Transportation and Logistics Data Lake

Enable transportation and logistics companies to break data siloes and create a single repository for their data from multiple sources. This fully managed, pay-per-use and scalable architecture removes the complexity and costs of managing and synchronizing on-premises databases, establishing a unique source of truth for the business.

1. Ingest data from multiple sources: historical, large DB with AWS Snowball, real-time data with Amazon Kinesis, ERPs data synchronization with AWS DataSync, and mobile app data with AWS AppSync.

2. Automatically extract, transform, and load raw data with AWS Lake Formation, discovering schemas and creating a data catalog for further data exploration.

3. Create an Amazon S3 data lake where your business-critical data can be stored, organized, and accessed in the most secure, resilient, and cost-effective way. Use Amazon EMR to gain fine-grained control over your extraction, transformation, and loading (ETL) jobs, and AWS Glue to manage your data catalog.

4. Turn raw data into actionable knowledge with Amazon Athena and Amazon QuickSight, disseminating key business intelligence insights across your organization. Use Amazon SageMaker to deliver machine learning (ML)-based predictive and prescriptive analytics.

5. Use AWS fully managed databases to deliver specific capabilities: Amazon Neptune graph database enables network optimization, Amazon Aurora runs queries five times faster than standard MySQL databases, and Amazon DynamoDB supports interactive mobile and web apps with real-time updates. Avoid data latency with live data stream ingestion from Amazon Kinesis.