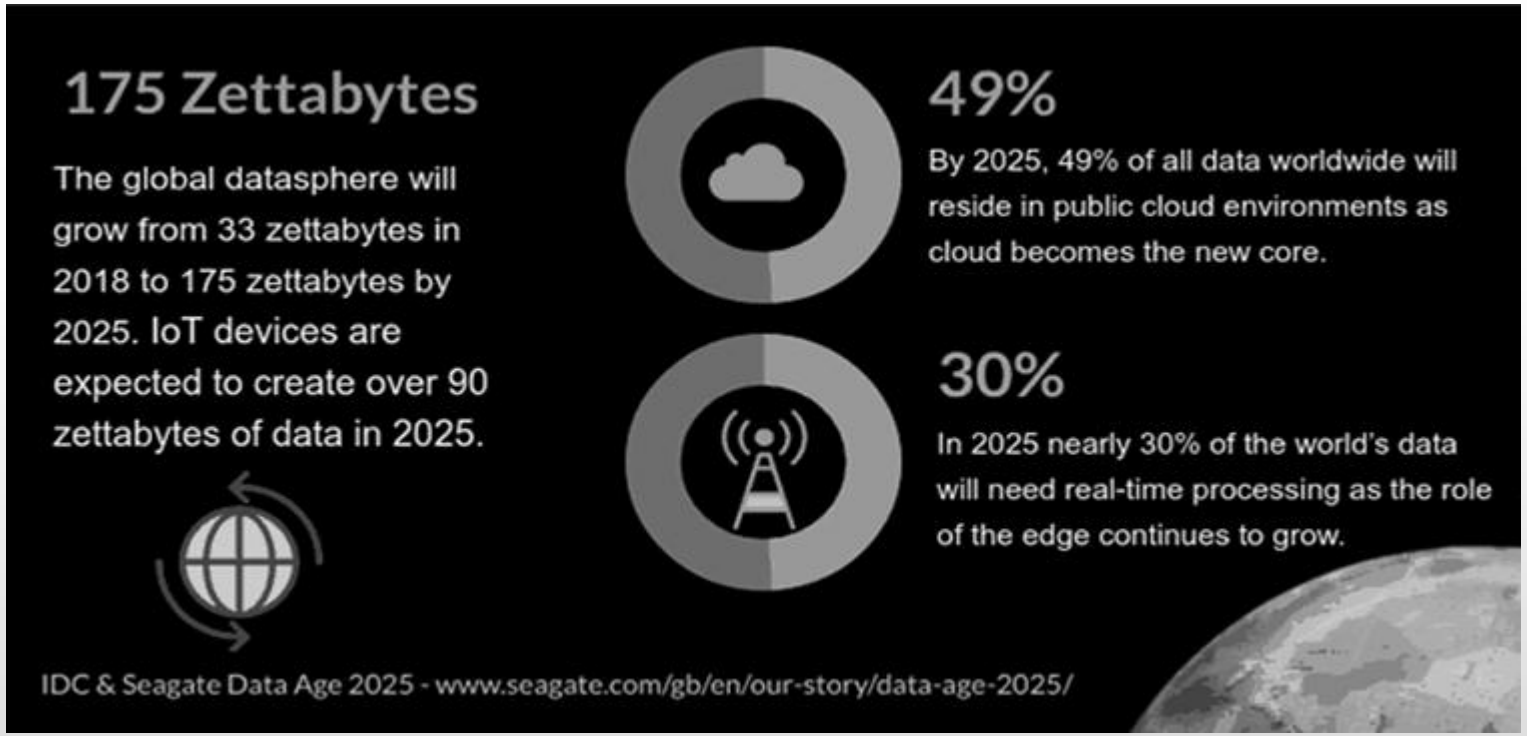

maxgeo
data integration & automation
26 NOVEMBER 2021 - Harare

overview

- the global datasphere
- benefits of automation
- maxgeo standard automation/integration
- maxgeo custom automation/integration examples
  - SQL to SQL
  - data warehousing
  - API to SQL
  - AWS S3 to SQL
- conclusion

the global datasphere

data volume and rate of accumulation



"In 2020, 64.2ZB of data was created or replicated, defying the systemic downward pressure asserted by the COVID-19 pandemic on many industries and its impact will be felt for several years. The amount of digital data created over the next five years will be greater than twice the amount of data created since the advent of digital storage.“ - Dave Reinsel, senior vice president, IDC's Global DataSphere.

