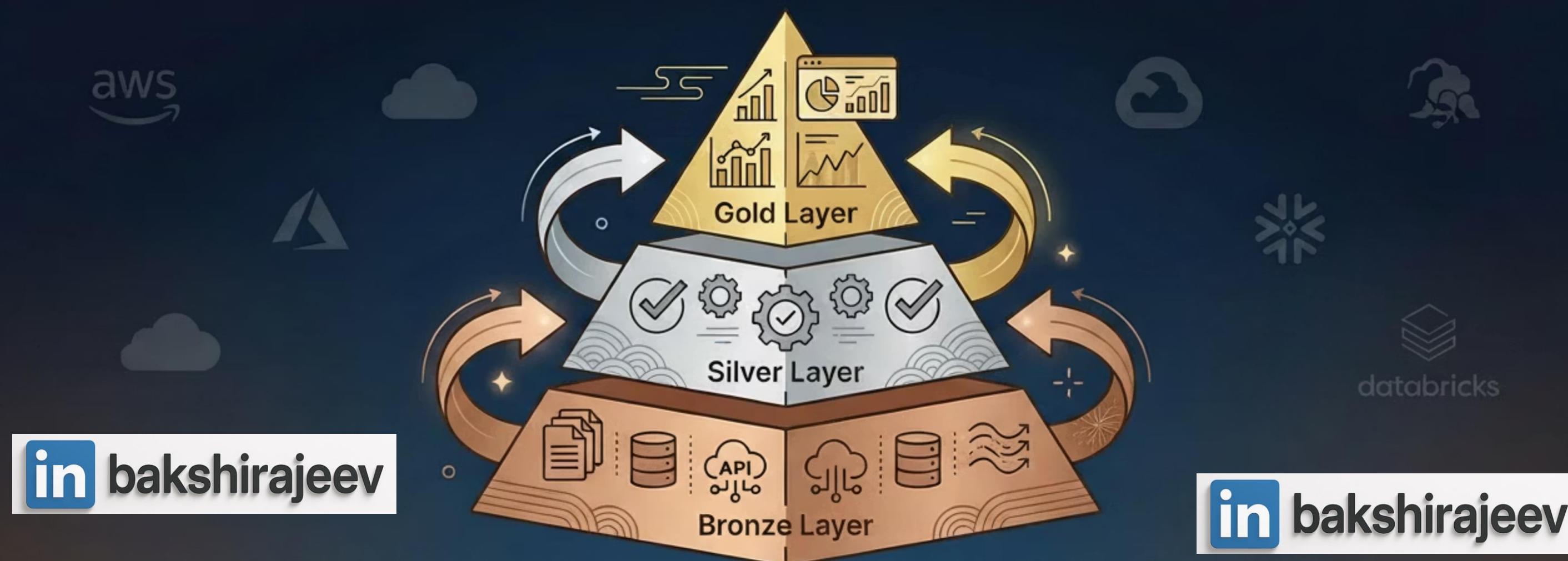


# Medallion Architecture: Enterprise Data Engineering Excellence



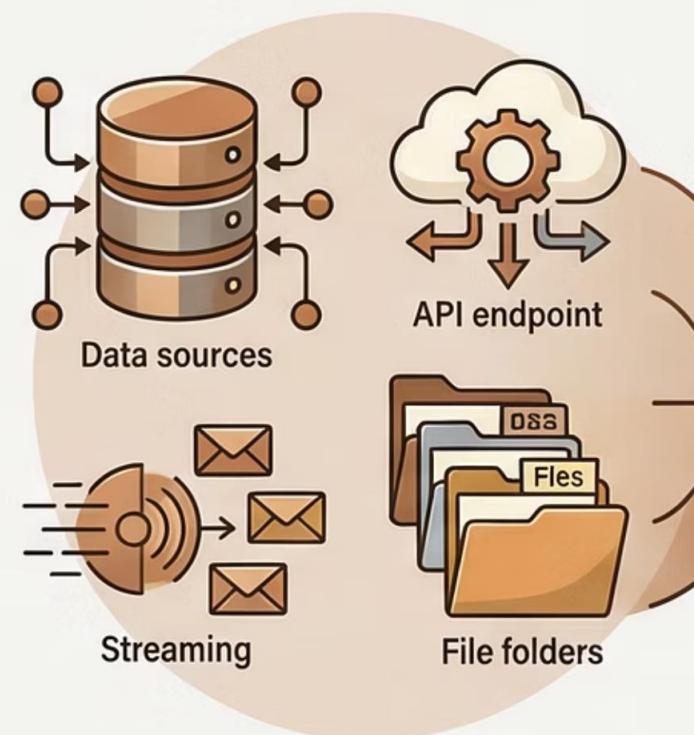
 **bakshirajeev**

 **bakshirajeev**

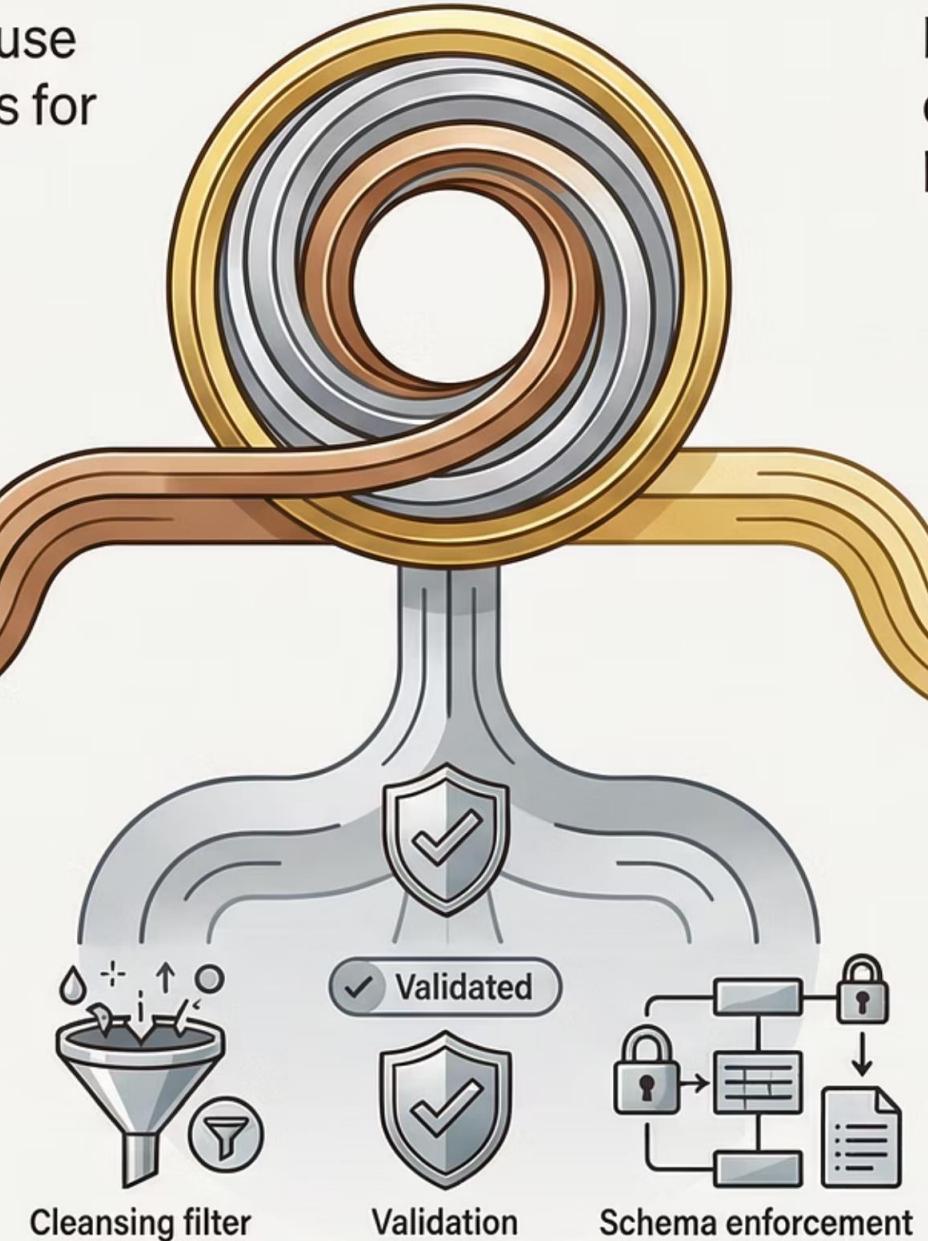
A Comprehensive Guide to Implementation Across Data Platforms

