

Amazon's Data Lake Project for Import Export Regulations

Speaker 1 (00:01)

Okay so this is a project that I worked on when I was entering at Amazon luxem this project was about

Speaker 2 (00:12)

About

Speaker 1 (00:14)

Data Let's say quality and completeness

Speaker 3 (00:19)

So here's the

Speaker 1 (00:20)

Background of the

Speaker 2 (00:21)

The entire thing

Speaker 1 (00:23)

So Amazon does a lot of import export between countries to you know maintain its supply chain and U knoc has just breakted from the EU so there were a lot of regulations that were Put up between the entire

Speaker 3 (00:42)

UK and you

Speaker 1 (00:44)

And Amazon had to file all their import export duties you know correctly between these 2 countries since it was new Amazon was catching up with you know all the import export the regulations and all the filings that they had to do so that they don't pay any fines or at least to reduce the amount of funds that they had to pay and to do this they were working with multiple brokers bit

Speaker 3 (01:11)

Green E and UK

Speaker 1 (01:13)

Like ups was one of the major brokers then they had DHL

Speaker 3 (01:17)

We were working with Shiva then ckb ck BD

Speaker 1 (01:23)

Charles in something yeah Charles can tell yeah Charles candle and then there was also

Speaker 3 (01:32)

Shiva I already said

Speaker 1 (01:34)

There was Kerry logistics and so on then there was an post Swiss post you know these were different brokers that they were working with To basically do the import export they were using them for their import export and these brokers

were filing the import export Paperwork for Amazon's input export Operations and my project here so Amazon was basically working on building a large data lake where they were collecting all this import export filing data from the brokers they were actually standardising and regularising the data that was coming in so based on the government regulations based on what the brokers were filing and there was a document called the S.A.

Speaker 1 (02:23)

single agreement document which was basically any broken had to file For the import export file right date is filing and Yeah so my role in this project was to talk to all the different external import export group is that Amazon was working with Understand how they are maintaining their data regarding the import export filings what are the fields what are the columns what are the data types

Speaker 2 (02:56)

So

Speaker 1 (02:57)

Was very deep dive into the data they were filing

Speaker 2 (03:00)

And

Speaker 1 (03:00)

And Amazon had a regularised more standardised data format that they were going to you know maintain in their

data lake so I had to speak to these external brokers understand their data talk about what how we want you know their data in our system and then I had to analyse the data that they already have and how we need it then I had to create basically business rules and technical data validation rules data check rules and data completeness rules to ensure that you know when we create AB to pipeline from their enterprise data systems to R and R enterprise data system the data was actually coming in you know more regularised and standardised in the way Amazon wanted it and if there was any issues then we create an alert to the broker's data team saying you know we are having some issues with the data that you sent and can you please regularise that our sent in a proper format or also we were writing transformation Scripts in our pipeline so if the data can actually be transformed from by

Speaker 1 (04:22)

us you know while being exchanged from there enter system to current system then we transform It on the way before we Put It into our system so my work here in this project was to analyse Deep dive understand the data from both the ends and then It was to write business rules and data validation check rules both precommit and postcommit check rules to regularise and standarze the data so that these rules can be you know kept a checkpoints in the bt data exchange pipeline and automatically the ETL process itself was process where the data were standardised on the go and you know stored into the Amazon data link so here if you see I mostly worked with quite a lot of external brokers getting on

calls with them getting on calls with their team explaining to them how we wanted the data explaining to them what they already have in their data analysing their data and telling what kind of changes we would like you know and especially there was a lot of data that was missing so we spoke them to

Speaker 1 (05:36)

see if we could get that information correctly you know because they were not having that data already or they were not at least extracting the data so we told them which document and where from where they could extract the data and if they could send us that data as well and so on and yeah so my role in this project as a business intelligence engineer was more on a business level and a data technical level where I was creating Jason schemas standarising schemas writing rules for the schema and writing you know rules for data validation data quality data completeness and all of these things and make Ensure that they were integrated into the bv data exchange pipelines so this was one project that I did in Amazon luxboon

Data validation proj

06:24

