Big Data Pipeline Architecture

linkedin.com/in/scgupta in

Collect

data from internal & external sources

Ingest

data through batch jobs or streams

Store

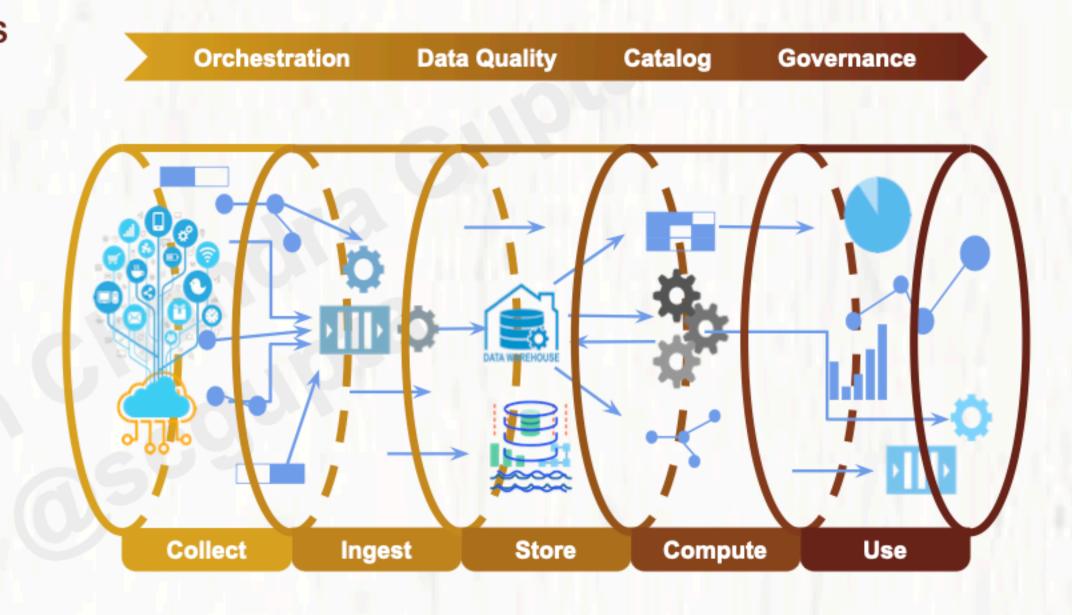
in Data Lake or Data Warehouse

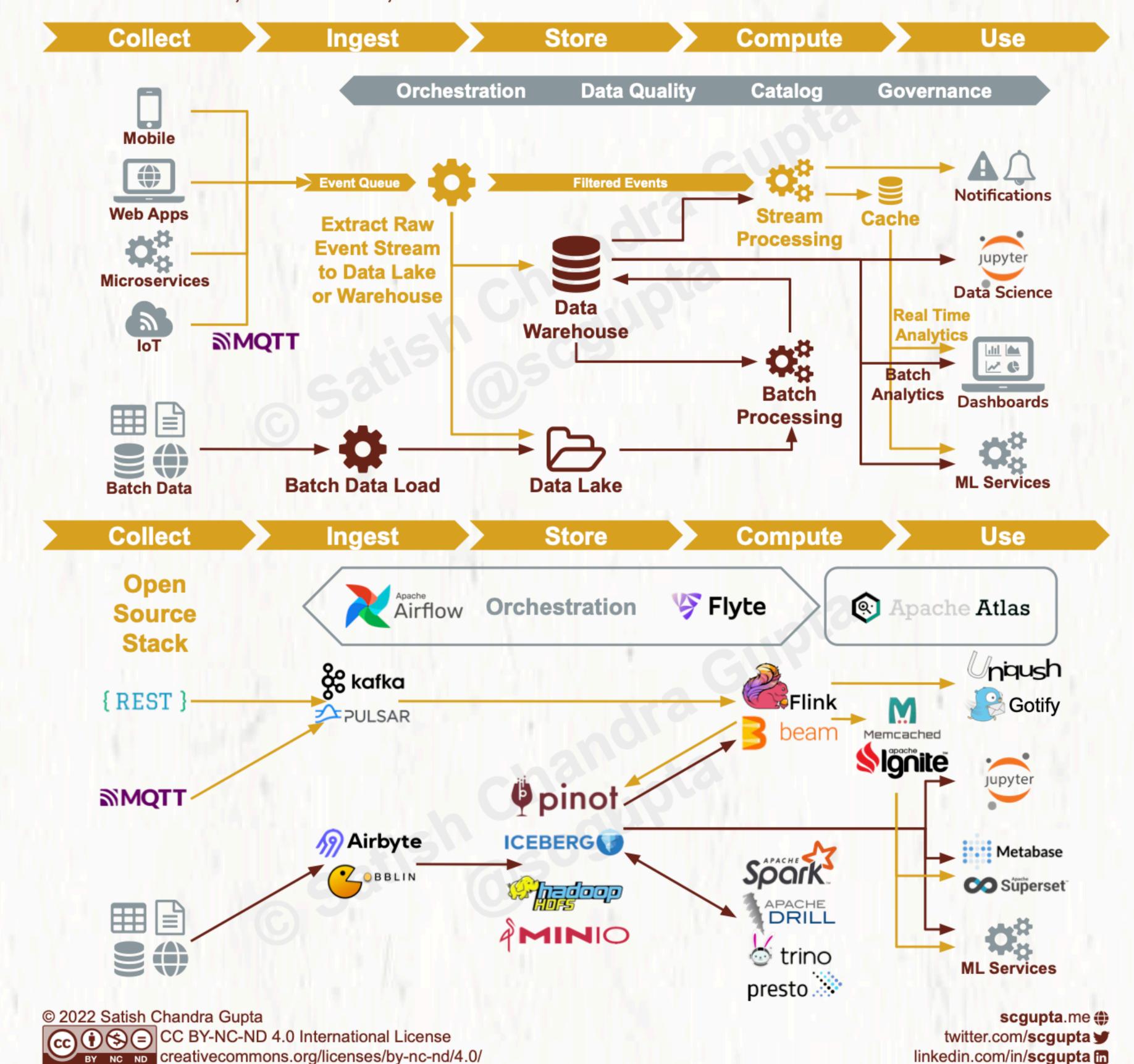
Compute

analytics aggregations, ML features

Use

it in dashboards, data science, ML





Big Data Pipelines on AWS, Azure, and Google Cloud 💁 ml4devs.com/big-data-pipeline aws Orchestration Glue Studio Data Pipeline Glue Catalog ₹. **Amazon SNS** Lambda **Kinesis Kinesis Data Streams** Data Analytics Elasti-Cache SageMaker RedShift **AWS** IoT Core QuickSight DataSync S3 Athena Glue ETL Lambda **EMR Collect Store** Compute Use Ingest **Orchestration Azure Data Factory** Oozie on HDInsight **Data Catalog Purview** Notification Azure **Event** Hubs **HDInsight Function** Hub **Azure Cache** for Redis Azure ML Synapse Azure IoT Hub Power BI Azure **Azure** Azure Data **Databricks Data Factory** Lake Storage Azure Synapse Function Collect **Store** Use Ingest Compute **Orchestration** Cloud Data Fusion **Data Catalog** Cloud Composer Dataplex Cloud Cloud Messaging Pub/Sub DataFlow **Function** Memorystore r on Cloud Vertex AI **BigQuery** IoT Core Looker Storage Transfer Cloud Storage Dataproc Service Cloud DataPrep **Function** © 2022 Satish Chandra Gupta scgupta.me (*) CC BY-NC-ND 4.0 International License twitter.com/scgupta NC ND creativecommons.org/licenses/by-nc-nd/4.0/ linkedin.com/in/scgupta in