# What is Medallion Architecture?

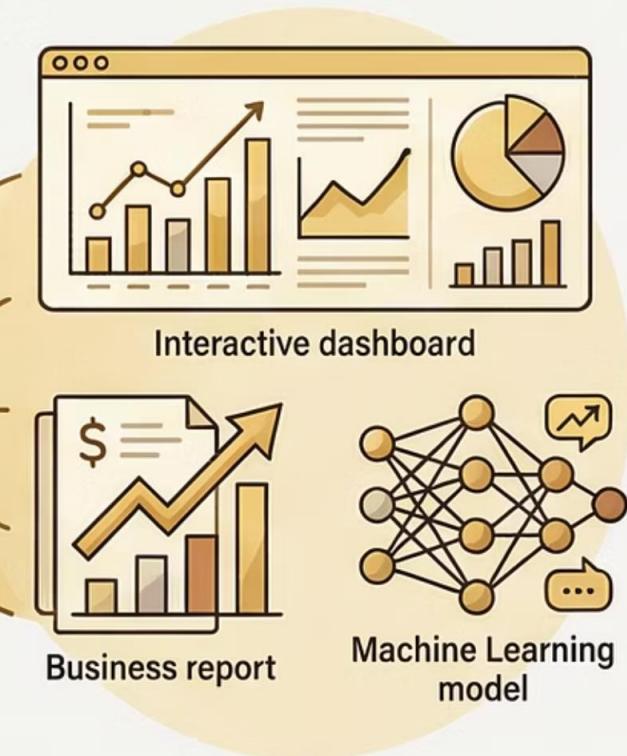
Data design pattern organizing lakehouse data through Bronze, Silver, Gold layers for progressive quality refinement



Ensures data lineage, quality governance, scalable analytics



Bronze stores raw data, Silver contains cleaned validated data, Gold delivers business-ready analytics



