VERITAS

Intelligent Data Migration to the Cloud

Using Veritas Information Studio to make data-driven decisions about your cloud migration strategy.

White Paper | June 2021

Contents

Executive Summary

Figure 1: 3340 Appliance

Organizations commonly have or are planning to implement a hybrid solution for storing their data. According to IDC[1], the amount of data transitioning to the public cloud will have an estimated 22 percent compound annual growth rate (CAGR) between 2019 and 2023. Major considerations about why to transfer data include cost optimization and overall efficiency. With diverse storage infrastructures, massive data growth and an ever-increasing set of compliance rules and regulations, organizations can find it overwhelming to decide on an optimal strategy for their data migration. An essential component in developing and tweaking this strategy is awareness about the type of information contained within their data. According to Veritas Databerg Research[2], 52 percent of all information stored in organizations is considered dark data whose value is unknown. You can't make strategic decisions about what to send to the cloud if you don't understand the risk, value and waste in your data. Those organizations that do migrate data without this deep understanding often end up with the expense of moving workloads back to on-premises storage. In fact, according to an ESG[3] study, 55 percent of organizations pulled workloads back from the cloud. And according to the same study, the top three reasons for doing so were security, cost and compliance issues that could be avoided if companies were able to do more complete due diligence by understanding details about the data being migrated. This is where Veritas Information Studio can help.

Regardless of where you are on your journey to the cloud, Information Studio can enable you to make intelligent, data-driven decisions about your migration strategy that help optimize cost and reduce risk.

Information Studio gives organizations visibility into information hidden within their data infrastructure and allows them to take informed action. It can provide details about information contained within files from within both primary and backup data that are stored on-premises and in the cloud.

With this information, you can confidently make data-driven decisions about what data can and cannot be migrated safely to the cloud. Plus, it can help reduce costs by providing insights to optimize your storage, including properly identifying relevant data to migrate, non-essential data to delete and the appropriate storage tier to use in the cloud. And after you complete your initial migration, it can help ensure your company's storage policies are followed by regularly monitoring your on-premises and cloud-based data to identify improper data placement.

How does one information studio help organizations

When creating a strategic cloud migration plan, organizations need to consider many business factors, including security, cost, business impact, service-level agreements (SLAs) and more. Once they've decided to move data to the cloud, organizations may make the common mistake of moving all their unstructured data from a particular data source or network share to the cloud. Too often they make these decisions without considering the different types of stored data they have, how moving everything can affect their risk and the impact it can have on regulatory compliance.

To make better decisions about what data—particularly unstructured data—to migrate to the cloud, organizations need to do the following:

- Gather details about the type of data they're considering migrating to the cloud.
- Look for ways to minimize cost.
- Ensure they maintain regulatory compliance.
- Reduce risk.
- Ensure data is protected appropriately.

Visualized detailed information about data

Information Studio identifies and retains important information about your data that it updates regularly using opt-in features. By making this data available on demand, it allows you to make quick decisions about your on-premises and cloud-based data storage as business needs change.

To make this process easy to visualize, Information Studio starts by capturing important information using both metadata scans and running classification. Then Information Studio displays this information in a user interface (UI) that administrators can filter to capture specific information to help you make data-driven decisions about your migration strategy.

Metadata Scanning

Information Studio gathers metadata information for files such as file name, file path, size, owner and create, access and modification time from both primary data and the Veritas NetBackup™ catalogue. It performs this scan on a user-defined schedule so it can provide up-to-date details that are always available.

Classification

Information Studio has a built-in classification engine that can find important information within your unstructured data. It comes with more than 700 preconfigured data classification patterns and more than 120 policies to identify common data privacy, corporate and regulatory compliance principles. Veritas regularly adds to and updates these patterns and policies to support new and changing compliance regulations. You can also create custom policies to detect files that contain specific information.

Information Studio's built-in policies can identify which files contain:

- Intellectual property
- Corporate compliance
- Industry-specific or government-regulated information
- Country-specific sensitive data
- Country-specific personally identifiable information (PII)

Beyond just identifying which files contain this type of information, Information Studio also captures pattern matches for the policies such as project names, credit card numbers or phone numbers.