*IoT = internet of things

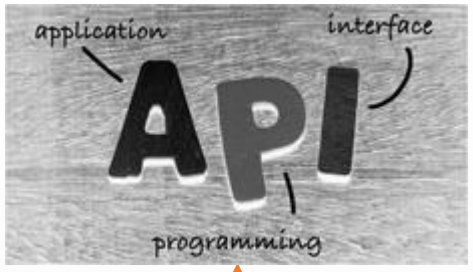
the global datasphere a changing landscape

sources

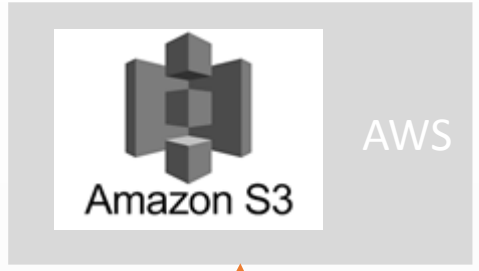
formats

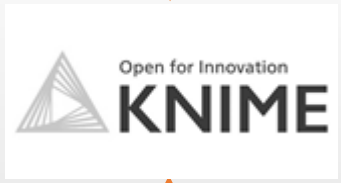
integration

database



or





the global datasphere

automation

- For these kinds of sources, we need to automate
  - the process of loading data 
  - the validation of data 
  - the reporting of data 

the global datasphere

the role of the DBA

- What is the impact on the role of the DBA?
- In the case of these more voluminous datasets, it is changing from a data input role to one of oversight of the automated tasks
- Dealing with exceptions identified by these processes and resolving them
- Finding better ways to analyze all this data through improved interpretation techniques

