Draft Document for Review August 31, 2023 9:26 am SG24-8541-00

Redbooks
ibm.com/redbooks

Cyber Resiliency with IBM Storage Sentinel and IBM Storage Safeguarded Copy

Axel Westphal
Christopher Vollmar
Daniel Thompson
David Green
Gerd Franke
Guillaume Legmar
Markus Standau
Nezih Boyacioglu
Thomas Gerisch
Vasfi Gucer

Storage

IBM

Redbooks

IBM Redbooks

Cyber Resiliency with IBM Storage Sentinel and IBM Storage Safeguarded Copy

April 2023

Note: Before using this information and the product it supports, read the information in “Notices” on page xiii.

First Edition (April 2023)

This edition applies to:

IBM Storage Copy Data Manager Version 15 2.2.19.0
IBM Storage Sentinel Version 1.1.2
Security Scan Engine Version 8.1.0 - Build 1.4
InterSystems IRIS Version 2022.1.1.374.0

Contents

Figures

Tables

Examples

Notices

This information was developed for products and services offered in the US. This material might be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, MD-NC119, Armonk, NY 10504-1785, US

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you provide in any way it believes appropriate without incurring any obligation to you.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to actual people or business enterprises is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks or registered trademarks of International Business Machines Corporation, and might also be trademarks or registered trademarks in other countries.

AIX®	IBM FlashSystem®	Redbooks (logo) ®
DS8000®	IBM Spectrum®	Resilient®
FlashCopy®	PowerVM®	Storwize®
HyperSwap®	QRadar®	Tivoli®
IBM®	Redbooks®	XIV®

The following terms are trademarks of other companies:

ITIL is a Registered Trade Mark of AXELOS Limited.

The registered trademark Linux® is used pursuant to a sublicense from the Linux Foundation, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis.

Microsoft, Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Red Hat, are trademarks or registered trademarks of Red Hat, Inc. or its subsidiaries in the United States and other countries.

VMware, VMware vSphere, and the VMware logo are registered trademarks or trademarks of VMware, Inc. or its subsidiaries in the United States and/or other jurisdictions.

Other company, product, or service names may be trademarks or service marks of others.

Preface

IBM® Storage Sentinel is a cyber resiliency solution for SAP HANA, Oracle, and Epic healthcare systems, designed to help organizations enhance ransomware detection and incident recovery. IBM Storage Sentinel automates the creation of immutable backup copies of your data, then uses machine learning to detect signs of possible corruption and generate forensic reports that help you quickly diagnose and identify the source of the attack. Because IBM Storage Sentinel can intelligently isolate infected backups, your organization can identify the most recent verified and validated backup copies, greatly accelerating your time to recovery.

This IBM Redbooks® publication explains how to implement a cyber resiliency solution for SAP HANA, Oracle, and Epic healthcare systems using IBM Storage Sentinel and IBM Storage Safeguarded Copy. Target audience of this document is cyber security and storage specialists.

Authors

This book was produced by a team of specialists from around the world .

Axel Westphal is an IT Specialist for IBM Storage Systems at the IBM European Storage Competence Center (ESCC) in Mainz, Germany. He joined IBM in 1996, working for Global Services as a System Engineer. His areas of expertise include setup and demonstration of IBM System Storage products and solutions in various environments. Since 2004, Alex has been responsible for storage solutions and Proof of Concepts conducted at the ESSC with DS8000®, SAN Volume Controller, and IBM XIV®. He has been a contributing author to several XIV and DS8000-related IBM Redbooks publications.

Christopher Vollmar is an IBM Certified Consulting IT Specialist (Level 3 Thought Leader) and Storage Architect who is based in Toronto, Ontario, Canada with the IBM Systems Group. Christopher is focused on helping customers build storage solutions by using the IBM Spectrum® Storage Software-Defined Storage (SDS) family. He is also focused on helping customers develop private and hybrid storage cloud solutions by using the IBM Storage Virtualize family and Converged Infrastructure solutions. Christopher has worked for IBM for more than 20 years across many different areas of IBM, and has spent the past 12 years working with IBM System Storage. Christopher holds an honours degree in political science from York University.

Daniel Thompson has been working in IT for more than 40 years. His specialty is data protection (Backup and Restore, Disaster Recovery, Business Continuity and Cyber Resiliency). He currently works in the Advanced Technology Group (ATG), IBM Technology, Americas.

David Green works with the IBM SAN Central team troubleshooting performance and other problems on storage networks. He has authored, or contributed to, several IBM Redbooks publications. He is a regular speaker at IBM Technical University. You can find his blog at *Inside IBM Storage Networking* at https://www.insidestoragenetworking.com/ where he writes about all things related to Storage Networking and IBM Storage Insights.

Gerd Franke is a Senior IT Architect in the EMEA Storage team of IBM Client Engineering. He is known as an expert for Storage strategy and architecture, focusing on Cyber Security and Resiliency, Hybrid Cloud Storage and Modern Data Protection. He often leads large and complex client projects and is a regular speaker at IBM conferences. Gerd has worked for IBM in various national and international roles for more than 30 years. He holds an engineer's degree in Electrical and Communications Technology.

Guillaume Legmar has been involved with IBM Storage solutions for more than 20 years. Currently, he is a part of the Montpellier Garage to develop demonstrations about IBM FlashSystem® and cyber resiliency. He is also a member of the IBM FlashSystem Beta and IBM Storage Virtualize Beta teams.

Markus Standau Markus Standau works for IBM Germany. He has more than 20 years of experience in the storage field in roles such as services, technical sales, and worldwide product management. He currently works in the storage sales acceleration team as the offering leader for Storage Virtualize, FlashSystems, and the Storage Control family in DACH. In his current role Markus work on the organization of various business partner and customer events in DACH, such as IBM Storage Strategy Days, the SVC/FlashSystem user group and more. He holds a degree in Computer Science from Baden-Wuerttemberg Cooperative State University. Markus is the co-author of several IBM Redbooks on IBM Spectrum Control and its predecessor, Tivoli® Storage Productivity Center (TPC).

Nezih Boyacioglu has 20 years of experience as an SAN Storage specialist and currently works for IBM Premier business partner Istanbul Pazarlama in Turkey. He has over 20 years in the IT arena. His IBM storage journey starts with Tivoli Storage Manager and tape systems and his main focus for last 10 years has been on IBM Storage Virtualize family (IBM SAN Volume Controller, Storwize®, and FlashSystem), and Storage Area Networks. He is an IBM Certified Specialist for Enterprise Storage Technical Support, Flash Technical Solutions, Virtualized Storage, and Storage Spectrum software.

Thomas Gerisch is an IT Specialist working at Client Engineering Storage EMEA, focused on helping customers to build valuable solutions using their IBM Storage assets. Joined IBM in 1999, he started as an instructor and specialist for open systems, before he became a storage expert at the IBM EMEA Storage Competency Center (ESCC) in Frankfurt, Germany. He has a lot of expertise in IBM storage development and testing. His current job role includes being the technical focal point for IBM customers running SAP HANA on IBM Storage. Thomas authored several IBM White Papers about SAP HANA and IBM storage environment.

Vasfi Gucer works as the Storage Team Leader on the IBM Redbooks Team. He has more than 30 years of experience in the areas of systems management, networking hardware, and software. He writes extensively and teaches IBM classes worldwide about IBM products. His focus has been primarily on storage, cloud computing, and cloud storage technologies for the last 8 years. Vasfi is also an IBM Certified Senior IT Specialist, Project Management Professional (PMP), IT Infrastructure Library (ITIL) V2 Manager, and ITIL V3 Expert.

Thanks to the following people for their contributions to this project:

James Munro, Martin Purkis
IBM UK

Brian Sherman, Pepe Lam, Dan Zehnpfennig, Matt Key, John Bernatz
IBM US

Michael Frankenberg
IBM Germany

Joseph Hand, Jim McGann
Index Engines

The IBM Advanced Technology Group (ATG) provided the environment and technical support used by the authors of this IBM Redbooks. Additionally one of the authors is a member of the ATG.

The ATG (https://www.ibm.com/support/pages/advanced-technology-group) supports IBM Clients and Partners by providing demonstration environments for a variety of IBM solutions, hosting Accelerate with IBM web events that cover a variety of technical topis of interest to our clients as well as workshops that allow clients to work with technical specialists on a variety of topics.

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an IBM Redbooks residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our books to be as helpful as possible. Send us your comments about this book or other IBM Redbooks publications in one of the following ways:

- Use the online **Contact us** review Redbooks form found at:

  ibm.com/redbooks

- Send your comments in an email to:

  redbooks@us.ibm.com

- Mail your comments to:

  IBM Corporation, IBM Redbooks
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- Find us on LinkedIn:

  http://www.linkedin.com/groups?home=&gid=2130806

- Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm

- Stay current on recent Redbooks publications with RSS Feeds:

  http://www.redbooks.ibm.com/rss.html

1

Introduction

In this chapter we introduce the concepts of cyber resiliency and we explain how cyber security is different from cyber resiliency. We will also give an overview of IBM Storage Sentinel, the currently supported applications and we introduce the use cases we are addressing. Finally, we will cover the IBM Storage Sentinel workflow.

This chapter has the following sections:

IBM Storage rebranding: In January 2023, IBM announced that the IBM Spectrum software defined storage products will be renamed to IBM Storage products. For example, IBM Spectrum Copy Data Management and IBM Spectrum Sentinel are renamed IBM Storage Copy Data Management and IBM Storage Sentinel. You will likely find documentation under both the Spectrum and Storage names for some time, and that documentation may refer to either or both naming conventions.

1.1 What is cyber resiliency?

All businesses are exposed to cyber attacks. These attacks often target applications that are critical to a business. Data or applications can be encrypted, stolen or both. Historically, disaster recovery and business continuity efforts focused on software and hardware failures. Businesses design redundancy into systems and storage, and use technologies such as backups and data replication to try to prevent the loss of data.

Many businesses are not prepared for, or are unaware of the extent of damage that a cyber attack can cause. They are also unaware of the costs of recovery from a cyber attack. Many of the businesses that do take steps to guard against cyber attacks focus efforts on prevention and not now to recover quickly from an incident.

Cyber resiliency is a measure of how well the applications and other infrastructure of a business can withstand events such as cyber attacks and natural disasters and still deliver business operations at a normal level. Cyber resiliency is critical for business continuity. It helps reduce financial losses and in some cases damage to the business's reputation. *A cyber-resilient company has a competitive advantage because of efficient and effective operations and can maintain or even grow business during a crisis if its competitors cannot*.

1.1.1 Cyber security versus cyber resiliency

A Cyber security framework (CSF), such as that by the National Institute of Standards and Technology (NIST) is broader than recovery. Recovery in the event of a ransomware attack can be very challenging, as key service data is increasingly fragmented across different data stores, both within and external to the organization. This complexity makes it even more difficult to identify the ransomware attack variant and the entry point of a corruption or encryption event, and then eradicate the risk of reinfection. It also increases the complexity of recovering and testing data to ensure parity and synchronicity across data stores, required to ensure that the service is recovered in a safe and correct sequence.

Organizations that are able to focus on cyber security as well as cyber resiliency, improve their ability to recover from corruption or encryption-based events and ransomware style attacks. The two principals working in concert provide, not only capability of mitigating the attempts to disrupt the business by bad actors, but also provide for the ability to quickly recover the data that supports the organization. Cyber resiliency can be seen as the way for an organization to recover, test, restore environments and in the face of a ransomware, data corruption or encryption event.

Figure 1-1 on page 3 shows the differences between cyber security and cyber resiliency.

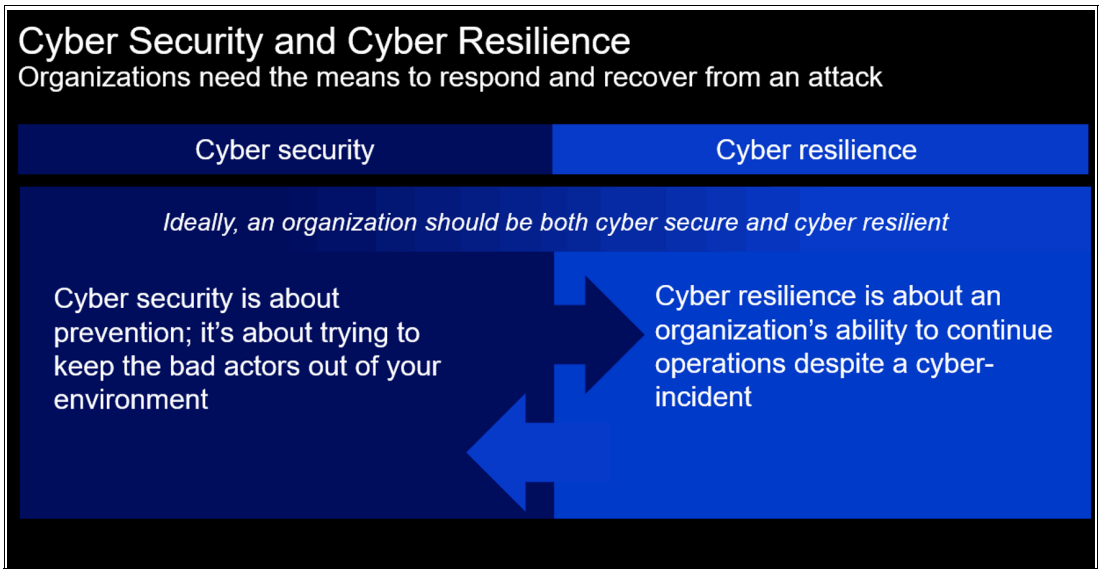


Figure 1-1 Cyber security and cyber resilience

Cyber security

Cyber security is the methods or practices that an organization uses to protect its systems and critical information from digital attacks. It is also known as Information Technology (IT) security. Cyber security measures are designed to combat threats against applications and networked applications. These threats can come from both inside and outside of organizations.

Cyber security includes IT intrusion detection and prevention, data loss or theft prevention, ransomware protection and protections for services running in the cloud. As organizations go mobile and employees work from mobile devices, mobile security is also a growing concern and must be included in a robust cyber security policy. Other concerns include minimizing potential damage from internal threat actors who already have privileged access to critical IT infrastructure.

Cyber security attacks often result in data theft, ransomware attacks or both. This data can include company secrets such as source code, or valuable customer and employee personal data.

Cyber resiliency

Where cyber security is focused on solely on intrusion detection and data loss prevention, cyber resiliency is a measure of an organization's systems to survive a disaster or cyber event and still allow the organization to function. Cyber resilience includes cyber security as a component. Cyber resilient organizations will be more secure than an organization that is not cyber resilient, but cyber resiliency is a measure of how well the organization functions during a cyber event and how quickly it can recover after an event (or a disaster event) occurs.

Cyber resiliency begins with a strategy or plan. This strategy identifies the critical assets that matter most to the organization and its stakeholders. Assets include the data or information that must be protected and is critical to the function of the organization, and the systems and services that matter most. The strategy must also include identifying the vulnerabilities and risks and organization faces.