Foundation for enterprise data platforms

# The Three-Layer Foundation

## Bronze

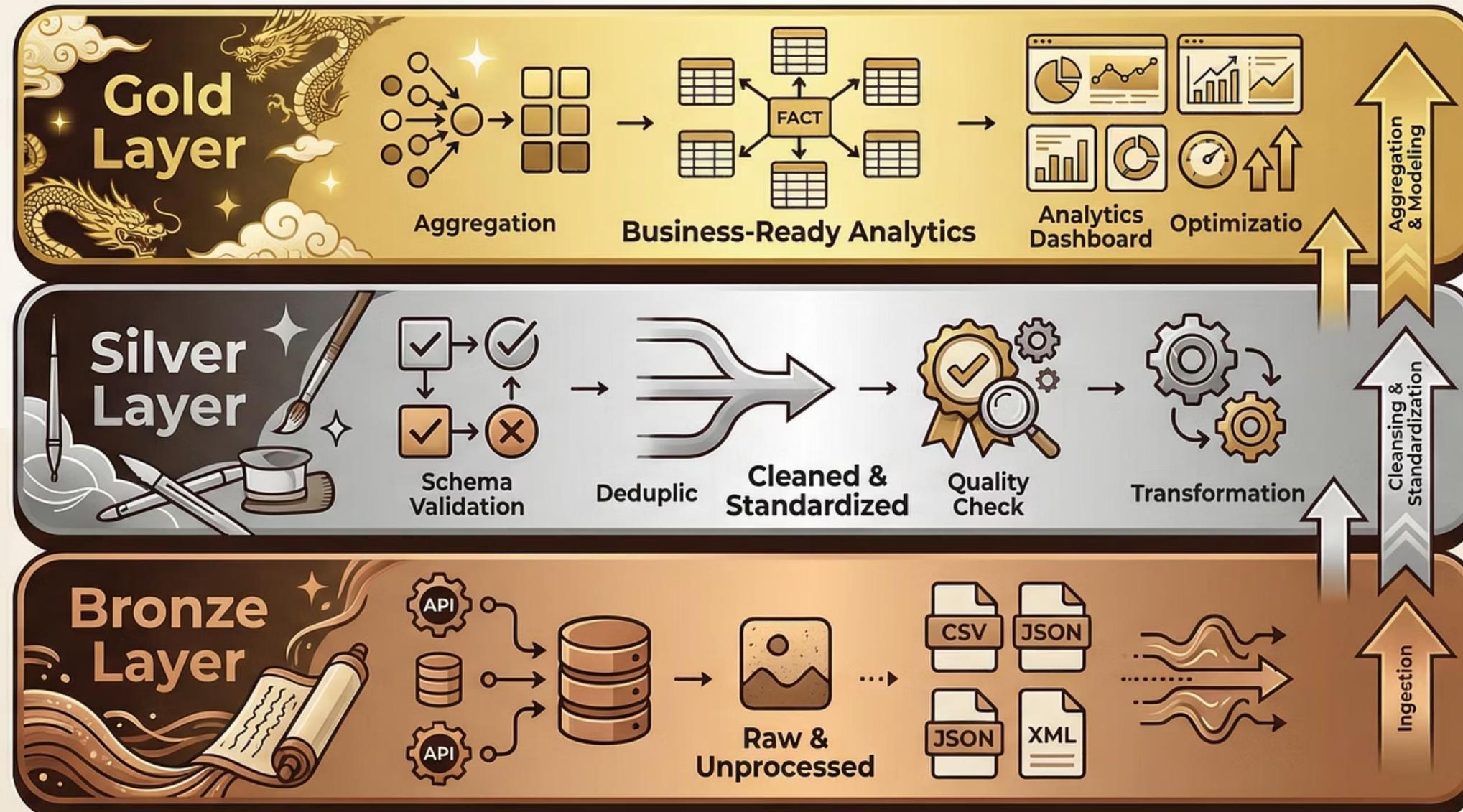
Raw data ingestion from multiple sources - APIs, databases, files, streams - stored in original format with minimal processing

## Bronze

Raw data ingestion from multiple sources - APIs, databases, files, streams - stored in original format with minimal processing

## Bronze

Raw data ingestion from multiple sources - APIs, databases, files, streams - stored in original format with minimal processing



## Gold

Business-ready datasets optimized for analytics, reporting, machine learning with aggregations and dimensional modeling

## Pipeline

Each layer builds upon previous, creating robust data quality pipeline

## Silver

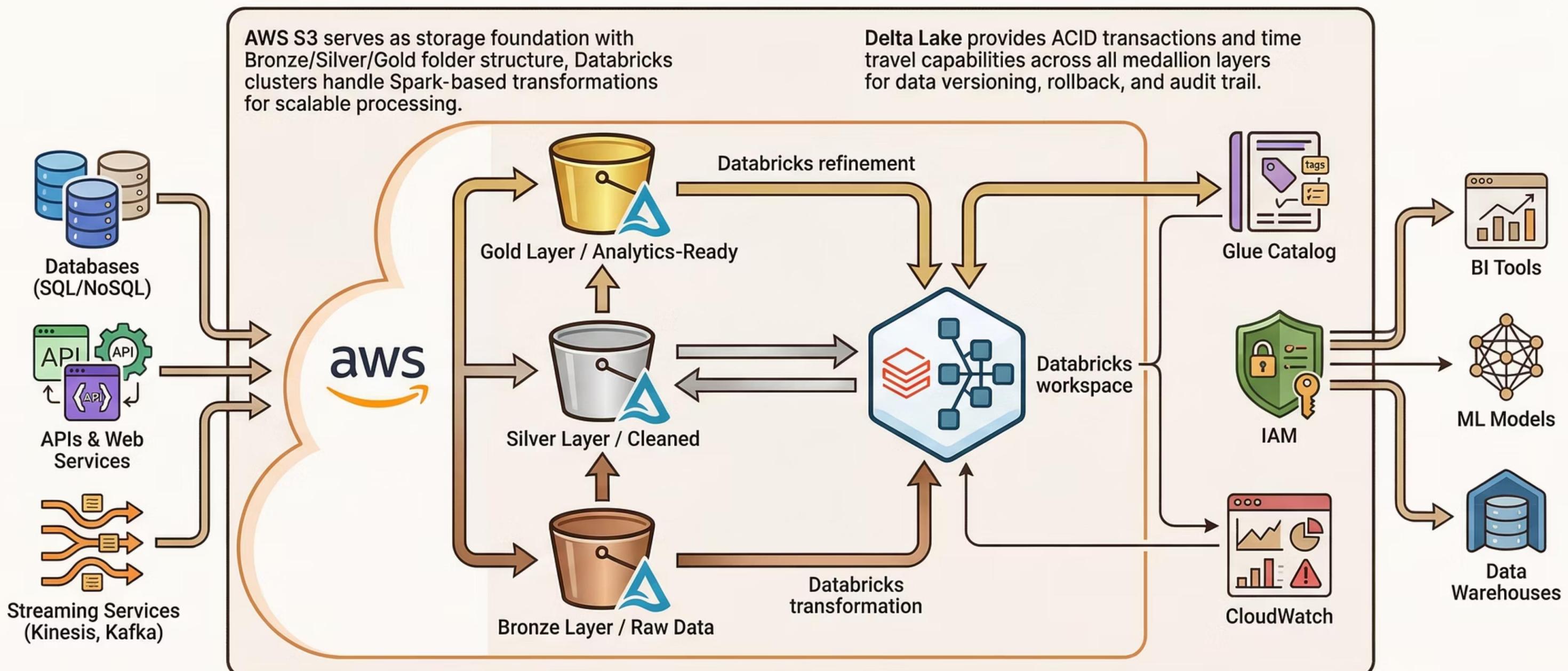
Cleaned, validated, standardized data with schema enforcement, deduplication, data quality checks applied

