Data is collected from multiple data sources across the enterprise, including software-as-a-service (SaaS) applications, edge devices, logs, streaming media, and social networks. Online web activity comes from web sites, social media platforms, emails, and online campaigns. Offline sources include purchase history and subscriptions—primarily customer relationship management (CRM) and 3rd party data.

Based on the type of data source, you can ingest the data into a data lake in AWS by using AWS Database Migration Service (AWS DMS), AWS DataSync, Amazon Kinesis, Amazon Managed Streaming for Apache Kafka (Amazon MSK), or Amazon AppFlow.

AWS Data Exchange can be used to integrate third-party data into the data lake.

You can also use AWS Lake Formation to enable unified governance, which helps you centrally manage security, access control (table, row, or column level security), and audit trails. It also enables automatic schema discovery and conversion to required formats.

AWS Glue extracts, transforms, catalogs, and ingests data across multiple data stores. Use AWS Glue DataBrew for visual data preparation and AWS Lambda for enrichment and validation.

Amazon QuickSight provides machine learning (ML) powered business intelligence. Amazon Redshift is used as a cloud data warehouse. Amazon SageMaker and AWS ML services can be used to build, train, and deploy ML models, and add intelligence to your applications. Amazon Redshift Spectrum and Amazon Athena have interactive querying, analyzing, and processing capabilities. Amazon Managed Service for Apache Flink is used to transform and analyze streaming data in real time.

Store unified customer profile information in Amazon OpenSearch Service (elastic search).

Build a single customer profile view with the help of identity resolution data coming from Amazon Neptune.

With Amazon API Gateway, you can expose developed APIs as microservices.

Activate the unified customer data and send it to internal and external parties.