

Building the ML Models for car components health status using various AWS Services

Maruti wants to build a Health Monitoring System for various components/parts of car manufactured to prevent defective parts to pass through to customers. This will help to reduce cost of vehicle call backs and subsequent repair.

This includes ingestion of 20 TB data from different cloud providers to AWS cloud and performing preprocessing on 20 TB once and then 2 TB monthly data. After ingestion to do analytics on data and build machine learning models and storing data in S3.

Maruti Suzuki India Limited (MSIL), a subsidiary of

Maruti Suzuki India Limited (MSIL), a subsidiary of

Suzuki Motor Corporation, Japan, is India’s largest passenger car maker. India’s first Company to produce and sell more than a million cars in India in a year, Maruti Suzuki is credited with having ushered in the automobile revolution in the country. The Company, formerly known as Maruti Udyog Limited, was incorporated as a joint venture between the Government of India and Suzuki Motor Corporation, Japan in February 1981.

**About Maruti**

**About Challenge**

**AWS Services used**

Amazon EMR Amazon Amazon API Amazon Simple Amazon Amazon AWS Glue Amazon

SageMaker Gateway Storage Service (S3) Lambda CloudWatch Quicksight



A



* Data collected & ingested from various sources for storing semi-structured data.
* Transformed old csv and json data into parquet format and did partitioning.
* Running Amazon EMR on incremental data daily, partitioning and storing it in parquet format for preprocessing.
* PCA (Principle Component Analysis) is applied on data set for feature reduction and selection.
* After that Machine Learning algorithms were used to predict car component health status using modeling techniques like Linear Learner, XGBoost etc.
* Did hyper parameter tuning for XGBoost using SageMaker Hyperparameter optimization (HPO)
* Deployed the ML model using SageMaker endpoint.

**Proposed Solution**







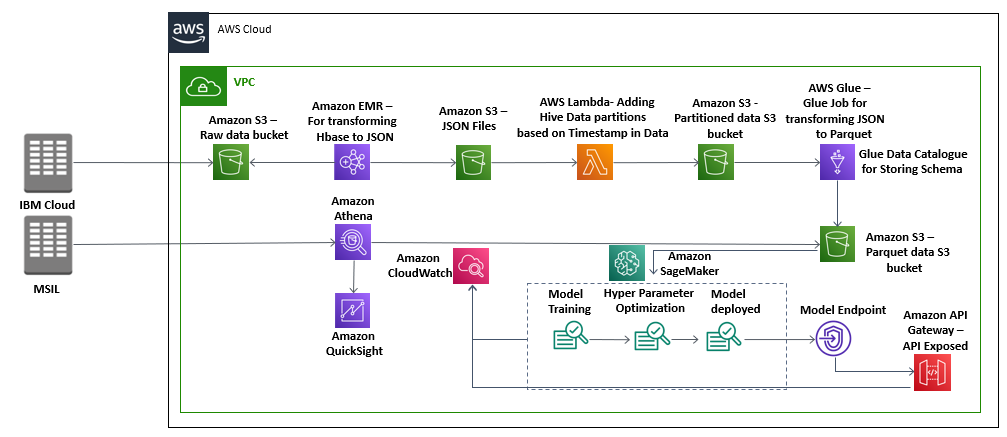
**Solution Outcome**

Increased profit due to fewer number of call backs rate for repair after sale

Reduction in production time due to timely detection of defective parts

Overall customer satisfaction increased which resulted in many intangible benefits like it helps to increase trust in brand which eventually leads to increased customer loyalty.

**Architecture Diagram**



**How AWS services helped in building the ML Model for car component health status**

**Amazon SageMaker**

Amazon SageMaker is used to create and manage Jupyter notebooks that were used to prepare and process data and to train and deploy the machine learning models.

Its high-power GPU Instance used for training of XGBoost Model which is optimized using Hyper Parameter Optimization Service followed by its deployment through SageMaker built-in deployment service

**Amazon Athena to query the partitioned data for insights**

It is an interactive query service that made it easy to analyze the output parquet data stored in Amazon S3 using standard SQL. It is serverless, so there is no infrastructure to manage, simply point to your data in Amazon S3, define the schema, and start querying using standard SQL and the results were delivered within seconds.

**AWS Glue to load partitioned data**

It is a fully managed extract, transform, and load (ETL) service that made it easy for us to prepare and load the processed, partitioned data for analytics. One can create and run an ETL job by simply pointing AWS Glue to the data stored in S3 bucket, which is then immediately searchable, query-able, and available for ETL.

**AWS Lambda to handle the backend API calls**

It helped to initialize and validate the input and acted as the backend of the whole task. AWS Lambda lets us run code without provisioning or managing servers. Also, it helped to connect with various AWS API’s to acquire various insights from the inputs.

**Amazon S3**

It is an object storage service that offers industry-leading scalability, data availability, security, and performance. In this solution it. It is used here tostore JSON and CSV raw documents and the output parquet files.

**Amazon EMR for data transformation**

It is a tool for big data processing and analysis, Amazon EMR is based on Apache Hadoop, a programming framework that supports the processing of large data sets in a distributed computing environment. Once the transformation of data is done then business analytics is done using Athena and Quick Sight

**Amazon Quick sight for visualization through dashboard**

It is a fast, cloud-powered business intelligence service offering by AWS which here made it easy to deliver insights related to the output data like TCU per day etc.

**About the Partner**

**MothersonSumi INfotech &Designs Ltd.**

MothersonSumi INfotech & Designs Limited (MIND), a SEI CMMI Level 5 IT services company and the IT back bone of Motherson group. MIND is a trusted technology partner to over 200 clients globally. Our value proposition is in our strength in specific Industry segments and years of experience in the areas of intelligent warehousing, Supply chain enablement, software application development, smart ERP customization, infra managed services, cloud, IoT & Analytics. MIND is serving customers in 41+ countries with a strong team of 1500+ professionals.