

Mahmoud K. Alsaffar

Senior Lead Data Engineer & Cloud Data Architect | AI/ML-Enabled Data Platforms | AWS

• Azure • GCP • Databricks • Snowflake

DOD & IRS + MBI Clearance Active

Email:

Phone:

With 12+ years of experience, I architect and scale secure, cloud-native data platforms that power enterprise transformation. I specialize in building Enterprise ecosystems across AWS, Azure, and GCP, and have led modernization Lakehouse architecture, Batch / Real-time Data Pipelines, and Data lakes across highly regulated industries.

I've delivered over 100 enterprise-grade pipelines and services using tools like Apache Spark, Databricks, Snowflake, Talend, and AWS Glue enabling high-throughput analytics, regulatory compliance, at scale. I design and operationalize end-to-end ETL/ELT/CDC/CDF pipelines and ML workflows, with strong proficiency in Python, SQL, Scala, and Java.

My domain expertise includes:

- **Modern Data Architecture:** Relational Databases (RDBMS), No Relational Databases (NoSQL), Data Warehousing, Big Data Architecture, Data Integration, Data Governance Lakehouse, Data lakes, Data mesh, Data Storage, Data Fabric and Data Virtualization
- **AI/Analytics integration:** Predictive Analytics, Natural Language Processing (NLP), ML, Anomaly Detection and Real Time Analytics
- **Governance & security:** Unity Catalog, IAM, Data Encryption, Data lineage, Data Masking, FedRAMP.
- **Streaming & Big Data:** Kafka, Spark Streaming, Hadoop, HDFS, Hive, Pig, Sqoop.
- **DevOps/DataOps:** GitHub Actions, Terraform, Kubernetes, CI/CD, Databricks CLI
- **Pre-sales & platform vision:** RFP leadership, stakeholder alignment, architecture reviews and SOW's.

Also, I am a strategic Data Architecture, I define Current-State, Transition-State and Target / Future-State Architecture, align data capabilities to business outcomes & requirements, and drive modernization across hybrid and multi-cloud ecosystems. I coordinate with C-level, engineering, and stakeholders to scale secure, governed solutions that accelerate time-to-insight and reduce operational risk.

- Conducted comparative evaluations of cloud tools (Databricks, Snowflake), guiding platform selection based on security, and scalability.
- Crafted executive-level technical narratives and roadmaps for cross-functional teams and non-technical stakeholders.
- I've contributed to enterprise data governance initiatives, driving adoption of domain-based data models, formalizing metadata and lineage policies, and aligning platform standards across programs.
- Drove platform innovation by developing reusable architecture blueprints and introducing governance workflows for secure data access.
- Automated operational workflows using CI/CD pipelines, Terraform, and Datadog, with a focus on resiliency, observability, and Ops maturity.
- Recognized as a thought leader and mentor, I fostered high-performing teams, lead architecture workshops, and shaped future-ready data platforms that are intelligent, secure, and business-aligned delivering measurable business outcomes.
- Defined module objectives, provided structured feedback, and supported junior engineers' growth through tailored development plans and QA reviews.

Skills Summary:

- **Cloud Platforms:** AWS (EC2, S3, Lambda, Redshift, DynamoDB, RDS, Glue, Kinesis, Route53, CloudWatch, Athena, IAM, KMS, VPC). GCP (BigQuery, Dataflow, Cloud Storage, Pub/Sub, Cloud Composer, Audit Logs).