The next part of cyber resiliency is the design. Design work chooses the controls, procedures, and training that are appropriate to prevent harm to critical assets. However, the design must

be practical. An impractical design that cannot or will not be implemented is not an effective one. The design work also should identify who has what authority to make decisions and act on them.

After the design is complete, the organization transitions to a test operational state. This tests where possible and closely monitors critical assets where it is not possible to test beforehand. The monitoring identifies when critical assets from the design phase are impacted by internal or external action. The design may be refined based on testing results.

After testing is complete the organization moves to an operational state. In this phase, the design has been deployed. There is still testing being done using controls to ensure that the operational state is effective and consistent.

From the operational state, an organization with a mature cyber resilient design will move to evolution. Environments are constantly changing with new threats and new technologies. Organizations will learn from incidents and how they recover from them. They will need to modify procedures, training, and even strategy as they learn.

IBM Storage Sentinel should be part of an organization's cyber resilience strategy.

1.2 Approaches to data resiliency

The ability to recover to a prior point in time relies on the availability of backups - either point-in-time, array-based snapshots (such as copies made of Primary Workloads) or written to backup applications and their repositories such disk, tape (or virtual tape (VTL)) or cloud (sometimes called Secondary Workloads). These recovery options require the availability of a system to enact recovery and assume the data being restored has not been compromised in some way. This assumption is generally false when faced with recovering from a ransomware level event where the system or data is locked/corrupted/encrypted.

Backup-up data (both Primary and Secondary copies), available for use in recovery scenarios, must be free of contamination to eradicate the risk of repeated attack and immediate reinfection. Even where the backup is written to immutable storage, the backup must be validated as being free of infection prior to supporting recovery. Restore from backups traditionally is limited to single system, single files, or relatively small volumes of data. They are not designed to restore mass volumes of data to multiple systems in a short space of time. Recovery from tape-based systems (or Virtual Tape Libraries) is limited by seek time, retrieval times and network bandwidth, restricting the ability to recover at scale and speed.

Even when the Disaster Recovery systems, and restored data, are available, confidence in running business services using the secondary site is typically low, either because it hasn't been tested sufficiently and/or due to differences between the primary and secondary configurations or their integration to other interfaced components. Testing is generally not representative of realistic failure scenarios and is typically based around site switching of single systems or single applications, which are quiesced and stopped gracefully first at the active site before being initialized at the recovery site. Often, the test at the secondary site does not go as far as processing transactions, or at best, for a limited period before switching back to the primary.

What can be leveraged from traditional Disaster Recovery (DR) testing is the priority and sequencing of applications, data and infrastructure dependencies, in support of the recovery of business-critical systems and applications and elements of Service Management. More mature traditional DR testing may have considered recovery from loss of data and restarting

systems, applications, and the associated business services to a prior point-in-time (RPO>0), even if the existing solution encompasses synchronous data mirroring. Complexities associated with microservices, distributed systems and the reconciliation of data across system boundaries, interfaces and with third parties, should also have been considered previously by traditional DR scenarios. Existing *Business Impact Assessments* (BIA) should also be leveraged in helping to determine appropriate recovery strategies for business services, including potential loss scenarios, prioritization, and service dependencies.

1.2.1 Data resilience in practice; not all data is created or valued equally

Some data is transactional and very volatile whereas other data is relatively static, however both types of data are important and support the organization in different ways. For example, in the Energy and Utilities sector trading applications generate highly active transactional data while seismic records are static, however both data types are regulated and could easily constitute the organizations Minimum Viable Business.

Both of those data types can be protected, but just in different ways. Figure 1-2 on page 5 highlights that while not all data needs the same level of protection because of factors like its activity, an organization can protect its data in various ways to support both its criticality and its activity.

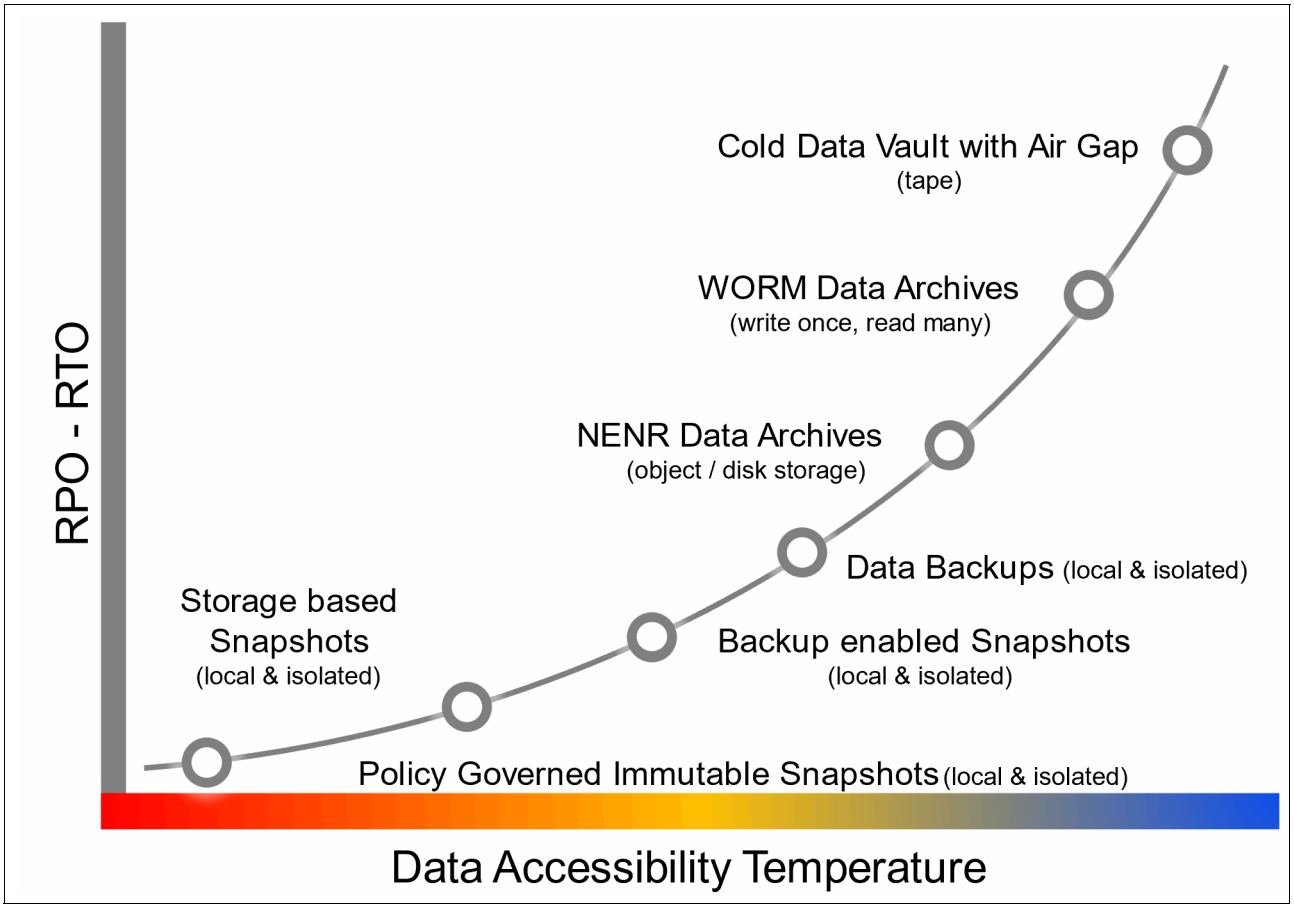


Figure 1-2 IBM's Point of View: Levels of data resilience

The most critical and highly transactional workloads generally need the lowest RTO and RPO and would need the quickest accessibility to the Cyber Resilient® copies using something like Policy Governed Immutable Snapshots. Whereas static data like seismic data or long-term

records that may have regulatory retention requirements still need to be protected, but given their nature and size may need a different level of protection since they are not changing. This could be achieved with something like a Cold Data Vault, potentially still using tape-based media.

1.2.2 Time to Recover

The ability of the organization to restore business services in a timely manner is key to success and in remaining within Impact Tolerance or in meeting the *Maximum Tolerable Period of Disruption (MTPD)*. There are additional considerations and therefore contributors to elapsed time of recovery from ransomware attack. This section identifies key considerations, which contribute to the elapsed duration of outage and recovery time.

The identification of the extent of the attack, including point of infection, nature of the attack, and blast radius (how widespread the damage) is critical and is therefore a key consideration in the time taken to recover. Understanding these elements will help determine the scope of recovery and what specific recovery actions are required. It will help determine what services are impacted and what services can continue as they are.

The number and nature of infected systems will have a direct bearing of the time to recover, as retrieval of data and restoration of data at scale will likely be limited by network bandwidth and the I/O capability of the source data repository, for example, the backup system. Whether recovery involves repairing a single file, dataset or the complete restoration of a data volume, complete system, or server farm will influence time that is spent on recovery. The bandwidth available from the source repository and its media will impact how quickly data can be transported the ability to restore data and recover systems at scale. Recovering large volumes of data from magnetic tape will no doubt take more time than *flashing* a disk subsystem volume.

There is a need to ensure that contaminated data is not reintroduced to the system as part of the recovery process following a ransomware attack. Eradication of risk of reinfection by ensuring that all trace of the infection has been removed prior to recovering *good data* and restarting services to existing systems must be undertaken. Recovery of services to an alternate, isolated system may be possible, if available, and executed concurrently to the disinfection of the primary system.

As traditional HADR solutions are not content aware, additional checks, validation, and technical solutions are required to restore service. Ideally, these validations and checks are made pre-event so that in the eventuality of using a prior copy of data to facilitate recovery, the integrity and useability of the data is already known. However, this may not always be the case and validation of data may be required during the recovery event itself, which will elongate the recovery time.

Where data has been restored to a prior point-in-time (RPO>0), there will be a need to roll forward or replay *good* transactions to ensure that a transactionally consistent state is returned. Prior to restarting business services, there will be a need to test the use of the restored data and its consistency. Reconciliation of data across multiple, interfaced systems and 3rd parties is also likely to be a key consideration and contributor of elapsed time during recovery. Reconciliation of data may be required intra-system (applications that are hosted on the same platform), inter-system (applications across disparate systems), with 3rd parties and across the eco-system in which the recovered system operates.

Regarding Figure 1-3 on page 7, technology solutions such as IBM Cyber Vault for the protection of Mainframe and Open Systems data, provides an ability to regularly backup data to immutable storage on the primary storage systems at either production and/or DR sites and provide validation of data in an air-gapped system. Having validated copies of data and

an environment to restore, inspect and enact recovery, significantly reduces the time to execute recovery and greatly improves the chance of a successful recovery than otherwise would be available.

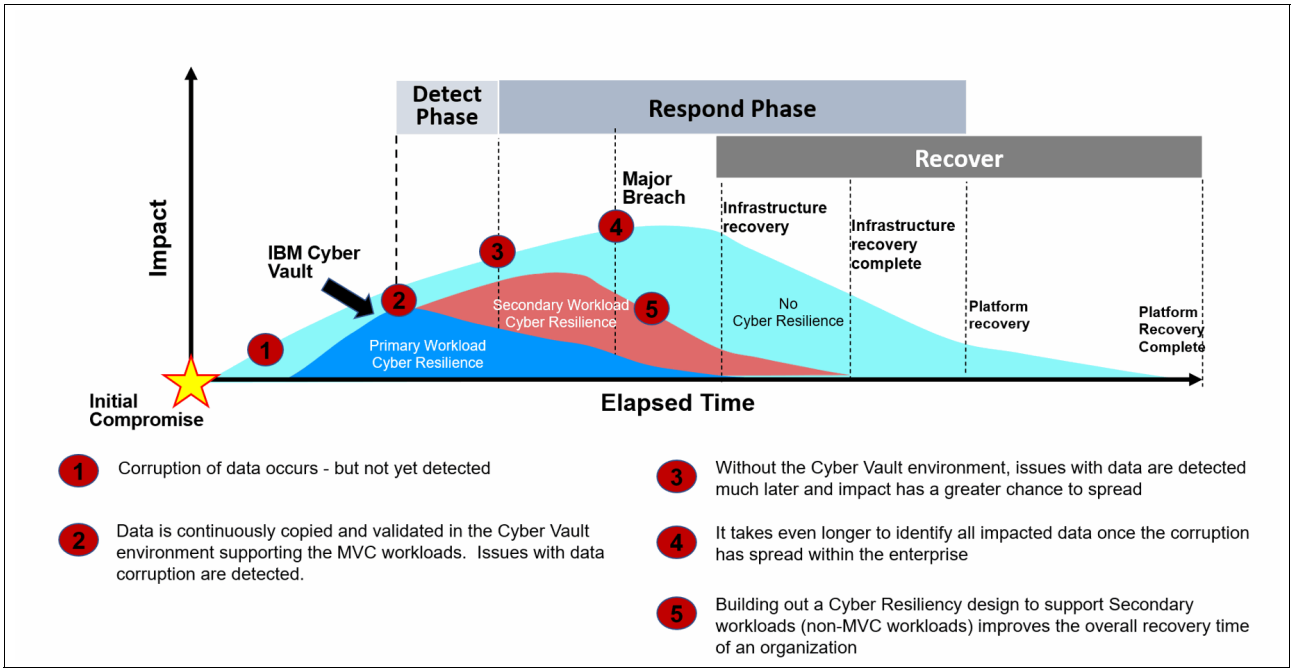


Figure 1-3 Recovery from ransomware attack timeline

1.2.3 Secondary Workload cyber resiliency

Looking at the large collections of systems and data that typically comprises an enterprise once an organization has prioritized its Minimum Viable Business (MVB) assets for rapid recovery, they are left with the reminder of their application catalog that may need to be addressed depending on the spread of the corruption or encryption. Building out additional capabilities for these workloads to support a data resiliency strategy accelerates the ability for an organization to recover the remaining workloads in a timely manner and returned to normal operations. Part of Secondary Workload strategy can include:

- Further prioritization of workloads for testing, recovery, and validation procedures.
- Regular backup copies being moved to alternate immutable data platforms.
- Moving copies from primary systems to secondary environments such as off the array.
- Including a random sampling of workload for full recovery and validation.

This leads to the ability to segment processes that support both the Minimum Viable Business as well as the remainder of the organizations workloads that are needed, see Figure 1-4 on page 8. It also provides for the ability to use different testing and validation methods, both types can carry automation, but where MVB workloads may have targeted testing and application-specific tools, the secondary workloads can take a more mass scale approach using more common set of generic tools. Proactively building an approach to support both Primary and Secondary recovery can mean accelerated recovery for both reducing overall organization disruption time.

