionality and ease of reporting.
- Experience as data modeller to ensure data profiling, source data analysis and representation of data in a structured format.
- Implementation of big data solutions using Databricks using AWS Glue jobs to ingest data from various sources such as Amazon S3, RDS, Redshift, or DynamoDB into a centralized data lake or data warehouse ensuring timely processing and availability of updated data for downstream analytics.
- Leveraged PySpark to perform data preparation tasks like cleaning, transforming, and structuring raw data into formats suitable for analysis
- Experienced in Python and SQL to develop workflows and data integration from different sources.
- Solid understanding and experience of Data Build Tool (DBT) to build pipelines.
- Functional domain experience involves Insurance, Banking and E-Com
- Effective in working independently and collaboratively in teams with experience in stakeholder interaction and requirement gathering to ensure seamless communication and objective fulfillment.
- Experience in marketing analytics using google analytics 4 to measure and optimize metrics for business growth.
- Experience with reporting metrics and KPIs in LOOKER for business analytics.

## PROJECTS AND INDUSTRY CATEGORIZATION

---

Client: Gt's living foods

Cloud Infrastructure Employed: Microsoft Azure/ Snowflake

Responsibilities:

- Create azure Data Factory pipelines to read files from blob storage.
- Impose custom schema and calculations on files read.
- Maintained and created data warehouse entities in Snowflake for reporting purposes in the form of a dimension model storing business activity
- Implemented complex pipelines in snowflake incorporating slowly changing dimensions to keep track of changes in data while providing seamless integrations with change data capture enabled to target candidate row-based processing.
- Optimized query performance using result caching, query pruning, and clustering keys.in snowflake while using materialized views for optimal resource allocation to do in memory calculations.
- Designed schemas using dimensional modelling (e.g. snowflake schema) to facilitate efficient querying and reporting, while implementing normalization and denormalization techniques to ensure granularity of data and balance performance and data integrity.
- Documented detailed instructions and inner workings specifying how each mechanism developed works for technical and non-technical stakeholder understanding which includes architecture diagrams, DDL, DML, comments in code and mapping sheets.

Client: Digicel

Cloud Infrastructure Employed: Amazon Web Services (AWS)

Responsibilities:

- Creating DWH model from ground up while working with Data Modellers in Redshift.
- Developing big data pipelines using Apache Spark on AWS, including PySpark-based Glue jobs for data transformations.
- Creating Lambda function-based pipelines to move files from landing to raw and curated zones.
- Implementing data partitioning and loading strategies in S3 and Redshift, leveraging Spark for efficient data processing.
- Adding remote scheduler to execute daily and load source tables in Redshift from new source files.
- Coordinating closely with the reporting team to include newly added KPIs for production dashboards.
- Documenting detailed instructions and inner workings for technical and non-technical stakeholders, including architecture diagrams, DDL, DML, code comments, and mapping sheets.

Client: Sugar Bear Pro

Cloud Infrastructure Employed: GCP Data Engineering

Responsibilities:

- Research and analysis based on business KPIs to be calculated from GA4 and supplementary integrations
- Set up custom queries to calculate data based on business needs from GA4 table and supplementary integrations and schedule queries to populate resultant tables - Importing tables into Looker for reporting.
- Working with google ads, Facebook ads, YouTube ads and other advertisement channels being used for targeted campaigns and stitching with GA4 google analytics data to extract calculated metrics to report on to drive business growth
- Documented detailed instructions and inner workings specifying how each mechanism developed works for technical and non-technical stakeholder understanding which includes architecture diagrams, DDL, DML, comments in code and mapping sheets.

Client: Mattress Firm

Cloud Infrastructure Employed: GCP Data Engineering

Responsibilities:

- Worked on OneStream, feedonomics and product master projects on mattress firm
- Worked on implementing existing utilities on Databricks while optimizing ETL pipelines.
- Assisted team mates by doing runtime research and analysis-based tasks to pin point data discrepancies in SQL and data ware house
- Experience with handling source files deposited on google cloud studio
- The source files on google cloud studio are then ingested via cloud function using triggers to ingest data into big query data warehouse and data is categorized and fragmented as key business identifiers such as sales data, product data, customer data.
- Creating sophisticated pipelines that get data from files and inserting them into staging and main tables using cloud function-based pipelines that adequately handle delta and differential.
- Writing store procedures that get data from different sources and inserting it into big query tables in DWH
- Pushing data generated by pipelines to graph DB (Neo4J) for visual representation via nodes and relationships
- Documented detailed instructions and inner workings specifying how each mechanism developed works for technical and non-technical stakeholder understanding which includes architecture diagrams, DDL, DML, comments in code and mapping sheets.
- Created DBT based workflows on top of Big Query and gradually incorporated DWH stored procedures into DBT models for execution.

Client: Cold Water Creek

Cloud Infrastructure Employed: Microsoft Azure

Responsibilities:

- Experience with handling ingestion and optimization of data via dynamic d365 finance and operations
- Utilization of business specific focal data categorized as sales, product, customer and other source identities in order to integrate business KPIs for data reconciliation.
- Creating end to end ETL pipelines to read source files from landing zone to curated zone in blob storage using notebooks in Microsoft Azure Data Bricks
- Configuring and synchronizing Databricks and Data Lake Storage
- Using PySpark and spark SQL for transformation, and delta handling on required source files
- Implementing stored procedures categorized as Lookups, Dimensions and Facts
- Documented detailed instructions and inner workings specifying how each mechanism developed works for technical and non-technical stakeholder understanding which includes architecture diagrams, DDL, DML, comments in code and mapping sheets.

## **CERTIFICATIONS**

- 
- DP-201 Designing an Azure Data Solution
  - IBM Big Data Foundations - Level 1
  - IBM Big Data Foundations - Level 2
  - IBM Hadoop Administration - Level 1
  - IBM Hadoop Data Access - Level 1
  - IBM Hadoop Foundations - Level 1
  - IBM Hadoop Programming - Level 1
  - IBM Spark - Level 1
  - DBT Fundamentals