# IATA Data Engineer Case Study

### DESCRIPTION

Imagine it’s a Tuesday and IATA’s Economics department (one of your main business stakeholders) is reaching out with a request. They just acquired a new dataset from an external provider and they need to load it into the data lake ASAP to perform an analysis across several datasets including this new one.

The request is quite urgent as the conclusions are needed for a meeting IATA’s Director General will have with the French Minister of Transportation at the end of the week.

After talking to your technical lead, she advised building an AWS Lambda based data pipeline to facilitate the industrialization of the process. She confirmed you will need to create a new Glue database and S3 bucket for the metadata and data.

### DELIVERABLE

The dataset in question is available here - <https://eforexcel.com/wp/wp-content/uploads/2020/09/2m-Sales-Records.zip>

The solution should be a collection of Lambda functions working together to fetch the file from the https endpoint, un-compress it, convert it to parquet, partition the data on S3 by the content of the Country field and define all relevant resources to allow then the querying of the final dataset via Athena. Original file should be archived under another prefix on the same bucket. The solution should be deployable via a CloudFormation template that contains all the relevant AWS resources.

Please share a public git repo with the code in advance of the presentation.

### EXPECTATIONS

You will present your work during a debrief session, with Q&A. The audience can be composed of a mix of technical and non-technical staff. You are expected to demonstrate your proficiency in raw data acquisition from various sources, and your data engineering skills in properly structuring and building datasets, and your industrialisation skills in designing, deploying and maintaining the full data processing pipeline.

Please feel free to organize your presentation in any way you feel best. There is no expectation for slides (unless it will help you). We will try to determine your familiarity with Python, relevant AWS services, how to break the logic into multiple Lambda functions and general data engineering practices.

Please elaborate on what things you would do differently in case you had more time.

The expectation is to deliver a solution using the defined services (Lambda, CloudFormation etc.) While proposing alternative solutions is acceptable, it is not expected, and it should be done only as an extra to the main solution.

The IATA team will also present their own solution to the case study at the end.

### DISCLAIMER

This is not a real-world scenario; it does not reflect actual IATA datasets or requirements. The solution to this case study has no real-world value to IATA except in determining the suitability of the candidate.