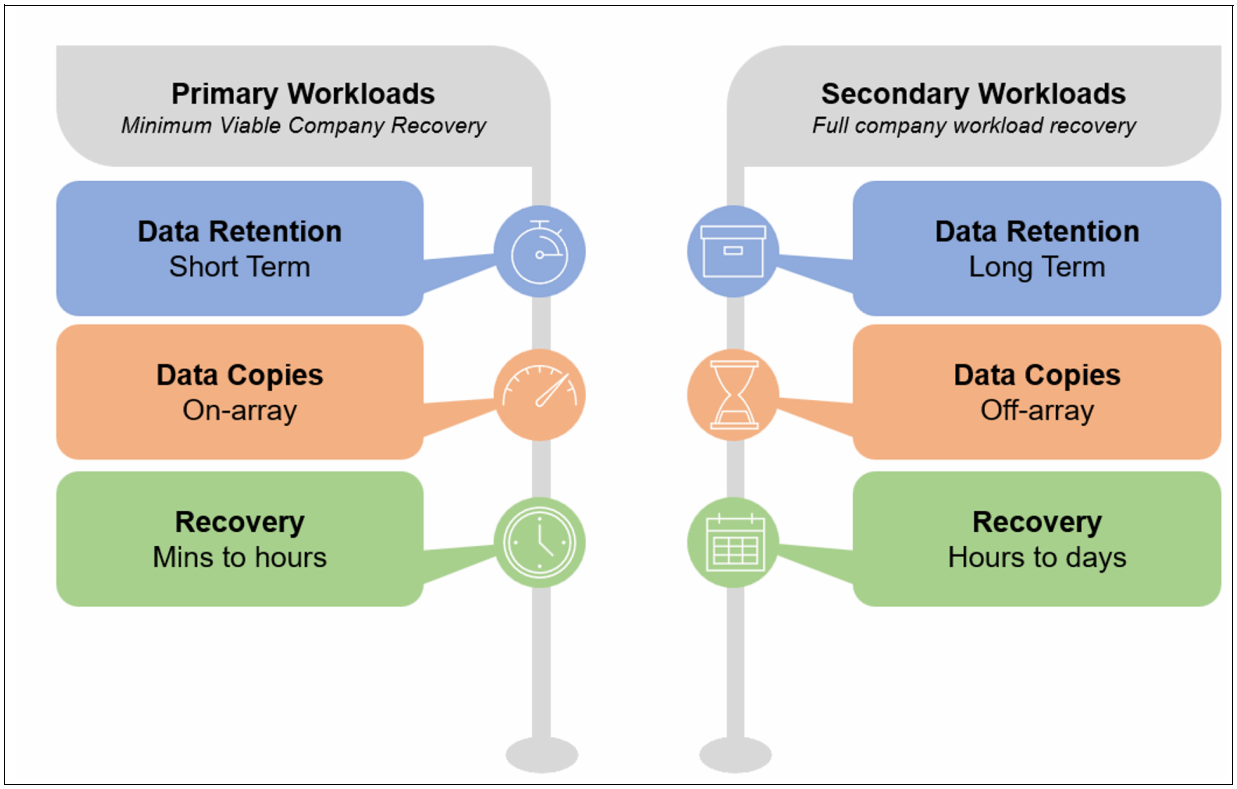


Figure 1-4 Segmentation of workloads

1.3 IBM Storage Sentinel overview

Organizations of all sizes in every industry that is threatened by increasingly malevolent ransomware and other cyber threats. Even with the strongest defensive measures, there is always a risk that threats will get past every barrier and penetrate an organization's information supply chain. In addition to the direct financial costs of lost business and recovery, these attacks can severely damage a company's brand, especially in industries where reputation is critical, such as financial services or healthcare.

Many organizations use some variant of a 30/60/90 policy for data backup and retention. Snapshots are captured hourly and/or daily with full backups generated every 30,60 and 90 days. Bad actors will exploit this by penetrating an organization’s cyber defenses, install malware and then leave it dormant for long enough to fully infect the business’s production and backup data. When the malware is triggered, it encrypts both production data plus all of the backup copies. In these situations the victim’s only alternative is to pay the ransom.

You can find a full report on cyberattacks and there effects on companies here:

IBM X-Force 2023 Cyber Security Report

IBM has also released a study on the impact of data breaches here;

IBM Cost of a Data Breach Report

What is ransomware?: Ransomware is an online attack that is perpetrated by cyber criminals or nation state-sponsored groups who demand a monetary ransom to release their hold on encrypted or stolen data.

A ransomware infection can be costly and disruptive if the only solution to return to normal business operations is to pay the cyber criminals' ransom. Statistics show, that only 50% of ransomware victims get back access to their data, even if the ransom was paid. One alarming trend is that cyber criminals now install malware and leave it dormant for 100 days or more before springing the trap. At that point, the malicious code has infected not only the target's production data systems and snapshots, but all of their backup copies, even if they use a $30 - 60 - 90$ *backup policy*. The victims have little choice but to pay up.

Ransomware attacks can use several methods to infect a device or network. Some of the most prominent malware infection methods include:

- **Phishing emails and other social engineering attacks:** Phishing emails manipulate users into downloading and running a malicious attachment (which contains the ransomware that is disguised as a harmless looking .pdf, Microsoft Word document, or another file), or into visiting a malicious website that passes the ransomware through the user's web browser.
- **Operating system and software vulnerabilities:** Cyber criminals often exploit existing vulnerabilities to inject malicious code into a device or network.

IBM Storage Sentinel (Sentinel) is a solution that is designed to help organizations detect ransomware and recover from cyber security incidents. It automatically creates immutable copies of data, then will use machine learning to detect possible corruption. It can generate forensic reports to help diagnose problems and find the source of an attack. Because it can isolate infected backups that you can identify which backup copies are verified and which are the most recent ones. This accelerates your time to recovery.

1.3.1 Supported applications

IBM Storage Sentinel supports the following applications.

SAP HANA

SAP HANA is a database that stores data in system memory instead of on disk. This enables processing data at speeds that are magnitudes faster than disk-based systems. This allows for advanced, real-time analytics to be performed on the data. SAP HANA can be used on premises, in the cloud or in a hybrid cloud and deployed in both locations. SAP HANA can apply machine learning and AI to data from multiple areas of a business. For example, it can integrate data from:

- Traditional documents - spreadsheets, contracts, and so forth
- Emails, website forms and other user experience documents
- Internet of Things (IoT) - such as data from sensors in warehouses or trucks, security sensors, RFID tags and the many types of sensors in all aspects of a business
- Mobile - data from the mobile devices of your customers and employees

SAP HANA can integrate and analyze the vast amounts of data that sits in data warehouses and provide no value unless it is analyzed to provide more customer value and increase business impact. The capabilities for SAP HANA include:

Backup

The following types of backup are available for SAP HANA:

- FlashCopy® NoCopy
- FlashCopy Incremental
- Global Mirror with Change volumes
- Safeguarded Copy

There is also an option to backup the log file.

Restore

Copy Data Services Manager creates a temporary volume from a backup, then mounts it to the original server for recovery.

Refer to Chapter 3, “Protecting Epic Cache and IRIS Databases with IBM Safeguarded Copy and IBM Storage Sentinel” on page 31 for more information.

Epic

EPIC is electronic healthcare record (EHR) software that covers all functions of healthcare operations. This includes patient records, patient engagement, billing, mobile, clinical data from medical tests, interoperability, specialist care, and even government regulations. EPIC uses two database technologies:

- An operational database that handles online transactions. This database runs Cache’ from Intersystems Corp.
- An analytical database that can run on either Microsoft SQL Server or Oracle.

Refer to Chapter 4, “Configuring IBM Storage Sentinel for SAP HANA” on page 51 for more information.

Oracle

In June 2023 (with IBM Storage Sentinel Version V1.1.4), IBM announced support for Oracle DB running on both Linux and IBM AIX®.

Further information can be found here.

An example of an Oracle Backup job is shown in Figure 1-5 on page 11.

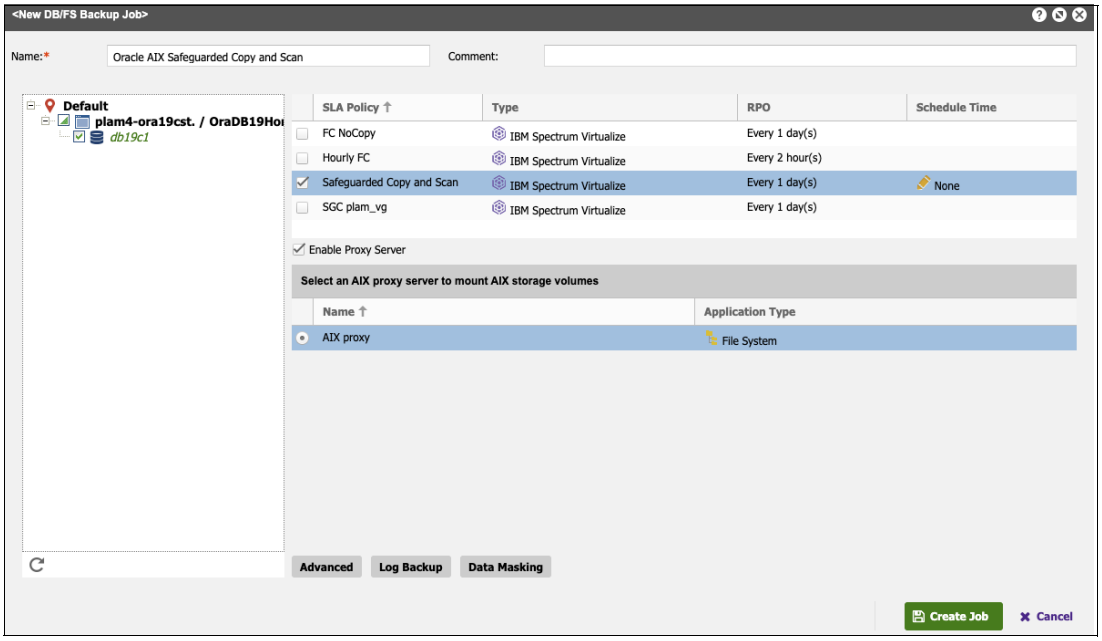


Figure 1-5 An example Oracle Backup job

The joblog shown in Figure 1-6 on page 11 shows an anomaly detected by IBM Storage Sentinel.

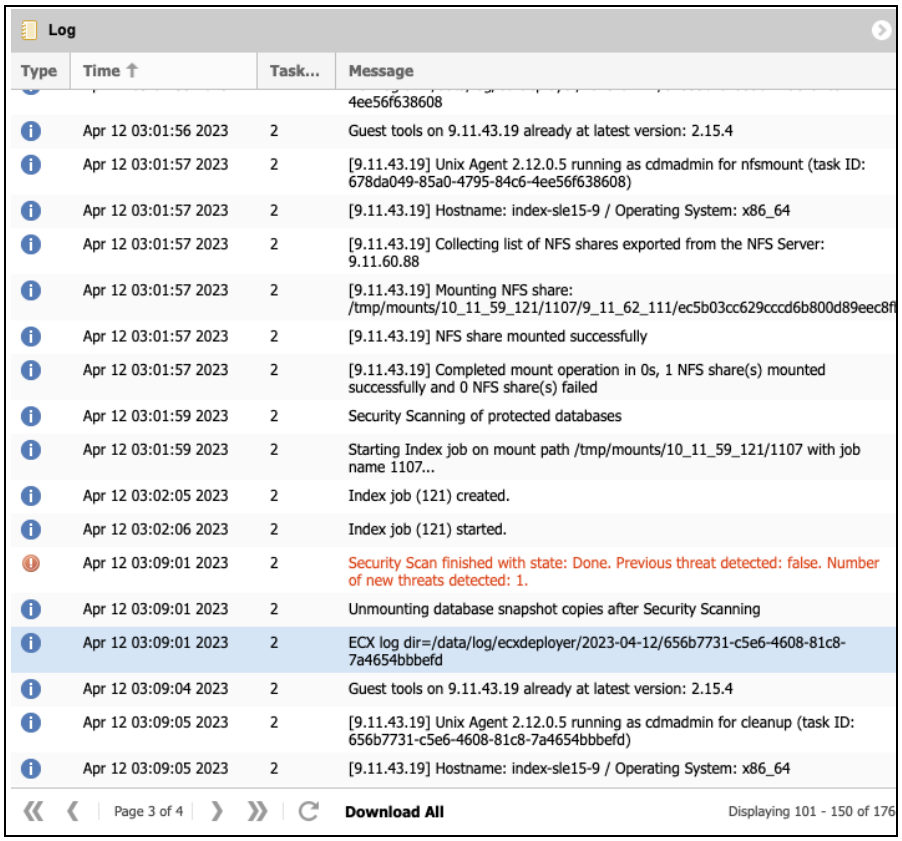


Figure 1-6 An example of a bad Backup joblog

Figure 1-7 on page 12 shows a good Backup joblog.

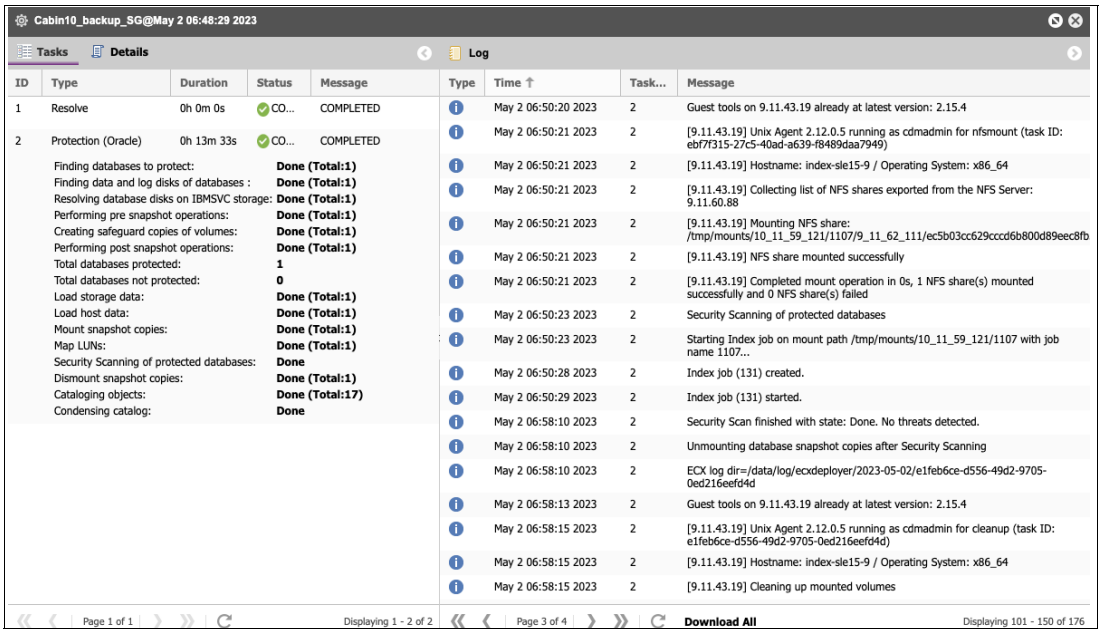


Figure 1-7 An example of a good Backup joblog

1.3.2 Use cases for Storage Sentinel

IBM Storage Sentinel can be used to:

- Detect ransomware in snapshots
- Protect your data from corruption
- Help recover after an attack

As mentioned earlier in this book, threat actors will often wait weeks or even months after ransomware is deployed to ensure that it infects all of a business's backups. IBM Storage Sentinel can detect ransomware in snapshots and other backups. It does this through a combination of detecting known patterns and using machine learning to extend capabilities to discover new patterns of infection.

