

# Leveraging Improved Speed & Immutability of a Data Lake

Customer Case Study - An automobile insurance company



## Business Need

The client's Enterprise Data Warehouse (EDW) solution was cumbersome, slow, and adhered to a stringent set of schemas. Their executive team partnered with Syntelli to provide a way to ingest data from various sources & house it in an easily accessible repository. Following this, they needed to transform the data, storing business-critical metrics & KPIs in data marts, which would then be utilized by their business users for analytics, visualization, & advanced data science activities.



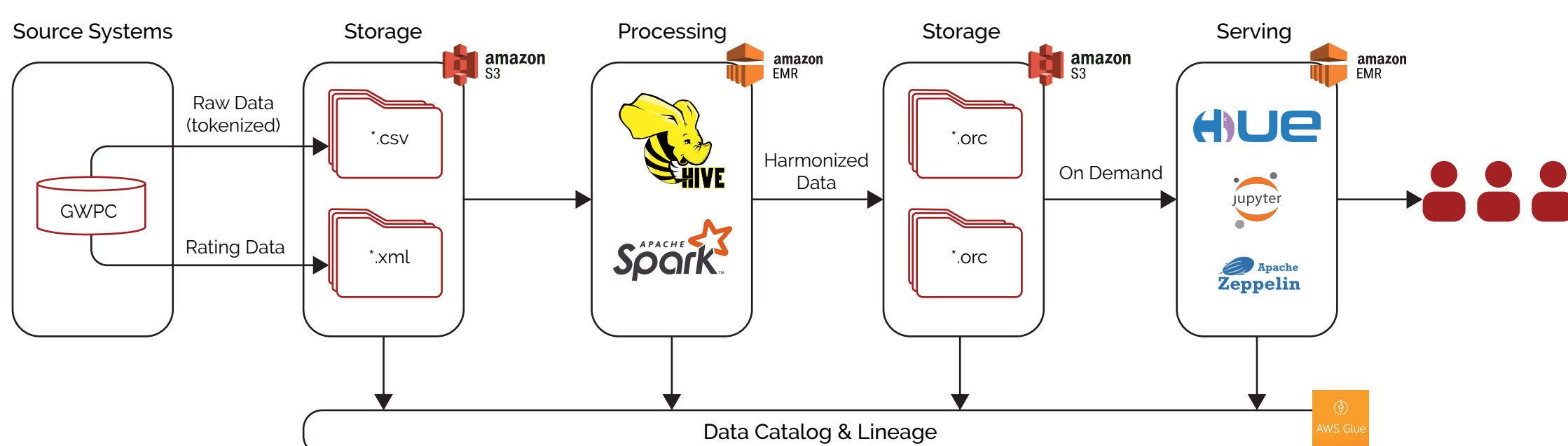
## Customer Challenges

- Very slow existing system; updating dashboards would take upwards of 19 hours
- Not easily accessible, required business users to perform multiple complex joins to find the right attributes
- Difficulty in maintenance
- Very strict schema; unable to adapt to the dynamic schemas inherent in the sources



## Solution Provided

- Modernization of data pipeline, leveraging the improved speed and immutability of a data lake instead of a traditional structured EDW.
- Automated and scheduled the process. End to end, and before extensive optimization, the process ran in less than 5 hours and was able to provide fresh data by the start of business.
- Rolled out this improved Data Strategy & Methodology across all business lines.
- Proposed a Lambda architecture which would allow initially 5-minute micro batching of data, and eventual real time streaming.



## Impact of Solution

- More flexible schemas & easy to consume for data science models in SAS & for visualization
- Easy to maintain
- Secure
- Reduced processing time
- Search capability on the metadata
- SLA under 5 hours End-To-End



## Technology Used



S3, EMR, glue, EC2, Athena, IAM Security



HiveQL, HDFS, Hue, Zeppelin, Spark, Scala

- Other: Crontab, Nifi, Talend, Airflow, PostgreSQL, Cassandra