

Raw Data [Client Specific]



... identifying the raw data from clients and means of retrieving such data

Data Location

- SFTP Server: A-EUW-MGF05-H3A
 - System: Landing Zone
 - File System: import directory
- App Server: A-EUW-GEN02-H3A
 - System: pControl Application
- DB Server: A-EUW-GEN03-H3A (GEN_PROD_DB)
 - System: SQL Database (110 GB for Generali) and Windows Server OS (per client)
 - Access: SSMS database client and remote desktop to A-EUW-GEN03-H3A\PROD\data\import (or \archive)

Data Identification

- raw input data are located in the **archive folder** inside the *DB server*
 - data are files in CSV and XSL format
 - file name formats are various depending on client, file name prefix and extension need to be same as defined
 - files can have data combined from more than 20 TPAs (3rd party admins)
- Generali provides their files in the below folder structure (subfolders of archive folder)
 1. asset_transaction
 2. asset_transactions_compare
 3. benchmark
 4. capital_flows
 5. external_prices
 6. financial_statement
 7. fx_rates
 8. hedge_ratio_data
 9. initialisation
 10. investment_holdings
 - U100_GIE_Holdings_20210625.xls
 11. nav_pack
 - NAV_PACKXXX.xlsx
 12. pending_settlements
 13. security_prices
 14. State Street
 15. static_manager
 16. StaticData
 17. tiralbalance
 18. unitised_cashflows
 19. unitonissue
 20. Viteos

Data Extraction

private subnets in client VPC

- **generali-production** to **generali-ai-data-subnet** private subnet within the h03 VPC in mshv1 account
 1. GTS to create a DB Server (Windows Server OS) in the generali-ai-data-subnet
 2. GTS to copy data from the archive folder inside DB Server in generali-production to archive folder inside the DB Server in generali-ai-data-subnet

private subnets between client VPC and mg VPC

- **generali-ai-data-subnet** to **mg-ml-private subnet**, which is a connection between private subnets from h03 VPC in mshv1 to mg-ml VPC in mg-ai environment
 - GTS to create a peering connection between the two VPCs
 - to allow resources in mg-ml-private subnet access the resources in the generali-ai-data-subnet
 - this is to enable querying the data in generali-ai-data-subnet and also transferring them to the mg-ml-private subnet for data analysis purposes

Data Filtering

- AI&DT to speak to Professional Services and GCS how to filter the input raw data for the selection of specific funds

Data Analysis

- AI&DT to setup Amazon SageMaker Studio with Data Wrangler or database administration tools (SSMS) with a connection to Amazon RDS (Aurora, PostgreSQL, etc.)

Next Steps

1. data extraction
2. data filtering
3. data analysis