Storage Sentinel can help protect data from corruption. It is designed to predict corruption and generate known good snapshots of data before they become corrupted or infected with malware.

Storage Sentinel helps recover after a ransomware attack. It can automatically generate reports listing the files or snapshots that were affected. This helps your organization identify clean copies of data that can be used to restore from.

1.3.3 IBM Storage Sentinel workflow

Figure 1-8 shows where Storage Sentinel fits in an overall cyber resilience strategy. The IBM *Cyber Vault Blueprint* identifies the two main phases of cyber security. The first phase is to protect your data. The second phase is recovering from a cyber attack. Storage Sentinel spans the protect and recover phases. It automates many of the tasks required for protecting data and can automatically identify safe recover points.

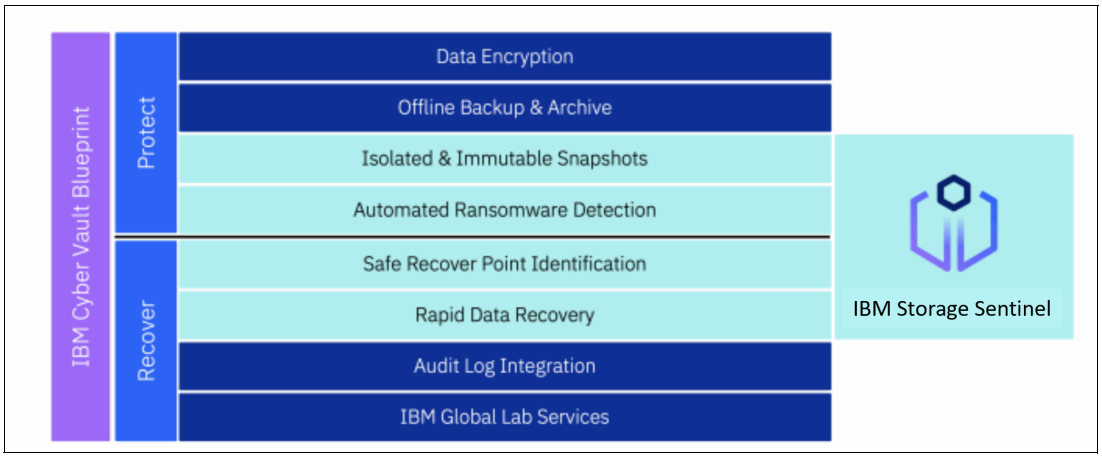


Figure 1-8 Storage Sentinel in Cyber Vault Blueprint

Figure 1-9 shows the key points on the Storage Sentinel Timeline for responding to an attack.

1. Before an attack begins, IBM Safeguarded Copy creates a series of immutable snapshots, which are proactively scanned for malware by IBM Storage Sentinel
2. Ransomware begins infecting production files / databases / systems with Safeguarded Copy, and those snapshots are protected and scanned proactively by Storage Sentinel.
3. Even as the attack is taking place, Storage Sentinel scans snapshots and analyzes changes, file extension mismatch, and other signs of data corruption.
4. Snapshots can now be assessed to confirm which have been verified to be free of malware and data corruption. This is the integrity review phase.
5. After the integrity review is complete, the most viable snapshot to restore from is identified.
6. The most viable snapshot is used to restore data to the production environment so the organization can resume normal business operations.

Figure 1-9 Storage Sentinel Attack Timeline

1.3.4 IBM Storage Sentinel components

IBM Storage Sentinel has the following components.

IBM Safeguarded Copy

IBM Safeguarded Copy is a feature of the DS8000, IBM FlashSystem, and IBM SVC storage systems that creates immutable snapshots of data to help protect against cyberattacks, malware, acts of disgruntled employees, and other data corruption. Safeguarded Copy uses the FlashCopy feature available on IBM Storage systems to create special FlashCopies that cannot be accessed by hosts. They cannot be mounted by or attached directly to a host. Instead, if recovery from a Safeguarded copy is required, another FlashCopy is created and that copy is presented to the host.

The IBM solution includes IBM Storage Sentinel. It complements IBM Safeguarded Copy by automatically scanning the copies that are created regularly by Safeguarded Copy looking for signs of data corruption introduced by malware or ransomware.

IBM Copy Data Management (CDM)

IBM Copy Data Management is a tool available from IBM that can manage Safeguarded Copy snapshots. It also has additional capabilities to catalog the existing copy data environment, including storage, virtual machines and applications. After it is deployed, it can significantly improve the IT team's ability to deliver key functions while dramatically reducing the cost of infrastructure and ongoing operations.

Anomaly scanning and detection

Cyber protection solutions are designed to protect from an attack in real-time. However, these solutions are not 100% effective and corporate data is corrupted daily. Scanning data for anomalies adds additional protection to these solutions. It detects and locates corruption that occurs when a successful attack makes it into the datacenter. Early detection of issues enables IBM Copy Data Management software to start fast application recovery. This minimizes downtime and flattens the data resiliency curve.

The scan software works by identifying corrupted files using statistics about files on the host being analyzed. It accomplishes this using a Machine Learning Model (MLM). The MLM is trained using real world malicious codes. In addition to identifying malicious code attacks, the anomaly scan software checks the integrity of databases to detect corruption of the internal database data. This corruption could occur because of an attacker, data corruption due to logical or physical causes, or damage at the disk/volume level, or as a flaw in the process in the creation of a snapshot or backup of the database.

The anomaly scan software examines existing database pages and allocation tables if they exist to ensure that all the allocated database pages are present and located in their correct position. In cases where some type of page data signature is available and/or enabled by the database administrator, such as a checksum or CRC, anomaly scan software recalculates the signature based on the current page contents and verifies it against the value found in the page header. Other ancillary fields are also verified within each page depending upon the database application. The anomaly scan software Machine Learning Model (MLM) has been designed to tolerate a small amount of database corruption that is commonly observed in production database systems to avoid excessive false-positive alerts.

Figure 1-10 on page 15 shows the five steps to cyber resilience.

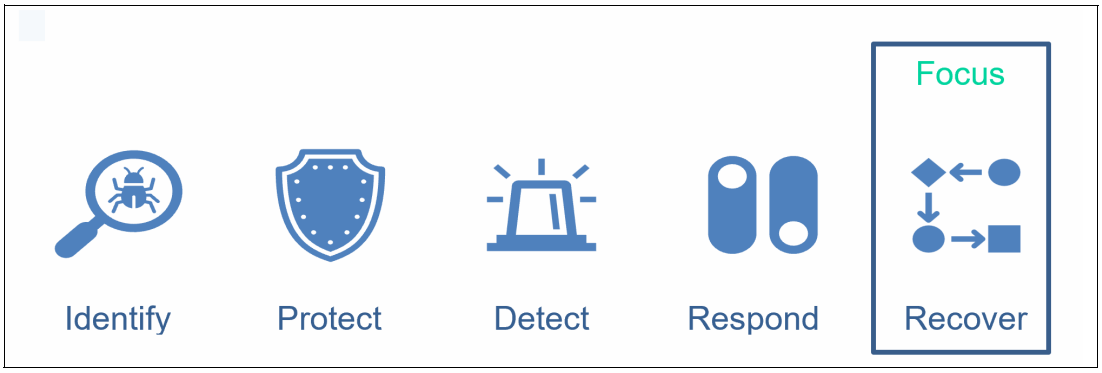


Figure 1-10 Five steps to cyber resilience

Orchestration and IBM Safeguarded Copy function

This chapter describes the IBM Safeguarded Copy feature, an integrated solution designed to create a logical air gap and protect critical data in the IBM FlashSystem product family. Safeguarded Copy creates immutable copies of the volumes and protect those copies for the duration of defined service level agreement (SLA) policy.

There are three methods to orchestrate the IBM Safeguarded Copy process.

- Safeguarded snapshot with internal scheduler
- Orchestration with IBM Storage Copy Data Management
- Orchestration with IBM Copy Services Manager

You can use both on the same storage system. As CDM is licensed on a TB basis, using it for volumes that require application-aware copies and/or anomaly scanning with IBM Storage Sentinel. The internal scheduler can be used for volumes where crash-consistent copies are sufficient.

This chapter has the following sections:

Important: When the term *Safeguarded Copy* is mentioned in this book, it refers to IBM FlashSystem Safeguarded Copy only. *The DS8000 Safeguarded Copy function is not supported by IBM Storage Sentinel or IBM Copy Data Management.*

2.1 Safeguarded snapshot with internal scheduler

Figure 2-1 shows the IBM Safeguarded Copy architecture.

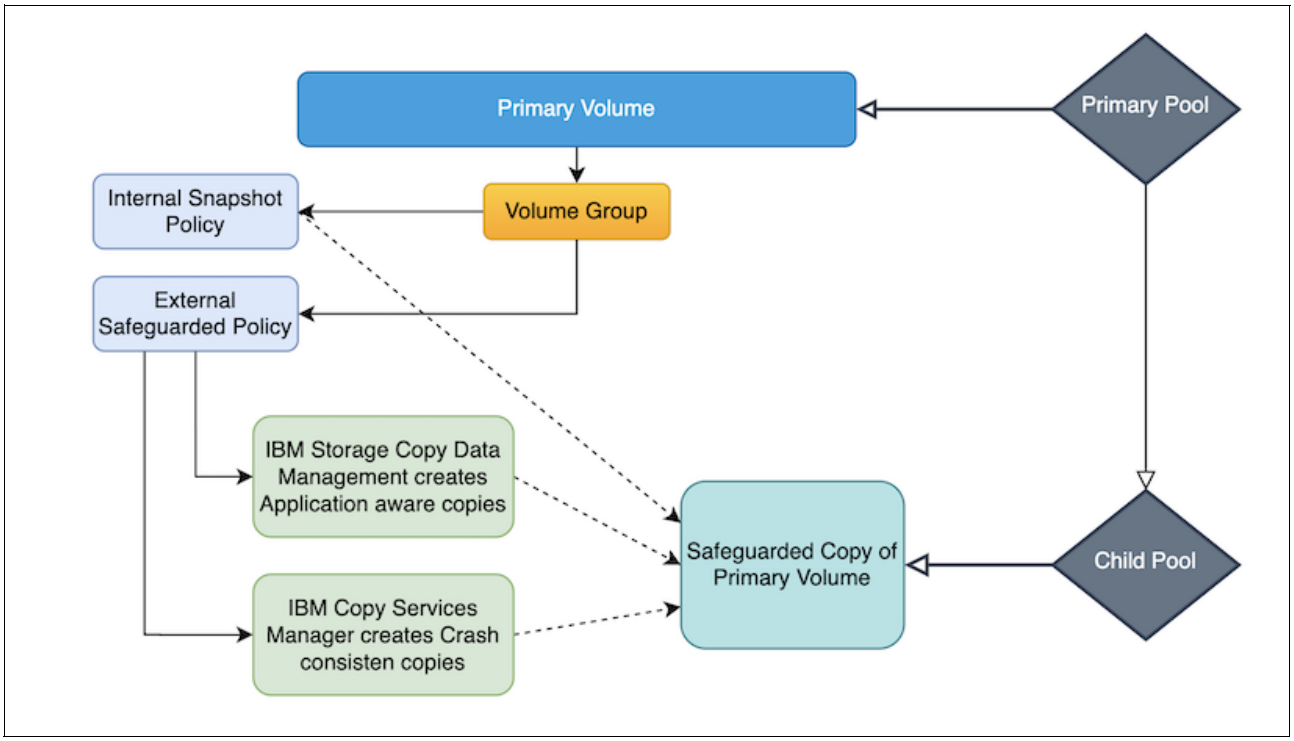


Figure 2-1 IBM Safeguarded Copy architecture

Safeguarded snapshot operates similar to snapshots and can use an internal scheduler. A new volume group can be created and pre-populated from another volume group's snapshot or Safeguarded snapshot. It inherits data, volumes, and volume groups from the snapshot.

Safeguarded copies managed by the internal scheduler create crash-consistent copies of primary volumes. For the application aware copies of primary volumes use IBM Storage Copy Data Management.

To create safeguarded copies using the internal scheduler the requirements are:

- **Safeguarded child pool**: Sub-pool under the primary pool designated as safeguarded.
- **Volume**: Volume must be located on the primary pool which contains safeguarded child pool.
- **Volume group**: The volumes should be assigned to this group to create safeguarded copies.
- **Safeguarded policy**: Policy for how often copies are created and how many days they are retained.

Figure 2-2 shows the volume group policy selection panel.

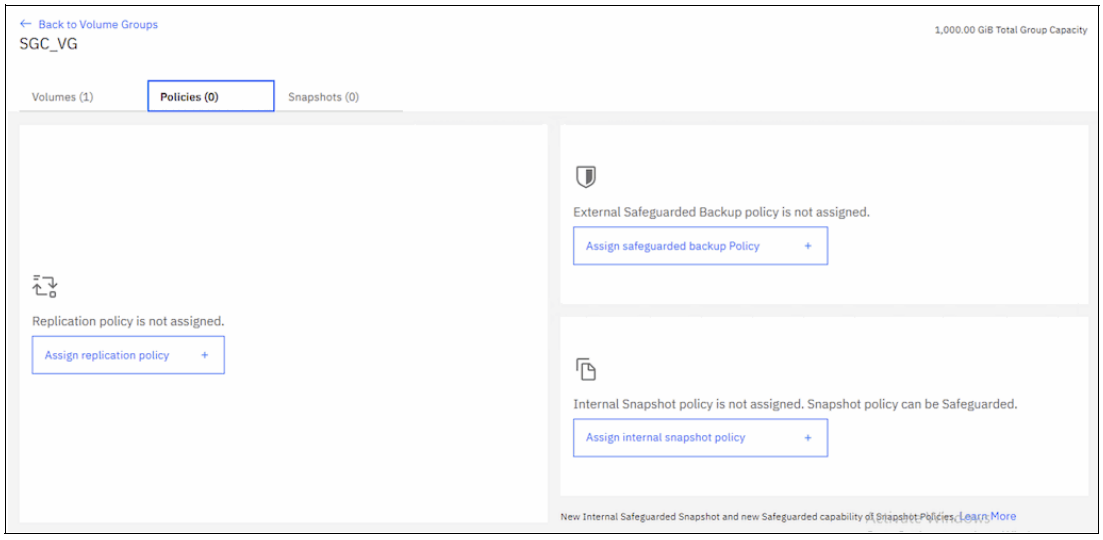


Figure 2-2 Volume group policy selection

Figure 2-3 on page 19 shows how to assign internal snapshot policy to the volume group.