## Gold

Business-ready datasets optimized for analytics, reporting, machine learning with aggregations and dimensional modeling

# Databricks Medallion on AWS

[in](#) bakshirajeev



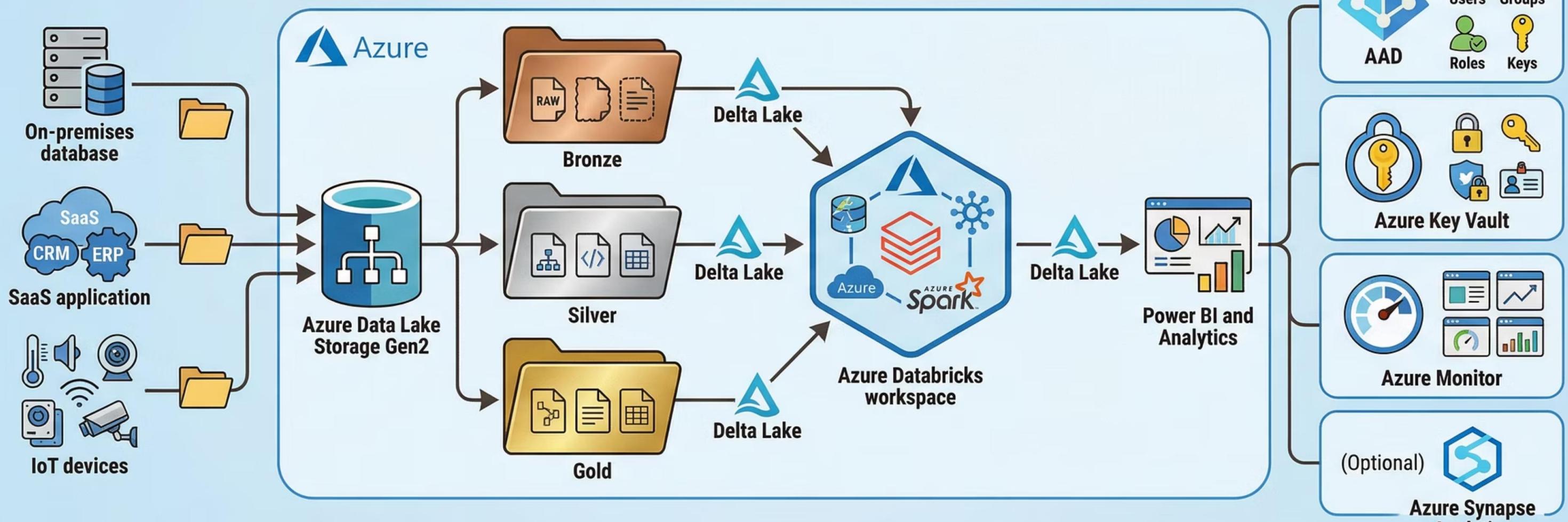
AWS Glue Catalog manages metadata and schema registry, enabling data discovery and governance across organization.

IAM controls access permissions and security policies for each architectural layer, ensuring compliance and data protection.

# Databricks Medallion on Azure

Azure Data Lake Storage Gen2 provides hierarchical namespace for efficient Bronze/Silver/Gold data organization, delivering superior performance optimization and cost management for petabyte-scale data

Azure Databricks workspace integrates seamlessly with Azure Active Directory for enterprise security, role-based access control, conditional access policies, and unified identity management