Mahmoud K. Alsaffar

Senior Lead Data Engineer & Cloud Data Architect | AI/ML-Enabled Data Platforms | AWS

• Azure • GCP • Databricks • Snowflake

DOD & IRS + MBI Clearance Active

Email: [REDACTED] | [REDACTED]

Phone: [REDACTED]

-
- **Cloud Security & Governance Tools:** FedRAMP-compliant architecture, IAM policies, encryption at rest/in-transit, tokenization, data masking, audit logging, role-based access control
 - **Big Data & ETL Tools Data Integration:** Databricks Auto Loader, Apache Spark, Spark Streaming, Kafka, Flink, Hadoop (HDFS, Hive, HiveMQ, Sqoop), AWS Glue, AWS Lambda, Snowflake, Informatica Cloud Services, Talend, SAS Viya, Apache Storm, Apache Flume, Apache Oozie, RESTful APIs, JSON Schema, SQL Server Integration Services (SSIS), MySQL, PostgreSQL, NoSQL (Cassandra, HBase), Datadog, Elasticsearch, Microsoft SQL Server, Parquet, Avro, JSON, XML.
 - **Languages:** Python, Scala, SQL, Java
 - **DevOps & Automation:** Terraform, CloudFormation, GitHub Actions, GitLab CI/CD, Databricks CLI, Docker, Kubernetes
 - **Modeling & Analytics:** Star/Snowflake Schemas, Dimensional Modeling, Tableau, Power BI, Microsoft Fabric
 - **Project Management & Agile Delivery:** Agile methodologies, sprint planning, business requirements analysis, solution documentation, stakeholder collaboration
 - **Industry Experience:** US Federal, Health & Life Insurance, Financial Services & Banking, Telecommunications, Energy, Retail & Restaurant Industries

Professional Experience:

Basecamp Consulting & Solutions

Lead Technical Architecture – Cryptocurrency Compliance | Remote / Washington DC | (06/2024 – 05/2025)

Client: IRS

Key Responsibilities & Contributions:

- Led the architecture for the Digital Asset Cryptocurrency Compliance 1099-DA system, aligning it with IRS regulatory requirements / Use Cases & User Stories, Data Strategies, and cloud security architecture across AWS EDP Platform.
- Facilitated workshops and technical sessions to define the current state, show transition needs, and lay out the target architecture for cryptocurrency compliance, incorporating security and compliance risk mitigation strategies which include strategy solution visions, Conceptual Target solution and roadmap, finally deliver the Enterprise Target Solutions with Target Solution Vision Architecture Blueprint.
- Worked on different technical designs related to compliance, de-anonymization, and filing processes, ensuring seamless integration with existing Criminal Investigation systems and adherence to FedRAMP frameworks.
- Developed system flow diagrams, system descriptions, and information flows for the target state, emphasizing data validation, encryption, and secure access control policies.
- Worked on compliance validation, TIN deanonymization, and filing processes with security in mind, ensuring the implementation of IAM controls, data encryption at rest/in transit, and role-based access models.
- Configured data pipelines using Python and Spark to simulate secure data flows, and collaborated on integration patterns using JSON schema, Kafka, and GitLab while applying governance guardrails and audit logging.
- Leveraged monitor container-based workloads for anomalies and endpoint protection within Kubernetes (EKS) and Docker environments.

Mahmoud K. Alsaffar

Senior Lead Data Engineer & Cloud Data Architect | AI/ML-Enabled Data Platforms | AWS

• Azure • GCP • Databricks • Snowflake

DOD & IRS + MBI Clearance Active

Email: [REDACTED] | [REDACTED]

Phone: [REDACTED]