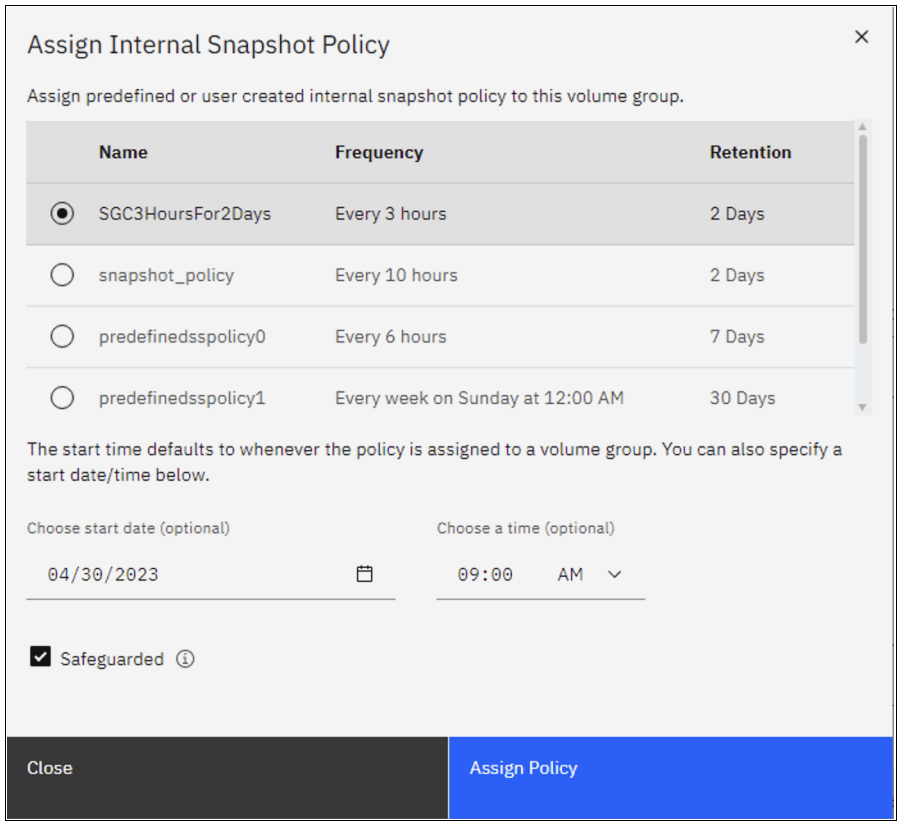


Figure 2-3 Assign internal snapshot policy to the volume group

Once a safeguarded snapshot policy is assigned to the volume group, the internal scheduler will automatically execute copy jobs at the specified times and safeguarded copies listed on volume group details. See Figure 2-4 on page 20 and Figure 2-5 on page 20.

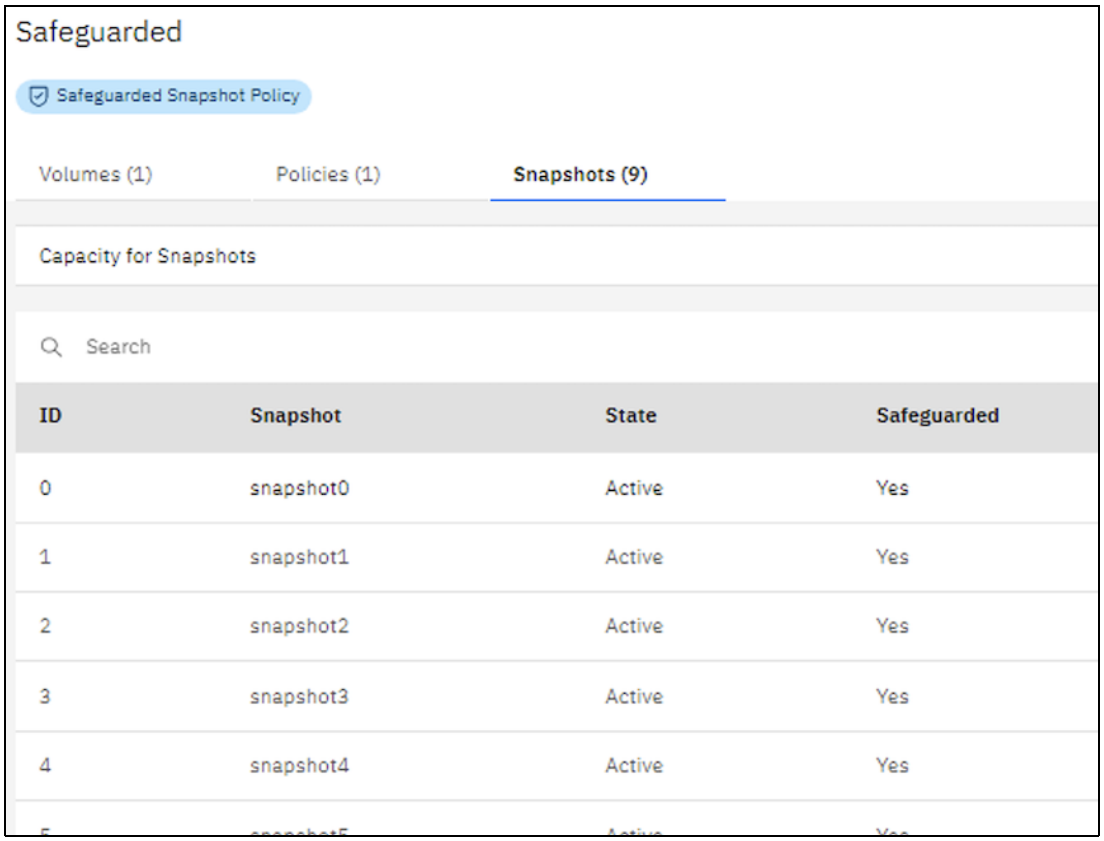


Figure 2-4 Safeguarded snapshots

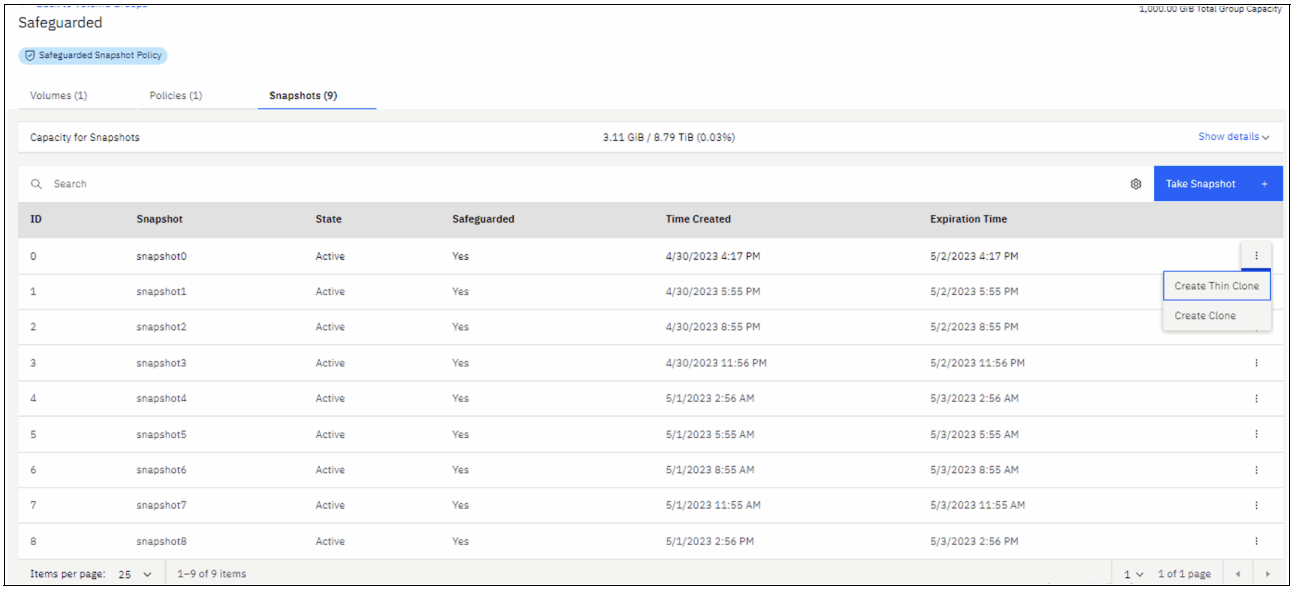


Figure 2-5 Recovering from safeguarded snapshots

To recover safeguarded copies, it is possible to create either thin copies or clones. Once a clone is generated from a safeguarded copy, it will be necessary to manually map the clone volume to the host in order to initiate the recovery process.

2.2 Orchestration for Storage Sentinel with IBM Storage Copy Data Management

The IBM Storage Copy Data Management workflow includes registering a provider, cataloging data, searching for objects, generating reports, and copying and using data.

2.2.1 Registering providers

Add providers, such as IBM Storage Virtualize based FlashSystem or SVC, application servers (SAP HANA, MS SQL, Oracle or VMware ESX) resources to the Inventory by registering them. Before registering providers, create a site to assign to your provider. A site is a user-defined grouping of providers that is generally based on location or department.

Registering IBM FlashSystem storage

Complete the following steps to register the IBM FlashSystem storage:

1. Click the **Configure** tab. In the Views window, select **Sites & Providers** and then, select the **Providers** tab.

2. In the Provider Browser window, select **IBM Spectrum Virtualize**.

3. Right-click **IBM Spectrum Virtualize**. Then, click **Register**. The Register dialog window opens.

4. Complete the fields in the dialog window. Select **New** to add credentials if they are not yet added through identities. We used a superuser account to register IBM FlashSystem storage. See Figure 2-6.

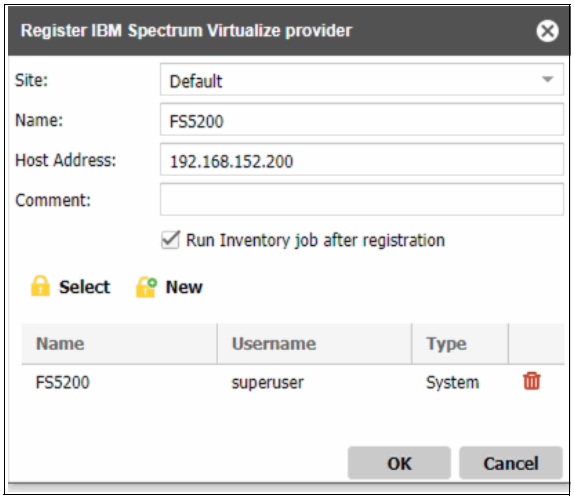


Figure 2-6 Registering IBM FlashSystem storage

5. Click **OK**. IBM Storage Copy Data Management first confirms that a network connection exists and then, adds the provider to the database.

Registering VMware vCenter

Complete the following steps to register the VMware vCenter:

1. Click the **Configure** tab. In the **Views** window, select **Sites & Providers** and then, select the **Providers** tab.

2. In the Provider Browser window, select **VMware**.

3. Right-click **VMware**. Then, click **Register**. The Register dialog window opens.

4. Complete the fields in the dialog window. Select **New** to add credentials if they are not yet added through identities. See Figure 2-7.

For the required VMware vSphere privileges see CDM documentation at IBM Docs.

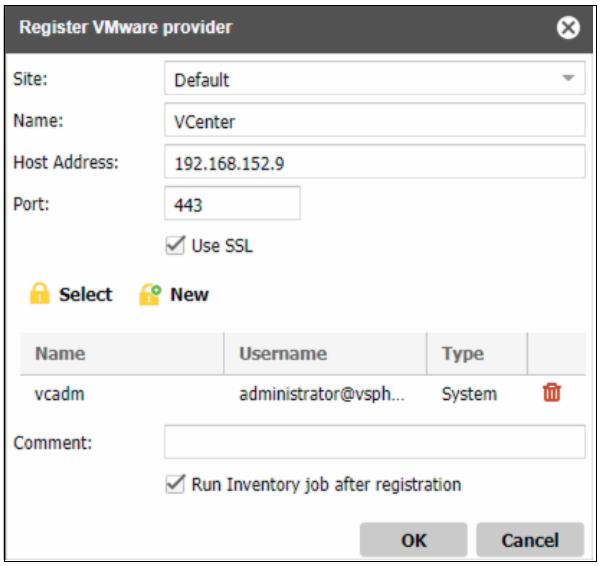


Figure 2-7 Registering VMware vCenter

Registering Sentinel Anomaly Scan Server

Complete the following steps to register the Sentinel security scan server:

1. Click the **Configure** tab. In the **Views** window, select **Sites & Providers** and then, select the **Providers** tab.

2. In the Provider Browser window, select **Security Scan Server**.

3. Right-click **Security Scan Server**. Then, click **Register**. The Register dialog window opens.

4. Complete the fields in the dialog window. Select **New** to add credentials if they are not yet added through identities. If the security scan server is virtual machine on VMware, select the pre-registered VMware VCenter and enter the credentials for Sentinel server. See Figure 2-8 on page 23.

