BHAVYA BIPIN GADA

bhavyagada.dev@gmail.com | (571) 771 5507

PROFESSIONAL SUMMARY

Senior Agentic Data Engineer with 5+ years of experience designing large-scale, cloud-native AI and data platforms across Google Cloud Platform (GCP), Microsoft Azure, and Amazon Web Services (AWS**)**. Combines deep technical expertise with a customer-obsessed, results-driven, and innovation-focused mindset to build trusted, compliant, and high-performance AI ecosystems for enterprise and government clients. Recognized as a two-time National Hackathon Winner and **Google** certified **Professional Data Engineer** and **Professional Cloud Architect, Databricks Certified Professional Data Engineer**, **AWS Solutions Architect Professional**, **Certified Kubernetes Administrator (CKA)** and **IAPP Certified Information Privacy Technologist (CIPT)**.

- Architected a multi-cloud data platform processing **1+ billion records monthly**, achieving **99.95% uptime**, **45% faster analytics**, and full **SOC 2 / HIPAA** compliance while reducing operational cost by **45%**.
- Re-engineered **20+ legacy ETL pipelines** into a modular **Dataflow** framework, cutting maintenance effort **70%** and enabling rapid feature releases from weeks to hours.
- Mentored engineering teams on **Data Mesh** and **Privacy-by-Design** principles; led cross-functional governance forums aligning data lineage, observability, and performance SLOs across product, SRE, and analytics domains.
- Delivered a **Retrieval-Augmented Generation (RAG)** and **LangChain + LangGraph** multi-agent AI system within six weeks, improving **Large Language Model (LLM)** recall accuracy **28%** and accelerating compliant enterprise AI adoption.
- Implemented **federated-learning** and **differential-privacy** pipelines with **zero audit exceptions** across global regions; continuously expands expertise in **Responsible AI**, **MLOps (Kubeflow)**, and **vector databases**, authoring enterprise documentations adopted across teams.

TECHNICAL SKILLS

Programming & Frameworks	**Python, SQL**, Java, JavaScript (Node.js / React), PHP (Laravel), Shell, YAML, FastAPI, Flask, Express.js, React + Redux Toolkit, Semantic Kernel, GraphQL, gRPC, WebSockets
Cloud & Infrastructure	• GCP: Vertex AI, **Kubeflow,** Cloud Run, Cloud Storage, GKE, IAM • Azure: **Databricks**, Machine Learning, Purview, Event Hubs, Microsoft Fabric, AI Search • AWS: S3, EC2, Lambda, Glue, **SageMaker,** Redshift, CodeDeploy, CloudWatch • **Datadog**
DevOps / MLOps / SRE	**Docker**, **Kubernetes** (CKA), **CI/CD, Terraform** + Sentinel, Helm, GitHub Actions, Cloud Build, Jenkins, Ansible, Kubeflow Pipelines, MLflow, DVC, SonarQube, Prometheus, Grafana, OpenTelemetry, Error-Budget Dashboards
Data Engineering & Storage	**Airflow / Composer**, Kafka, **Pub/Sub, Dataform,** dbt, Datastream CDC, Great Expectations**, BigQuery**, PostgreSQL, MySQL, Delta Lake, Parquet, Avro, Redis, CockroachDB, Neo4j, Pinecone (Vector DB), FAISS
Artificial Intelligence & ML	TensorFlow, PyTorch, scikit-learn, XGBoost, Prophet, LangChain, LangGraph, LangSmith, Retrieval-Augmented Generation (RAG), Large Language Models (LLMs), Prompt Engineering, Federated Learning (FedAvg), Differential Privacy (DP-SGD), Homomorphic Encryption (CKKS), Secure Multi-Party Computation (SMPC), Semantic Kernel, LLM Fine-Tuning, Prompt Engineering, MCP (Model Context Protocol)
Security & Compliance	HIPAA, HITRUST, SOC 2, ISO 27701, GDPR, PHIPA, EU-AI, NIST AI RMF, GCP DLP, Azure Purview, AWS Macie, spaCy PII Detection, Tokenization & Masking, VPC-SC, CMEK / HSM Encryption, OAuth 2.0, JWT, RBAC, OIDC, Zero-Trust Workload Identities
Analytics & Visualization	Looker, Power BI, Tableau, Grafana, Streamlit, Chart.js, **WCAG 2.2** & Section 508 Accessibility Auditing
Collaboration & Delivery	**Agile** / DataOps / DevSecOps, Sprint Planning, UAT, RCA, Confluence, Jira, Notion, Slack, **ServiceNow**, Miro, Technical Documentation, BRD / FRD Authoring, Runbooks, Mentorship

Education Degree: Master of Science in Information Systems
University Name: University of Maryland Baltimore County

PROFESSIONAL EXPERIENCE

myRocky.ca | Baltimore, MD July 24 – Present

Agentic Data Engineer – AI Privacy (PHI)

