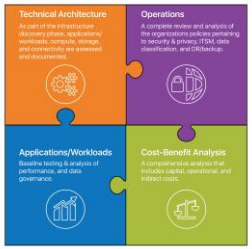
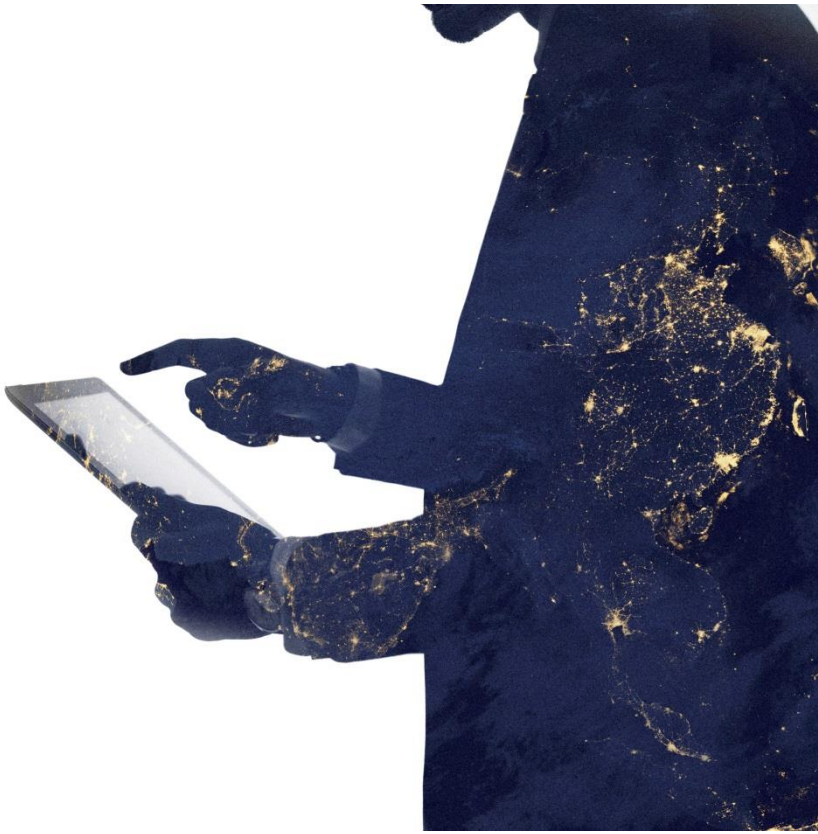
ADVISORY SERVICES

CLOUD READINESS ASSESSMENT

WG ADVISORY SERVICES





This report was prepared by WG Advisory Services for SPi Health and Safety on December 23rd , 2017

The information in this report is part of a Cloud Readiness Assessment framework developed by WG Advisory Services Inc. As part of the infrastructure discovery phase, applications/workloads, compute, storage, and connectivity are assessed in order to establish a baseline. Within the Technical Architecture phase, Storage is analyzed to determine: Capacity Utilization, Unstructured Data Lifecycle Management, and Performance Tiering.

DATA ASSESSMENT – ANALYSIS OF UNSTRUCTURED DATA

DATA ASSESSMENT - An analysis was performed on December 18th, 2017 on the following data volumes (as provided by Fujitsu): **E:\DATA\PARTAGE**

