Data Platform Architecture

Medical Processing Company $\overline{\nabla}$ AWS Glue Catalog [Metadata] aws Data Lake Medical Facilities Glue Data Crawler 2 AWS Database Migration Service Kinesis Kinesis Spark 6 Transformed Redshift Data Firehose QuickSight Streams ibūd 🗥 AWS ibürl 🗥 Glue Cloudwatch Glacier Athena **Data Sources** Ingestion Storage Storage Processing Serving

- Amazon Kinesis Firehose can be used to ingest Medical Facilities Data in real-time
- On-premises data warehouses and OLTP data can be migrated using AWS Data Migration Services (DMS)
- 3 S3 can be used to store data with infinite scalability using Apache Hudi
- S3 glacier can be used to backup the data with infinite scalability
- AWS Glue Data Crawler uses crawlers to discover the schema of the data stored in S3/Redshift and catalog it with rich metadata
- AWS Glue can be used to perform data processing jobs at scale periodically using the open-source Apache Spark framework
- 7 The transformed data using Glue is stored in specific S3 buckets using the Apache Hudi format.
- The processed data can be stored in AWS Redshift (Data Warehouse) for efficient data queries at scale
- 9 Using Quicksight data from Redshift (or with Athena) can be visualized efficiently
- AWS Athena can be used to query the raw or transformed data in the