**Data Virtualisation - The Enterprise Strategy to Unlock Big Data Potential and Simplifying Bigdata Projects**

According to Gartner, 60% of all big data projects fail, while as per Capgemini, 70% of big data projects struggle to be profitable. One thing is for certain that big data projects are challenging!  There are numerous factors at play in this case, rather than just one. For instance, some big data storage and processing technologies are difficult to use and big data products are frequently used incorrectly.

However, organisations continue to launch big data projects in order to take advantage of the possible business benefits promised by the big data proponents despite these discouraging results.

**The Key Concepts In Data Virtualisation**

Big data projects can be made easier with the support of data virtualisation. However, it won't solve every issue, but if used correctly, it will unquestionably increase the likelihood that big data projects will be successful. Here are a few good use case examples.

Big data is stored in plain files in enormous quantities. This makes it challenging for non-tech-savvy business users to access it. Big data can be made available to a wider business audience and to many well-known BI tools by using data virtualisation servers, which can obscure this complexity. They do include big data in an accessible, basic virtual table.

Big data may be developed remotely all over the world by international businesses. For instance, factories, manufacturing plants, and stores may all produce data. For reporting and analytical reasons, the volume of data that each remote site generates could be too much to replicate to a single location. To put it another way, big data may be hard to move. Virtual tables that hide the source of the data can be defined with the help of data virtualisation. To business users, it will appear as though all the data is stored in one source. Instead of moving all that large amounts of data to a single location for processing, the data virtualisation server moves the processing to the remote sites.

High-end, transactional NoSQL products are used to store big data in many projects. The majority of these products are built and optimised for handling huge volumes of transactions. Unfortunately, they neglect their analytical and reporting skills in favour of transaction processing. Data virtualisation makes it simple to cache and transfer NoSQL product data to a quick analytical platform. Transactional data may be made quickly and easily accessible for reporting and analytics in this way with little effort.

Big data is not always stored in systems that make it possible to define and document it. As a result, business and technical metadata is frequently absent. A data virtualisation server enables the definition of metadata for many types of data sources in the form of definitions, descriptions, and tags.

The final use case explains the rapidly developing field of big data technologies. Data consumers must be disconnected from data producers and stores in order to fully utilise any new technology. For instance, it will be more difficult to move an application to a new, promising technology if the more apps are dependent on a certain data storage technology. Applications and reports can be created regardless of the data storage system being used with the help of data virtualisation. It will make switching to a new, quicker technology simpler. Organisations become less reliant on cutting-edge technologies that could easily become outdated in our fast-paced environment due to data virtualisation.

**How Data Virtualisation** **Can Be Useful For Businesses?**

In conclusion, businesses aim to create big data systems. It's vital for them to become more data-driven, and for the majority of them, it's an aspect of their digital transformation. However, experience has proven that creating big data systems is challenging. A significant chance of failure exists. Although data virtualisation servers can't address every issue, there are certain crucial ones that they can address to streamline big data projects and make the most of an organisation's big data investment. Big data may be simple to use owing to data virtualisation.

Organisations must consider how they will manage their big data repositories and give their users speedy access as they rapidly develop data lakes and other big data structures. Without it, the repositories run the risk of developing into "big data silos," which would complicate rather than simplify the information landscape.

Big data fabric becomes useful in this situation. A "platform that accelerates insights by automating ingestion, curation, discovery, preparation, and integration from data silos" is how Forrester describes a big data fabric. Certainly, you must consider how you will build your big data repository, but even more significantly, you must consider how your users will access and use this data in order to acquire those business insights.

Solution approach proposed for one of customers heavily using bigdata technical stack using Denodo as Data Virtualisation tool.

Graphical user interface, diagram

Description automatically generated

Denodo reference architecture consists of:

* A layer for ingesting data referred to as "Data Collection" in the standard architecture.
* A processing and persistence layer referred to as "Data Storage and Analysis" which goes beyond merely being Hadoop-based that incorporates more conventional data processing and storage methods, including a data warehouse.
* A layer of data virtualisation that consists of data access as "data services," data access through the Data Catalog, data integration and transformation with improved performance optimisations for big data situations, and data security, governance, lineage, etc.