Information Visualization

To develop a migration plan, you should start by visualizing the important information hidden within your data infrastructure. The powerful views built into Information Studio's UI of data gathered from both your cloud and file native data sources as well as your NetBackup catalogue will give you a head start (see Figure 1).

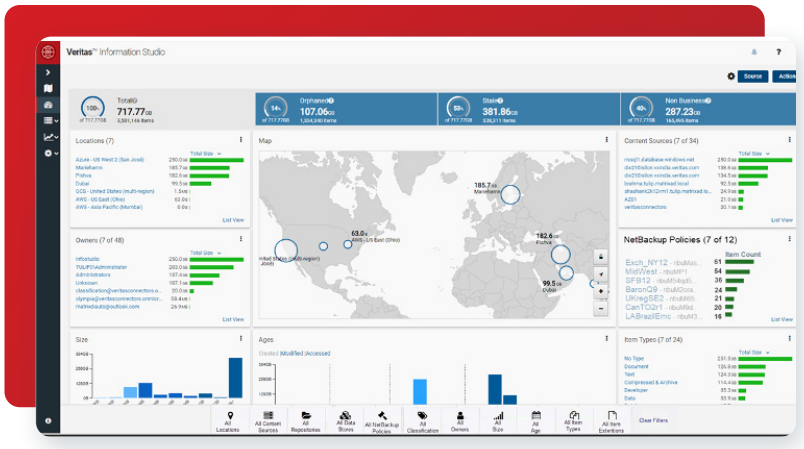

Figure 1. Information Studio lets you visualize data about your data infrastructure and filter it to generate specific reports.

- Locations (including a map view)
- Content sources
- Repositories
- Datastores
- Owners
- Size
- Age
- Item types
- NetBackup policies (supports Microsoft Windows, standard, Network Data Management Protocol [NDMP], VMware and Hyper-V policy types)
- Item extensions
- Classification tags
- Classification rules
- Classification sources

Information Studio groups the information it captures into the following pertinent categories that it displays in the UI:

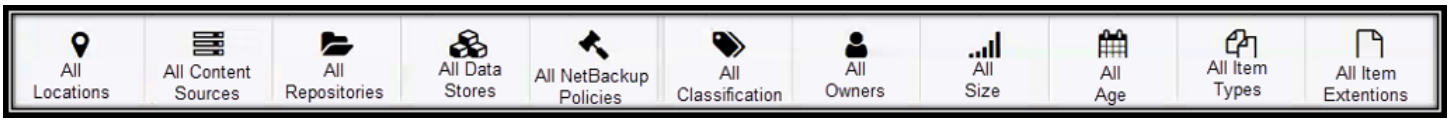

Figure 2. Filters available to dynamically change the views in Information Studio's UI.

As shown in Figure 2, you can also filter the views by any number of these categories. As you apply filters, the views change dynamically.

For example, you can filter for all items containing regulated data like Sarbanes-Oxley (SOX), Payment Card Industry Data Security Standard (PCI-DSS) and Health Insurance Portability and Accountability Act (HIPAA) information so you can ensure primary and backup data is stored on-premises. Filtering also helps you ensure data that is more than 120 days old and does not contain intellectual property is stored in the cloud.

Users can query and filter the data using APIs or Information Studio's web-based graphical user interface (GUI). The results of this filtering are presented within the web GUI and/or a comma-separated variable (CSV) or SQLite file report. The report contains a list of files or items and their associated metadata that meet the user-specified filters. You can use the reports generated by Information Studio to develop an intelligent data migration strategy and perform filtering and reporting automatically on a schedule you set.

Optimize Costs

There are several ways to optimize costs when migrating data, including deleting non-essential data and choosing appropriate tiers within the cloud to store data.

Delete non-essential information

Before migrating data to the cloud, it's important to first identify the value of your existing data so you can determine what is relevant and should be retained and what data is not worth keeping.

