

KHUSHI SINGH

Buffalo, NY

716-416-1029

khushisingh.021103@gmail.com

linkedin.com/in/khushisingh03

github.com/ksingh38-git

Summary

Data Professional with 3+ years of experience designing, building, and optimizing **data pipelines** and **ETL/ELT workflows** across cloud platforms. Strong background in **SQL**, **data modeling**, **data warehousing**, and **data quality**, with hands-on experience migrating on-prem systems to the cloud and supporting enterprise analytics. Experienced in **Python-based ingestion pipelines**, cloud-native data systems, and integrating downstream AI and analytics applications.

Experience

University at Buffalo

Feb 2026 – Present

Research Assistant

- Designed and deployed ETL pipelines to ingest and process multi-year hourly time-series datasets (8,000+ records per source annually) across live sensor streams and EPW climate files using Python and Pandas.
- Implemented **schema normalization**, **data validation**, and **transformation workflows** across **10+ time-series columns**, storing curated datasets in **Parquet format** for optimized **batch processing** and **analytical querying**.
- Built **hourly and day-of-year aggregation pipelines** to compute **climate bias and performance metrics** across heterogeneous datasets.

Minedco Holdings, Inc.

May 2025 – Aug 2025

AI Software Engineering Intern

- Engineered a **Python-based ingestion pipeline** to collect and store **500+ posts per user** (posts, media, metadata) in **Azure Blob Storage**.
- Implemented **Facebook OAuth authentication** and **Graph API-based ingestion** to securely retrieve user content for cloud-based processing.
- Implemented **on-demand, user-triggered workflows** that fetch data and curate memories only when users click **Generate Memories**.
- Integrated **LLM services** into a production pipeline using structured **prompt engineering** to generate scrapbook chapters, captions, and **personality summaries** from post text and engagement metadata, enabling a customer-facing feature with automated **PDF export**.

Deloitte US India (USI)

May 2021 – Aug 2024

Analyst – Data Engineering

- Designed, developed, and maintained **ETL data pipelines** using **SQL**, **PL/SQL**, and **BigQuery** to support enterprise-scale financial datasets.
- Built and optimized **data transformation workflows** within **BigQuery data warehouse**, improving pipeline performance from **6 hours to 25 minutes**.
- Executed **on-prem to cloud data warehouse migration**, performing **schema design**, **data modeling**, **data validation**, and **reconciliation** across 85+ analytical assets.
- Developed **analytical views and data models** within the **Enterprise Data Hub (EDH)** to enable downstream reporting and analytics consumption.
- Performed **production pipeline monitoring**, **root cause analysis**, and **data quality checks** to ensure reliability and consistency of enterprise datasets.

Projects

Airline Delay Prediction & Analysis — *Python, SQL*

- Cleaned, engineered, and analyzed **1M+ flight records** across multiple features to create structured datasets for delay analysis.

Transportation Safety Analytics — *Python, SQL*

- Processed **1M+ NYC collision records**, creating datasets to identify high-risk locations and safety patterns.

Skills

Programming: SQL, PL/SQL, Python

Data Engineering: ETL/ELT, Data Pipelines, Data Modeling, Data Warehousing, Data Quality, OLAP Concepts

Databases: Oracle, BigQuery, Snowflake

Cloud & Orchestration: Azure (Blob Storage, Functions), Apache Airflow, Cloud Data Pipelines

Analytics & Tools: Incorta, Excel, Git, Jira, REST APIs, OAuth2, Power BI, Data Validation

Education

University at Buffalo — M.S. Computer Science & Engineering

Aug 2024 – Dec 2025

GPA: 3.63 / 4.00

St. Xavier's College — B.Sc. Information Technology

Jul 2017 – Nov 2020

GPA: 8.06 / 10.00

Certifications & Awards

Oracle Database SQL Certified Associate — Oracle University

Issued May 28, 2024