intertrust®

White Paper

Platform

Data-driven digital transformation through agile data operations

Contents

Introduction

The move to the digital world is accelerating. Whether working, gaming, shopping, or consuming digital content, we are spending more of our time online.

As digital activity increases, organizations need to come up with data-driven transformation strategies and invest in newer agile data management technologies. They must also consider the right strategy around implementing a cloud data platform. There are numerous issues to consider in making the crucial decision around which platform to adopt.

Different platform types have different tradeoffs. The best platform should be one that maintains data security and data rights, gives your organization the flexibility needed to quickly implement and adapt data operations, and avoids vendor lock-in.

Managing data as an asset

The world's data is growing at an exponential rate. The amount of data created, captured, copied, and consumed worldwide is expected to grow from around 59 zettabytes (ZB) in 2020 to around 149 ZB in 2024. (Figure 1).[1]

Data as an asset is still in the "early adoption" phase, meaning it can be a competitive differentiator for organizations as they focus on digital transformation. Creating the right data strategy—and selecting the best platform to support it—is key to making full use of this opportunity.

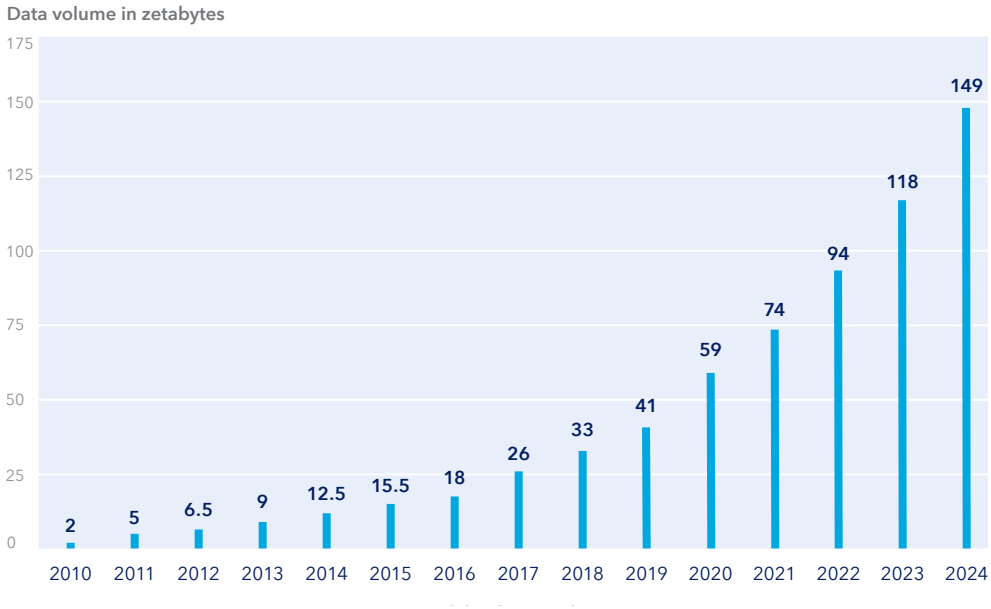
At a high level, mining a potentially huge volume of data serves two purposes: helping business leaders make better decisions (analytic use cases) and building data intelligence into customer-facing applications (operational use cases). This foresees the rise of massive, complex systems built around data—where the primary business value of the systems comes from data analysis. Acquiring useful insights derived through big data analytics is a major driver of digitization and automation of workflows.

Therefore, the ability to analyze vast amounts of both structured and unstructured data to gain insights, often in real time, is what underpins these systems and, accordingly, most digital transformation efforts.

Many enterprises already deal with highly complex IT systems, often spread across multiple geographies and component systems. In addition to core ERP (enterprise resource planning), they use specialized software to manage pricing, demand planning, customer relationship management (CRM), human resources, point-of-sale, and a multitude of other mission-critical business functions. These siloed and distributed IT systems make data management more challenging, and mission-critical as well.

1 Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2024, Statista Feb 5, 2021

Figure 1.
Total data volume worldwide, 2010-2024
Source: Statista



Data also comes from numerous internal and external sources. Metadata management, data quality, data catalogs, and maintaining security and data rights management are all essential for accessing this data.

While ensuring data security (protecting data from internal and external threats) is a given, organizations must also provide data privacy (complying with regulations on data processing).

Following numerous scandals such as Facebook and Cambridge Analytica, privacy issues have become more important for the general public and of course, regulators. Regulations like Europe's GDPR (General Data Protection Regulation) and U.S. state regulations such as the CCPA (California Consumer Privacy Act) introduced new data tracking and security constraints. With data integrity, consistency, availability, and usability critical at every stage of the data pipeline, companies need simple solutions to control data.