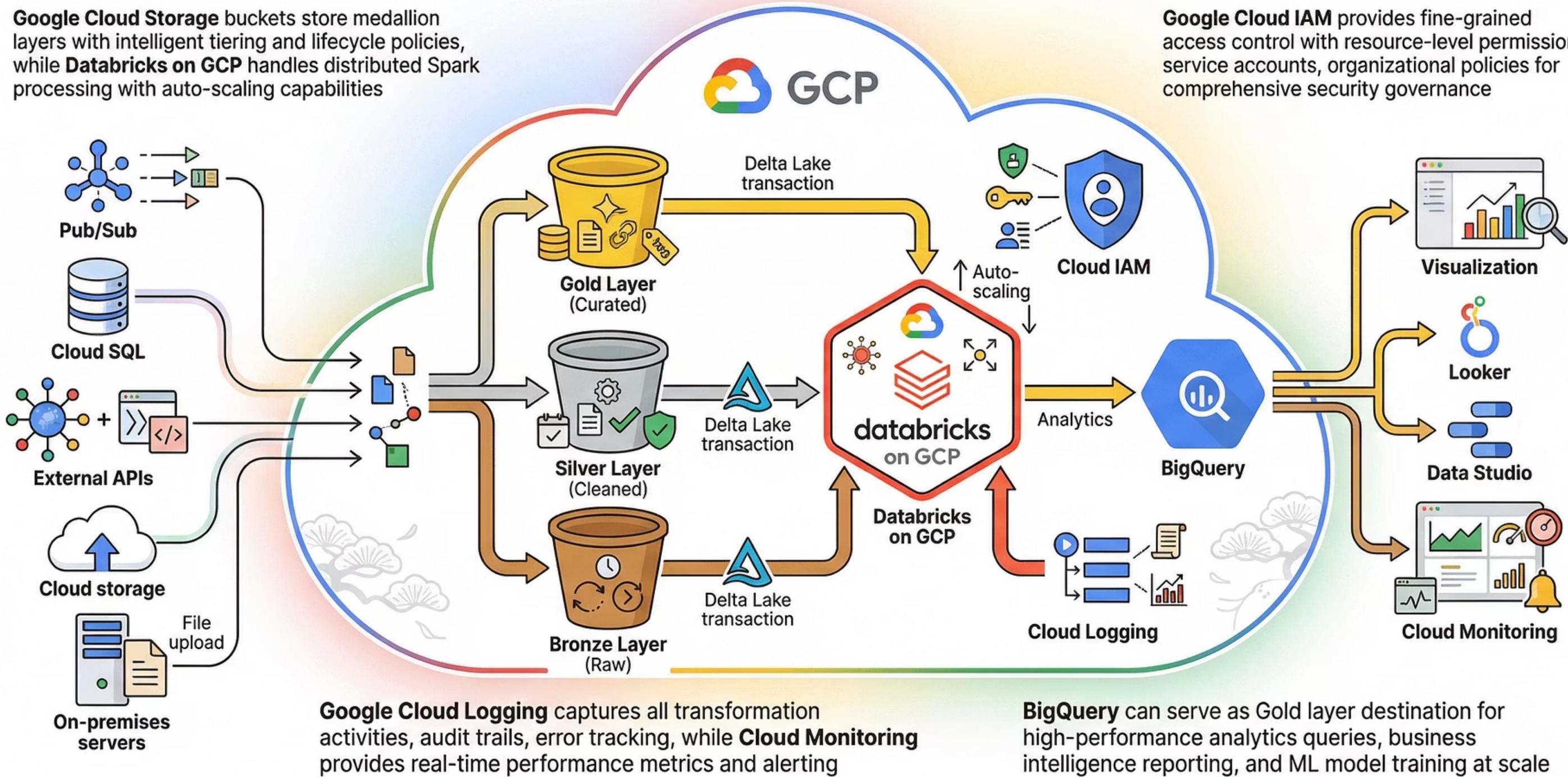
Azure Key Vault manages secrets, connection strings, credentials securely, while Azure Monitor provides comprehensive logging, performance monitoring, alerting for all transformation activities

Delta Lake on ADLS enables ACID transactions, schema evolution, time travel, and data versioning across all medallion layers

# Databricks Medallion on GCP

[in](#) [bakshirajeev](#)

Google Cloud Storage buckets store medallion layers with intelligent tiering and lifecycle policies, while **Databricks on GCP** handles distributed Spark processing with auto-scaling capabilities

