REPLY
GLUE

EMPLOYING YOUR DATA AS AN ASSET

Data is an Asset

Valuable assets such as gold, must be mined, refined, assayed and stored safely. But, as the kings of old discovered, storing gold in the vaults of the castle does not make the owner richer as it provides nothing to daily life in the kingdom. In fact, its collection and protection are a drain on the economy. To become richer the holder of the gold needs to use it; to invest, to create value using it through trade and internal investment.

Traditionally enterprises treat data like the kings' gold. They centralise it, bringing it in from the sources to a central database at great cost and effort. When it is in the centre the data is refined and assay through transformation routines so that it conforms to an enterprise canonical form, and some may argue it is put under lock and key and is as difficult to access as the gold in the king's vault.

The analytical data users have to go to this central vault from which they get quality data, but it may not give them the value they need because it has been heavily refined and assayed: transformed into a centralised data model, but in the past, it was good enough.

Return on Investment

There are many benefits of Data as a Service, including the ability to assign a monetary value to your data assets. However, there are further considerations and concepts to consider to truly harness the value of your data and issues to be resolved to achieve a good return on your investment in data.

The centralised approach for data warehouses has been highly successful using the old relational technology, but it isn't scaling. It is no longer good enough for the business to hold vast reserves of data; that data, like the gold of old, must be put to work. But, enterprises are finding that that the move from traditional data warehouses to Big Data and Cloud has not given the expected return on investment because it is much more difficult to put the data to work than of old.

Glue Reply proposes that there are two key issues with using the centralised Data Warehouse approach to satisfy modern Analytics requirements in the Cloud

1. Architecture
2. Knowledge

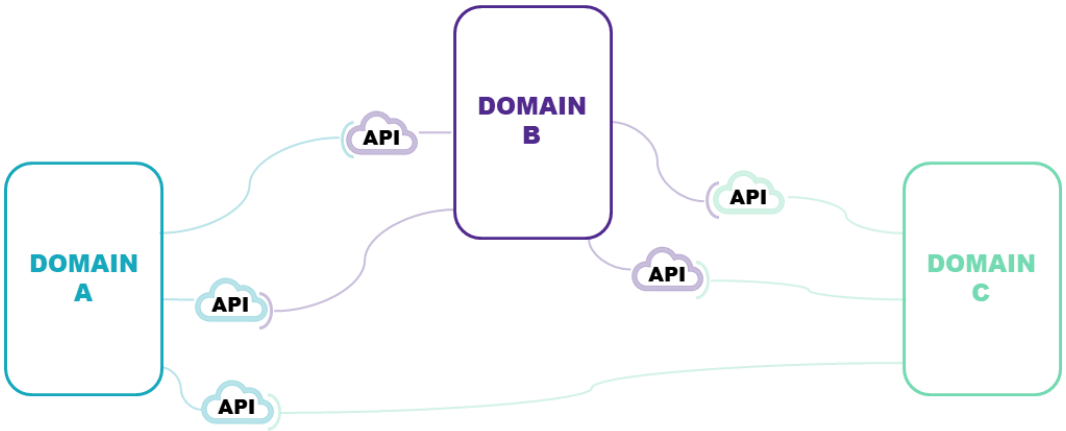
Let's look at these two issues in more detail.