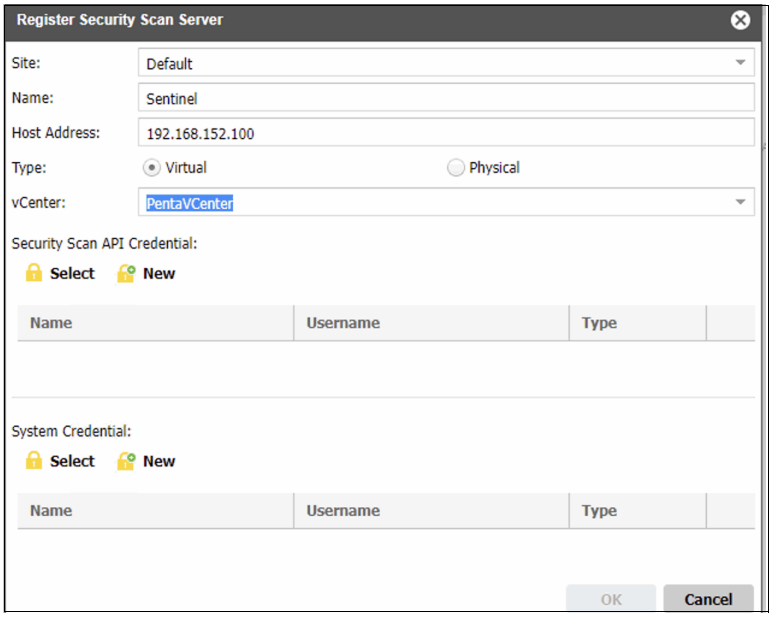


Figure 2-8 Registering Sentinel Security Scan Server

Registering Oracle Database Server

Complete the following steps to register the Oracle database server:

1. Click the **Configure** tab. On the **Views** pane, select **Sites & Providers**, then select the **Providers** tab.

2. In the Provider Browser pane, select **Application Server**.

3. Right-click **Application Server**. Then click **Register**. The Register Application Server dialog opens.

4. Select **Oracle** as the Application Type and complete the fields in the dialog window. See Figure 2-9.

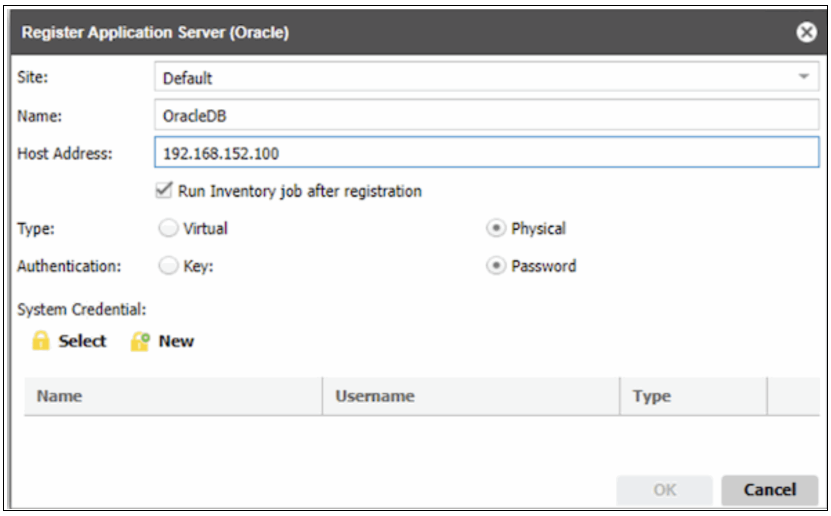


Figure 2-9 Registering Oracle Database Server

Registering SMTP Server

Complete the following steps to register the SMTP server:

1. Click the **Configure** tab. In the **Views** window, select **Sites & Providers** and then, select the **Providers** tab.

2. In the Provider Browser window, select **SMTP**.

3. Right-click **SMTP**. Then, click **Register**. The Register dialog window opens.

4. Complete the fields in the dialog window. See Figure 2-10 on page 24.

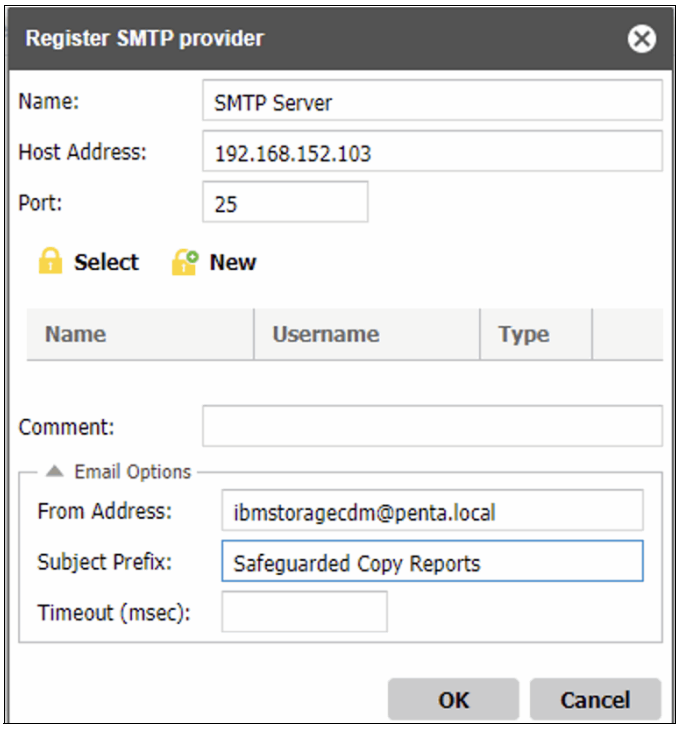


Figure 2-10 Registering SMTP Server

2.2.2 Configuring SLA policies

The use of SLA policies allows for the customization of templates by administrators for the primary processes involved in the creation and utilization of Backup jobs. These policies configure copy types, destinations, and parameters that can be reused in future Backup jobs.

Note: For more information on SLA policies refer to 4.4.2, “Service Level Agreement (SLA) policies” on page 60.

During the configuration of a Backup job, suitable SLA Policies will appear in the job creation wizard, which are tailored to the specific type of Backup job being created.

1. Click the **Configure** tab. On the **Views** pane, select **SLA Policies**. The All SLA Policies pane opens.

2. In the All SLA Policies pane, click **New**. The New SLA Policies pane opens.

3. Select a type of policy to create based on your storage provider. Select **IBM Spectrum Virtualize** to create an IBM Backup policy.

4. Add a sub-policy (SLA Policy) to an IBM Spectrum Virtualize SLA policy.

a) Select the source icon and define the recovery point objective to determine the minimum frequency and interval with which backups must be made. In the Frequency field, select Minutes, Hourly, Daily, Weekly, or Monthly, then set the interval in the **Interval** field. The lowest available frequency is five minutes.

Note: If changes are made to the frequency and interval of an SLA Policy, those modifications will impact all job schedules that are linked to it.

b) Click **Add Safeguarded Copy**.

c) In the Associated Safeguarded Volume Group pane, expand the storage device and select the volume group that you want to be back up as Safeguarded Copy. Any volume that you want to back up as safeguarded copy must belong to a volume group. If it is not a member of any of these groups it will not back up as a Safeguarded Copy.

Note: The Associated Safeguarded Volume Group lists only those volume groups that have the Safeguarded Copy policy applied on the storage array side.

d) In the Options pane, set the Safeguarded Copy sub-policy options.

Keep Snapshots

After a certain number of snapshot instances are created for a resource, older instances are purged from the storage controller. Enter the age of the snapshot instances to purge in the Days field.

Name

Enter an optional name to replace the default FlashCopy sub-policy name displayed in IBM Storage Copy Data Management. The default name is Safeguarded Copy0.

FlashCopy Volume Prefix

Enter an optional label to identify the FlashCopy. This label is added as a prefix to the FlashCopy name created by the job.

Tip: FlashCopy labels must contain only alphanumeric characters and underscores.

Perform Security Scan

You must enable this and select your security scan servers. This allows to scan for every backup number you have specified.

e) Enter a name for the new sub-policy (SLA Policy).

f) Click **Finish**. See Figure 2-11 on page 26.

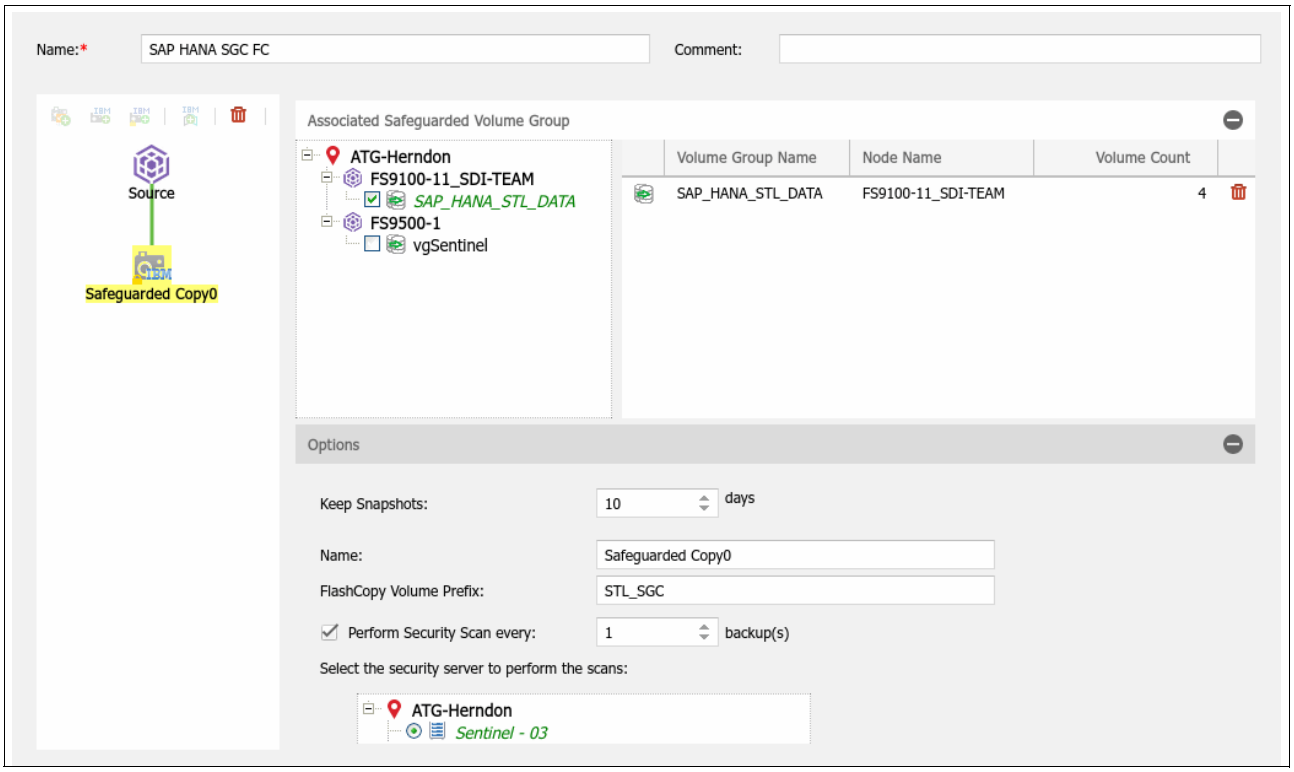


Figure 2-11 Configuring SLA policy and Sentinel Scan frequency

2.2.3 Creating backup jobs

The backup jobs create a copy of your selected applications and dependent volumes, according to rules defined in a SLA policy. See Figure 2-12 on page 27.

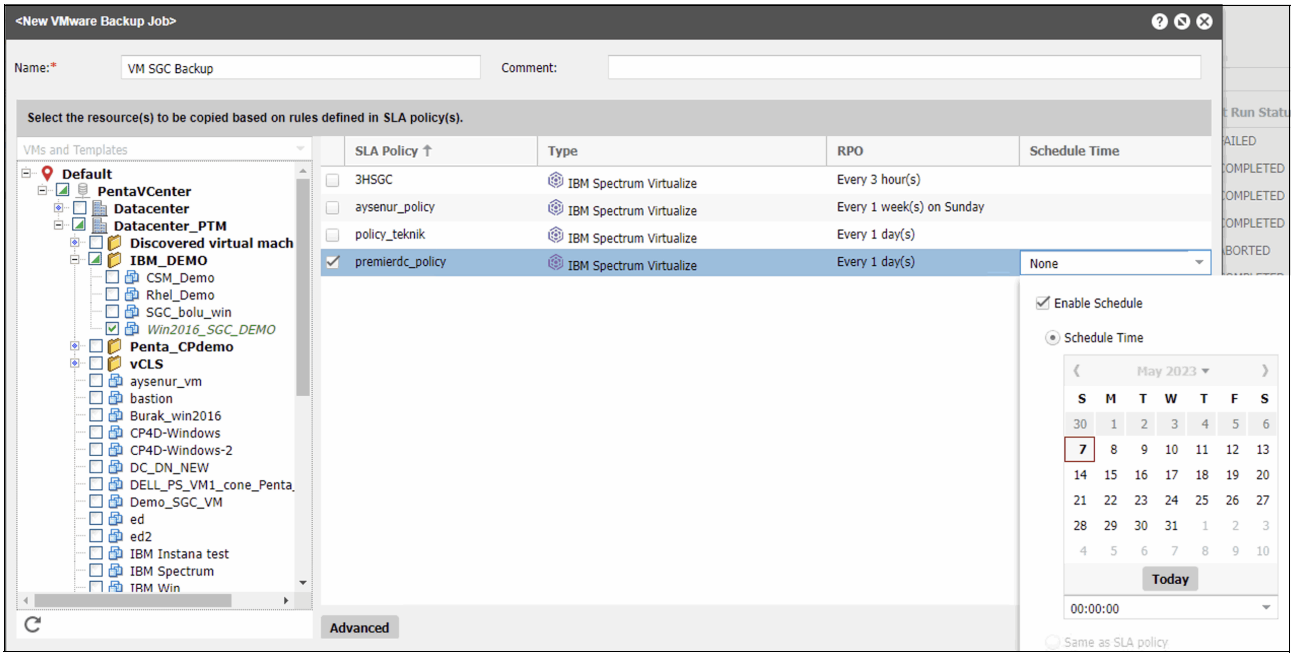


Figure 2-12 Creating backup job with schedule

SAP HANA backup procedure described in Chapter 4, “Configuring IBM Storage Sentinel for SAP HANA” on page 51.

2.2.4 Restore and recovery jobs

IBM Storage Copy Data Management leverages Copy Data Management technology for testing and cloning use cases, instant recovery, and full disaster recovery.

There are multiple methods to regain access to your safeguarded copies for each application.

For VMware environment, you can choose **Instant Disk Restore**, **Instant VM Restore** or **Instant VM Restore (Long Distance)**. See Figure 2-13.

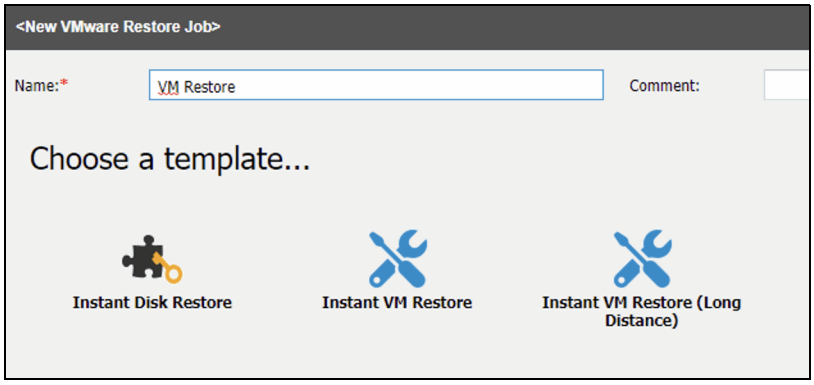


Figure 2-13 VMware restore options

For MS SQL databases, you can choose **Instant Database Restore** or **System DB Restore** for Standalone and MSCS Failover Clusters. For MS SQL Always on configuration, you can choose **Instant Disk Restore**, **Instant Database Restore** or **Instant Seeding**. See Figure 2-14 on page 28.

Enter a meaningful name and comment, select to perform an **Instant Disk Restore** or **Instant Database Restore**. An Instant Disk Restore mounts the file systems to the target host but does not try to define or start the database. An Instant Database restore will define and start the database to an database instance.

Detailed information and requirements for MS SQL databases see IBM Storage Copy Data Management Microsoft SQL Server Support FAQ at IBM Docs.