benefits of automation

Reduced Risk

Data transfer is secure and accurate

User self-help

Users can subscribe to the reports and dashboards they want to see

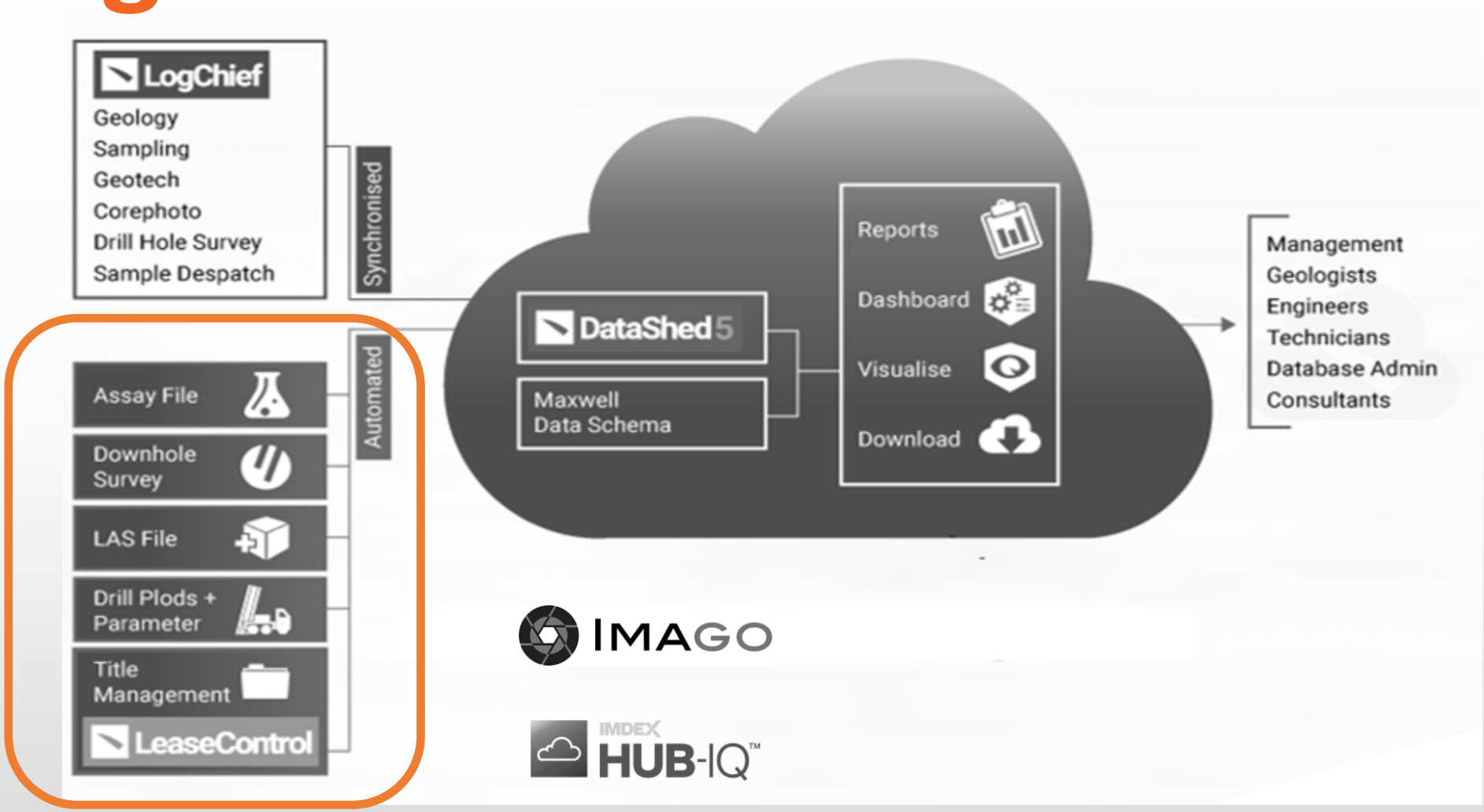
Efficiency

Once set up, data transfer is fast and efficient

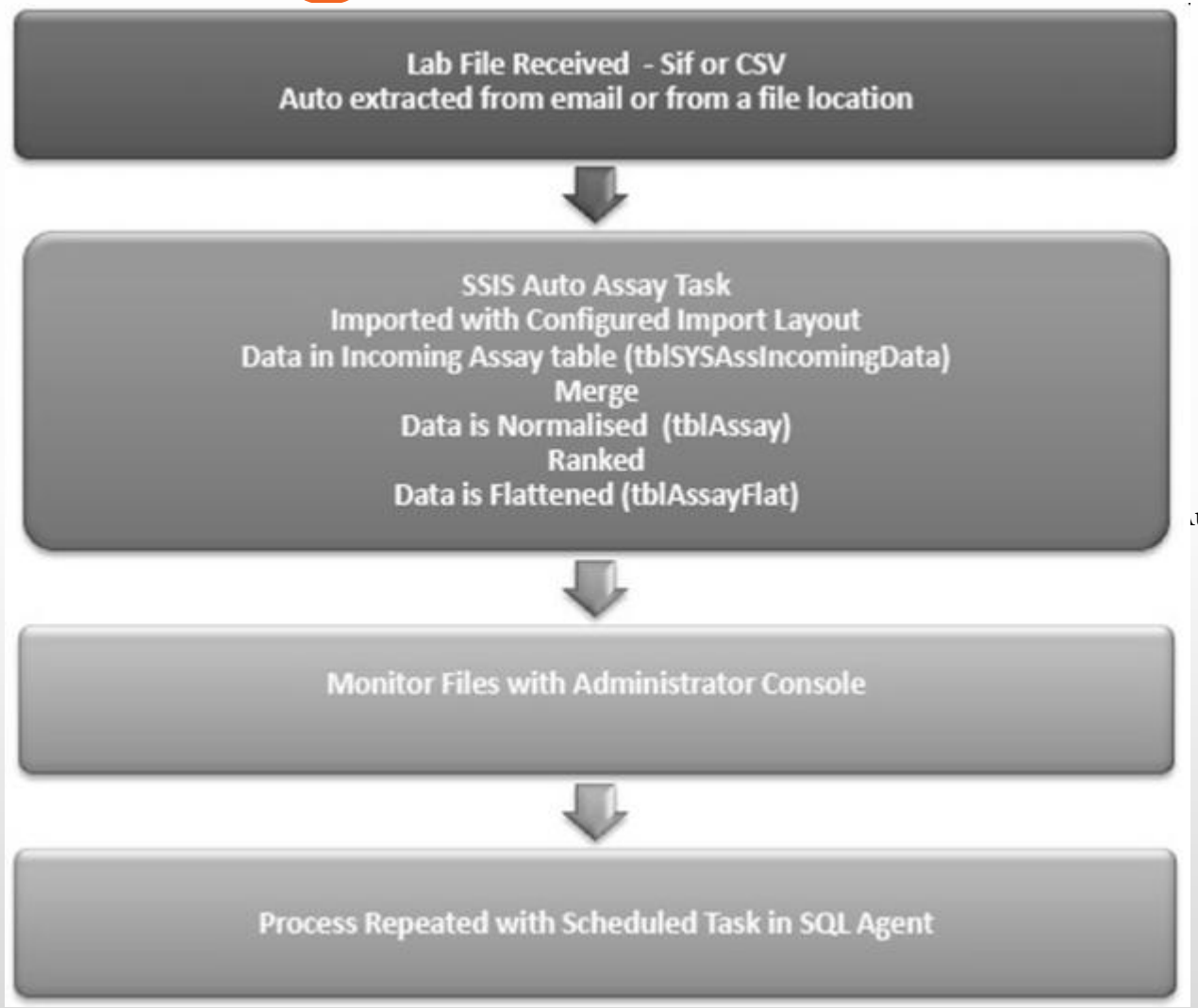
Integration

Bring data from various sources together to provide real-time consolidation

maxgeo automation



maxgeo automation



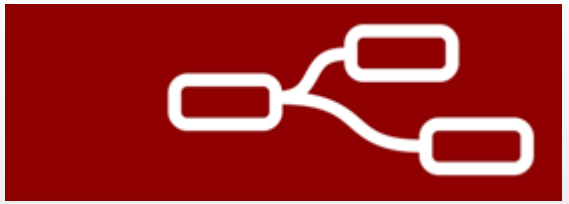