The total amount of utilized disk for this volume: **2.49 Terabytes**

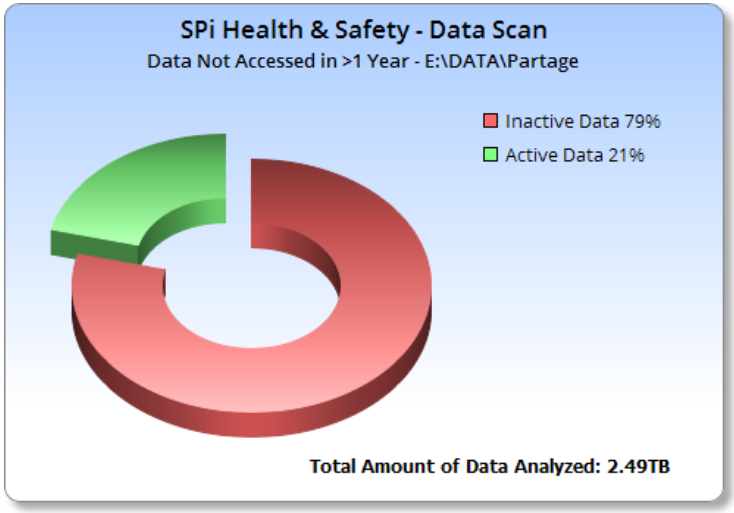
CRITERIA		ASSESSMENT
DATA ACCESS Analysis of last access dates using a threshold of >1 year.		The assessment determined that a substantial percentage of data has not been accessed in more than 1 year (79%). It is also worth noting that an additional 48% of data has not been accessed in over 2 years. This has resulted in a large amount of "stale" inactive data residing on production storage.
DATA MODIFICATION Analysis of last modification dates using a threshold of >1 year.		The assessment determined that a substantial percentage of data has not been modified in more than 1 year (81%). It is also worth noting that an additional 67% of data has not been modified in over 2 years. This has resulted in a large amount of "stale" inactive data residing on production storage.
FILE TYPE Analysis of file types by group types		The bulk of the files located on the assessed volume were determined to be Office & Multimedia filetypes. There were some unknown files (other) flagged and a deeper dive into the actual filetypes being stored on the volume may be recommended at a future date.
CAPACITY MANAGEMENT Review of storage occupation as compared to available disk		There appears to be no formal capacity management in place. The assessed volume is currently at 90% utilization.

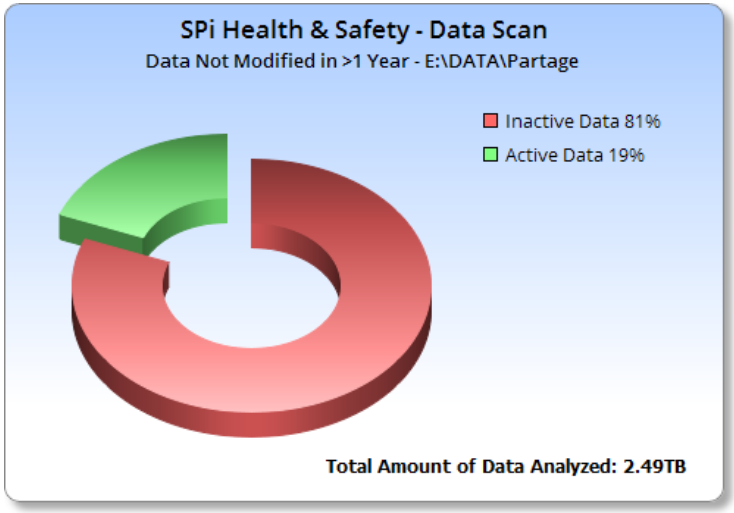
RECOMMENDATIONS

- Start the process of moving "stale" data to lower cost disk. This will free up expensive primary storage and also provide the organization with a clear understanding of production storage requirements. This will also have an immediate effect on backup and Disaster Recovery procedures.
- Consider utilizing Object-based storage - By implementing an object-based cloud storage solution the organization will realize the following benefits:
  - Data protection for archived files
  - Ability to access archived data quickly and efficiently
  - Lower costs for primary storage
  - Lower costs for future disaster recovery implementation
- Good records management includes both backup and archiving. However, while these terms are often used interchangeably, it is important to distinguish between them when considering a records management process. Backup is used for data recovery, while archiving is used for preserving and retrieving data in the event of a disaster, inquiry or litigation. In simple terms, think of backup as short-term and archival as long-term.
- Develop a formal data archiving strategy
- An archiving strategy will help ensure the organization is meeting regulatory requirements

Data Assessment – Highlighted Graphs

The graphs below illustrate the substantial amount of Inactive "stale" data residing on production storage at the managed service providers datacenter. As outlined in the recommendations section of the report card, it is critical for SPi to develop a strategy that will move inactive data (as defined by the organization) to lower cost, secure storage.





Backup & Archiving

Good records management includes both backup and archiving. However, while these terms are often used interchangeably, it is important to distinguish between them when considering a records management process. Backup is used for data recovery, while archiving is used for preserving and retrieving data in the event of a disaster, inquiry or litigation. In simple terms, think of backup as short-term and archival as long-term.