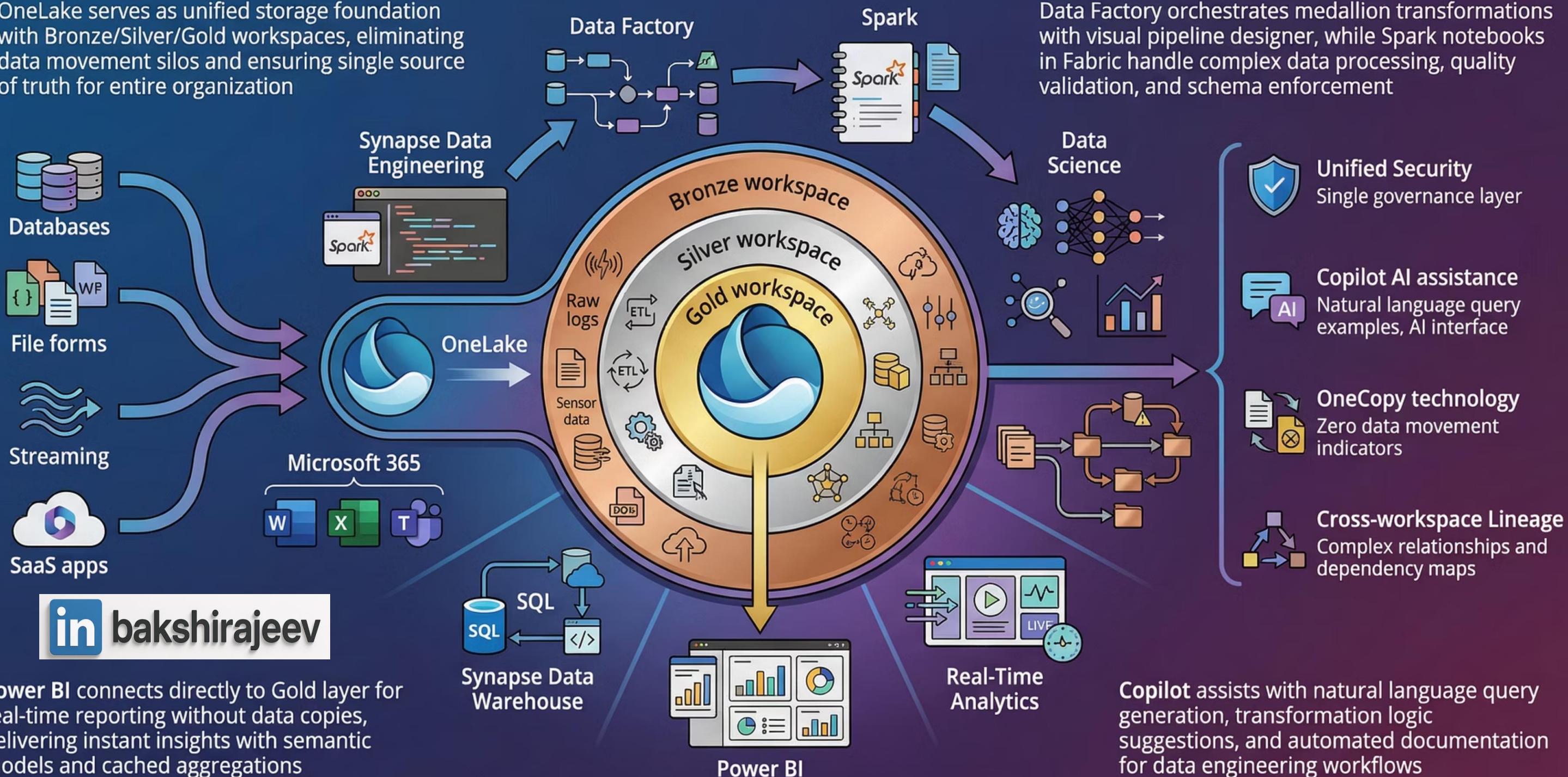


**Google Cloud Logging** captures all transformation activities, audit trails, error tracking, while **Cloud Monitoring** provides real-time performance metrics and alerting

**BigQuery** can serve as Gold layer destination for high-performance analytics queries, business intelligence reporting, and ML model training at scale

# Microsoft Fabric Implementation

OneLake serves as unified storage foundation with Bronze/Silver/Gold workspaces, eliminating data movement silos and ensuring single source of truth for entire organization

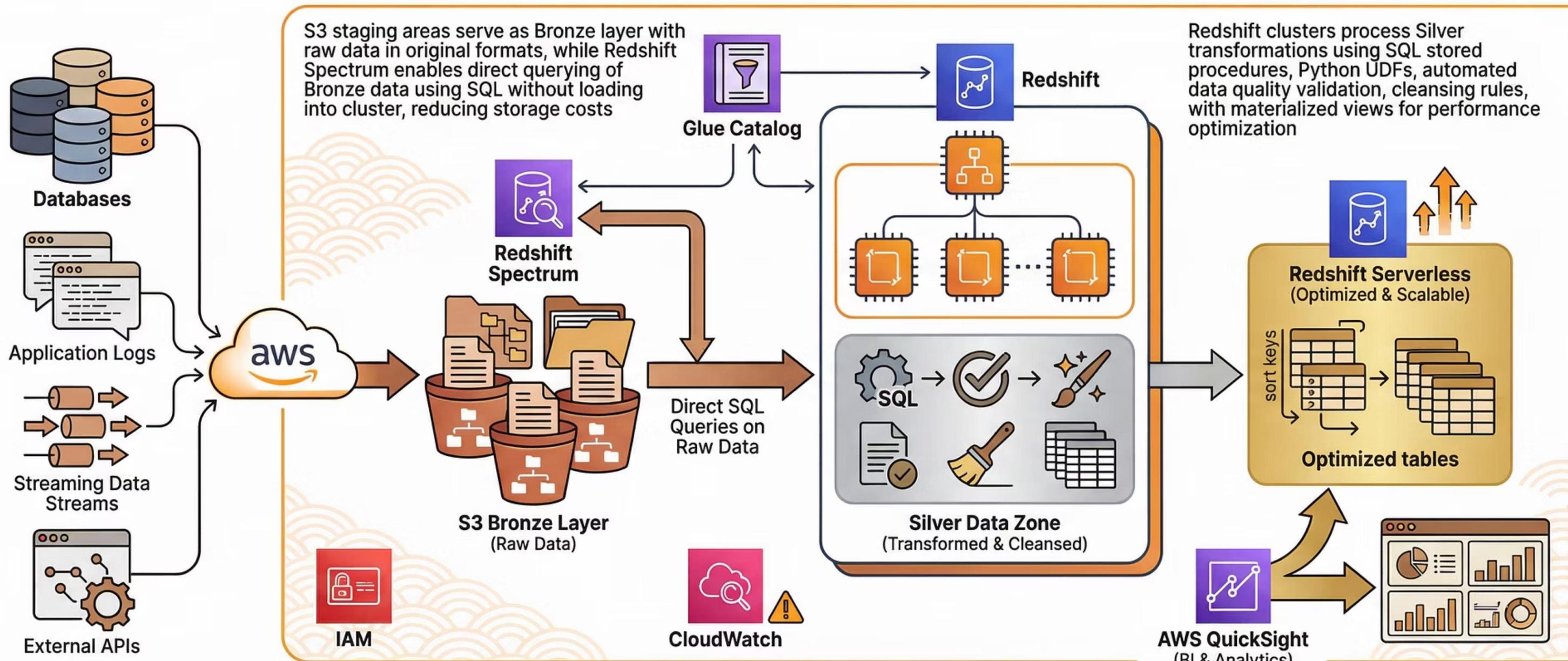


Power BI connects directly to Gold layer for real-time reporting without data copies, delivering instant insights with semantic models and cached aggregations

Data Factory orchestrates medallion transformations with visual pipeline designer, while Spark notebooks in Fabric handle complex data processing, quality validation, and schema enforcement