Enterprise infrastructure options—pros and cons

To support their digital transformation strategies, companies are adopting different cloud infrastructures to meet their needs. Public, on-premises (on-prem), and hybrid cloud infrastructures all have different advantages and disadvantages, detailed in Table 1.

Table 1.
Different cloud infrastructure options.

Model	Pros	Cons
Public cloud	• Shared scalable infrastructure resources • Multi-tenant architecture reduces costs • Minimal configuration complexity and maintenance required • Highly available and scalable • Pay-as-you-go or usage-based subscription models	• Less control over data security • Blackbox feel—usually impossible to get under the hood to troubleshoot problems • Maintenance windows or new feature availability is outside your control • The physical location of the data may be difficult to control
On-prem center	• Dedicated resources and infrastructure • Highest level of control and security • Resource access limited to a single secure private network • Maintenance windows or new feature rollout are usually 100% in your hands	• Maintenance of the physical and software computing infrastructure is the responsibility of the enterprise • System Admins are required, either by direct hire or through service providers
Hybrid cloud	• Greater flexibility with the best of both worlds of cloud and on-prem • Balance of control, performance, and scalability • Performance—ability to address network bandwidth concerns and local computational efficiency • Security—can support strict security demands • Cost efficiency—since you are only paying for portions of the public cloud you need	• High responsibility and burden to manage and provide visibility of security measures • The mix and matching of services can make diagnosing problems harder e.g., is this failure a result of on-prem or a cloud service the system is using?

Meeting the challenges

Cloud computing has been around for some time, but businesses still struggle with challenges as they move to cloud-based infrastructure.

These challenges are especially true for data, and include lack of central control, rising/unpredicted costs, infrastructure complexity, security and compliance, and true scaling.

The Intertrust Platform™ is a robust solution to these challenges. A neutral, "mix and match" platform for governing access to multi-party data, the Platform facilitates secure collaboration between internal stakeholders as well as third-party service providers. It enables organizations to manage distributed data at scale.

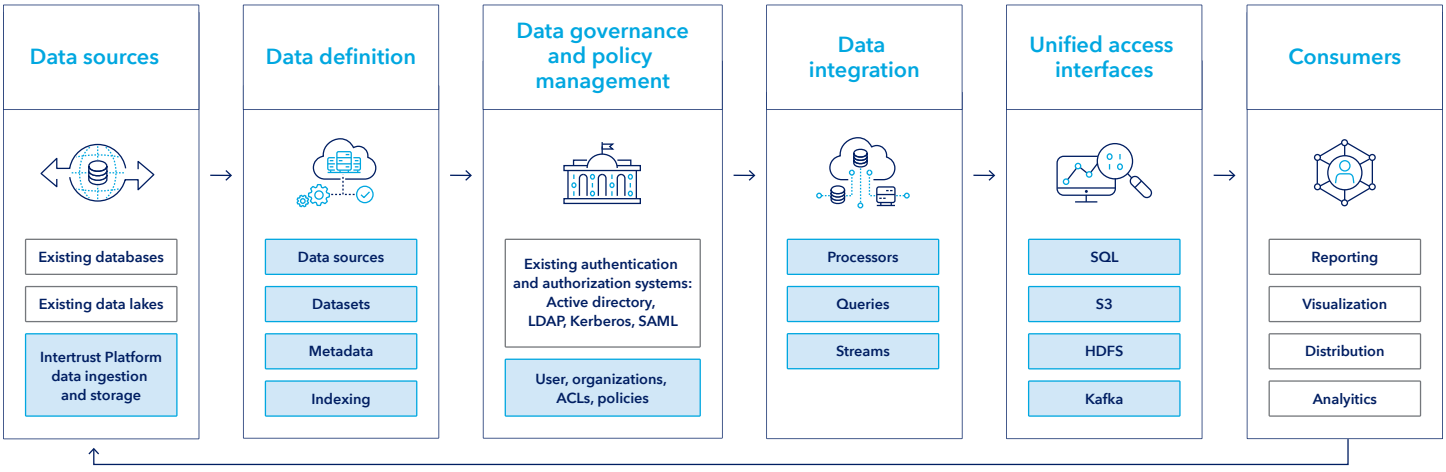
Here are some critical challenges that need to be carefully considered as part of any cloud or hybrid strategy—and how the Intertrust Platform addresses them.

