Bhavya Bipin Gada

+1 (571) 771 5507 | bgada1@umbc.edu | in/bgada | github.com/bhavyabgada | https://portfolio.bhavyabgada.dev/

Summary

Data Engineer with 2+ years of experience designing, building, and operating secure, high-throughput data pipelines in AWS-based, regulated enterprise environments. Proven track record of owning production systems end-to-end, optimizing CDC ingestion and distributed processing to scale data volumes while meeting strict accuracy, latency, and SLA requirements. Strong background in SQL and Python, cloud-native data platforms, and SDLC-driven delivery.

Skills

- **Data Engineering & Platforms**: Python, SQL, Java, Airflow, Kafka, AWS S3, Glue, EMR, Lambda, Redshift, BigQuery, Databricks, Snowflake
- **Cloud & Infrastructure**: AWS, Google Cloud Platform, Azure, Docker, Kubernetes, Terraform
- **Workflow Orchestration & Data Quality**: Airflow, dbt, Great Expectations
- **Monitoring & Operations**: CloudWatch, Grafana, Prometheus, CI/CD, Incident Response
- **Analytics & Visualization**: Power BI, Tableau, Looker
- **Delivery, SDLC & Enterprise Tools**: ServiceNow, Agile / Scrum, Jira, Production Support
- **Data Security & Compliance**: HIPAA, GDPR, SOC 2, PHIPA, PCI DSS
- **Certifications**: AWS Certified Solutions Architect – Professional, Google Cloud Professional Data Engineer, Google Cloud Professional Cloud Architect, Databricks Certified Professional Data Engineer, Certified Kubernetes Administrator (CKA), Certified Information Privacy Technologist (CIPT – IAPP)

Work Experience

UPS (Contract)   Jul 2024 - Present
Data Engineer

- Owned and operated mission-critical payroll and financial reporting datasets, supporting downstream analytics, audits, and compliance workflows with strict accuracy, timeliness, and SLA guarantees.
- Designed, developed, and maintained scalable CDC-based ingestion pipelines on AWS using S3 as the system of record, Glue and EMR for distributed processing, and Redshift for analytical storage.
- Re-architected CDC ingestion workflows after identifying foundational throughput and contention issues in existing jobs, enabling reliable processing of 10× higher data volumes while remaining within SLA.
- Introduced multi-threaded and parallelized ingestion patterns within Glue and EMR jobs, reducing end-to-end data delivery latency and improving pipeline predictability under peak load.
- Optimized CDC handling logic to correctly process late-arriving changes, high-frequency updates, and replay scenarios, reducing reprocessing overhead and data inconsistency in downstream datasets.
- Implemented schema evolution controls and versioned data contracts to safely accommodate upstream source changes without breaking downstream analytics.
- Tuned Redshift workloads through distribution styles, sort keys, and incremental load strategies to improve query performance and stabilize reporting during high-concurrency usage.
- Built incremental transformation pipelines using SQL and Python, balancing batch window sizes and resource allocation to meet delivery SLAs while controlling compute cost.
- Implemented automated data quality validations, freshness SLIs, and reconciliation checks at multiple pipeline stages to prevent silent data corruption and downstream reporting defects.
- Integrated CloudWatch metrics, logs, and alarms to monitor pipeline health, execution latency, failure rates, and data freshness, enabling rapid detection of production issues.
- Served as primary on-call engineer for data pipelines, performing deep root-cause analysis on production failures and implementing durable fixes to prevent recurrence.
- Partnered closely with upstream source-system teams and downstream analytics consumers to investigate discrepancies, validate assumptions, and ensure consistent end-to-end data delivery.
- Supported CI/CD-driven pipeline deployments using infrastructure-as-code and SDLC best practices, ensuring safe promotion of changes across development, UAT, and production environments.

