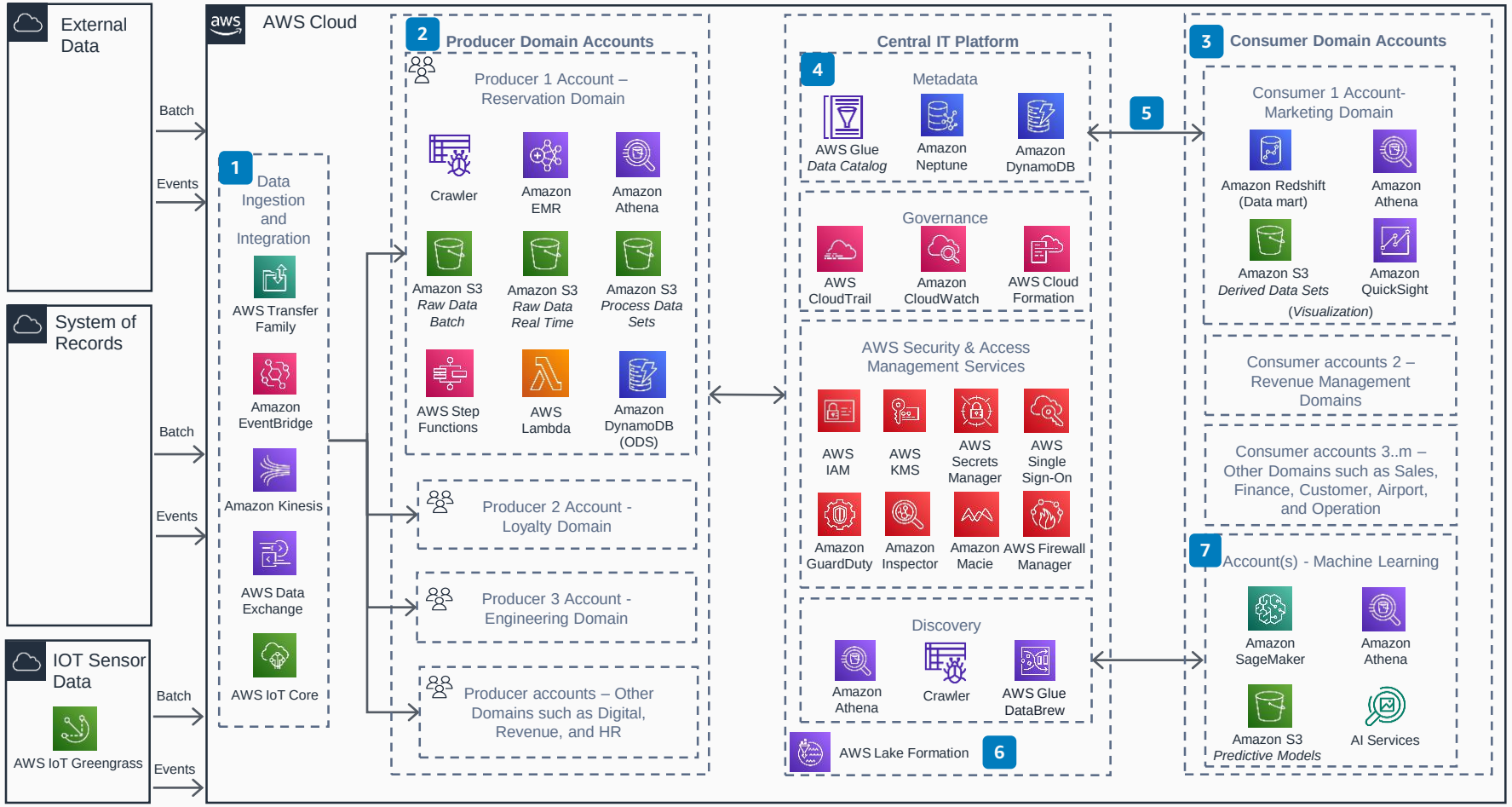
Implementing Travel & Hospitality Data Mesh

The travel and hospitality industries are facing new challenges when generating, accessing, and analysing data at scale. Use this new approach to build a data platform that serves both operational and analytical needs by using domain owned design, maintained data properties, open data standards, purpose-built databases, and extensible serverless architecture. This data delivery platform architecture helps relieve and eventually replace the on-premises data platform load, leading to cost savings and an agile environment.



1. Data flows into AWS through batch processing, real-time data, Secure File Transfer Protocol (SFTP) and IoT sensors.

2. Data sources are managed by the Business Domain. Producers use organization-level blueprints to provide core services such as security, governance, and open standards. Producers build the Operational Data Store using **Amazon DynamoDB** and **Amazon DynamoDB Accelerator (DAX)** to provide consistent single-digit-millisecond latency performance.

3. Consumer Domain uses process data sets from multiple producer domains based on business needs. Consumers build a data mart using **Amazon Redshift** and analytics using **Amazon QuickSight**.

4. Metadata is managed via multiple services. **AWS Glue** is used for data cataloguing. Data lineage is stored in **Amazon Neptune**. Data contracts are stored in **DynamoDB.**

5. Consumers use online, text-based search and filtering of data sets recorded in the central data catalog to collect data. They can search for relevant data by name, contents, sensitivity, or other any other custom labels you have defined.

6. **AWS Lake Formation** provides centralized management of security, governance and auditing policies including fine-grain permissions. It also enables automatic schema discovery and conversion to the required format.