utomation of assay loading, QC checks, de-surveying, significant intercepts and expor

Batch Status	Sample Count	Samples Merged	Results Count	Results Merged	Active QC Policy
Accepted	150	150	278	278	IRR_QCPolicy_Au
Accepted	17	17	30	30	IRR_QCPolicy_Au
Unmerged Assays	28	0	1369	0	IRR_QCPolicy_Au
Accepted	248	248	465	465	IRR_QCPolicy_Au
Accepted	250	250	483	483	IRR_QCPolicy_Au

custom integration

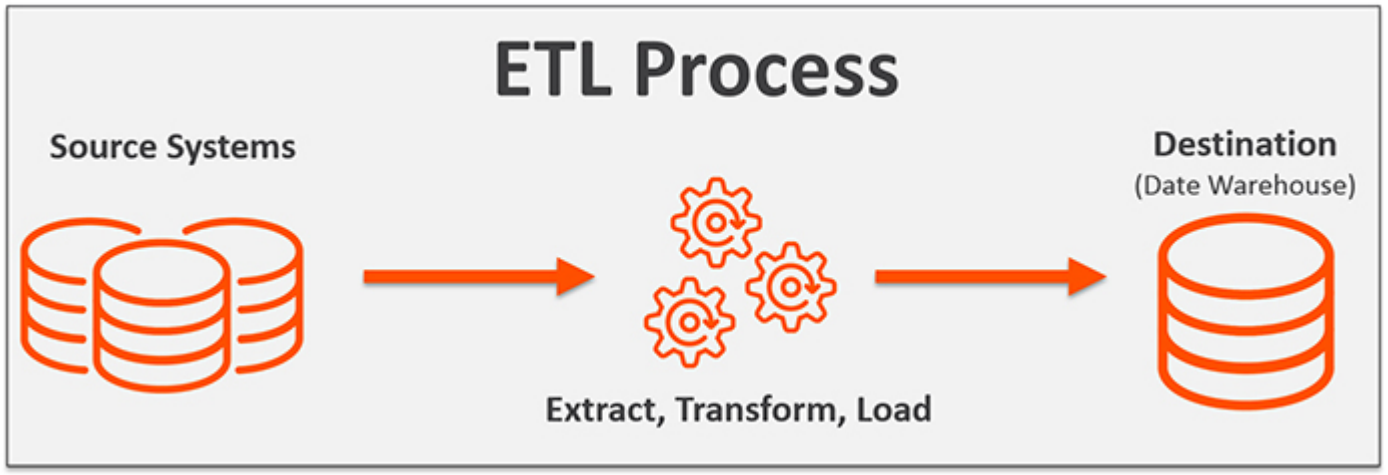
A large variety of code-heavy to code-light tools available

