AWS Architecture for Big-Data Analytics on Streaming Data

Channels

Mobile / Web

Wearables

Echo

AWS Cloud

8 CloudWatch

9 Amazon SNS

Lake Formation

1 Amazon Kinesis Data Streams

2 Amazon Kinesis Data Firehose

Amazon S3   3   Glue

4 EMR

5 Amazon S3

QuickSight

6 Athena

7

---
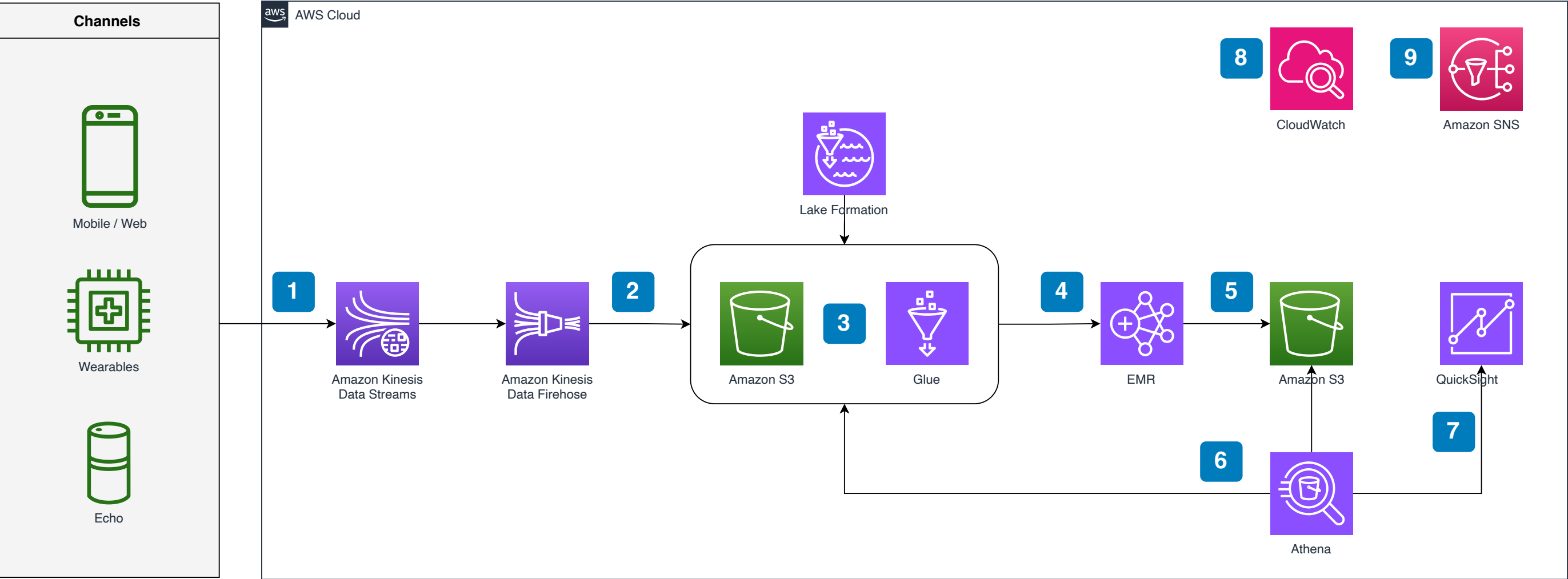
1 Data streaming from devices, websites, social media.

2 Data ingestion using Kinesis.

3 Lake formation using Glue crawler and data catalogue, and data storage on S3. Glue performs any preprocessing, e.g. conversion to parquet.

4 Data input from S3 to EMR. EMR performs big-data analysis using Apache Spark, PySpark etc.

5 Storage of the analysed data to S3.

6 Athena runs analytic queries on data stored in S3 before or after analysis.

7 Loading Athena query results to QuickSight for visualisation using dashboards.

8 CloudWatch monitors the AWS resources in real-time and stores logs.

9 Notification to users using SNS.