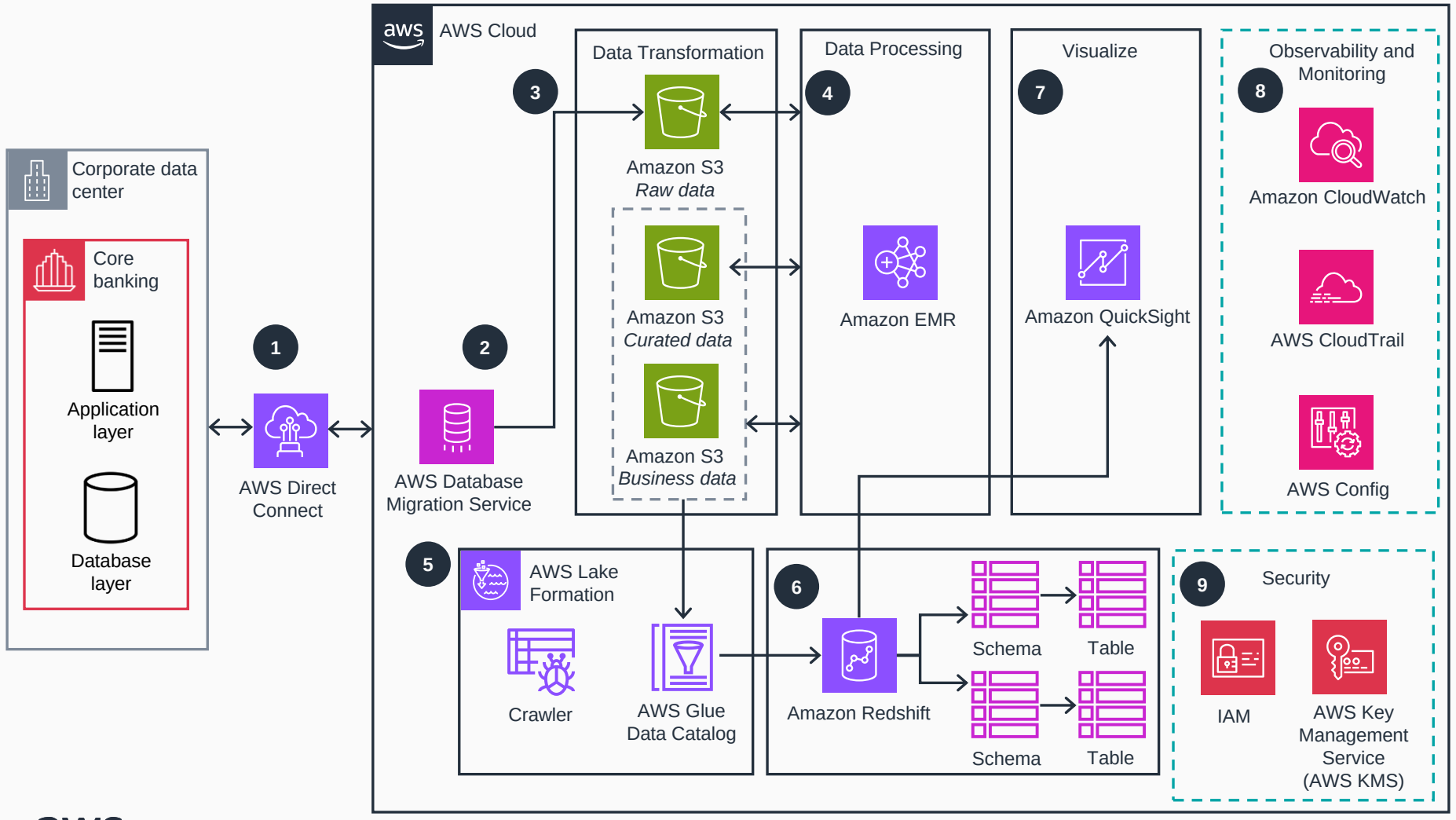
Guidance for Core Banking Data Lake on AWS

This architecture diagram shows how to securely replicate and transform core banking data on AWS and unlock actionable insights on that data. These insights can help credit unions gain a deeper understanding of their members, drive revenue growth, and enhance data-driven decision making. This slide shows Steps 1-7.