Data migration

Migrating data to the cloud is the top challenge for two main reasons:

1. Moving data in different formats from various sources is tedious, expensive, and needs a long-term strategy to work with the ever-evolving types of new data.
2. Many of the current processes for managing and protecting data are highly integrated into legacy on-premises solutions.

Figure 3.
Technical benefits of the Intertrust Platform.



1 Enterprises need to provide uniform access to existing data stores such as SQL databases, NoSQL databases, Kafka queues, and file based storage systems. This can the achieved with the Intertrust Platform, eliminating the need to consolidate data into a single data warehouse or duplicate data across an organization.

2 Virtual datasets may be accessed using common ODBC, JDBC, and Spark interfaces.

The Intertrust Platform includes a secure data virtualization layer that connects to different data sources, integrates all the information, and publishes it as a dataset. This enhances data accessibility and makes specific bits of information readily available for reporting, analysis, and decision making. There is no need to copy data from one location to another or ingest data into the Intertrust Platform. Figure 3 shows some of the technical benefits of the Platform's data virtualization layer.

The Intertrust Platform acts as a window to connected data sources, where access to the underlying data is controlled through the secure data virtualization layer. Fine-grained access controls can be defined even if the source databases only support limited access control. For example, row-level access control can be defined on relational databases that natively don't support such controls.

The Platform also provides uniform governance across all data coming from different data sources that can be distributed among various cloud providers. Virtual datasets that span multiple databases can be easily queried on the fly via cross-database JOINs with row- and column-level access control. Restrictions can also be defined to further limit access based on dynamic attributes or parameters of individual user accounts, their groups, or organizations.

The secure execution environment provided by the Intertrust Platform can preserve end-to-end governance during compute operations. It provides an environment where programs that access data can be governed according to the data provider requirements. Secure execution environments are uniquely positioned to provide a high level of data control for various types of computational software.

The environment also enables enterprises to run third-party programs without exposing data to the third party or others. At the same time, the intellectual property of third-party software providers can be protected, since their programs can also be governed by the Platform.

The security framework of the secure execution environment is built into the Intertrust Platform's overall data management and application execution environment, making it simple to use for enterprises.

Conforming with data security and privacy regulations

Enterprises need to monitor all data that enters and exits their systems—continuously checking, scanning, and classifying data while in motion. With massive volumes of existing enterprise data and live data continuously streaming in, the robustness of the encryption and validation systems becomes very important. Systems that can provide for compliance with an ever-increasing number of data privacy regulations while allowing for mutual collaboration with partners at scale are also essential.

Since violations of data privacy regulations such as the GDPR and CCPA can come with extremely serious consequences, it is worthwhile to look further at these challenges.

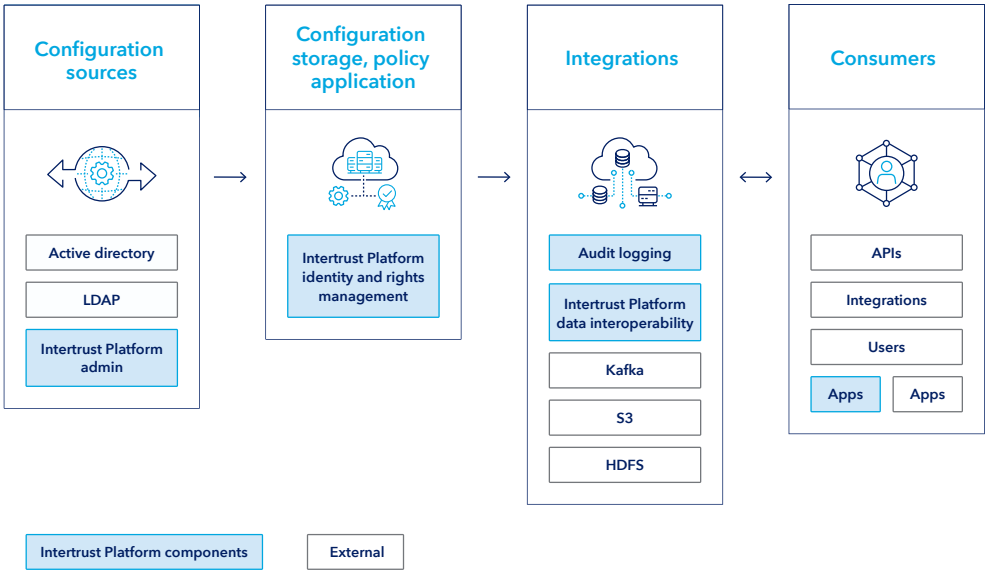
To comply with data privacy regulations, some of the things organizations need to focus on are:

