

Jack Sargeant

From: David Mason
Sent: 13 February 2024 10:56
To: Jack Sargeant
Cc: Jamie Atrill; Daniel Whitney
Subject: RE: Data Transfer Project

Hi Jack,

Please note that the decision to transfer the test data to the UK LabVIEW database is to harmonise the UK test data storage method with the other global sites. The global IT team carried out a project last year to copy the data overnight from all the LabVIEW databases for all sites into a global database for global BI dashboard reporting etc. The Global database is under control of the Spellman US IT team and if cloud storage can be used this will be their responsibility to implement.

Kind Regards,
Dave

From: David Mason
Sent: Friday, February 9, 2024 10:31 AM
To: Jack Sargeant <jsargeant@spellmanhv.co.uk>
Cc: Jamie Atrill <jatrill@spellmanhv.co.uk>; Daniel Whitney <dwhitney@spellmanhv.co.uk>
Subject: RE: Data Transfer Project

Hi Jack,

Thanks for your email, I agree your proposed updates to SSATS and look forward to seeing how it progresses and evolves. Please keep up the correspondence with the global automation team and IT when necessary. Our team will support and advise you as required.

It's an important point you raise about traceability for transfers, we have seen issues with this before so I'm keen to see your solution.

Kind Regards,
Dave

From: Jack Sargeant <jsargeant@spellmanhv.co.uk>
Sent: Friday, February 9, 2024 9:26 AM
To: David Mason <davem@spellmanhv.co.uk>
Cc: Jamie Atrill <jatrill@spellmanhv.co.uk>; Daniel Whitney <dwhitney@spellmanhv.co.uk>
Subject: Data Transfer Project

Hi Dave,

I am e-mailing you to confirm some of the details in regard to our outstanding project to transpose test data from our SSATS database to the LabVIEW database.

From what I understand, the US are requiring that all test data globally is to be collated centrally. This is twofold. Firstly, test results or "Overall Verdicts" derived from test data are to present alongside other serial number specific information in Glovia (ERP system), And secondly the US require such to enable data driven initiatives outlined in the last global meeting, by means of data lakes and cubes.

Whilst this both completely feasible and possible, I have identified a few areas of concern. Firstly, it appears that data is written to our SSATS database in two different locations, SSATS itself and Squirrel. To ensure all data remains dynamic, and with the interest of utilising Power BI for ourselves, I would suggest that we make the transposition of data dynamic, rather than scheduled. This will mean that we will more than likely need to trigger movement events at the point of data inception to the SSATS database. I am proposing the development of an ETL (Extract Transform Load) tool, driven by a API (Application Programming Interface). This will mean we can use the same stack for both data write points, and only have to work with one project rather than developing bespoke solutions for each. Whilst it would make sense to drive data transposition using triggers on the SQL database, I personally feel, and I am sure the rest of the team does, that the volatility and configurations of our databases has historically is cause for concern, and I would recommend that we take this action into our own hands, rather than leverage IT.

There are a couple of more things I would like to add, that I feel will bring both pertinence to the project, and the data we collect. Firstly, I believe that data sanitisation is a crucial element here. We have historically witnessed the poor, erratic state of the data written to our database, and the transposition would be a fantastic opportunity to sanitise, clean and reinforce all data collected. This will enable us to create far more accurate and influential data visualisations. It will also give us far more flexibility within Power BI, without having to articulate its own resource heavy sanitisation features. I also feel that we will require some form of tracability for transfers, a receipt almost, that will enable us to see what data has/has not been transferred. Whilst it would be possible to simply cross compare both datasets, doing this will **not** flag data that was unable to be moved. Cause of which could be anything from network problems, or nomenclature derived issues. In recording this, we could see what problems are arising, what kind of problems are arising, and allow us to fix them where appropriate. This data could also be written into a web application so we can see statuses, or even be used in appropriate data visualisations.

If there are any other requirements of this project, or areas we should be conscious of, please could you let me know, so we can convene and plan appropriately. If also appropriate, please could you direct me to members of the global test team for requirements we may have overlooked, or not currently be aware of.

The project will be outlined using a Project Initiation Document, once I am happy that all relevant requirements have been identified. I will then proceed to distribute such to all stakeholders. Following your approval.

Regards,
Jack