

ACCELERATE DATA MODERNIZATION

Eliminate security, privacy and compliance resistances in modernizing data infrastructure

WHAT IS DATA MODERNIZATION?

Data modernization is an umbrella concept that involves recognition and adoption of newer systems that will make collection, storage, consumption and utilization of data more effective and efficient. This could imply fundamental changes at many levels of the data lifecycle for example, data that is rarely accessed could be migrated into a cheap low throughput cloud store, or data that is concurrently queried might be migrated into a highly optimized data warehouse. The possibilities of data modernization are many, but at the core lies the fundamental requirement of physically moving the data from one repository to another.

DATA MODERNIZATION CHALLENGES?

The journey to completely modernize your data involves many challenges. These include understanding the multitude of dimensions on which a positive ROI can be achieved. There may be certain areas which might need a higher investment, either up front or overtime, but in most case the net gain is positive. Almost always, the utilization of data brings a lot of positive value, while security, risk and compliance can pose a barrier. The core of this problem lies in the fact that data has to be migrated onto a newer infrastructure, whether on-premises or in the cloud. So how does one mitigate this resistance in order to embrace data modernization.

STEP 01

OPTICS

Prior to any migration, it is imperative to gain optics into the sensitivity of the data. This means scanning through all the legacy data sources and creating actionable insights that satisfy some of the fundamental questions that can help either calculate the risk of migration, or enforce controls and remediation that will ensure the risk is mitigated prior to migration.

There are 3 main reasons why Dataguisse is a perfect solution to achieve this step:

- Breadth of support – Dataguisse can scan for sensitive data across most Relational and NoSQL Databases, Data Warehouses, all flavors of Hadoop and File Shares. This is imperative because every legacy system that contains data needs to be scanned for comprehensive optics into exactly what sensitive data is contained in these repositories. Moreover, due to the inherent characteristics of these platforms, data may be structured, unstructured or semi-structured, all of which are support by Dataguisse.
- Multiple Reporting Dimensions – The fundamental data points to be reported are the sensitive data found, exactly where it is found, and how much is found. Dataguisse can provide comprehensive reports around this information. The migration can also be highly dependent on other insights like age of data, identity information etc. Dataguisse reports on many of these, specifically targeted for a pre-migration report.
- Scalability – The Dataguisse solution is built robustly with natively built connectors for every platform and every file type. This is extremely important for scalability, because such a design ensures minimum overhead while scanning multiple Terabytes and Petabytes of data. Moreover, this ensures that there is no coding needed and the solution can run automatically out-of-the-box.

REMEDIATION

Once there are clear insights into the sensitivity of data, Dataguisse solution can automatically remediate sensitive data, either as part of the same discovery process or a separate remediation process.

02 STEP

There are 2 main approaches to this:

- Data Masking – This is an irreversible process, which provisions the data typically for a lower environment like dev/test or analytics. This allows the data to be easily migrated as the sensitive parts of it are anonymized, but the resulting datasets still maintain full referential integrity and useful analytical value.
- Element-level Encryption and Decryption – This is a reversible process, where the data is encrypted prior to the migration and only certain roles have the ability to surgically decrypt certain sensitive elements, based on a RBAC and ACL defined in the Dataguisse product. The decryption is typically done on-the-fly to ensure there is no plaintext movement of data upon access.

DEPLOYMENT?

The process of Data Modernization needs to be as low touch and efficient as possible, because every cost involved in it can damage the overall ROI. This means that mitigating the security, privacy and compliance resistance needs to be done without any overheads. This means that there needs to be flexibility in the design and deployment of the Dataguisse product.

Which is where there are 3 ways of doing this:

- Fully Hosted Product – In this deployment model, Dataguisse is installed near the source of the migration, so it can easily scan and remediate the data. This can be easily installed and connected to the data, and the product is scanning and/or remediating data at rest.
- Embedded Library – In this deployment model, a library is embedded in the data stream, so any of the Dataguisse processes are taking place as the data is flowing through the migration pipeline.
- Hosted API Gateway – In case of a serverless architecture, the Dataguisse processes can be invoked via its APIs, which may be hosted in the customer environment or offered as a SaaS application running from the Dataguisse cloud.

CONCLUSION

Dataguisse plays a key role in enabling secure data migrations which is the key requirement in Data Modernization. Whether it be a legacy on-premise system to another repository in the same data center, or whether it may be migrating an on-premise app to the cloud, Dataguisse helps ease any resistance from the security, privacy or compliance to ensure complete compliance with flexible data infrastructure.