- The ability to access data from many different file types to gain an integrated view and understanding of what personal data they hold
- Identifying and extracting personally identifiable information (PII) from structured and unstructured data sources
- Capabilities needed to enforce governance policies, monitor data quality, and manage business terms across the organization
- Role-based data masking and encryption technologies to secure potentially sensitive information, as well dynamically blend data without moving it
- Interactive reports to identify users with access rights, files, data sources, and types of PII detected
- The ability to perform audits

In public clouds, security starts with authentication of the user who consumes the cloud resources. Cloud services provide additional controls, but often need experienced security architects who understand the specifics of these controls, which can also differ between cloud service providers. Since enterprise system architectures need to be built around these services, this can be a point of vendor lock-in in addition to the costs involved.

Intertrust Platform enables IT professionals to build on top of existing enterprise authentication systems, enriching them with granular access control to the unified catalog of all objects entered into the platform (like data sources, datasets, computation clusters, API clients and applications, etc.). Figure 4 shows a technical overview of the Platform's access control feature.

Figure 4.
Technical overview of the Intertrust Platform's access control feature.



1 All requests for data via the Intertrust Platform data interfaces are subject to the rights and permissions set by the data owner. Through these permissions, access to data is either granted or denied to specific accounts, groups, or orginizations. Additional restrictions may be set on dataset rows or columns, enabling granual control over which attributes are exposed.

2 Virtual data-sets may be accessed using common ODBC, JDBC, and Spark interfaces.

With the Intertrust Platform, enterprises have access to an intuitive user interface where data and resource security can be managed from one place.

Data privacy regulations may require the storage of personal data within specific geographical regions. To comply with these regulations, enterprises may have to store user data within local regions, which adds overhead to existing systems.

Intertrust Platform's data virtualization functionality allows regulated data to physically remain in the regulated geography yet be made available for necessary data operations. For enterprises that desire more fine-grained geographical control, Intertrust has a partnership with InCountry, a global data infrastructure provider that specializes in helping enterprises comply with differing data privacy regulations. InCountry accomplishes this by providing separate databases for regulated data and the environment to manage this data.

With the Intertrust Platform, enterprises have access to an intuitive user interface where data and resource security can be managed from one place. For example, object access, roles, and permissions, or programs that run in the Intertrust Platform's secure execution environment can be defined in the Platform. The Platform also ensures that deployment of these programs is done only after obtaining valid authorization after authentication. The data that these programs access is controlled and can be set up to never leave the system. The external API calls from these programs can also be secured through the use of access tokens. Network access for these programs can be controlled to the IP address and port levels as well as limited to scheduled time windows.

The key benefit to enterprises is that all the security framework and access controls can be easily configured through the tools built into the platform. Programs can be deployed into the platform with few configurations, and changes to security policies can be dynamically applied. All this is done through very simple control panels without a lot of integration effort or the need of skilled resources.

Since analytics and machine learning programs can be run within the secure execution environment, the resulting output data from their operations can be indexed back into the Platform. Enterprises can then craft new data access rules for freshly created and indexed data to make a subset of it available internally and to the partners without the raw data leaving the enterprise's systems. The vendors who develop these programs can also guard their intellectual property rights. The Platform establishes trust between the enterprise and software vendors by creating transparent access rules and audio records as proof of rule enforcement.

With the Platform controlling access to data through its data services, access to data can be configured to filter out sensitive information. If programs do need access to sensitive data, privileged access can be granted with the secure execution environment, where data egress can be restricted.

Intertrust Platform complements traditional data lakes and ware-houses, adding operational functionality and simplicity to the deployment.

Lack of centralized management

Lack of centralized management is a challenge when data is spread across multiple clouds and platforms, and literally dozens of tools. The solution is having a system that is fully integrated into a single data management tool that works across on-premises and different cloud infrastructures. This ensures the health and protection of data at all times, massively reducing the drain on human resources.

The Intertrust Platform provides the ability to integrate multiple sources and expose them in a uniform manner. It provides a user interface to manage these sources, as well as a rich API set so enterprises can customize interfaces and designs as desired. Additionally, the Platform can be white-labeled and modified to reflect the interface and branding of the enterprise.

When looking for a centralized cloud data management solution, enterprises may look at data lakes or data warehouse vendors. While these technologies may have their place, the Intertrust Platform has a number of advantages to consider (Table 2). The Intertrust Platform can work with each of the available options of data lakes and warehouses (on-prem/cloud) to help augment the setup into an agile platform.

Table 2.
Different cloud infrastructure options.