-
- Designed and implemented scalable Snowflake data models and pipelines to ensure compliance
 - Used Erwin Data Modeler to design, document, and maintain enterprise-level logical and physical data models supporting warehousing and governance requirements.
 - Applied Kimball methodology to build enterprise data marts, including conformed dimensions and slowly changing dimensions.
 - Developed and validated complex ETL workflows using Databricks auto loader workflow for data ingestion, transformation, and loading part of transactional data from various sources into Snowflake for masking capabilities.
 - Optimized Snowflake environments by leveraging advanced features like materialized views, clustering keys, data shares, clones, and resource monitors for cost-effective performance.
 - Implemented robust data validation rules and conducted performance tuning of Snowflake SQL queries to improve pipeline efficiency and integrity.
 - Applied data masking techniques in Snowflake to protect sensitive information.
 - Contributed to the organization's data mesh strategy by enabling domain-oriented data ownership and creating interoperable Snowflake data products with standardized schema and lineage.
 - Leveraged Tableau deliver self-service analytics and executive dashboards, enabling data-driven decision-making across business units.
 - Designed star/snowflake schemas and optimized DW models using Microsoft SQL Server.
 - Established enterprise-wide metadata cataloging and lineage tracking using Databricks Unity Catalog, enhancing data discoverability, governance, and secure access across domains.
 - Familiar with modern orchestration tools including Airflow and Dagster for managing data pipelines.
 - Utilized Datadog for real-time cloud infrastructure monitoring, service uptime tracking, and alert configuration to ensure high availability.
 - Familiarity with LLMOps workflows, including prompt engineering, fine-tuning, and operational deployment of LLMs.
 - Designed hybrid batch and real-time pipelines integrating Kafka (Red Panda), Spark Streaming, and Snowflake.
 - Architected and implemented API-based data integration pipelines using RESTful services and event-driven frameworks to enable real-time and batch data processing.
 - Driving platform cost savings by optimizing query performance, Table Partitioning, implementing auto-scaling resources, and configuring usage monitors in Snowflake and AWS.
 - Researched and incorporated new Snowflake capabilities.
 - Partnered with data science teams to operationalize ML pipelines and enable LLM workflows using Databricks.
 - Collaborated with account executives and delivery leads to conduct technical discovery sessions, capturing client pain points and aligning them to secure cloud-native solutions.
 - Designed reference architectures and technical solution briefs for client proposals (RFP/RFQ/RFI), highlighting multi-cloud security architecture and continuous compliance posture.
 - Owned and delivered end-to-end technical content for RFP responses and client proposals, showcasing architecture roadmaps, ROI benefits, and solution blueprints for Databricks
 - Facilitated pre-sales workshops and assessments with clients to identify modernization and cloud security opportunities, recommending CSPM tools and governance frameworks.
 - Managed & mentoring a team of 6 engineers and architects, guiding them on both technical design and cloud security implementation, conducting sprint ceremonies, and mentoring compliance-focused development.

Mahmoud K. Alsaffar

Senior Lead Data Engineer & Cloud Data Architect | AI/ML-Enabled Data Platforms | AWS

• Azure • GCP • Databricks • Snowflake

DOD & IRS + MBI Clearance Active

Email:

Phone:

-
- workshops, fostering talent growth and promoting a culture of continuous learning.
 - Conducted interviews for cloud engineers and architects.

Tech Stack & Tools: AWS (Glue, Redshift, Lambda, S3, IAM, CloudWatch) Databricks (Unity Catalog, Workflows, CLI) Snowflake · Kafka (Red Panda) Spark Streaming · Informatica Cloud · Python · SQL · GitLab CI/CD · Docker · Kubernetes (EKS) REST APIs · JSON Schema · Tableau · Erwin Data Modeler · Microsoft SQL Server · Airflow.

Booz Allen Hamilton Inc.

Lead Data Engineer – AWS | Remote / Washington DC | (04/2022 – 06/2024): Led architecture for cloud-based IRS solutions, aligning enterprise systems with data strategy goals. Conducted detailed requirements gathering, helped workshops, and led sprint reviews to ensure prompt delivery.

Client: IRS Enterprise Data Platform | (04/2022 – Sep 2023): The Enterprise Data Platform (EDP) is a centralized data product platform serving the IRS, enabling integration of domain-level datasets into business-friendly, linked, and explorable data products. It supports advanced analytics, compliance, and cross-functional data consumption via a secure, cloud-native architecture.