# Amazon Redshift Implementation

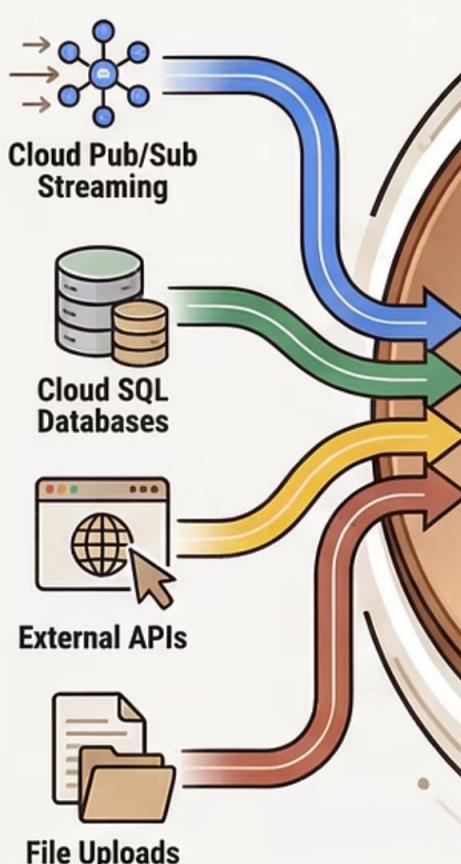


Redshift Serverless scales Gold layer workloads automatically based on demand, eliminating capacity planning, while concurrency scaling handles burst analytics queries

AWS QuickSight provides business intelligence dashboards, ML-powered insights, embedded analytics, connecting directly to Gold layer for real-time reporting

# Google BigQuery Implementation

Cloud Storage buckets maintain Bronze raw data with automatic lifecycle management, tiered storage classes for cost optimization, and long-term retention policies, with external BigQuery tables for direct SQL access.



**Bronze Layer**  
Foundation: Google Cloud Storage Buckets



**BigQuery Serverless Architecture**

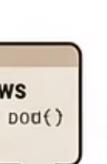


BigQuery



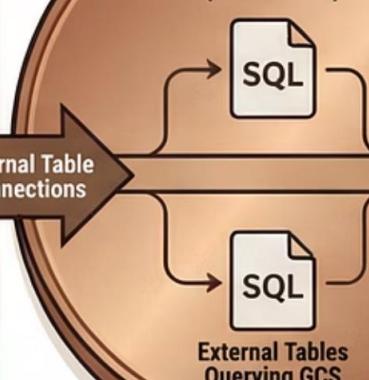
SQL Views  
`code = nos()`  
`, cverel`

**BigQuery Serverless Architecture**

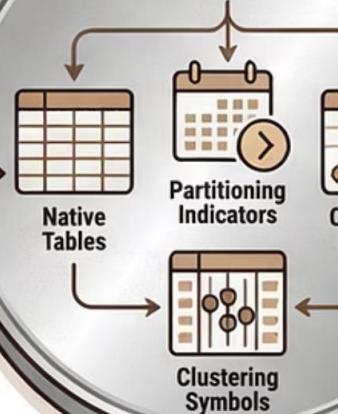


Scheduled Queries

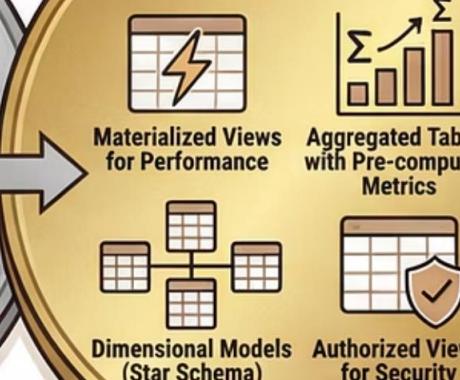
**Bronze Dataset (Raw Data)**



**Silver Dataset (Enriched Data)**



**Gold Dataset (Optimized Data)**



BigQuery datasets organize Silver and Gold layers with table partitioning for date-based queries, clustering for filter optimization, delivering massive query performance improvements and significant cost management.



Looker / Data Studio Visualization

**Cloud Dataflow**  
ETL Job Icons

Layer-to-Layer Transformations

Apache Beam Pipelines

Auto-scaling

Exactly-Once Processing

**Cloud Composer (Apache Airflow)**

Complex Pipelines

Dependency Management

Error Handling

Monitoring

Cloud Dataflow handles ETL transformations between layers with Apache Beam pipelines, auto-scaling for variable workloads, exactly-once processing semantics for data accuracy.

Cloud Composer orchestrates complex multi-step data pipeline workflows, dependency management, error handling, monitoring, and alerting using managed Apache Airflow.