Cloud data lakes (DL)	Data warehouses (DWH)	Intertrust Platform environment
Store data in raw form in the cloud.	Store data in a structured format in a dedicated cloud. They can also be hybrid in nature.	Virtual view of data that could be either in a data lake, data warehouse, or other existing data stores. Can connect to unstructured, semi-structured, and structured data without any data loss, can also transform, reformat, and unify this data for easy analysis. Intertrust Platform complements traditional DL/DWH and adds operational functionality and simplicity to the deployment.
DL are suitable for bespoke applications and more advanced data processing needs.	DWH are designed to quickly and easily generate insights from core business metrics, usually with SQL.	One-stop shop for most any type of analytics processing or operational metrics development as data can be processed from any connected source without moving it.
If DL are poorly managed, they quickly accumulate huge amounts of uncontrolled data, much of it often useless. It becomes unclear where the data came from and when, how relevant it is, and whether it can be used for analysis.		Provides data governance capabilities, so that the data can also be operated at its current location without moving it.
At their core, DL are a collection of a huge volume of raw data. By continuously ingesting disparate pieces of customer data from a variety of sources in a DL, organizations often have no clue what sensitive or useful information they have and how it is being combined.	DWH are huge volumes of structured data. As such, they are expensive to maintain.	The Intertrust Platform addresses data requirements for both analytical use cases and operational use cases. For analytical purposes, the data used is generally structured, often coming from disparate databases located in different physical locations. The Platform provides a unified view of this data. It provides a UI for accessing and governing from a unified view. For operational purposes, the Platform provides a secure execution environment where data in unstructured and semi-structured formats can be used by AI and ML learning based systems. Therefore, the Platform can effectively replace DL and DWH for these applications.
Compared to DWH, DL are less expensive. Still DL/DWH pricing is based per access request.	Costly to setup and run due to the following costs: • Per GB/TB storage cost • Monthly maintenance • Skilled resources • Tooling cost	The Intertrust Platform reduces costs by allowing data to be operable in its original locations. Storage and access costs remain essentially the same and data repositories can be maintained with current resources. This also simplifies data protection and governance. The Platform enables operators to create virtual instances of data authorized for particular users. Authorized users can then quickly connect to various data sources, regardless of where the data resides. The authorization layer also allows very granular settings for access rights. For example, a user from one department can expose a subset of a virtual data set to other departments, external users, etc. Since data resides in its original location, no additional data protection efforts are needed. In essence, the Intertrust Platform provides most of the functionality of DL and DWH. It also brings additional features such as fine-grained data governance and secure execution environments to safely run third-party programs.
DL/DWH tend to become virtual dumping grounds for any and all business-generated data. This means data protection and data governance in these platforms must be carried out at enormous scale. Everything, from discovery and classification, to access controls and security, to policies and auditing, needs to be re-invented to contend with the sheer volume of data stored and the velocity of its generation and usage.		

Intertrust Platform helps streamline the operational costs involved in the creation of a data collaboration ecosystem.

Rising IT and planning costs

Demands on IT resources are increasing, but budgets may not change at the same rate. With the availability of scalable storage options and the huge proliferation of devices producing data, data becomes scattered across multiple systems and vendors, requiring skilled resources to manage its movement to the cloud. On top of these costs, managing these diverse systems becomes an overhead on the existing system operations.

Intertrust Platform helps streamline the operational costs involved in the creation of a data collaboration ecosystem by bringing users closer to the data and by reducing friction that can impede collaboration between multiple users. It simplifies data access and identity management across diverse datasets, and integrates it in a single place regardless of the location of the data. This helps the organization significantly reduce planning costs for data integration. More and more users can access relevant data views in the shortest possible time within the proper security frameworks. This leads to a reduction in operational, legal, and transactional costs and risks. It also empowers end users such as data stewards and data operations engineers to make data-driven decisions, all within a thoroughly optimized cost structure.

Conclusion

Agile data operations practices are intrinsically tied to any organization's digital transformation path.

Selecting the right cloud data infrastructure is also key. To stay competitive, companies must be flexible, avoid vendor lock-in, embrace new technologies and practices, and adopt the right security framework. Partnering with a neutral trusted intermediary is a viable strategy for managing internal and external data consumption. To that end, Intertrust Platform facilitates secure collaboration and enables companies to make data-driven business decisions, confidently and securely.

intertrust®

Building trust for the connected world.

Learn more at: intertrust.com/platform
Contact us at: +1 408 616 1600 | dataplatform@intertrust.com

Intertrust Technologies Corporation
400 N McCarthy Blvd, Suite 220, Milpitas, CA 95035