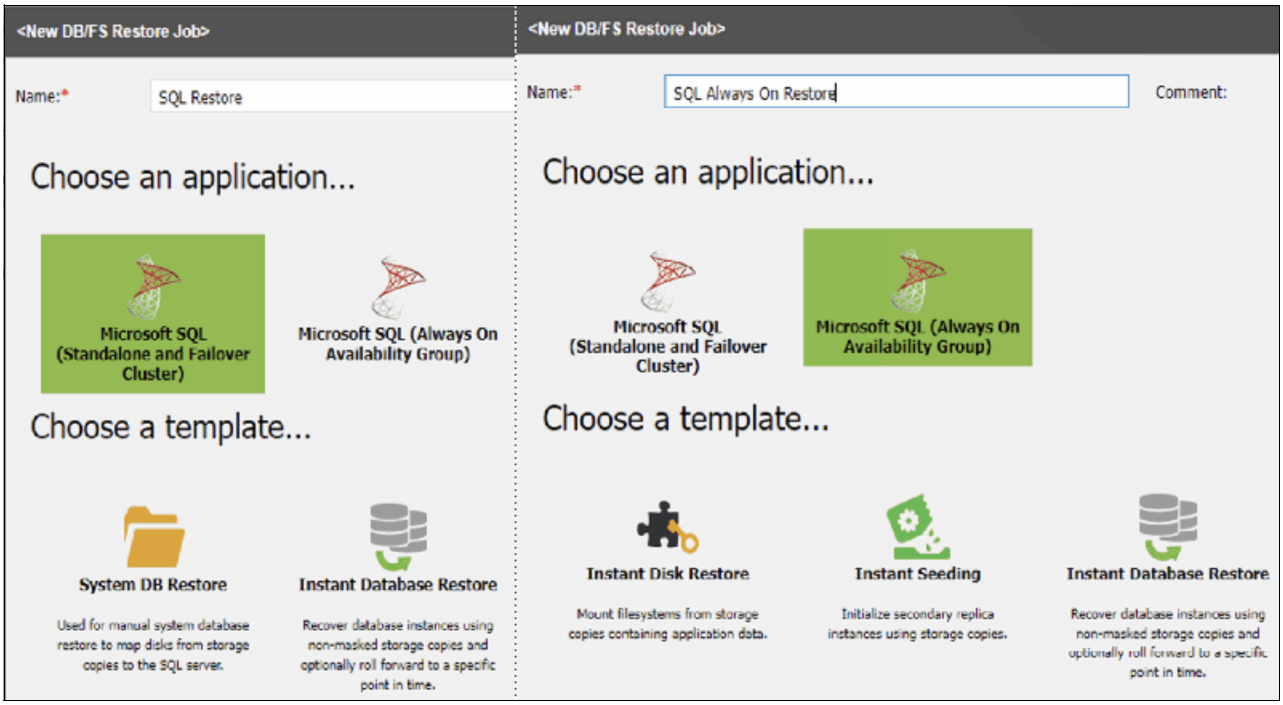


Figure 2-14 SQL restore options

For Storage (Spectrum) Virtualize volume based restore requirements, you can choose **Instant Disk Restore** or **Restore Volumes**.

- Instant Disk Restore option creates clone of your safeguarded copy and instantly mounts it to your hosts to read, control or copy your required files.
- Restore Volumes option restores chosen Safeguarded Copy over the original volume.

  See Figure 2-15.

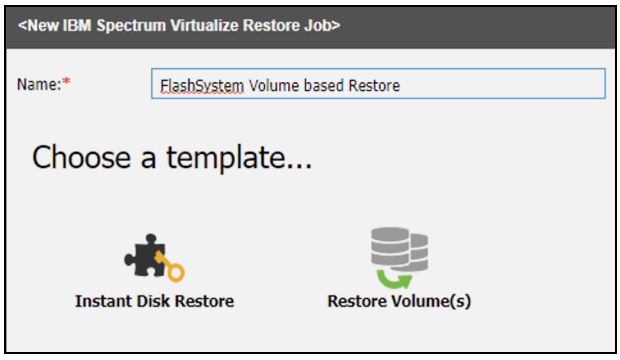


Figure 2-15 Storage (Spectrum) Virtualize Volume based restore options

For Oracle databases, you can choose **DevOps**, **Instant Database Restore** and **Instant Disk Restore**. See Figure 2-16.

Detailed information and requirements for Oracle databases see IBM Storage Copy Data Management Oracle Database Support FAQ at IBM Docs.

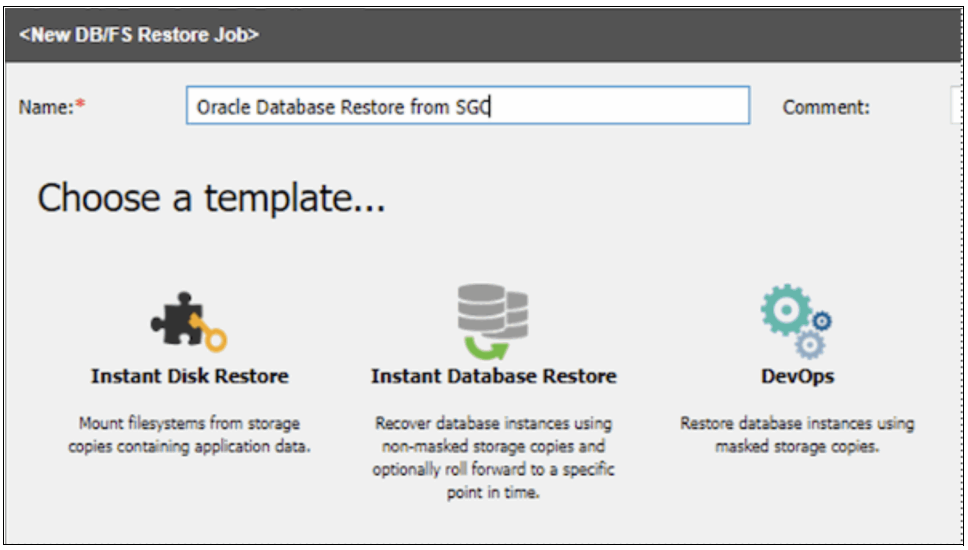


Figure 2-16 Oracle restore options

Performing an SAP HANA instant database restore job described in “Performing an SAP HANA instant disk restore job” on page 65.

2.2.5 Pre-script and post-script

Prescripts and postscripts are scripts that can be run before or after Backup and Restore jobs run, both at a job-level and before or after snapshots are captured. A script can consist of one or many commands, such as a shell script for Linux-based servers or Batch and PowerShell scripts for Windows-based servers.

Scripts can be created locally, uploaded to your environment through the Scripts pane, then applied to job definitions. In a Windows environment, if your application supports Volume Shadow Copy Service (VSS), the Backup job triggers the VSS application quiesce logic if the make these VMs application/file system consistent option is enabled when creating the VMware Backup job. However, for applications that don’t support VSS, or on Linux virtual machines, pre and post snapshot scripts can be used to quiesce your application for the snapshot backup.

Note: If adding a script to a Windows-based File System job definition, the user running the script must have the *Log on as a service* right enabled, which is required for running prescripts and postscripts.

Uploading a script

Perform the following steps for uploading a script.

1. Click the **Configure** tab. On the **Views** panel, select **Scripts**.

2. Click **Upload**. The Upload Script dialog opens.

3. In the Script field, browse for a local script to upload, then click **Open**.

4. Enter an optional comment, then click **OK**. The script appears on the Scripts panel and can be applied.

3

Protecting Epic Cache and IRIS Databases with IBM Safeguarded Copy and IBM Storage Sentinel

This chapter describes creating application-consistent, immutable snapshots of Intersystems Cache or IRIS databases and then having them scanned with IBM Storage Sentinel to detect possible malware corruption.

This chapter has the following sections:

3.1 Introduction

This chapter describes creating application-consistent, immutable snapshots of Intersystems Cache or IRIS databases and then having them scanned with IBM Storage Sentinel to detect possible malware corruption.

One of the largest uses of Intersystems databases are those that run the Epic health records system that utilizes the Intersystems database technology. For this document, we have chosen to refer to these as Epic databases, but there is no difference in this context. The Epic health records management solution is an industry leader in this space. Since cyber criminals often target healthcare organizations, the Epic databases were the first applications supported by IBM Storage Sentinel.

This chapter will describe the following:

- The currently supported configurations and how the backup, scanning and recovery flow will differ across those configurations.
- Registering storage, VMware, Sentinel and Epic database components.
- Defining SLAs and running backups.
- Running a recovery of Epic databases.

3.2 Supported configurations for IBM Storage Copy Data Management and IBM Storage Sentinel for Epic Databases

As always, check your application, OS, platform and storage against the current requirements documentation for IBM Storage Copy Data Management and IBM Storage Sentinel before implementation. The current manuals and other documentation may be more up to date than this book due to the passage of time. The following are the supported configurations as of the publication of this IBM Redbooks for IBM Storage Copy Data Management (CDM) v2.2.19 and IBM Storage Sentinel (Sentinel) 1.1.2. *To repeat, the product documentation is the official source, and you should double-check any design against that documentation before implementation*.

The supported Intersystems database applications are:

- Intersystems Cache 2015, 2016, 2017, 2018 or later.
- Intersystems IRIS 2021, 2022 or later.

Storage Copy Data Management and Storage Sentinel support the Epic databases being hosted in vSphere virtual machines or on physical hosts. We will give details on exactly how each configuration will operate in this chapter.

For virtual machines we support vSphere 6.5 and 6.5.x levels, v6.7 and 6.7.x levels and, v7 and v7.0.x levels.

For Intersystems Databases running within virtual machines, we support Red Hat Enterprise Linux (RHEL) 6.5 or later or CentOS 7.0 or later. It should be noted that Intersystems documentation states that CentOS is only supported for dev/test environments. See Figure 3-1.

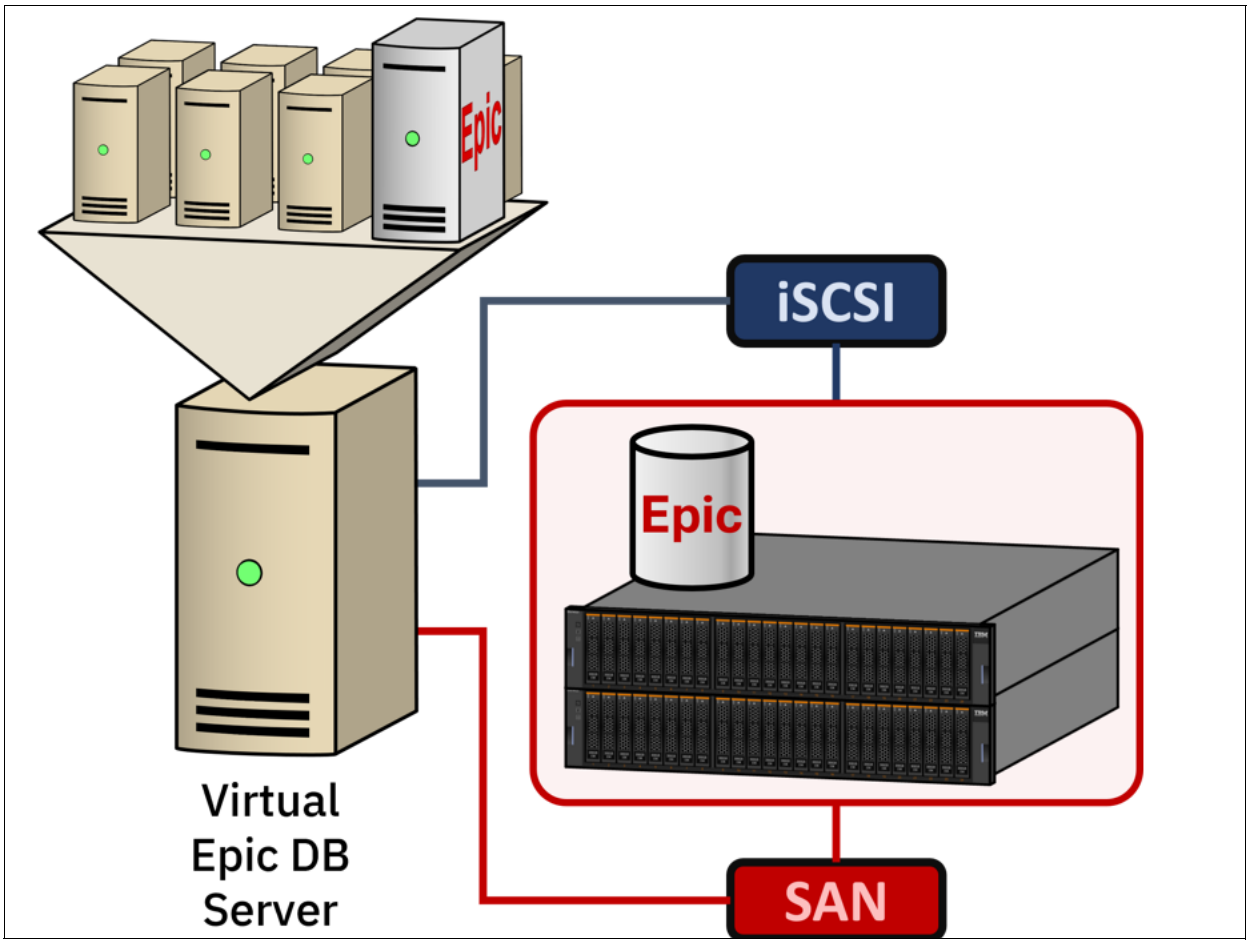


Figure 3-1 Virtual Epic DB Server

Physical machines on the Power platform (little-endian) are supported running AIX 6.1 TL9 or later, AIX 7.1 or later, AIX 7.2 or later and AIX 7.3 or later. *IRIS was tested only on the AIX 7.x operating system versions*.

Physical machines on the x64 platform are supported running RHEL 6.5 or later or CentOS 7 or later. It should be noted that Intersystems documentation (https://docs.intersystems.com/irislatest/csp/docbook/DocBook.UI.Page.cls?KEY=ISP_technologies#ISP_platforms) states that CentOS is only supported for dev/test environments. See Figure 3-2 on page 34.

