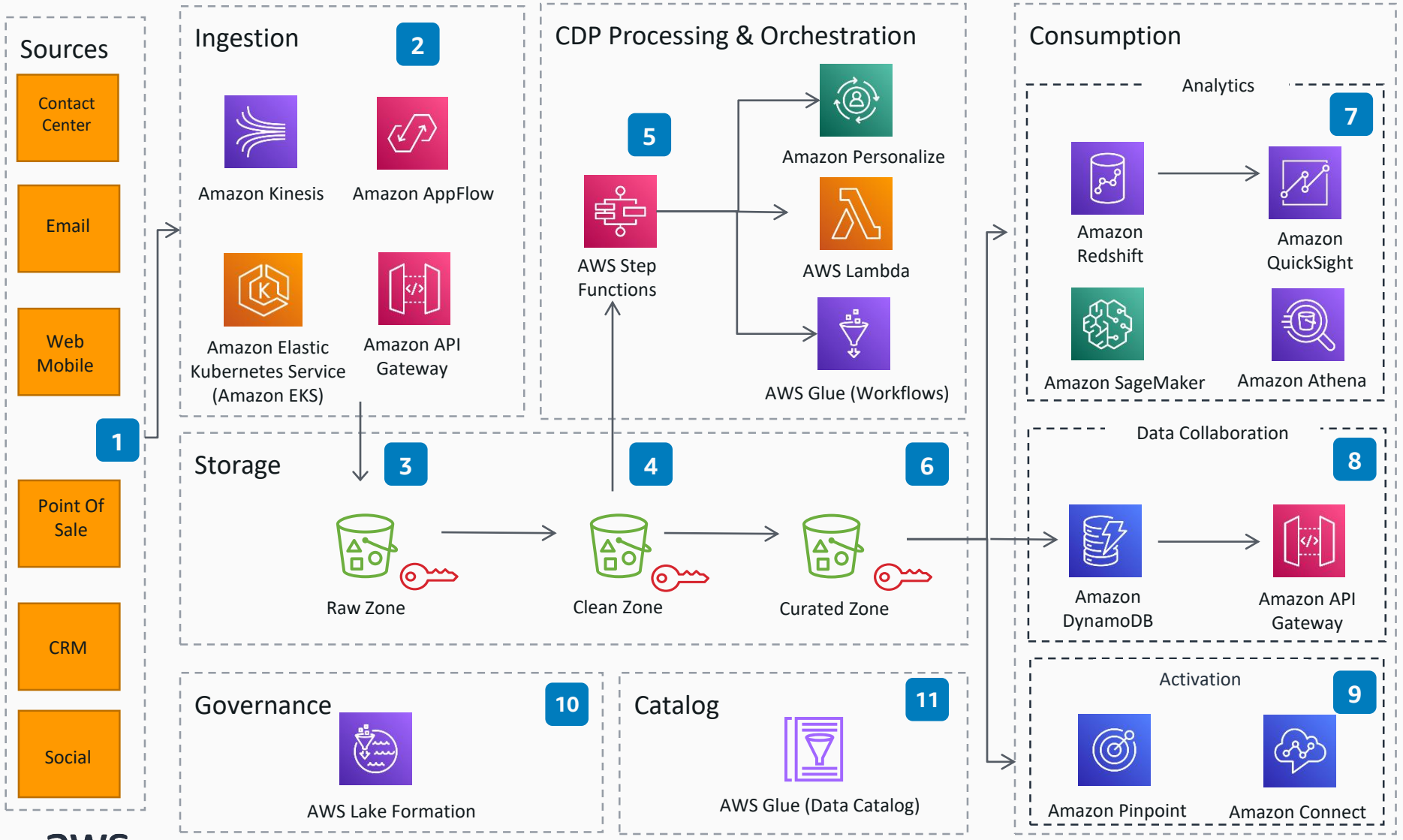
Guidance for Customer Data Platform on AWS

This guidance shows best practices for building a customer data platform covering data ingestion, identity resolution, segmentation, analysis and activation.



1. Source systems of customer data include customer interactions, clickstreams and call center logs.
2. Data from customer touchpoints is ingested into the marketing customer data platform (CDP) data lake using **Amazon Kinesis**, **Amazon AppFlow**, **Amazon EKS** and an **Amazon API Gateway**.
3. Ingested data is sent - in its original, immutable format – to an **Amazon Simple Storage Service (Amazon S3)** Raw Zone bucket
4. Raw data is then transformed into efficient data formats - such as Parquet or Avro – and moved to a Clean Zone **Amazon S3** bucket.
5. CDP processing and pipeline orchestration is conducted using purpose-built data processing components and transformation libraries through **AWS Step Functions** and then **Amazon Personalize**, **AWS Lambda**, and **AWS Glue**.
6. Data in the **Amazon S3** Curated Zone is now ready for post-CDP-processing consumption and is organized by subject areas, segments, and profiles.
7. The analytics layer uses **Amazon Redshift**, **Amazon QuickSight**, **Amazon SageMaker** and **Amazon Athena** to natively integrate with the Curated Zone for analytics, dashboards, ad hoc reporting, and ML purposes.
8. Customer data is then aggregated across platforms and published using customer APIs for consumption using **Amazon DynamoDB** and an **Amazon API Gateway**.
9. **Amazon Pinpoint** and **Amazon Connect** are used to activate multiple customer channels such as mobile push, voice, and email for targeted marketing communications.
10. Using **AWS Lake Formation**, fine-grained access controls can be enforced on catalog tables, columns, and rows on the data lake.
11. The resulting catalog in **AWS Glue** helps you manage both business and technical metadata, with versioning, at scale.

Reviewed for technical accuracy August 9, 2022

AWS Reference Architecture