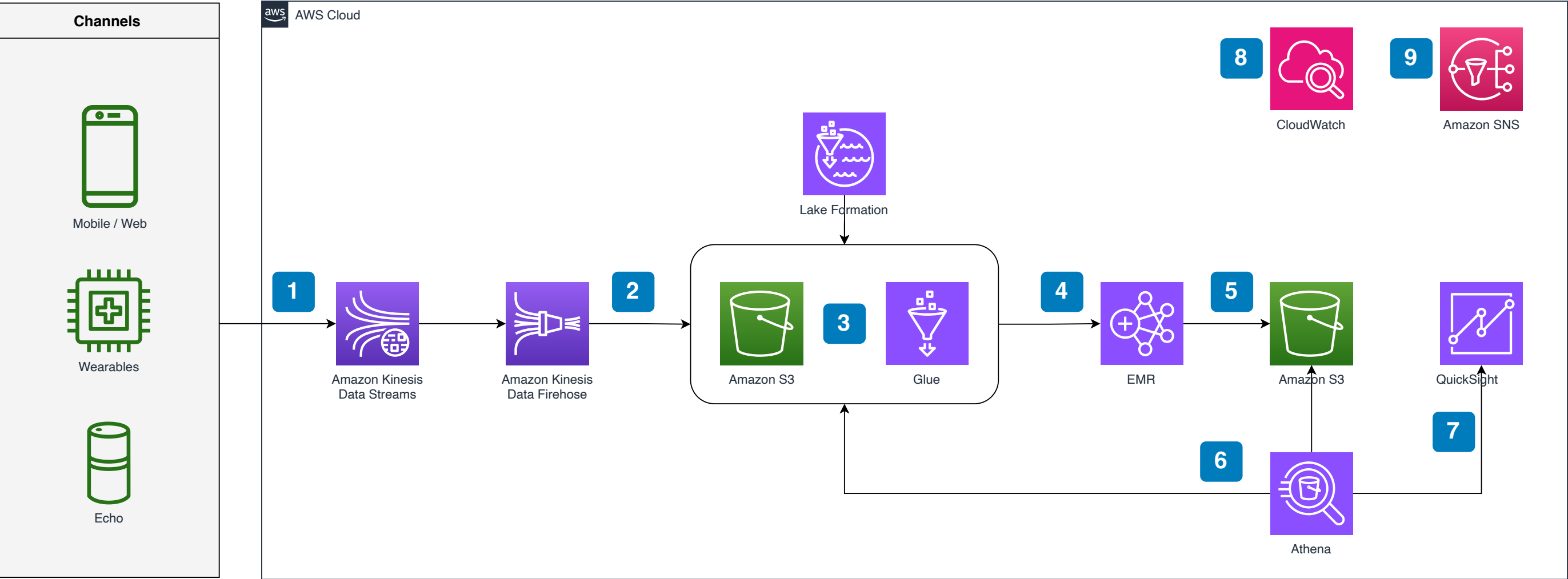


AWS Architecture for Big-Data Analytics on Streaming Data



- 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.