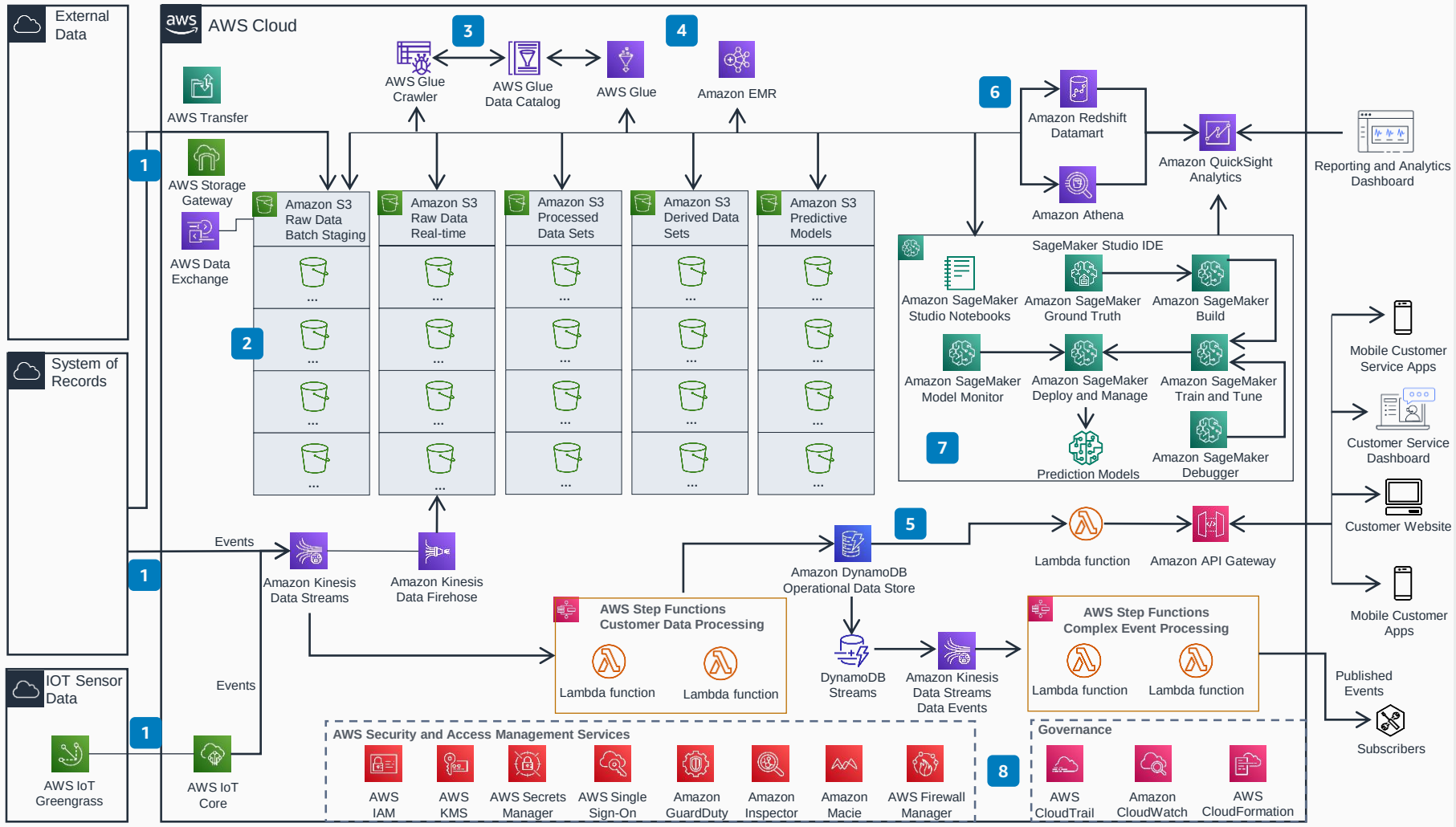
7. ML (Machine Learning) is performed using **Lake Formation**. Use **Amazon SageMaker** to provide standard AI/ML models for customer segmentations and lifetime value. Other AI services such as **Amazon Personalize** can be utilized to get actionable insight.

Reviewed for technical accuracy August 3, 2021

AWS Reference Architecture

Implementing Travel & Hospitality Data Mesh

A view of all of the services in the data mesh and how they connect and work together on AWS.



1. Data flows into AWS through batch processing, real-time data, SSH File Transfer Protocol (SFTP) and IoT sensors.
2. Provide staging for ingesting all batch and real-time data using cost effective storage classes in **Amazon Simple Storage Service (Amazon S3)**.
3. An **AWS Glue** crawler creates the table metadata in the data catalog. The **AWS Glue** data catalog is an index of the location, schema, and runtime metrics of your data. Use the information in the data catalog to create and monitor your ETL (extract, transform, and load) jobs.
4. Use open standards to build the data lake using the same data as the operational data platform. Use a read pattern schema to make the raw data and curated data readily available seamlessly for all user roles using a workflow such as Apache Airflow.
5. Use purpose-built databases such as **Amazon DynamoDB** and serverless architecture to deliver microservices and events to the operational data store.
6. Build all reportable data sets in **Amazon S3** and leverage **Amazon Redshift** and **Amazon Athena** for analytics. Optionally, for heavily used analytics, build data marts in **Amazon Redshift**. For ad-hoc requirements, use **Athena** to analyse data in the data lake with standard SQL.
7. Use **Amazon SageMaker** to provide standard AI/ML models for customer segmentations and lifetime value. Other AI services such as **Amazon Personalize** can be utilized to get actionable insight.
8. As a best practice, use a multi-account strategy to provide resources and security isolation for workloads. This helps with categorization and reducing blast radius.

Reviewed for technical accuracy August 3, 2021

AWS Reference Architecture