Architecture

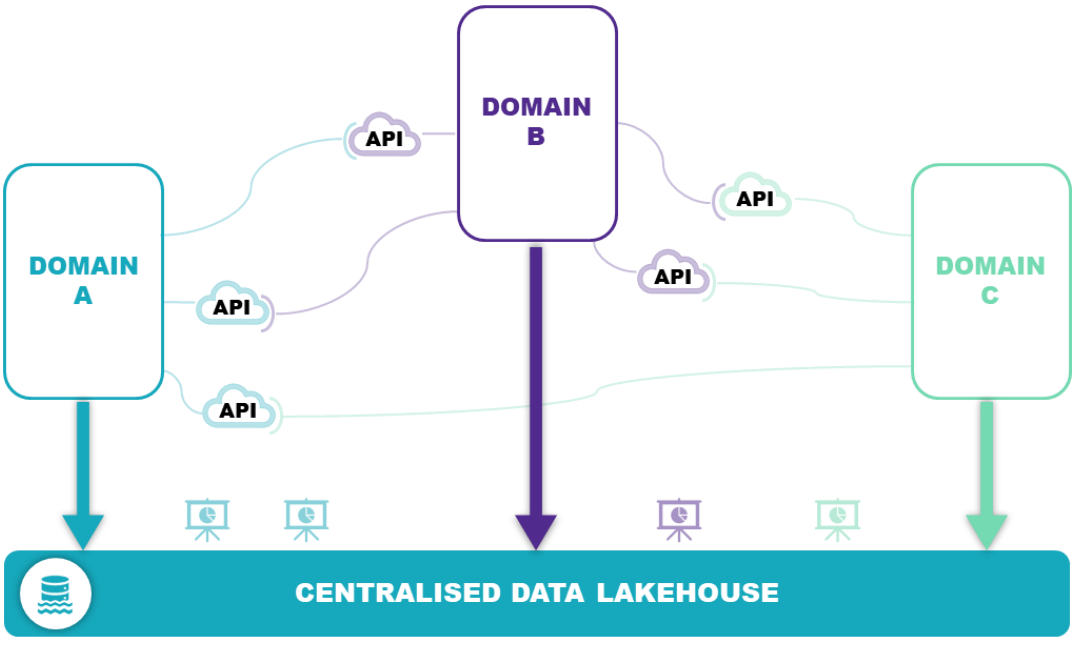
There are two very different IT architectures co-existing within the Enterprise

1. Operational Architecture: Application agnostic, implementation agnostic architecture where the domains offer operational services (APIs) for operational needs
2. BI/Analytics Architecture: Centralised monolithic technology specific where the central technologic functions offer either access to the raw data in a data lake or highly refined data fact often only in a specific format (Blob/XML/JSON/AVRO/CSV/PDF or PowerBI Report)

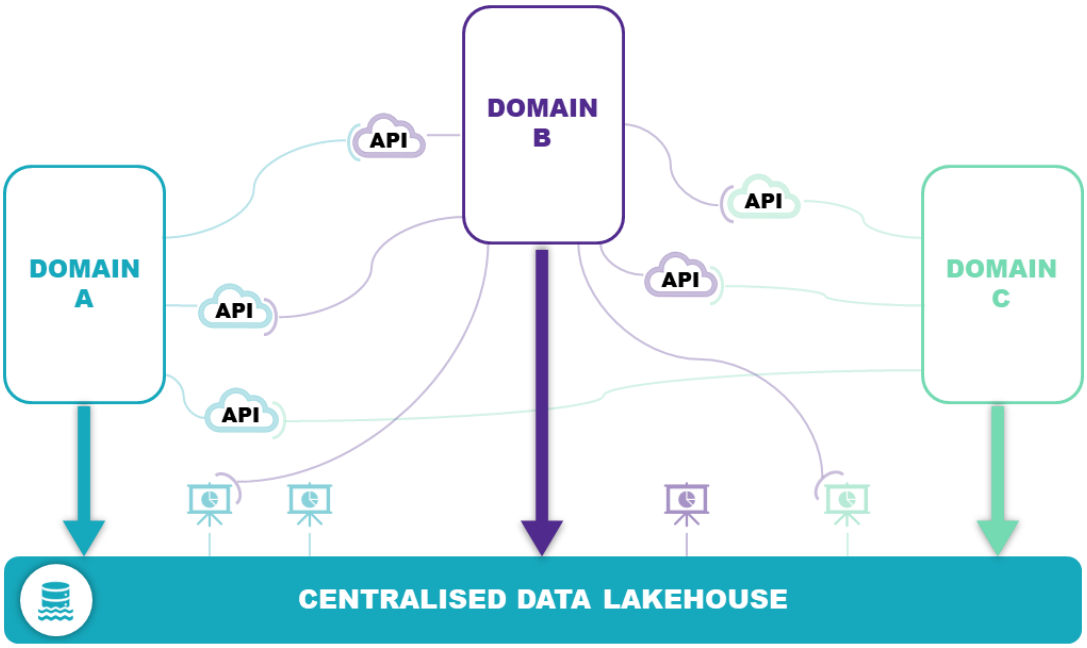
The Operational Application architecture allows the underlying technology, data sources, structures and methods to be hidden from the users of the services provided by the web-based APIs. This approach allows access to data and functions without knowledge of how the underlying systems work and for these systems to change without impacting the users of the services they provide.