It's not unusual to find a large majority of data being stored contains useless information. Information Studio identifies non-essential

Figure 3. Information Studio identifies stale and non-business data and its associated size and cost.

data and groups it into categories of stale and non-business data, making it easy to recognize the size of this data and the cost of retaining it (see Figure 3).

Deleting data is one way to reduce your storage costs. However, you can't just delete data without understanding the type of data you're deleting, limiting who has authority to delete files and keeping an audit trail of this activity.

Information Studio makes it easy to filter for this data and provide the information needed to make evidence-driven decisions to defensibly delete data. Authorized users can identify unnecessary files like those that are stale or non-business-related within your data infrastructure and request that Information Studio delete these files and retain an audit log of the action.

This is what is considered a defensible deletion because valid factors such as the data type or age were determined for each and deemed unnecessary. Deleting such files before migration helps reduce migration time and the overall cost of storing this data in the cloud.

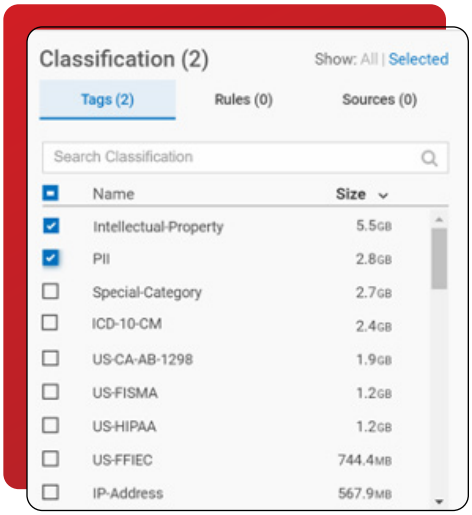

Figure 4. Filter options to find files containing intellectual property, personal and regulated data.

Consolidate data from local shares onto Microsoft Office 365

Many companies are looking to reduce costs and consolidate their unstructured data by migrating it to Office 365. However, business processes often restrict storing certain types of data like personal data, regulated data or intellectual property in the cloud. To gain the benefits of moving to Office 365 and adhere to business policies, you need to know what files contain this type of information. Information Studio includes classification policies that can identify this information. You can then apply filters to identify what data to migrate to Office 365 and what stays on-premises (see Figure 4).

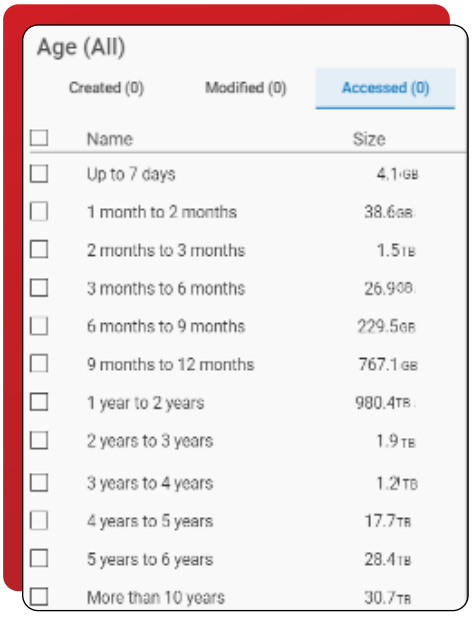

Figure 5. Filter data by last accessed time.

Determine the optimal cloud storage tier

Most cloud storage vendors support several different backup storage locations. It's important to be able to make informed decisions about what data should be stored where to make the most cost-efficient use of your cloud storage. By looking at access history, Information Studio can determine the last access time for your data. As shown in Figure 5, you can add a filter in Information Studio's UI for different time ranges to determine what files meet these criteria and use this information to make appropriate decisions about where to store data of different ages. For example, you could store:

- Dynamic data that has been accessed in the last 7 days **on-premises**
- Data last accessed 7–30 days in a **standard storage tier**
- Data last accessed 30–365 days in a **nearline storage tier**
- Data last accessed >365 days in a cold or **archive storage tier**

You can filter data by age both at the time of migration and at regular intervals after migration when you're considering moving data between tiers.

