Bria Yasir

Data Solutions Architect | Cloud-Native Platforms | Real-Time Data Engineering

briayasir@gmail.com | (215) 650-3625 | Dayton, OH 45401, USA | linkedin.com/in/bria-bushra-yasir-364686398/

briayasir.github.io/

Summary

Experienced and results-oriented Data Solutions Architect with 10+ years of expertise in designing and implementing scalable, cloud-native data platforms across healthcare, finance, and enterprise industries. Adept at architecting high-performance lakehouse architectures, real-time data pipelines, and advanced ETL/ELT workflows utilizing technologies such as Databricks, Palantir Foundry, Snowflake, Apache Kafka, and Apache Spark. Proven track record in leading data modernization initiatives, integrating multi-cloud environments (AWS, Azure, GCP), and delivering data-driven strategies that enhance operational intelligence and business outcomes. Strong background in data governance, security, and compliance (HIPAA, GDPR, SOC 2), focused on building reliable and actionable data products. Demonstrates expertise in leveraging real-time analytics, machine learning, and data observability while providing technical leadership, optimizing cloud architectures, and mentoring teams in DataOps and modern data engineering practices.

Technical Skills

Programming & Scripting
Python (ETL automation, PySpark, data validation, APIs), SQL (advanced query optimization, CTEs, window functions), Scala, Java, Bash/Shell scripting, JSON, YAML, Node.js (Backend development, REST APIs, data services for ETL/ELT integration)

Data Engineering & Processing
Apache Spark, PySpark, Databricks, Apache Flink, Apache Kafka, Spark Streaming, Structured Streaming, Apache Beam, Apache Airflow, Luigi, AWS Glue, Azure Data Factory, Google Dataflow, Palantir Foundry

ETL / ELT Development
dbt, Talend, Apache NiFi, Informatica, SSIS (legacy package handling, migration planning to modern frameworks), Matillion, DataStage, Azure Synapse Pipelines, AWS Glue Jobs, Great Expectations (data validation)

Data Architecture & Modeling
Lakehouse Architecture, Data Vault, Kimball & Inmon methodologies, Star & Snowflake schemas, Domain-Driven Design, Bronze–Silver–Gold architecture, Dimensional Modeling, Semantic Layer Design

Data Warehousing & Storage
Snowflake, Databricks Delta Lake, Amazon Redshift, Google BigQuery, Azure Synapse Analytics, Apache Hive, Presto, Trino, Parquet, ORC, Avro, Apache Iceberg, Palantir Ontology & Data Products

Real-Time & Batch Data Pipelines
Apache Kafka, AWS Kinesis, Azure Event Hubs, Apache Flink, Spark Streaming, Google Pub/Sub, AWS Lambda, Debezium (CDC), StreamSets

Databases (SQL & NoSQL)
PostgreSQL, Oracle, SQL Server, MySQL, MongoDB, Cassandra, DynamoDB, Neo4j (Graph), HBase, Elasticsearch, InfluxDB (time-series)

Cloud Platforms & Services
Microsoft Azure (Data Factory, Synapse Analytics, Databricks, Blob Storage, Event Hubs, Logic Apps, Key Vault)
Amazon Web Services (S3, Glue, EMR, Lambda, Redshift, Kinesis, Athena, CloudFormation)
Google Cloud Platform (BigQuery, Dataproc, Dataflow, Pub/Sub, Cloud Composer, Cloud Functions)

Data Governance, Quality & Security
Apache Atlas, Alation, Collibra, Great Expectations, Data Lineage, Data Cataloging, RBAC/ABAC, PII/PHI masking, encryption, audit logging, Palantir governance workflows, HIPAA, GDPR, SOC 2

Automation, DevOps & CI/CD
Terraform, Pulumi, Ansible, Jenkins, Azure DevOps, GitHub Actions, GitLab CI/CD, Docker (containerization for ETL pipelines and modernization of SSIS workflows), Kubernetes (AKS, EKS, GKE), Infrastructure as Code (IaC), Observability & Monitoring (Prometheus, Grafana)

Analytics & Machine Learning Enablement
BI-ready data marts, feature store design, MLflow, model deployment pipelines, Scikit-learn, TensorFlow, Power BI, Tableau, Looker, Qlik Sense, AI/ML integration with Databricks

Healthcare Data Engineering Expertise
FHIR, HL7, EHR/EMR integration, HIPAA-compliant data pipelines, clinical & claims data processing, patient outcome analytics, predictive healthcare modeling, real-time IoT data processing (sensors & devices), risk stratification pipelines

Collaboration & Leadership
Agile & Scrum methodologies, cross-functional stakeholder management, solution architecture design, technical mentorship, cloud migration strategy, DataOps implementation, documentation & knowledge sharing

Professional Experience

Data Solution Architect 07/2022 – Present
Contour Software
- Spearheaded the design and architecture of real-time healthcare data platforms, leveraging HL7/FHIR, IoT telemetry, EHR systems, and Palantir Foundry to deliver scalable, governed, and compliant data solutions for clinical environments.
- Architected high-performance, low-latency streaming data pipelines using Kafka, Flink, and Palantir Foundry data workflows, optimizing the processing of millions of patient records with real-time insights.
- Implemented and optimized Databricks-based lakehouse solutions (Delta Lake, Spark ML) integrated with Palantir Foundry, enabling predictive analytics, AI-driven intelligence, and patient risk prediction capabilities.
- Directed the migration of legacy data systems to Azure Synapse, Palantir Foundry, and cloud-native architectures, increasing system processing efficiency by 40% while reducing operational costs.
- Developed and instituted enterprise-wide data governance frameworks using Collibra, Monte Carlo, and Palantir governance capabilities, ensuring data quality, lineage, compliance, and full platform visibility.
- Collaborated with security teams to integrate HIPAA, SOC 2, and GDPR controls into platform architecture by implementing encryption, tokenization, role-based access controls, and secure data governance standards.
- Led cross-functional teams in the end-to-end execution of data migration and modernization strategies, ensuring minimal downtime and seamless transition of business-critical datasets into modern data ecosystems.
- Guided engineering teams in adopting DataOps practices, CI/CD automation, Palantir operational workflows, and cloud-native technologies, improving delivery speed, scalability, and deployment reliability.
- Fostered innovation through the adoption of modern data architecture frameworks, Palantir-enabled data products, and cloud-native solutions to enhance platform flexibility and long-term growth.
- Mentored junior engineers and provided technical leadership across data architecture, cloud modernization, streaming platforms, and enterprise data engineering best practices.