- Architected a **cloud-agnostic**, GCP-native healthcare data platform (BigQuery, Dataflow, Composer, Vertex AI, Kubeflow) provisioned with Terraform for cross-cloud portability, enabling unified analytics on **regulated PHI** and achieving 99.95 % SLA reliability and full HIPAA compliance.
- Designed dual-layer ingestion (on-prem EHR → GCP) using Datastream, Kafka, and Pub/Sub, ensuring exactly-once delivery and schema-evolution safety across regions.
- Implemented bitemporal data modeling for prescriptions and encounters, enabling auditable time-travel analytics and reproducible regulatory backfills.
- Optimized **Composer DAGs with Pub/Sub** event triggers, reducing pipeline latency 35 % and enabling fault-tolerant streaming to ingest from on-prem Oracle and EHR systems.
- Integrated Cloud Storage lifecycle policies and **CMEK encryption**, ensuring secure retention of healthcare data under HIPAA and PHIPA standards.
- Built hybrid Dataflow + Kubeflow Pipelines orchestrated through Composer, reducing incident MTTR 60 % via automated checkpoint recovery and rollback.
- Developed **LangChain + LangGraph** pipelines for doctor–patient dialogue summarization, improving LLM recall by 28 % while preserving PHI isolation through clean-room data embeddings.
- Deployed federated-learning workflows with TensorFlow Federated + Kubeflow, applying FedAvg and Differential Privacy (DP-SGD) to eliminate raw-data sharing and maintain < 2 % model drift.
- Integrated Dataform, dbt, and Great Expectations to automate validation and freshness SLIs, sustaining 99.95 % data accuracy across all pipelines.
- Automated compliance-as-code using Terraform Sentinel + Cloud Build, generating audit evidence per deploy and removing 100 % manual SOC 2 / HIPAA checks.
- Led FinOps and Observability initiatives via BigQuery**, Looker,** and Grafana, tracking lineage and cost to keep quarterly spend variance < 3 %.
- Developed and optimized multiple Looker dashboards using Materialized Views and Derived Tables for performance efficiency, while mentoring analysts on LookML best practices, version control, and coding standards to ensure scalable, maintainable analytics delivery.
- Partnered with Privacy, SRE, and Analytics teams to embed Privacy-by-Design governance, raising platform reliability from 98.5 % → 99.95 % and improving cross-team transparency.

Ardent Privacy | Baltimore County, MD Dec 2021 – Apr 2023

Privacy Data Engineer

- Engineered secure, event-driven data and compliance pipelines powering Ardent Privacy's PrivacyOps platform, supporting global healthcare, finance, and defense clients while maintaining 99.9 % uptime and zero audit exceptions.
- Designed and deployed **multi-tenant architectures** in CockroachDB and BigQuery, enforcing regional partitioning and data-residency rules to ensure cross-border GDPR and HIPAA compliance across the U.S., EU, and India.
- Implemented Kafka + Pub/Sub ingestion **for DSAR, consent, and audit** events, achieving sub-second policy propagation and exactly-once delivery semantics across distributed microservices.
- Built modular privacy microservices in Node.js and Python, containerized with Docker and deployed to GKE via Helm, integrating automated provisioning and **guardrails** through Terraform + Sentinel for **reproducible IaC**.
- Architected policy-tagged ETL pipelines in Airflow and Dataflow, embedding data classification, masking, and tokenization directly into DAG execution layers to automate access control and risk mitigation.
- Developed **custom PII/PHI discovery** APIs by combining GCP DLP with spaCy NLP, improving detection precision to 95 % + and automating risk-tier labeling at ingestion.
- Deployed federated-learning workflows (TensorFlow Federated) with **differential-privacy noise injection**, enabling distributed analytics without raw-data exposure while keeping model drift < 2 %.
- Integrated model-lineage, dataset-versioning, and bias-detection metrics into privacy pipelines, delivering audit-ready traceability across ingestion, fine-tuning, and inference.
- **Automated DSAR request pipelines** (verification → extraction → deletion → logging), cutting manual processing 65 % and achieving < 72 h SLA compliance across tenants.
- Codified AI-governance and compliance pre-checks in Terraform Sentinel and Airflow hooks, automating policy validation for datasets and ML deployment workflows.
- Implemented retention, consent, and pseudonymization logic directly within pipelines via policy-based validation jobs, ensuring full alignment with GDPR Articles 5–35 and the HIPAA Security Rule.
- Generated RoPA, DPIA, SOC 2, and ISO 27701 audit evidence packs from operational metadata, sustaining zero non-conformities across independent assessments.