This strongly contrasts with the architecture of Centralised Data Lakehouse where the centralised analytic functions don't integrate into the operational architecture and require in-depth knowledge of the Cloud technology, tools and data structure to use them. Additionally, despite the Data Lake approach it a slow and difficult process to add new data sources or change existing ones.



Is it possible to take the same approach with Analytics Services as with Operational Services to resolve these problems?



Some issues with trying to do this include:

- The short lifetime that these data services may have as individual data services may be required only for a single campaign or experiment.
- The available bandwidth of centralised roles required to building these services.
- Providing the analytic results in many different formats or modes. While Operational Services can exclusively mandate XML or JSON the users of analytics services will almost certainly want the same information in multiple formats such as CSV, PDF, Blob, SQL table etc.

As a result, the analytics remain stuck with the centralised architecture which is, as it has always been, to build a rigid set of reports and to allow the users to build their own analytics by trawling the data lake themselves much to the long-term dissatisfaction of the users and inefficiencies and inconsistencies this can bring about.

Knowledge

The modern centralised data store (Big Data Warehouse, Lake, Lakehouse) requires technical knowledge of Big Data and/or Cloud technology to build it and maintain it. On the other hand, to get the best use out of the data held there requires knowledge of the data semantics, syntax and the business that created the data - this causes tension and dissatisfaction amongst the three parties involved:

1. The data providers. The providers get frustrated because they can't change the structures or meaning of the data they provide without the considerable time and cost involved in amending the data pipelines and ETL

2. The data consumers. They can't get the analytics data they want in the timescale they need, unless it is by trawling through the data lake and building their own analytic models. The data scientists may love this, but the general business community don't.

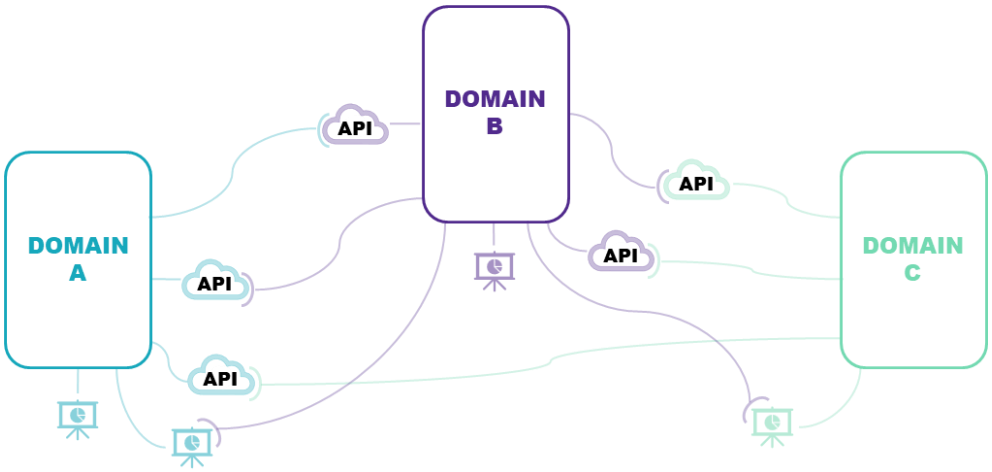
3. The technicians managing the central data platform. They are technical Big Data, Cloud and pipeline specialists; they're not business or data experts and so sitting in the middle between the providers and consumers it is inevitable that they don't understand what the other two want, especially because they don't have