Most are proprietary systems and there's a cost involved



Node-RED

extract **t**ransform **l**oad

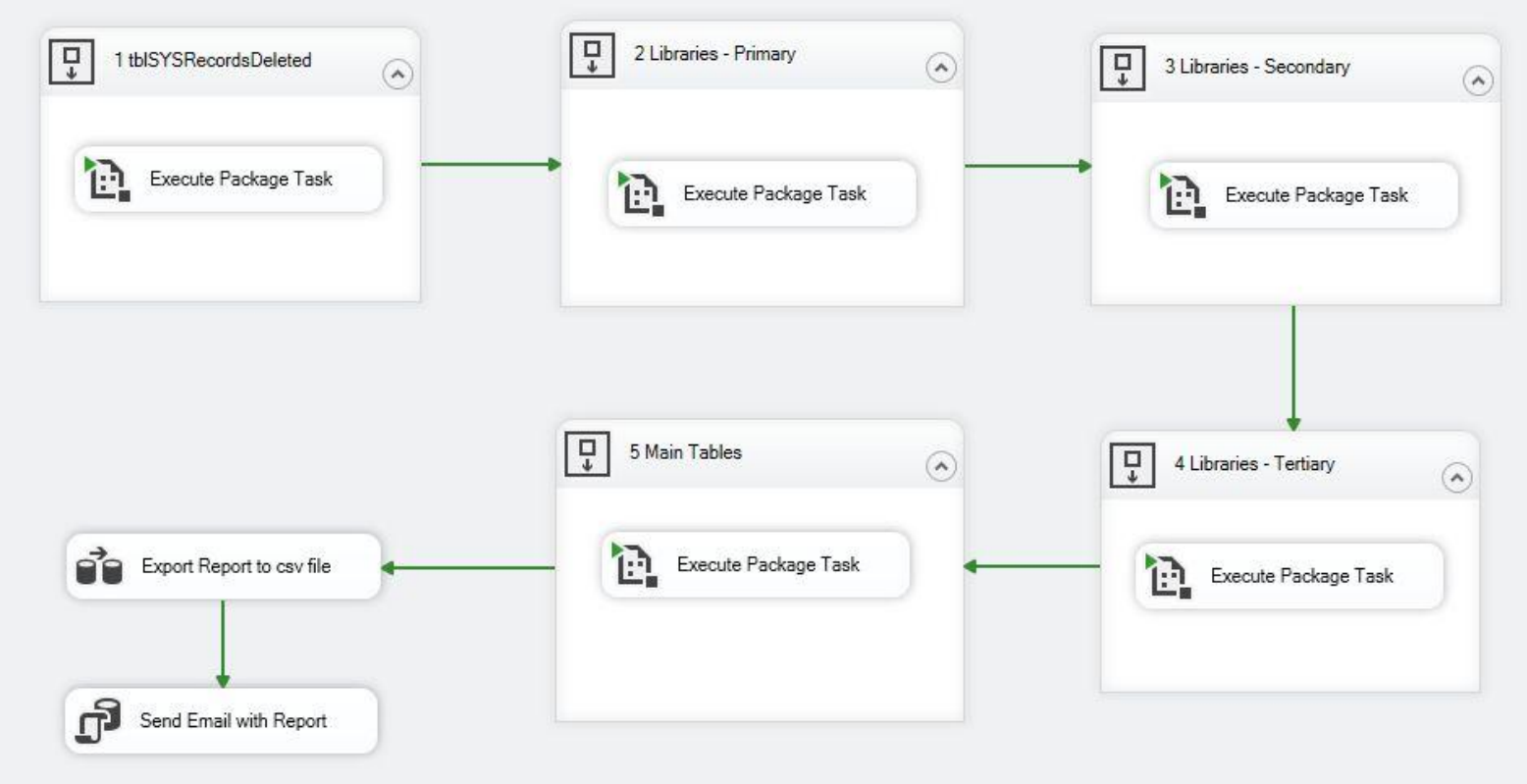


In essence what integration packages do is to:

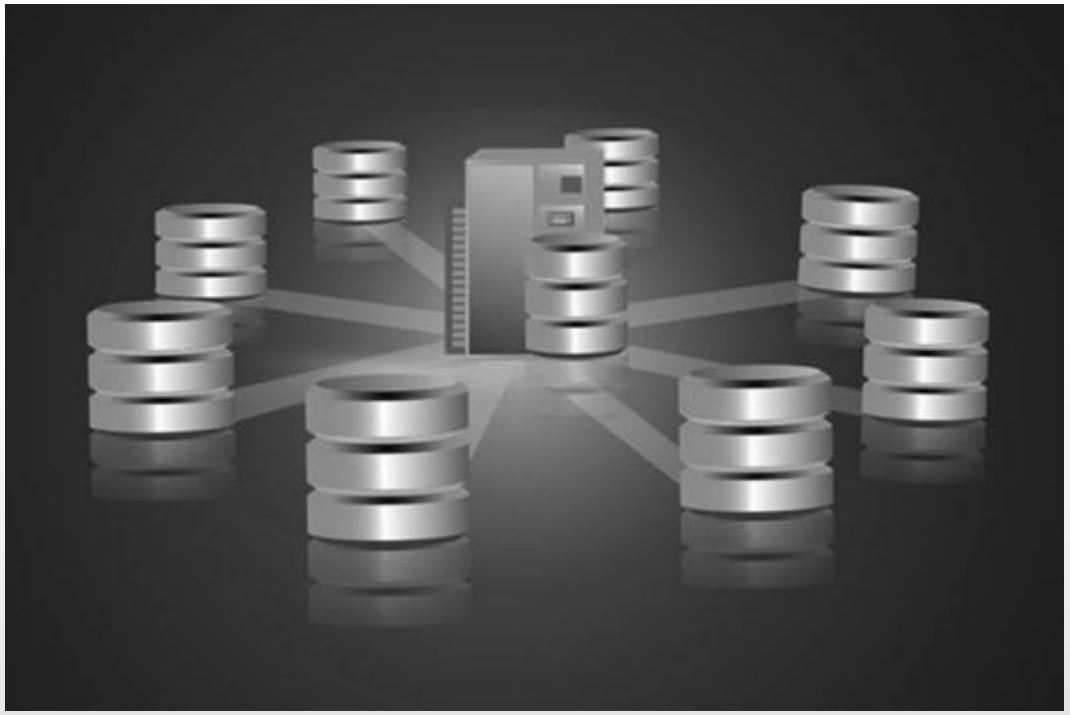
- EXTRACT data from a/various source(s)
- TRANSFORM data
- LOAD data to a/various destination(s)

example – SQL to SQL

- MS SSIS is good for basic integration inside of the Microsoft environment on older platforms
- Handles flat files and database connections well
- Somewhat outdated (SSIS 2016)
- Can integrate with modern sources via 3rd party applications or else coding in C#
- Integration jobs can be scheduled with SQL Agent
- Being replaced by Power Automate