Specifically, **backup** is a snapshot or picture of the state of the data before it disappeared or was destroyed, with the data periodically overwritten as it changes. In the backup process, a copy of data at a specific point in time is created in case something should happen to the original. Therefore, in the case of a failure, the data can be reconstructed from that time. This concept is very similar to backing up a Word document either automatically or with the "Save" button on a computer, except that the backed up data is stored on a remote device. Fixed disk storage devices are typically used for backup because of their speed and ability to provide instant access to the data on the disk.

On the other hand, ***archiving*** is long-term and unalterable. Used for compliance or disaster recovery, redundancy and physical separation are crucial to effective archiving. Archived files can be kept for decades, usually at an offsite location, and two or more remote copies are better than one to properly safeguard this data.

Contrary to popular belief, archive systems aren't junk technology kept in service so they can perform the lowly task of storing old data no one cares about. They're part of an ecosystem of storage systems no more or less critical than your nimblest solid-state drive or hard disk drive hybrid array placed behind your most mission-critical transaction processing system. Without the archive system, the transaction system can't function cost effectively or at peak efficiency.

Still, some draw a distinction between *deep archive* and *active archive*. The former is what we have traditionally considered an archive to be: a collection of data with historical business value or specific regulatory or legal retention requirements that's seldom if ever accessed. Deep archive systems, because of their limited re-reference rates, are designed with specific attention to the container that will be used to hold data for later use and to the technology to which that container is written -- specifically, to the longevity of the media, including its interoperability with future data access technologies. It does no good to have a 20-year-old archive of kept technology that can no longer interface with contemporary servers or operating systems. That's a big concern of deep archiving.

Active archive, by contrast, refers to storing data that, once written, changes very infrequently. However, it's data that may be read a lot but not modified, so it presents a different set of storage requirements than either read-, write- or modify intensive primary storage systems or read-, write-, modify, never deep-archive systems. Think video: Once the television episode is recorded, it won't be modified, but it might be replayed (read) many times. The video file is archival data, but it's still active (read).

This type data is ideal for Object-based storage.

Object Storage Defined

With *block storage*, files are split into evenly sized blocks of data, each with its own address but with no additional information (metadata) to provide more context for what that block of data is. You're likely to encounter block storage in the majority of enterprise workloads; it has a wide variety of uses (as seen by the rise in popularity of SAN arrays).

Object storage, by contrast, doesn't split files up into raw blocks of data. Instead, entire clumps of data are stored in an *object* that contains the data, metadata, and the unique identifier. There is no limit on the type or amount of metadata, which makes object storage powerful and customizable. Metadata can include anything from the security classification of the file within the object to the importance of the application associated with the information. Anyone who's stored a picture on Facebook or a song on Spotify has used object storage even if they don't know it. In the enterprise data center, object storage is used for these same types of storage needs, where the data needs to be highly available and highly durable.

Object storage works very well for unstructured data sets where data is generally read but not written-to. Static Web content, data backups and archival images, and multimedia (videos, pictures, or music) files are best stored as objects. Databases in an object storage environment ideally have data sets that are unstructured, where the use cases suggests the data will not require a large number of writes or incremental updates.

Geographically distributed back-end storage is another great use case for object storage. The object storages applications present as network storage and support extendable metadata for efficient distribution and parallel access to objects. That makes it ideal for moving your back-end storage clusters across multiple data centers.

[1]Data Lifecycle Management (DLM)

Data Lifecycle Management (DLM) is a policy-based approach to managing the flow of an information system's data throughout its lifecycle – from creation and initial storage, to the time it becomes obsolete and is deleted, or is forced to be deleted through legislation.

DLM products attempt to automate processes involved, typically organizing data into separate tiers according to specified policies, and automating data migration from one tier to another based on those criteria. As a rule DLM stores newer data, and data that must be accessed more frequently, on faster, but more expensive storage media. Less critical data is stored on cheaper, but slower media.

[1] A proper DLM product provides software for data archiving, retention, regulatory compliance, storage consolidation, backup and recovery optimization, and capacity management across a tiered-storage infrastructure. The software redefines storage management by enabling organizations to profile the criticality of their data within its lifecycle to automate discovery, classification, and placement on the most appropriate storage resource. The software operates across heterogeneous DAS, NAS, and SAN environments without the deployment of agents.