    - Knowledge of the meaning of the data provided
    - Knowledge of the services required by the consumers

Data Products: Employing Your Data Asset

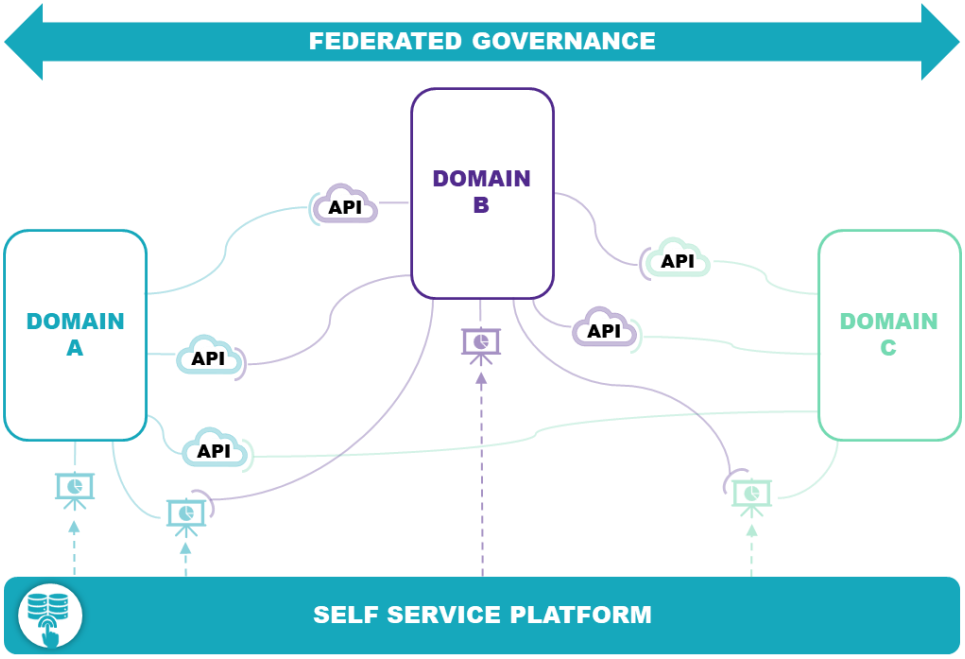
Glue Reply, and others including Zhamak Dehghani, the author of "Data Mesh, Delivering Data-Driven Value at Scale" (https://www.oreilly.com/library/view/data-mesh/9781492092384/), propose that the data asset should be deployed and employed in the business as Analytical Data Products. The Data Products are defined by the owning domain subject matter experts in response to demands from the analytics users, cutting out the centralised technical functions.

This approach aligns the Analytical Architecture with the Operational Architecture. The Data Domains are made responsible for the creation and management of the Analytical Data Products, just as they are for the Operational Services. In the end-state the Data Domains that create the data manage the Analytical Data itself as part of the Data Mesh architecture rather than centralising it.



The Data Products are created quickly and easily by the data experts in the domain that created the data, rather than technology experts. The Data Products have a managed life-cycle (birth-in, service, death) and the data itself is offered in a wide range of formats. The Data Product provides the application agnostic, implementation agnostic functionality delivered by Operational Services.

Glue Reply believes a critical part of the solution, and the first step towards it, is the Self-Service Platform, supported by a Data Marketplace, which allows the non-technical Data Owners to create and manage Data Products quickly and simply without the involvement of technical staff while the platform itself enforces governance policies such as privacy and access control.



In this way the data asset in the Data Product is an integrated component of the Enterprise Architecture and becomes part of the revenue generation process rather than being locked away in a vault.

For more information on how we help our client's create Self Service Data Platforms, implement a Data Marketplace and to further understand our Accelerated Discovery proposition as well as associated frameworks, contact us by visiting https://www.reply.com/glue-reply/en/contact-us.