Key Responsibilities & Contributions:

Cloud Data Architecture & Platform Engineering

- Led the design and implementation of a modern Lakehouse architecture on Databricks, applying the Bronze/Silver/Gold pattern for both batch and streaming pipelines.
- Integrated IRS on-prem systems with AWS cloud, enabling hybrid data pipelines and scalable ETL workflows.
- Modernized legacy IRS systems, including batch mainframe workflows, into event-driven cloud pipelines using Spark + Kafka.
- Designed and documented shared architectural capabilities, including workspace setup, access controls, and automation pipelines across IRS domains.
- Orchestrated Glue + Lambda pipelines for event-driven data processing.
- Participated in PI planning and Scaled Agile Framework (SAFe) ceremonies as part of IRS agile transformation.
- Streamlined team delivery processes and enforced agile practices to meet critical timelines.
- Familiar with AWS DynamoDB design patterns, including partition key strategy, global secondary indexes (GSIs), and conditional writes for operational data systems.
- Participated in legacy-to-cloud modernization efforts for large-scale federal systems, mapping mainframe batch processes to event-driven streaming pipelines using Spark and Kafka.
- Designed normalized data models and operational schema structures to support real-time transactional systems using Databricks and AWS.

Databricks Development & Data Engineering

- Developed and optimized PySpark ETL pipelines to ingest and transform structured/semi-structured data from legacy RDBMS, Kafka (Red Panda), and REST APIs.
- Deployed and orchestrated jobs using Databricks Workflows, implementing CI/CD automation through GitHub Actions and Databricks CLI.
- Leveraged Delta Lake features (time travel, schema evolution, and partitioning) for scalable and audit-compliant data storage.
- Implemented Databricks Audit Framework to monitor job executions, logging, and exception handling across environments.

Mahmoud K. Alsaffar

Senior Lead Data Engineer & Cloud Data Architect | AI/ML-Enabled Data Platforms | AWS

• Azure • GCP • Databricks • Snowflake

DOD & IRS + MBI Clearance Active

Email: [REDACTED] | [REDACTED]

Phone: [REDACTED]

Security, Governance & Data Management

- Integrated Unity Catalog for fine-grained access control, lineage tracking, and metadata governance, aligning with FedRAMP and internal audit standards.
- Best practices applied in data modeling, quality assurance, and security, supporting cross-domain governance.
- Automated partitioning and management of Business Master File (BMF) tables to support high-volume workloads.
- Integrated open-source data quality frameworks into enterprise ETL pipelines to automate validation of schema, nulls, uniqueness, and referential integrity checks across domains.
- Built internal tools and CI/CD hooks for validating data quality at ingestion and transformation stages, improving SLA adherence and trust in data assets.
- Designed and implemented data lineage tracking by integrating metadata extraction into ETL jobs and surfacing lineage through Unity Catalog and custom-built dashboards.
- Partnered with Legal and Security teams to define and enforce data handling policies, integrating encryption standards, fine-grained access control, and privacy redaction rules within data pipelines.

Stakeholder Collaboration & Solution Delivery

- Conducted technical workshops, sprint reviews, and requirement sessions with IRS product owners and domain leads.
- Delivered solution designs, POCs, and documentation for tools including Tableau dashboards and Informatica metadata integration.

Tech Stack: AWS (EC2, Lambda, S3), Databricks (Delta Lake, Workflows, Notebooks, Unity Catalog) PySpark / Spark SQL / DataFrames, GitHub Actions / CI/CD · Glue ETL · Snowflake · Tableau · Informatica · Agile / Scrum.

Client: HHS ASPE (Assistant Secretary for Planning and Evaluation within the US Department of Health and Human Services) | SAS Viya Lead Strategy Technical

Architecture | (09/2023 – 06/2024): Led the modernization of ASPE's legacy on-premises SAS environment by transitioning to SAS Viya on AWS Cloud, enabling improved model performance, secure operations, and scalable data analytics across diverse data sources for daily predictive processing.

Key Responsibilities & Contributions:

Cloud Strategy & Technical Requirements

- Conducted comprehensive technical assessments and captured business capability needs to guide the design and deployment of SAS Viya in AWS/ Azure Cloud.
- Worked directly with program analysts and technical teams to gather and translate user and system requirements into actionable system architecture.

System Design & Deployment

- Designed the target cloud architecture, including networking, security, and compute configurations, to meet federal ATO (Authorization to Operate) requirements.
- Authored and delivered a detailed System Design Document (SDD) addressing infrastructure, access control, encryption, logging, and resilience features.
- Facilitated and supported the deployment of SAS Viya within a secure Azure environment, enabling scalable and performant data modeling workflows.