- Designed Prometheus + Grafana dashboards for pipeline freshness, DSAR throughput, and cost SLOs, integrating real-time alerts into Slack and PagerDuty for proactive SLA governance.
- Tuned BigQuery with partitioning, clustering, and materialized views, reducing query-scan costs ≈ 35 % and latency ≈ 40 %.
- Partnered with privacy, risk, and ML-governance teams to align controls with the EU AI Act, GDPR Recital 71, and NIST AI RMF, enabling regulatory traceability for AI-driven systems.
- Authored Privacy-by-Design engineering playbooks and incident-response runbooks, institutionalizing privacy engineering practices across product and DevOps teams.
- Delivered measurable impact: 100 % DSAR SLA compliance, zero audit exceptions, ≈ 60 % reduction in manual compliance cost, and ≈ 25 % infrastructure cost savings.

University of Maryland – Baltimore, County | MD Apr 2020 – Nov 2021

Software Engineer & Graduate Assistant

- Engineered accessibility and SEO upgrades for **UMBC.edu** and affiliated academic sites, achieving **WCAG 2.2 / Section 508 compliance** and improving Lighthouse Accessibility Scores from **78 → 97**.
- Led accessibility audits with **axe-core, Pa11y, Monsido**, and manual screen-reader testing (NVDA / VoiceOver), remediating hundreds of ARIA, color-contrast, and keyboard-navigation issues.
- Automated HTML / CSS / JS compliance checks via custom **PHP utilities** and scheduled scan reports, reducing manual QA overhead **40 %**.
- Partnered with IT and content teams using **JIRA** to triage and verify fixes across multiple web properties, ensuring consistent accessibility before release.
- Mentored graduate students on CNN, LSTM, and Transformer models; authored faculty-adopted guides on optimization and evaluation metrics (F1, RMSE, sMAPE).

Virtuals Design Labs – Mumbai, India Apr 2020 – Nov 2021

Full Stack Software Engineer (Multiple Clients)

- Architected multi-tenant SaaS on **Laravel + React** with layered Repository/Service/Controller design, RBAC, per-tenant schema isolation, and config-driven feature flags; shipped admin portals, reporting, and partner APIs used by 40 K + monthly users.
- Built SPA dashboards with Redux Toolkit, form engines, route guards, and accessibility standards; implemented code-splitting, lazy routes, and granular memoization to cut TTI and improve Lighthouse performance by ≈ 35 %.
- Implemented **WebSockets** (Laravel Echo + Socket.IO / Pusher) for live notifications, queue status, chat/helpdesk, and operational dashboards; added SSE fallbacks for restricted networks to maintain real-time responsiveness.
- **Exposed REST APIs** (OpenAPI / Swagger) for partners and internal gRPC endpoints using Swoole async workers for low-latency jobs (pricing, reconciliation, document parsing), reducing cross-service call latency by ≈ 40 %.
- Tuned **PostgreSQL with composite and partial indexes**, optimized Eloquent via eager loading and query caching, and used Redis for caching, queues, and rate-limits; sustained 5 K + requests / day at low p95 latency.
- Built an API-key gateway with usage metering, per-plan throttles, idempotency keys, and soft quotas; delivered **usage-based billing** (Stripe / Razorpay webhooks + reconciliation jobs) that grew partner API revenue from $25 K → $150 K.
- Standardized background jobs (emails, PDF rendering, ETL) via Redis queues, delayed jobs, and DLQ retry rules; added idempotent handlers with request hashing to prevent duplicate effects.
- **Enforced TDD** with PHPUnit / Pest (unit + feature), Postman / Newman contract suites, and k6 / JMeter load profiles; wired PHPStan (level 7–8) and ESLint pre-commit hooks; maintained high-value coverage on authentication, payments, and exports.
- Containerized services with **Docker**; implemented CI/CD using GitHub Actions → AWS CodeDeploy (blue / green) with smoke tests, schema migrations, and feature-flag rollouts; achieved mean rollback < 5 minutes.
- Centralized logs and traces via CloudWatch + Sentry with correlation IDs and structured JSON logs; built latency / error dashboards and alerting (p95 SLIs, error budgets) to detect regressions before incidents.
- Applied **OWASP controls** (CSP headers, strict cookies, CSRF, input validation), JWT / OAuth 2.0 for partner access, environment-scoped secrets management, rate-limiting, and HMAC-signed webhooks for financial events.
- Integrated Stripe / Razorpay payments, Twilio OTP / voice, Google Maps geocoding, Gmail / SES SMTP, and AWS S3 storage with presigned URLs and lifecycle rules; added PDF pipelines for invoices and certificates.
- Built an internal React component library (tables, forms, charts) with Storybook-style documentation; standardized Zod validation and reusable error boundaries; improved developer velocity and reduced UI regressions.
- **Authored runbooks** for release steps, incident response, and hot-fix flows; scheduled cron / queue health checks and backup verification jobs; implemented maintenance windows with user messaging.
- Served as the primary **cross-stack debugger for CORS**, session, and webhook sequencing issues across Laravel, React, and vendor services; reduced production incident resolution time by > 50 %.