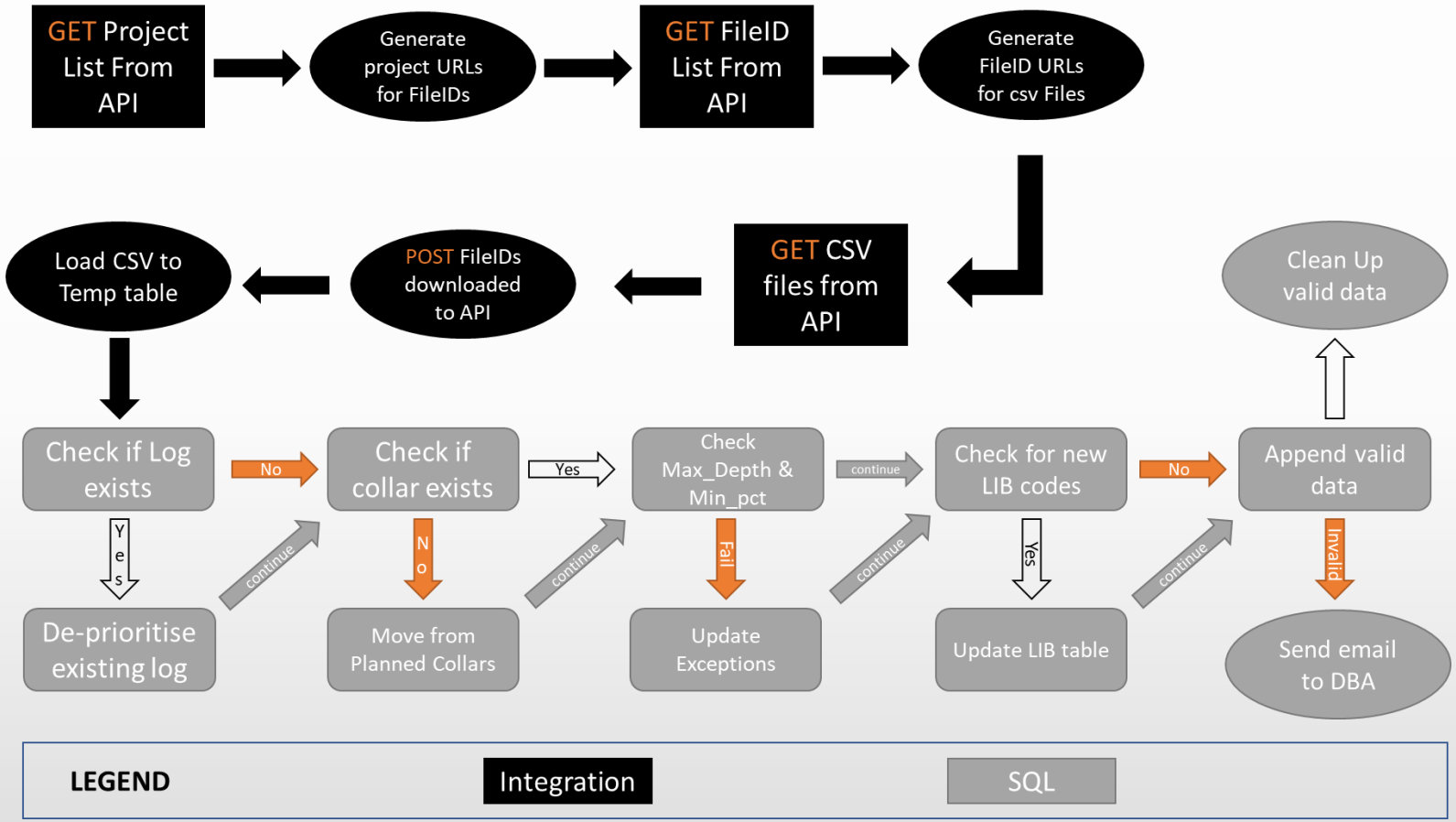
example – data warehousing



- Suited to large organisations with multiple source databases such as mine sites
- Automate the consolidation of data
- A data warehouse's architecture and design is geared towards data analysis and extraction instead of data creation and validation
- Simple SQL Server SSIS packages can be scheduled using SQL Agent to extract data from the various sources at scheduled intervals, to do any necessary data transformation and load to the central data warehouse.

