SURBHI BHOR

BENGALURU, KA | SURBHI.BHOR.IN@GMAIL.COM | linkedin.com/in/surbhi-bhor | github.com/surbhi-bhor | PORTFOLIO

SUMMARY

Data and BI engineer with 3+ years building data pipelines, cloud warehouses, and BI systems across retail, consulting, and enterprise domains on AWS and Azure. Delivered Spark-based ETL, dimensional models, and data quality frameworks that reduced processing time and infrastructure costs while improving reporting reliability, spanning the full pipeline lifecycle from raw ingestion to dashboard delivery. Experienced in leveraging AI-assisted tooling (Cursor, Claude, GitHub Copilot) to accelerate script refactoring, data transformation, and pipeline debugging.

EDUCATION

University of Illinois at Urbana-Champaign – Master of Science in Information Management (GPA: 3.9/4.0) **Aug 2022 – May 2024**

University of Mumbai – B.E. in Electronics Engineering **Aug 2016 – Oct 2020**

EXPERIENCE

Data Engineer | Calimontex Technologies – Austin, TX **Jan 2026 – Aug 2026**

- Refactored enterprise ETL workflows consolidating POS, inventory, and product data into a unified warehouse, shortening pipeline runtime by **30%** via incremental processing.
- Designed a star-schema dimensional data model separating transaction-level sales from snapshot-based inventory, standardizing core metrics like net sales, margin, and stockout rate across reporting.
- Reverse-engineered legacy SQL using agentic coding tools (Cursor) to build data lineage and reconcile conflicting metric definitions into a single source of truth, enabling self-serve metadata lookup.

Business Intelligence Engineer | Amazon Grocery (Whole Foods Market) – Austin, TX **Dec 2024 – Aug 2025**

- Engineered retail analytics infrastructure with Spark SQL processing multi-million-row datasets from S3 and internal data lake sources, cutting ad-hoc reporting cycles by **71%** and bringing dashboard refresh time under **1 minute** for inventory and campaign decisions.
- Delivered Redshift data marts with AWS Glue transformation pipelines, powering QuickSight dashboards for Parcel Returns conversion analysis, Seasonal inventory optimization, and Supplier Marketing ROI tracking that drove **$5.9M** in incremental revenue.
- Optimized Spark transformations with partitioning, broadcast joins, and resource tuning, improving query performance by **25%** and saving **~$20K** in annual AWS compute costs; automated flat-file ingestion via AWS Lambda and S3.
- Applied AI-assisted tools (Claude, Amazon Q) to accelerate SQL optimization, pipeline debugging, and incident triage, reducing root-cause analysis time by **65%** and helping cross-functional teams respond **3x faster**.

Data Engineer | Bright Mind Enrichment and Schooling – Austin, TX **Aug 2024 – Dec 2024**

- Centralized fragmented volunteer outreach and donation records from multi-chapter intake forms and spreadsheets into a single analytics warehouse, lowering monthly reporting effort from **~20h to under 2h.**
- Published dashboards connecting Street Care app interaction logs and volunteer activity to operational KPIs, giving 10+ chapter leads self-serve reporting and visibility into impact across **12,600+** individuals aided.

Data Engineer Intern | Boston Consulting Group (BCG) – New York, NY **Jun 2023 – Aug 2023**

- Built batch processing pipelines in Azure Data Lake, transforming transactional records through staging and curated Delta layers using Medallion Architecture and Slowly Changing Dimension (**SCD**) logic, eliminating **75%** of data discrepancies.
- Modeled data for a **15+** table domain, replacing a manual, serialized ERD review process with shared, real-time design collaboration.
- Developed Change Data Capture (**CDC**)-based ETL workflows in Azure Synapse for full and incremental loads, backed by a PySpark validation framework enforcing record-level quality checks that prevented anomalies from reaching production layers.

Software Development Engineer in Test (SDET) | Capgemini – Mumbai, India **Sep 2020 – Jun 2022**

- Spearheaded serverless validation pipelines using Node.js, AWS Lambda, Step Functions, and API Gateway, building automated middleware that raised code coverage to **70%** across **7 microservices**.
- Programmed scalable JavaScript automation test suites validating backend REST API endpoints on AWS, reducing sprint-over-sprint issue resolution time from **2+ days to under 6 hours**.
- Instrumented pipeline observability via CloudWatch and Power BI dashboards, diagnosing root causes to **halve** Mean Time to Resolution (MTTR) and curb recurring failures.
- Configured a JSON-driven test framework integrated into Azure DevOps CI/CD pipelines with automated email alert triggers, accelerating integration release cycles by **40%** with fewer downstream bug tickets.

TECHNICAL SKILLS

- **Languages and Frameworks:** SQL, Python, Apache Spark (PySpark, Spark SQL), Pandas, JavaScript, Node.js
- **AWS Cloud:** Redshift, Glue, Lambda, S3, QuickSight, Step Functions, API Gateway, CloudWatch, DynamoDB
- **Azure Cloud:** Synapse Analytics, ADLS Gen2, Blob Storage, Data Factory, Delta Lake, Azure DevOps
- **Data and BI:** PostgreSQL, MySQL, dbt, Apache Airflow, Dagster, Tableau, Power BI, Metabase, MS Excel, Git, Postman

PROJECTS

Personal Content Intelligence Pipeline | Dagster, dbt, Python, PostgreSQL, Metabase, Ollama - Repo

- Created a self-hosted ELT platform ingesting data from 4 APIs into PostgreSQL, orchestrated with Dagster and modeled in dbt with 16 models, 80 data tests, incremental facts, idempotent loads, and pipeline guardrails.
- Integrated Flask and Metabase as serving layers, with local Llama 3.2 text-to-SQL for natural-language analytics and personalized recommendation feeds over the warehouse.