Adhere to compliance regulations

There are all kinds of regulations on data, including country and regional data privacy regulations as well as financial, healthcare and various government regulations. Each regulation stipulates specific requirements that apply for both on-premises and cloud-based data storage. Sometimes, these stipulations mean you need to store data in Information Studio comes with over 700 preconfigured data classification patterns and 120 built-in classification policies to identify common data privacy and regulatory compliance principles.

Using the built-in classification engine, Information Studio can search within files to see if they contain any regulated data identified by the classification policies.

You can use this information as part of your migration strategy to ensure the data you move to the cloud is stored appropriately in adherence with data compliance regulations. For example, before you migrate data, filter for all files containing personal data that hasn't been accessed in over three years and have Information Studio delete these files to reduce your risk. Alternatively, you can search for all files that contain credit card numbers or healthcare information and ensure that if these files are migrated to the cloud they are stored and secured appropriately.

You can also enforce ongoing adherence to these policies using Information Studio reports to identify when these policies are broken so you can make repairs. You can perform filtering and reporting automatically on a schedule you set.

Understand data locality

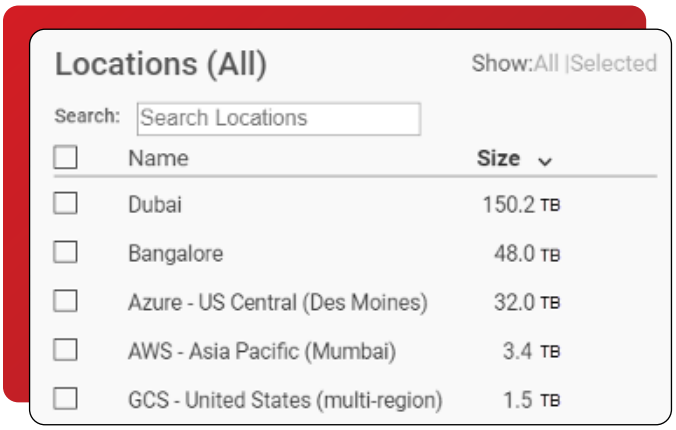

Figure 6. Add filters for specific locations and combine with other filters such as personal data from a specific region to check for data locality of personal or sensitive data.

Some data privacy regulations such as the EU's General Data Protection Regulations (GDPR) require personal data to be located within the same region as the person whose data it is unless there's a valid business
reason to store it elsewhere. This requirement applies to both on-premises and cloud-based data storage and also if the cloud storage is replicated to another site. To meet this requirement, you must be able to identify files containing personal or sensitive data from a variety of countries and determine where the files are located. Information Studio's built-in classification policies and patterns can identify personal and sensitive data about residents from over 35 countries.

To make sure you are complying, you can filter for personal or sensitive data and then add another filter to exclude data stored in that same region (see Figure 6). You should investigate the remaining files.

Information Studio's UI will display details about these files including a map view of where they are located, information about their age, the type of datastore they're contained within and more. You can then export the list of items and have someone evaluate and ensure there's a valid reason to keep these files where they are or rectify the situation by moving or deleting the files.

Another option is to add an additional filter for older age ranges. These items have a higher probability of no longer being required and you should consider deleting them to reduce risk.

Ensure appropriate protection

By identifying files with regulated data, you can ensure your backup data is stored appropriately in WORM with proper retention, on-premises or if business policies allow, securely in the cloud. Information Studio can not only identify files containing the regulated data, but also apply NetBackup policy filters to identify which files are backed up by which NetBackup policy (see Figure 7). You can use this information to validate the proper protection and storage is used for this regulated data.