Ardent Privacy   Jul 2023 - Jun 2024
Privacy Data Engineer

- Designed and operated compliance-centric data pipelines powering enterprise privacy operations (DSAR, consent, audit, retention) across healthcare, finance, and government clients.
- Built event-driven ingestion pipelines using Kafka and AWS-native messaging patterns to process high-volume privacy events with exactly-once semantics and end-to-end auditability.
- Developed and deployed containerized data services using Docker and Kubernetes, enabling isolated, repeatable processing of regulated datasets across environments.
- Implemented policy-aware ETL workflows using Python and SQL on Glue- and EMR-style distributed processing frameworks, enforcing purpose limitation, minimization, and lawful-basis controls at execution time.
- Embedded privacy controls directly into pipelines, including data classification, masking, tokenization, and pseudonymization aligned with GDPR Articles 5–35 and HIPAA Security Rule requirements.

- Designed multi-tenant data models with tenant isolation, region-based partitioning, and residency enforcement to support cross-border compliance across the U.S., EU, and India.
- Integrated automated DSAR workflows (identity verification → data discovery → extraction/deletion → audit logging), consistently meeting sub-72-hour SLA commitments.
- Implemented retention and deletion enforcement using policy-driven jobs and lifecycle rules, ensuring defensible data disposal across structured and semi-structured datasets.
- Built audit-ready metadata pipelines to generate ROPA, DPIA, SOC 2, and ISO 27701 evidence directly from operational data and system logs.
- Integrated LLM-assisted data discovery and classification workflows using controlled prompts and policy-tagged context to accelerate DSAR processing while preventing sensitive data leakage.
- Applied token-level filtering, clean-room embeddings, and strict prompt isolation to safely enable AI-assisted summarization and metadata extraction over regulated datasets.
- Supported fine-tuning and evaluation of NLP models on privacy-labeled data under strict training/inference separation and governance controls.
- Monitored production pipelines using CloudWatch-style metrics and alerts, proactively identifying SLA violations, backlog growth, and policy enforcement gaps.
- Performed root-cause analysis on compliance- and AI-impacting incidents, implementing durable fixes to prevent recurrence and sustain zero audit exceptions.
- Collaborated closely with legal, privacy, security, and engineering stakeholders to translate regulatory and AI-governance requirements into production-grade technical controls.

University of Maryland, Baltimore County   Sep 2022 - Jun 2023
Software Developer & Graduate Assistant
- Engineered and maintained production web systems for university-wide platforms, supporting academic and administrative workflows used by thousands of users.
- Led accessibility remediation initiatives to achieve WCAG 2.1 and WCAG 2.2 compliance, aligning with Section 508 and ADA requirements.
- Conducted detailed accessibility audits using axe-core, Pa11y, Monsido, and manual screen-reader testing (NVDA, VoiceOver).
- Remediated complex accessibility defects involving ARIA roles, focus management, keyboard navigation, color contrast, and semantic HTML structure.
- Implemented automated accessibility testing pipelines integrated into CI workflows to detect regressions before release.
- Partnered with IT, compliance, and content teams to validate accessibility fixes and ensure standards adherence across multiple releases.
- Developed backend integrations and data-backed services using Python and SQL, supporting reporting and operational use cases.
- Authored documentation and runbooks detailing accessibility standards, testing procedures, and remediation patterns for long-term maintainability.
- Mentored graduate students on applied data and machine learning concepts, emphasizing data preparation, evaluation rigor, and reproducibility.

Virtuals Design   Apr 2020 - Aug 2022
Full Stack Software Developer
- Designed and developed backend services for multi-tenant SaaS platforms supporting thousands of daily users and data-driven business workflows.
- Built and maintained Java-based backend components and services for transactional processing, data validation, and integration with downstream systems.
- Developed RESTful APIs using Java, Python, and SQL to support data ingestion, transformation, and reporting use cases.
- Implemented asynchronous processing patterns using message queues and worker-based execution to improve throughput and system resilience.
- Designed relational data models and optimized SQL queries through indexing, query plan analysis, and transaction boundary tuning.
- Implemented batch and incremental processing logic for reporting and reconciliation workflows, ensuring data consistency under concurrent access.
- Applied structured logging and exception handling to diagnose production issues involving concurrency, data integrity, and integration failures.
- Supported CI/CD pipelines with automated testing and controlled deployments, reducing production risk during releases.
- Collaborated with cross-functional teams to debug complex backend issues spanning application logic, database behavior, and external integrations.

Education

University of Maryland Baltimore County
MS inInformation Systems   *Maryland, USA*