Senior Data Engineer 11/2018 – 06/2022
VentureDive
- Engineered and optimized multi-cloud data platforms on AWS, Azure, and GCP, ensuring secure, scalable, and high-performance solutions across diverse industries, including healthcare and finance.
- Developed and automated high-throughput streaming pipelines with Kafka, Flink, and Kinesis, enabling the real-time processing of billions of events, ensuring data accuracy and minimal latency.
- Created and maintained modular ETL/ELT pipelines using dbt and Airflow, driving efficiency and reducing infrastructure costs by optimizing data workflows and pipeline management.
- Built advanced analytics solutions by integrating Snowflake and Looker, delivering real-time dashboards that provided insights to support business operations and executive decisions.
- Collaborated with stakeholders and cross-functional teams to develop and implement scalable pipeline designs that aligned with evolving business needs and market trends.
- Implemented data governance and quality controls, leveraging metadata management, lineage tracking, and automated validation frameworks using Great Expectations.
- Optimized data workflows through partitioning, workload tuning, and distributed processing, significantly improving system reliability and reducing operational costs.
- Led the integration of secure data sharing mechanisms, implementing encryption, access control, and compliance features across all cloud environments.

Big Data Engineer 03/2016 – 10/2018
NorthBay Solutions
- Designed and implemented a unified cloud-native lakehouse architecture using S3 and Delta Lake, consolidating multi-source data into a single analytics platform for efficient business insights.
- Developed and optimized scalable ETL pipelines with NiFi, Talend, and AWS Glue, enabling seamless data ingestion from diverse sources such as APIs, databases, and SaaS applications.
- Built and fine-tuned predictive models using Apache Spark, enhancing forecasting accuracy and enabling smarter, data-driven operational decisions.

- Established proactive data observability frameworks with ELK and Prometheus, improving real-time monitoring and preventing potential SLA breaches.
- Implemented advanced storage optimization techniques, including partitioning and lifecycle management, to improve system performance while reducing operational costs by 20%.
- Collaborated with cross-functional teams to deliver optimized, analytics-ready datasets that powered reporting, machine learning, and business intelligence use cases.
- Mentored junior engineers, fostering a culture of continuous learning, best practices in cloud data engineering, and CI/CD automation.

Data Engineer 01/2014 – 02/2016

CodeNinja

- Designed and managed an enterprise data warehouse solution with SQL Server, supporting critical financial reporting and operational analytics.
- Modernized legacy ETL processes by transitioning from SSIS to Python-based pipelines, improving scalability and processing speed by 35%.
- Developed and automated data validation frameworks using Python, ensuring consistency and reliability across multiple data sources, improving data accuracy.
- Built real-time, interactive dashboards using Tableau, providing actionable business insights and KPI monitoring.
- Worked closely with stakeholders to understand business requirements, translating them into scalable data models and ETL solutions for improved decision-making.
- Optimized SQL queries and database performance, enhancing reporting speed and ensuring high-quality performance metrics for reporting workflows.

Projects

Real-Time Healthcare Intelligence Platform

Tech Stack: Databricks, Apache Spark, Kafka, Delta Lake, Azure Data Lake, MLflow

- Engineered a real-time data platform to process HL7/FHIR, IoT, and clinical data streams, enabling near real-time patient monitoring and insights.
- Built scalable streaming pipelines using Kafka and Spark Structured Streaming, reducing event-to-insight latency significantly.
- Designed a lakehouse architecture (Delta Lake) to unify batch and streaming data for BI and ML workloads.
- Enabled predictive analytics use cases by delivering ML-ready datasets and integrating model tracking with MLflow.

Enterprise Multi-Cloud Data Modernization Program

Tech Stack: Palantir Foundry, Apache Spark, Databricks, Azure Data Lake, Kafka, Python

- Designed and implemented enterprise-scale data workflows using Palantir Foundry to streamline ingestion, transformation, and governed delivery of business-critical data products.
- Built and optimized scalable batch and real-time pipelines leveraging Spark, Kafka, and Databricks, enabling low-latency analytics and operational intelligence across distributed data environments.
- Established end-to-end data governance, lineage, and observability capabilities to improve data quality, accelerate decision-making, and support secure enterprise analytics initiatives.

Streaming Data & Fraud Detection Analytics System

Tech Stack: Apache Kafka, Apache Flink, Snowflake, Python, AWS Lambda, S3

- Designed and deployed a high-throughput streaming architecture to process transactional data for fraud detection and risk analysis.
- Built real-time event processing pipelines using Kafka and Flink, enabling instant anomaly detection and alerting.
- Integrated Snowflake for low-latency analytics and reporting, supporting business-critical decision-making.
- Developed Python-based validation and monitoring frameworks to ensure data quality, reliability, and system observability.

Certificates

- Certified Data Engineer – Google Cloud Professional
- Databricks Certified Data Engineer Professional
- AWS Certified Solutions Architect
- Certified Data Management Professional (CDMP)

Education

Punjab University

Bachelor of Science in Computer Science