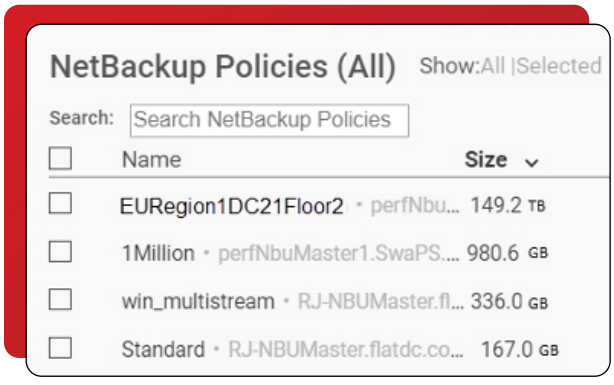

Figure 7. Filter files by NetBackup policy to ensure they're appropriately protected.

Reduce risk

When it comes to reducing risk, it's important to be aware of what data is being sent to the cloud. Business policies often dictate what type of data can safely be sent to the cloud and what should stay on-premises. It's common to see policies restricting personal or sensitive data and intellectual property from being stored in the cloud. To help organizations adhere to these policies, Information

Studio classifies data that can be filtered to identify the location of files containing this type of data (see Figure 8). You can then use this information to identify what data to migrate and what stays on-premises.

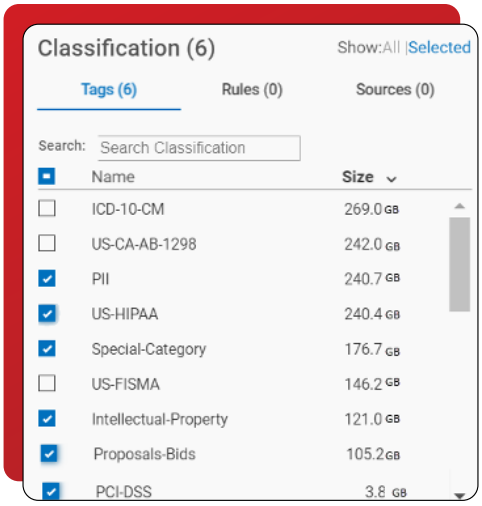

Figure 8. Filter for type of information such as personal or sensitive data or intellectual property.

On the other hand, if organizations allow storing some or all of this information in the cloud, you can use the information from Information Studio to ensure proper storage, security and retention measures are used. For example, it could validate if PCI-DSS data is stored on immutable storage with the legally required retention set.

Because Information Studio regularly updates this information from both on-premises and cloud-based data sources, it can help with ongoing enforcement by identifying when these business policies are broken so you can repair them. For example, some companies may require that sensitive data in both primary and backup copies is only stored on-premises on certain encrypted immutable storage devices. However, it's not uncommon for users to take a copy of this information and store it on their OneDrive shares. NetBackup policies may also store backup data in the cloud containing this sensitive data. Information Studio can help you find these rogue copies, identify their owners and take remedial actions. Information Studio reports can show how you discovered the bad behavior and then how you resolved it. You can perform filtering and reporting automatically on a schedule you set, which can help show regulators the effort you're making to properly protect this information.

SUMMARY

Information Studio provides organizations with the right information they need to make the data-driven decisions necessary to intelligently migrate data to the cloud. With Information Studio, they can decide what data to safely migrate and ensure it is stored appropriately while adhering to regulations and keeping their total cost of ownership (TCO) low.

This process starts with Information Studio's built-in scan and classification capabilities, which uncover the important information about your data including age, location, data privacy and regulatory-specific insights. It displays this information in an intuitive UI that presents map and graphical views you can dynamically update by adding filters to meet specific requirements. Information Studio also includes an action framework that can delete files that present unnecessary risk or generate a report you can use within your data migration strategy and to ensure ongoing compliance with regulations or business policies.

1. IDC ESS Workload Tracker

2. Veritas Databerg Research, https://www.veritas.com/news-releases/2020-04-21-veritas-technologies-projects-dark-data-to-waste-up-to-6-4-m-tons-of-carbon-dioxide-this-year

3. ESG Research Report, "Data Storage Trends in an Increasingly Hybrid Cloud World," March 2020

VERITAS

2625 Augustine Drive
Santa Clara, CA 95054
+1 (866) 837 4827
veritas.com

For global contact information visit:
veritas.com/company/contact

V0139 07/21