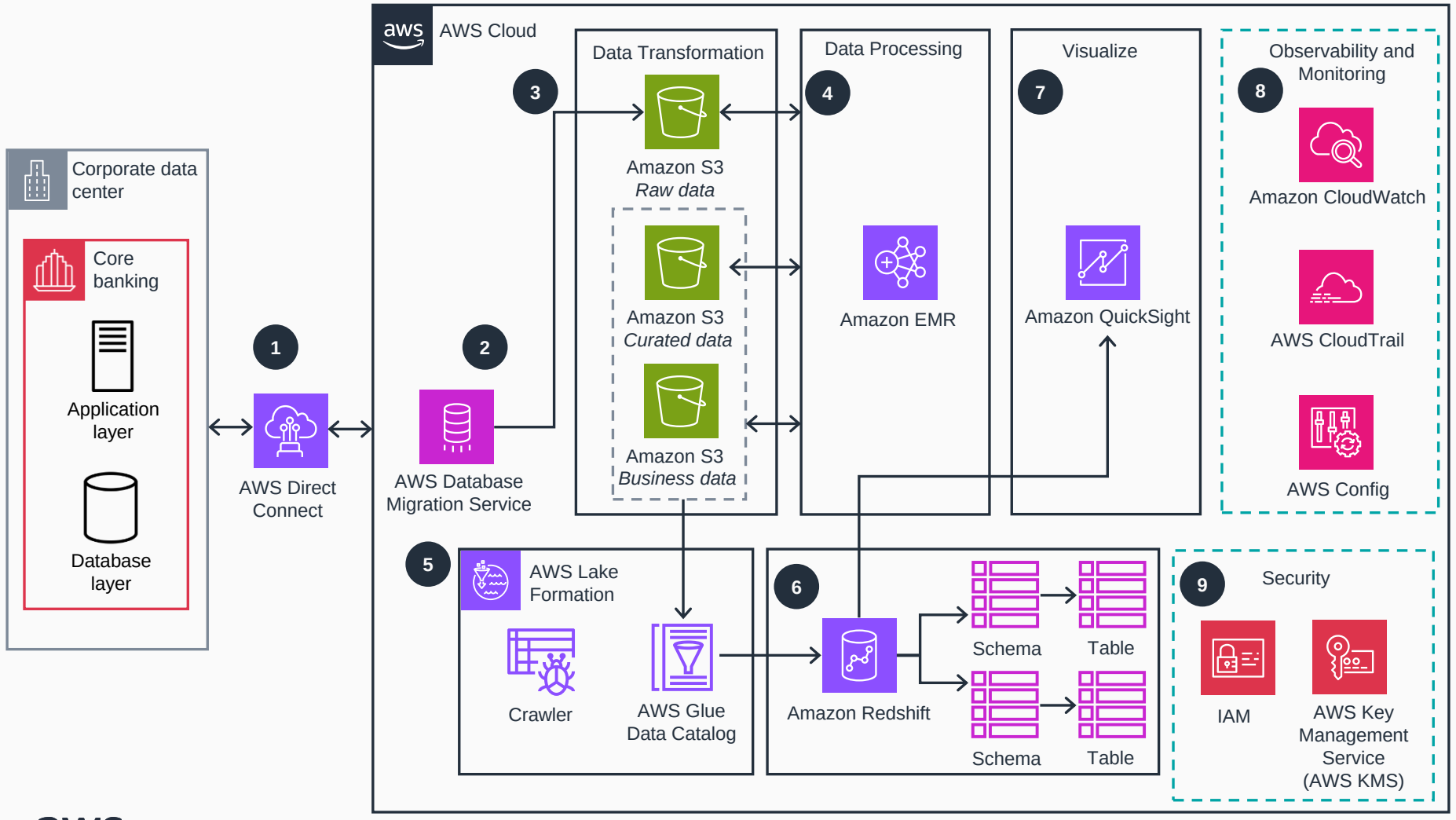
1. **AWS Direct Connect** provides transparent and resilient connectivity by connecting core banking data from customer data centers to the AWS Cloud.
2. **AWS Database Migration Service (AWS DMS)** migrates and replicates data from the on-premises core database to the AWS Cloud.
3. **Amazon Simple Storage Service (Amazon S3)** migrates and replicates data. The source database remains fully operational during the migration, minimizing downtime to applications that rely on the database.
4. An **Amazon EMR Serverless** cluster is created with Apache Spark, an open-source, distributed processing system used for big data workloads to transform your data.
5. **AWS Lake Formation** centrally governs, secures, and shares your data, while allowing you to easily manage permissions. **AWS Glue** crawlers scan data in your data lake, classify it, extract schema information from it, and store the metadata automatically in the **AWS Glue Data Catalog**.
6. **Amazon Redshift** uses SQL and machine learning (ML) features built into the service to analyze structured and semi-structured data across your data lake at scale.
7. Visualize your data using **Amazon QuickSight**, a fast, easy-to-use, serverless business analytics service that makes it easier to build visualizations, perform ad-hoc analysis, and quickly get business insights from your data, anytime, on any device.

Reviewed for technical accuracy June 19, 2024

AWS Reference Architecture

Guidance for Core Banking Data Lake on AWS

This architecture diagram shows how to securely replicate and transform core banking data on AWS and unlock actionable insights on that data. These insights can help credit unions gain a deeper understanding of their members, drive revenue growth, and enhance data-driven decision making. This slide shows Steps 1-7.



1. **AWS Direct Connect** provides transparent and resilient connectivity by connecting core banking data from customer data centers to the AWS Cloud.
2. **AWS Database Migration Service (AWS DMS)** migrates and replicates data from the on-premises core database to the AWS Cloud.
3. **Amazon Simple Storage Service (Amazon S3)** migrates and replicates data. The source database remains fully operational during the migration, minimizing downtime to applications that rely on the database.
4. An **Amazon EMR Serverless** cluster is created with Apache Spark, an open-source, distributed processing system used for big data workloads to transform your data.
5. **AWS Lake Formation** centrally governs, secures, and shares your data, while allowing you to easily manage permissions. **AWS Glue** crawlers scan data in your data lake, classify it, extract schema information from it, and store the metadata automatically in the **AWS Glue Data Catalog**.
6. **Amazon Redshift** uses SQL and machine learning (ML) features built into the service to analyze structured and semi-structured data across your data lake at scale.
7. Visualize your data using **Amazon QuickSight**, a fast, easy-to-use, serverless business analytics service that makes it easier to build visualizations, perform ad-hoc analysis, and quickly get business insights from your data, anytime, on any device.

Reviewed for technical accuracy June 19, 2024

AWS Reference Architecture