IAM



Cloud Monitoring

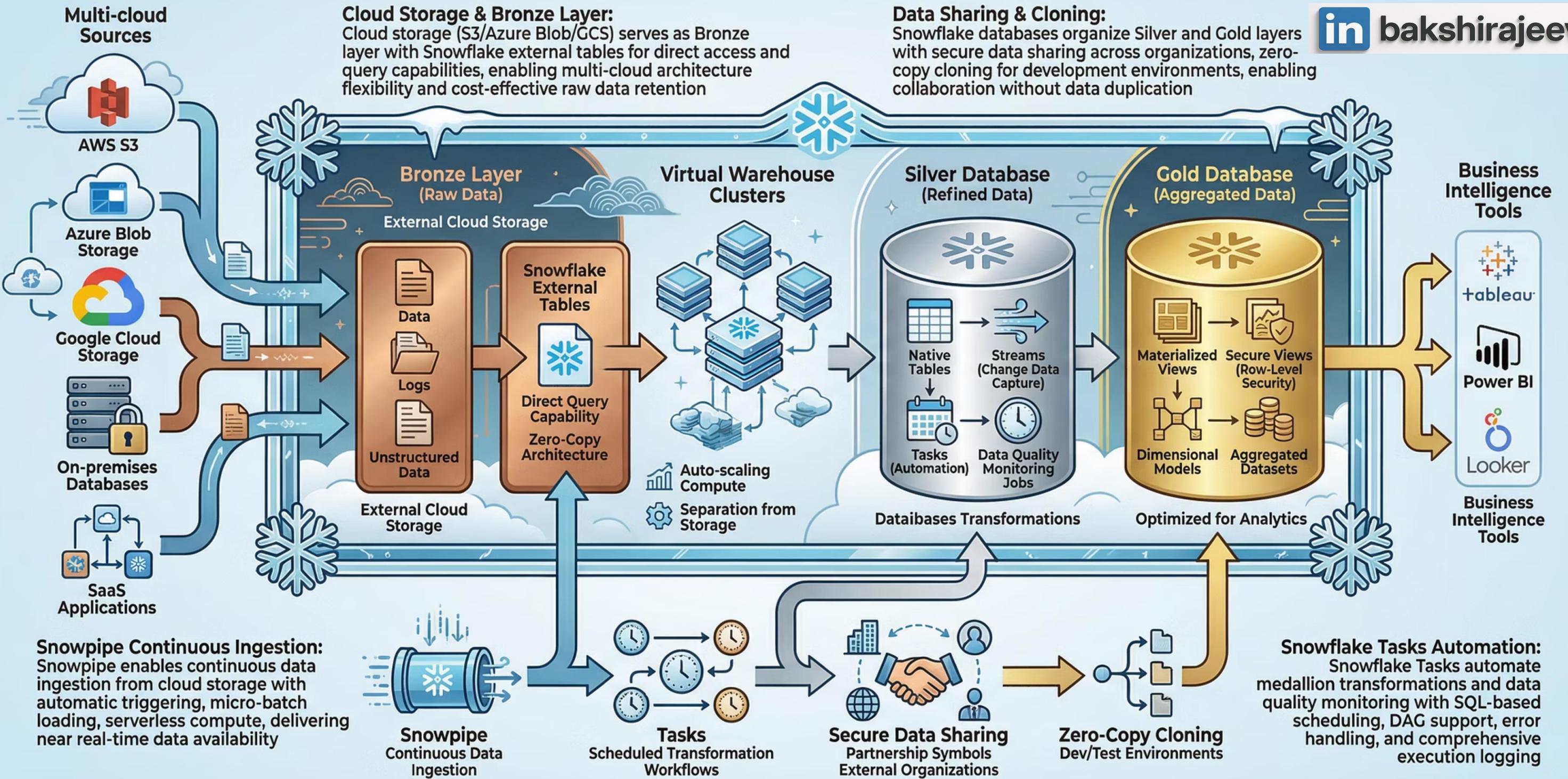


Data Catalog

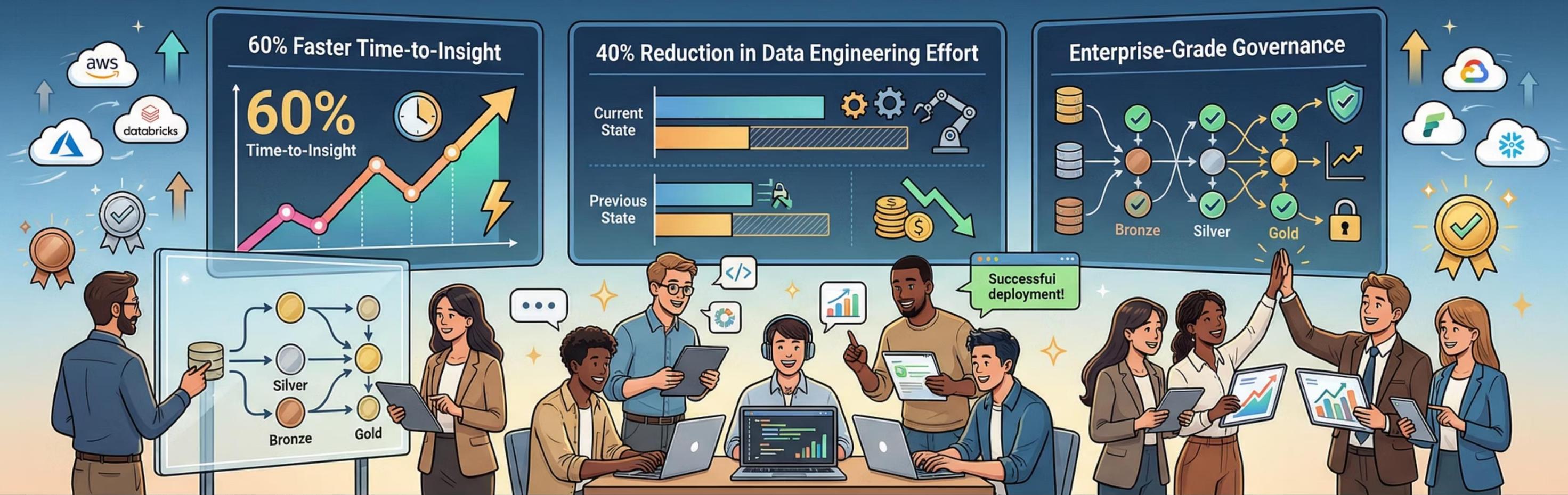


bakshirajeev

# Snowflake Implementation



# Enterprise Success with Medallion Architecture



**Medallion Architecture delivers measurable business value:**  
60% faster time-to-insight through structured data layers, accelerating analytics and decision-making capabilities.

**40% reduction in data engineering effort** via standardized transformation patterns, reusable components, and automated quality validation across platforms.

**Enterprise-grade governance** with complete data lineage, quality monitoring, compliance tracking, and unified security across all cloud platforms.

Implement Medallion Architecture to Transform Your Data Platform

 **bakshirajeev**