Mahmoud K. Alsaffar

Senior Lead Data Engineer & Cloud Data Architect | AI/ML-Enabled Data Platforms | AWS

• Azure • GCP • Databricks • Snowflake

DOD & IRS + MBI Clearance Active

Email:

Phone:

Vendor Collaboration & Solution Integration

- Served as a liaison between HHS and multiple technology vendors, including SAS, Stata, and the cloud hosting provider, to align system capabilities with licensing, support models, and cloud deployment best practices.
- Evaluated integration approaches for hybrid data access and computational performance across analytical tools.
- Familiarity with SAP integration approaches in cloud environments

Operations & Administration

- Delivered ongoing system administrative support, managing day-to-day cloud environment operations, user provisioning, monitoring, and maintenance.
- Helped define operational controls and support protocols to ensure availability, security, and compliance in the production environment.

Tech Stack & Tools: SAS Viya on AWS / Azure · AWS EC2 / IAM / VPC · Stata · Cloud Security (ATO / NIST controls) Linux System Administration · System Design Documentation (SDD) Vendor Management · Agile / Client Workshops.

PricewaterhouseCoopers (PwC)

Client: Verizon | Google Cloud Architect (GCP) - Manager | Remote / Chicago, IL | (01/2022

– 04/2022): Led Verizon's enterprise cloud modernization initiative, focusing on migrating its big data infrastructure from on-premises Hadoop to Google Cloud Platform (GCP). Delivered architecture blueprints, stakeholder alignment, and technical prototypes to accelerate Verizon's analytics and infrastructure transformation.

Key Responsibilities & Contributions:

Cloud Strategy & Migration Architecture

- Designed the comprehensive migration strategy for legacy Hadoop clusters (HDFS, Hive, HiveMQ, Sqoop) to GCP services, ensuring alignment with security, scalability, and business SLAs.
- Created GCP reference architecture with standardized data ingestion, processing, and storage pipelines using BigQuery, Cloud Storage, Pub/Sub, and Dataflow.

Technical Prototypes & POC Development

- Built and delivered multiple Proofs of Concept (POCs) to validate cloud-native replacements for on-prem services, including Hive-to-BigQuery translation, data pipeline optimization, and schema evolution strategies.
- Demonstrated streaming ingestion use cases using Pub/Sub and Dataflow, showcasing real-time analytics enablement in BigQuery.
- Developed reusable Terraform modules for infrastructure provisioning and CI/CD templates to support production readiness.

Demo Delivery & Stakeholder Enablement

- Conducted live demos for technical teams and leadership to showcase the performance, cost-efficiency, and scalability of GCP-native data solutions compared to legacy systems.
- Facilitated technical deep-dives, architecture whiteboarding, and walkthrough sessions for Verizon's engineers and decision-makers.
- Provided hands-on knowledge transfer sessions and onboarding materials for Verizon's internal GCP teams.

Mahmoud K. Alsaffar

Senior Lead Data Engineer & Cloud Data Architect | AI/ML-Enabled Data Platforms | AWS

• Azure • GCP • Databricks • Snowflake

DOD & IRS + MBI Clearance Active

Email:

Phone:

Leadership & Collaboration

- Led architecture planning workshops, working directly with Verizon's cloud, data governance, and compliance teams to align migration path and implement milestones.
- Defined best practices for access control, data governance, and network policies in multi-project GCP environments.

Tech Stack: Google Cloud Platform (BigQuery, Dataflow, Pub/Sub, Cloud Storage, Composer)
Hadoop (HDFS, Hive, Sqoop) Terraform · Python / SQL · IAM, VPC, Cloud Audit Logs · CI/CD (Cloud Build) Agile / Demo-Driven Development

Additional Experience:

Accenture LLP

Manager | Chicago, IL | (08/2015 – 12/2021): Led large-scale cloud migration projects, transitioning enterprise systems to AWS, GCP and Azure, implementing data pipelines, and ensuring end-to-end architecture integration.