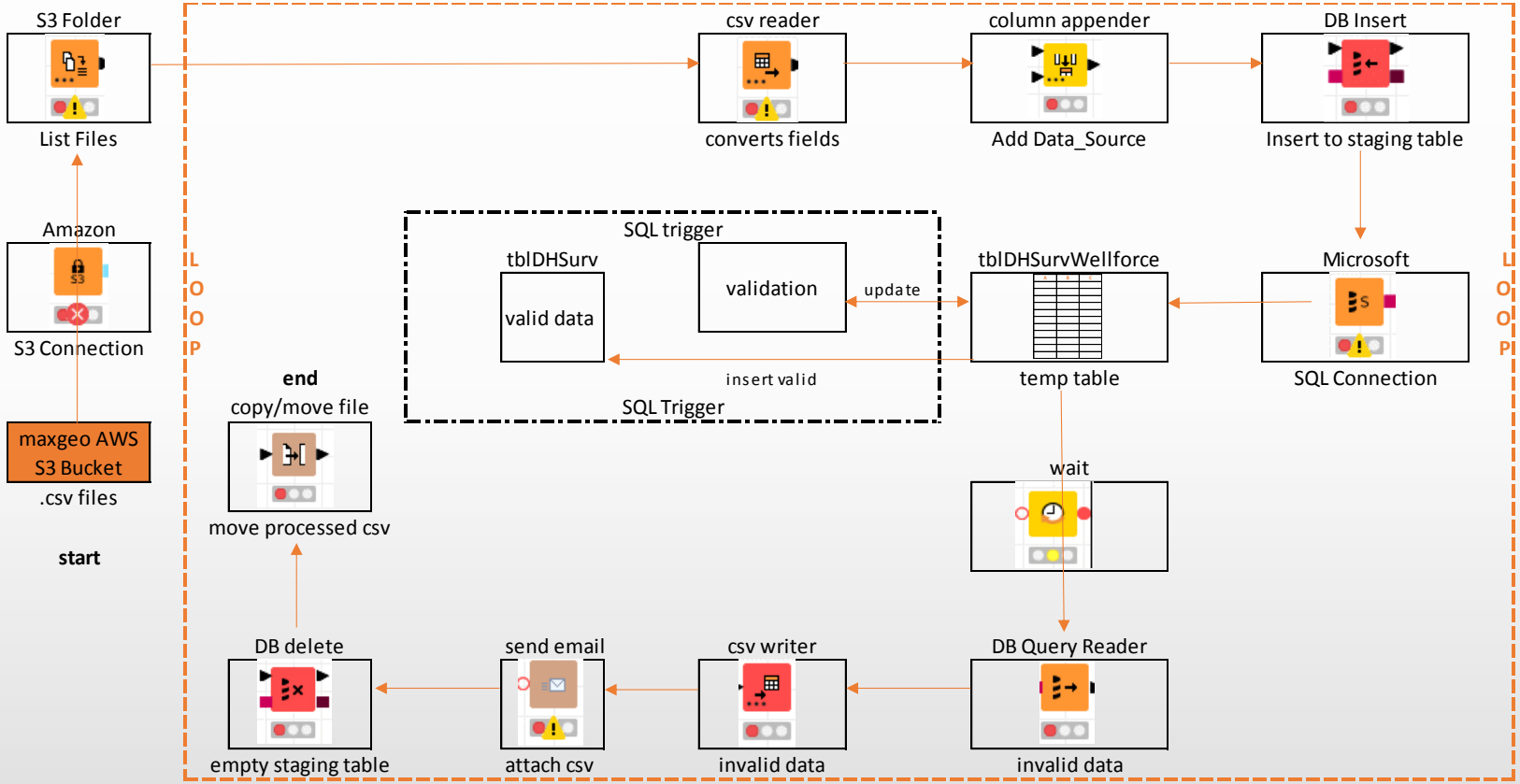
example – API to SQL

- Most APIs are of type REST (REpresentational State Transfer)
- They work by sending requests for information and receiving responses
- Data is retrieved through various GET endpoints and uploaded via POST endpoints
- Each API is unique and provides different endpoints such as GET, POST and DELETE (tools to work with) to the client application
- Once loaded into a temporary table in SQL, the data is validated inside of SQL making use of triggers and stored procedures.
- The Integration software is the middleman that queries the API and extracts what is available and passes it to the database



example – AWS S3 to SQL

- S3 buckets inside Amazon Workspaces (AWS) are becoming a popular way with vendors to deliver data
- In this case the challenge is being able to connect across the Microsoft / Amazon divide and extract data from the cloud to an on-premises SQL database
- Once again, the integration software is the middleman that scans the bucket, extracts what is available and passes it to the database



conclusion

The global datasphere is evolving and presenting new data management challenges

maxgeo

the solutions you need

sales@maxgeo.com

esnyman@maxgeo.com