Directed cross-functional teams to build real-time analytics and big data solutions, enhancing data-driven decision-making.

Projects Highlights:

- **US Department of Defense – Airforce (AWS Data Lake) Washington, D.C., Lead Enterprise Architecture:** Developed the AWS Data Lake architecture, integrating enterprise solutions, cost estimation, and data governance processes.
- **Internal Accenture Project (CO2 Emissions Data Management), Lead Data Architecture, Chicago Illinois:** Spearheaded the digital transformation by designing centralized data storage for CO2 emissions, implementing real-time analytics.
- **Boston Green Energy - Lead Data Architect – AWS Cloud | Boston, MA:** Led a cross-functional team to migrate on-prem databases to AWS, designing the architecture for data pipelines and implementing batch data loads for SQL Server databases. Managed the overall ETL implementation and data modeling, ensuring seamless integration with AWS cloud services. Managed distributed engineering teams across U.S. and offshore delivery centers, ensuring follow-the-sun delivery and agile alignment.
- **Hewlett-Packard (HP), Enterprise Architecture Practitioner - San Francisco, CA:** Designed and implemented a real-time data analytics solution using Kafka (Red Panda) and Spark, improving data ingestion and processing for high-volume transactions.
- **Combined Health Insurance, Lead Data Architect Modeler – Azure Data Lake, Chicago Illinois:** Managed Azure Data Lake migration, leading offshore teams in data modeling and ETL processes. Oversaw the integration of multiple data sources into a unified data warehouse, improving data accessibility and reporting.
- **Technical Product Manager (TPM) – GCP Data Engineering:** Led delivery of GCP-based data engineering solutions for Google's internal teams, focused on big data architecture and pipeline design. Oversaw development of batch and streaming pipelines using BigQuery, Dataflow, Pub/Sub, and Cloud Storage. Worked with Google engineers to build scalable, production-ready pipelines for real-time and large-volume data processing. Supported pipeline optimization, schema design, and integration patterns for high-throughput analytics. Facilitated planning, technical reviews, and documentation to ensure alignment with Google's engineering standards.

Mahmoud K. Alsaffar

Senior Lead Data Engineer & Cloud Data Architect | AI/ML-Enabled Data Platforms | AWS

• Azure • GCP • Databricks • Snowflake

DOD & IRS + MBI Clearance Active

Email: [REDACTED] | [REDACTED]

Phone: [REDACTED]

-
- **Al Rajhi Bank (Banking & Financial Services), Lead Architecture - Saudi Arabia:**
Lead project big data architectures, enabling legacy systems integration with cloud data lake solutions strategy.
 - **Lowe's HDP Hadoop Cluster (Retail Sector), Lead Big Data Engineer, Charlotte North Carolina:** Hadoop Tech Lead, Managed Hadoop testing efforts and ETL processes, overseeing data ingestion and technical documentation for end-to-end data pipelines.
 - **Fanny May - Hortonworks Data Platform (Financial Sector), Senior Data Engineer, Washington, D.C.:** Provided Level 2 support for Hortonworks Data Platforms, ensuring system stability, monitoring, and issue resolution.
 - **Royal Bank of Canada (Banking Canada), Data Architecture, Toronto Canada:**
Developed big data architecture strategies, advising on technology options, and managing data migration and ETL processes.

CSC-Computer Science Corp.: Hadoop Consultant (2014 – 2015): Led Hadoop installations and configurations for multiple clients, focusing on MapReduce jobs and Hive queries.

3SNetwork Inc.: Hadoop Engineer (2012 – 2014): Managed Hadoop clusters, implemented data processing workflows using Apache Oozie, and automated data management for end-to-end processes.

Education

- **Master of Science in Computer Science**
Roosevelt University, Chicago, IL | **Dec-2012**
- **Bachelor of Science in Software Engineering**
Al Mansur University College, Iraq | **Sep-2007**