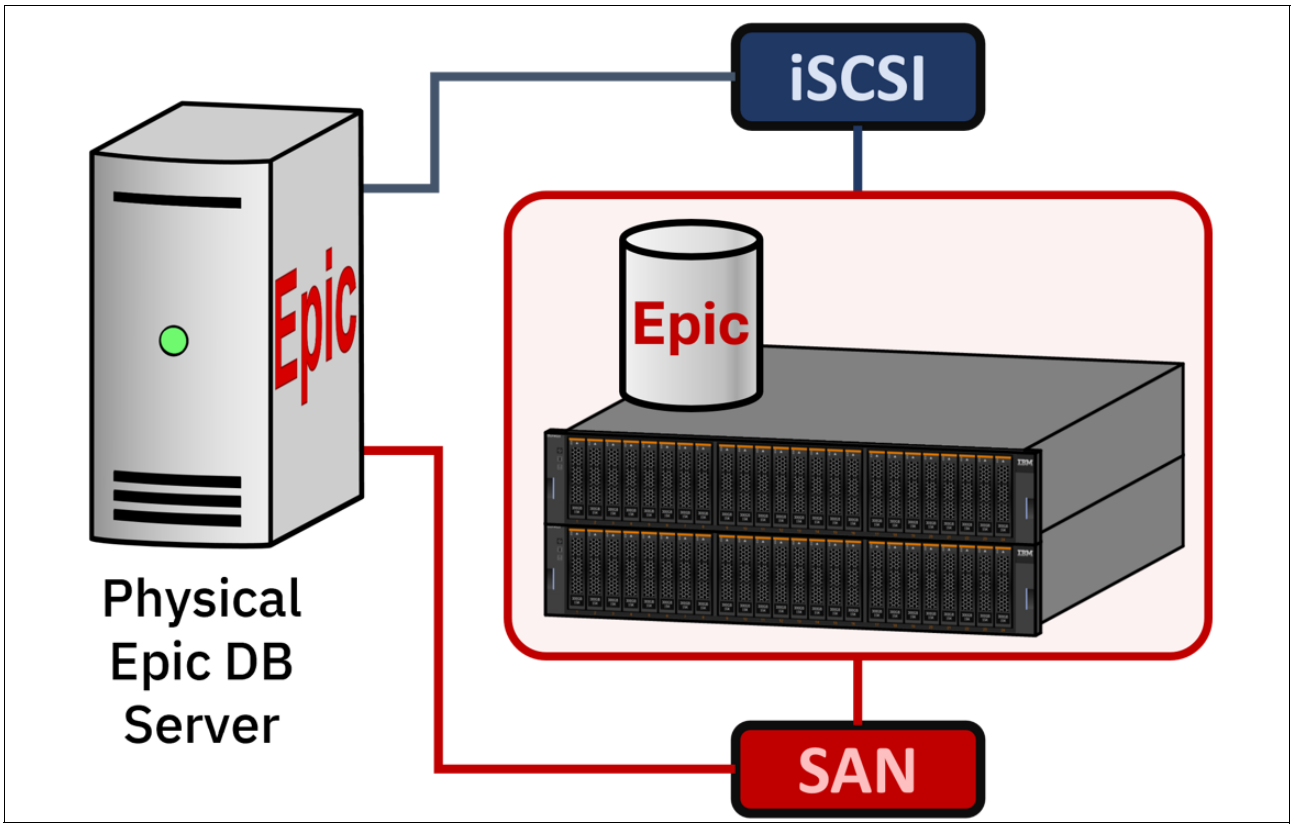


Figure 3-2 Physical Epic DB Server

3.3 IBM Storage Sentinel server platform choice

Sentinel can be run either within a virtual machine or a physical machine. If choosing to use a physical machine, one can only scan Safeguarded Copies taken of Epic databases also running on a physical machine. If choosing to use a virtual machine to host Sentinel, one can scan Safeguarded Copies of BOTH physical and virtual machines.

3.3.1 Supported storage configurations for virtual Epic database servers

Although CDM supports both IBM and Pure storage for protecting Epic databases, this publication only covers IBM Storage Virtualize or IBM FlashSystems storage that provides Safeguarded Copy (SGC) functionality. I*n the current version, CDM does not support SGC Volume Group Snapshots (also known as SGC V2.0) but rather supports SGC V1*.

There are 3 different supported configurations for taking Safeguarded Copies of Epic database volumes and then scanning them with Sentinel. As mentioned in 3.2, “Supported configurations for IBM Storage Copy Data Management and IBM Storage Sentinel for Epic Databases” on page 32, to scan the databases volumes of a virtual machine, Sentinel needs to also be running as a virtual machine. (There is one exception which we will cover in 3.3, “IBM Storage Sentinel server platform choice” on page 34).

Use virtual disks in a vSphere VMFS datastore

The first supported configuration is to configure the Epic VM to use virtual disks (VMDK) stored in a VMFS datastore. As with all configurations discussed in this publication, the

storage must be a supported IBM Storage Virtualize/FlashSystems storage. We support volumes shared via the SAN or over iSCSI. When running in this configuration, the volumes containing the datastore are snapshotted in the storage controller. If you have other VMs sharing this datastore, there is no way to exclude them from the snapshots, as all volumes containing the VMFS datastore must be protected at the same time. It is recommended to limit how many other VMs share this datastore to reduce waste. Before the Sentinel VM can scan the Epic DB volumes, CDM will orchestrate the mounting of a copy of the snapshotted datastore and mapping the VMDK files to the Sentinel server. See Figure 3-3 on page 35.

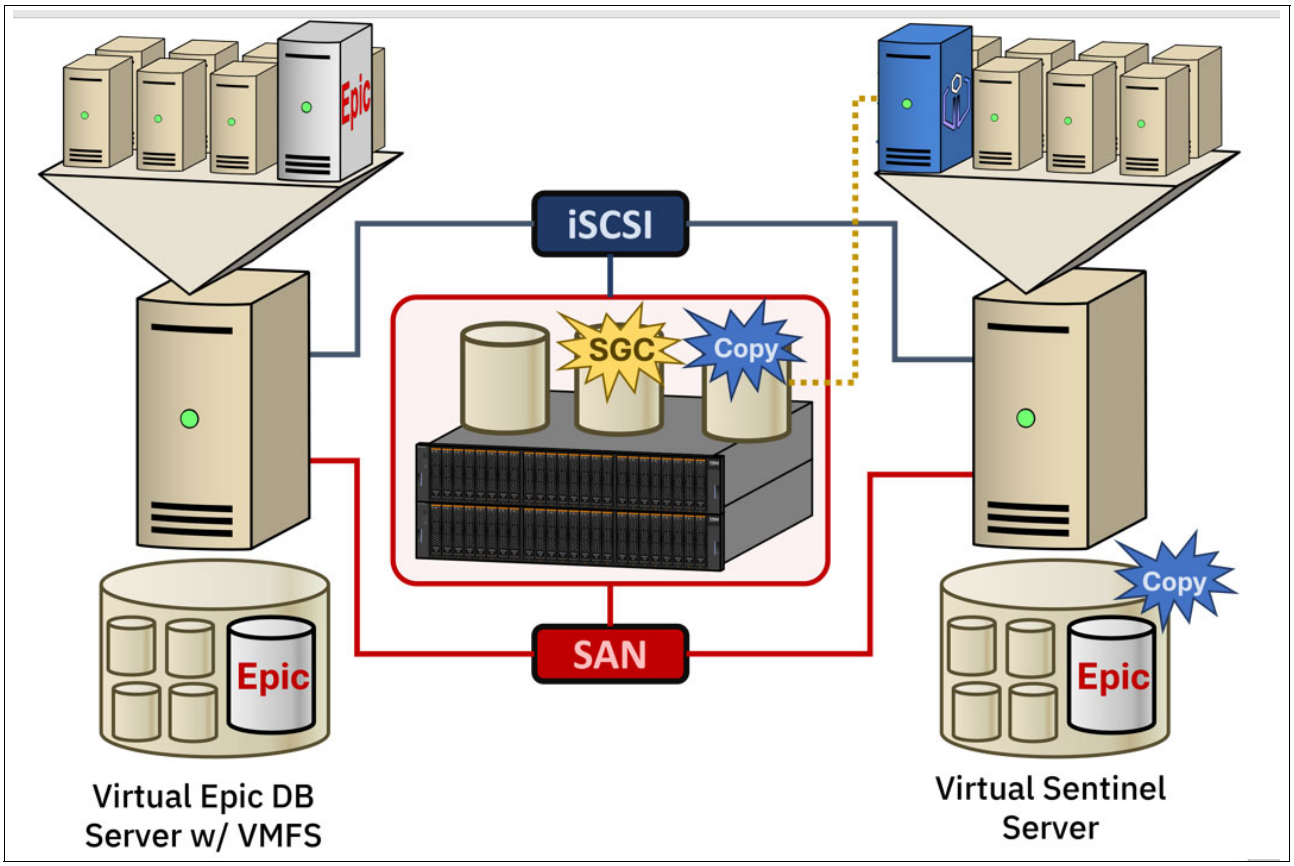


Figure 3-3 Virtual Epic DB server with virtual disks in a datastore

A backup job workflow will perform the following high-level tasks:

> **Note:** There are many more tasks within this workflow, but we are not going into that level of detail in this chapter.

1. Quiesce the Epic database.
2. Take a Safeguarded Copy of the volumes containing the vSphere datastore.
3. Unquiesce the Epic database.
4. Create a copy of the Safeguarded copies.
5. Create a datastore on vSphere using these copies.
6. Modify the Sentinel VM configuration and mount the virtual disks for scanning.
7. Scan the copy of the Epic DB.

8. Reverse the process to dismount the volumes, change the Sentinel VM configuration, remove the temporary datastore and destroy the copies of the Safeguarded Copy volumes.
9. Flag the recovery point created by this backup job as having passed or failed the scan and raise an alert if it did fail.

Note: IBM does not currently support mixing VMDK disks and physical raw device mapped volumes with CDM or Sentinel.

Use physical raw device mapped (pRDM) volumes

It is very common that an organization will chose to use pRDM volumes mapped directly to a VM rather than virtual disks on a VMFS datastore. This will provide some performance benefits and prevent having to create an isolated datastore. When using a VM configured with pRDM disks to contain the Epic databases, Sentinel must also be running within a VM. See Figure 3-3 on page 35.

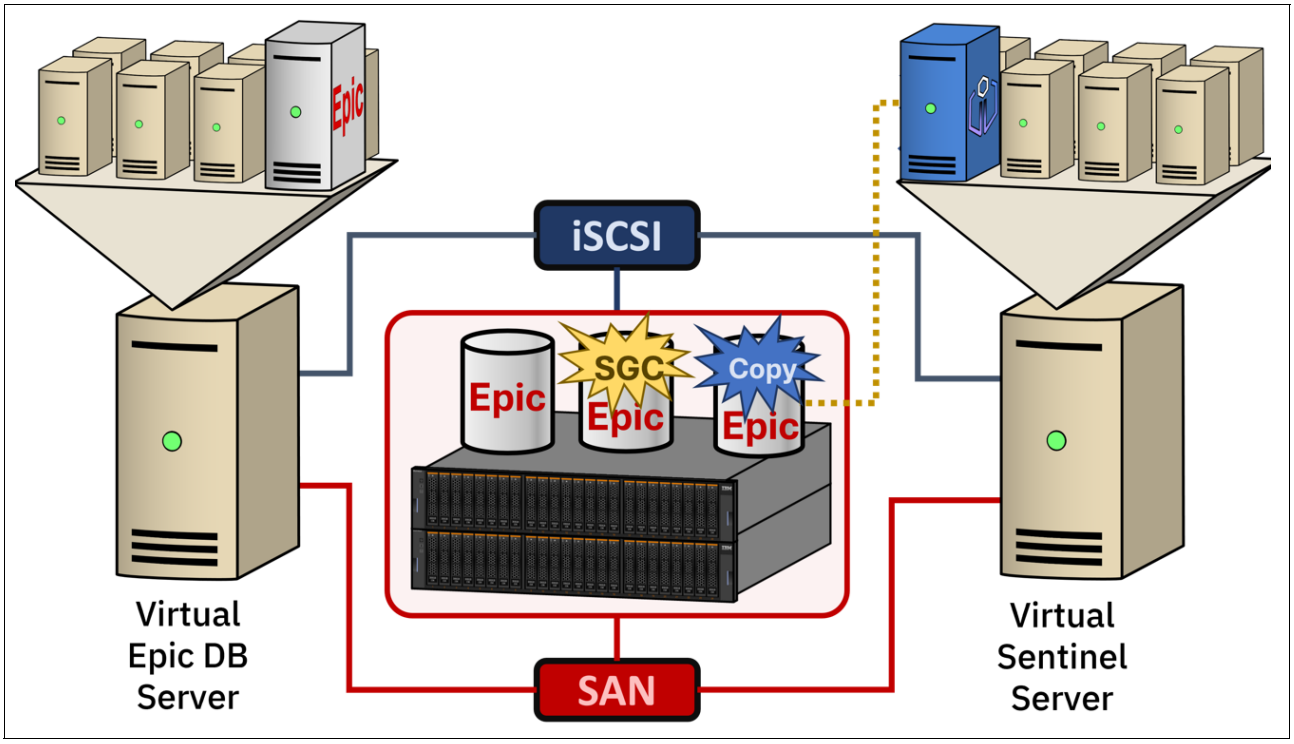


Figure 3-4 Virtual Epic DB server with either pRDM or iSCSI volumes

A backup job workflow will perform the following high-level tasks:

Note: There are many more tasks within this workflow, but we are not going into that level of detail in this chapter.

1. Quiesce the Epic database.
2. Take a Safeguarded Copy of the pRDM volumes.
3. Unquiesce the Epic database.
4. Create a copy of the Safeguarded copies.

5. Modify the Sentinel VM configuration and mount those copies as pRDM disks for scanning.
6. Scan the copy of the Epic DB.
7. Reverse the process to dismount the volumes, change the Sentinel VM configuration and destroy the copies of the Safeguarded Copy volumes.
8. Flag the recovery point created by this backup job as having passed or failed the scan and raise an alert if it did fail.

Note: IBM does not currently support mixing VMDK disks and physical raw device mapped volumes with CDM or Sentinel.

Use volumes shared over iSCSI directly to the virtual machine

This configuration refers to creating volumes and sharing over iSCSI directly to the virtual machine, not to creating a VMFS datastore on iSCSI disk. In this configuration the iSCSI volumes are not managed by the vSphere virtualization layer (although the I/O is going over the vSphere virtual networks). When registering the Epic DB server to CDM, you will register it as a physical machine, not as a virtual machine. This is the one configuration where a physical Sentinel server can scan the Epic DB hosted in a VM.

A backup job workflow will perform the following high-level tasks:

Note: There are many more tasks within this workflow, but we are not going into that level of detail in this chapter.

1. Quiesce the Epic database.
2. Take a Safeguarded Copy of the iSCSI volumes.
3. Unquiesce the Epic DB.
4. Create a copy of the Safeguarded copies.
5. Mount the iSCSI volumes to the Sentinel VM.
6. Scan the copy of the Epic DB.
7. Reverse the process to dismount the volumes and destroy the copies of the Safeguarded Copy volumes.
8. Flag the recovery point created by this backup job as having passed or failed the scan and raise an alert if it did fail.

Note: IBM does not currently support mixing iSCSI volumes with other configurations in a single machine.

3.3.2 Supported storage configurations for physical Epic database servers

CDM and Sentinel support physical Epic database servers with volumes mapped from the Storage Virtualize/FlashSystems controllers over iSCSI or a Storage Area Network (SAN). These physical Epic DB servers can be scanned with either a physical Storage Sentinel server (Figure 3-5 on page 38) or a virtual Storage Sentinel server (Figure 3-6 on page 38).

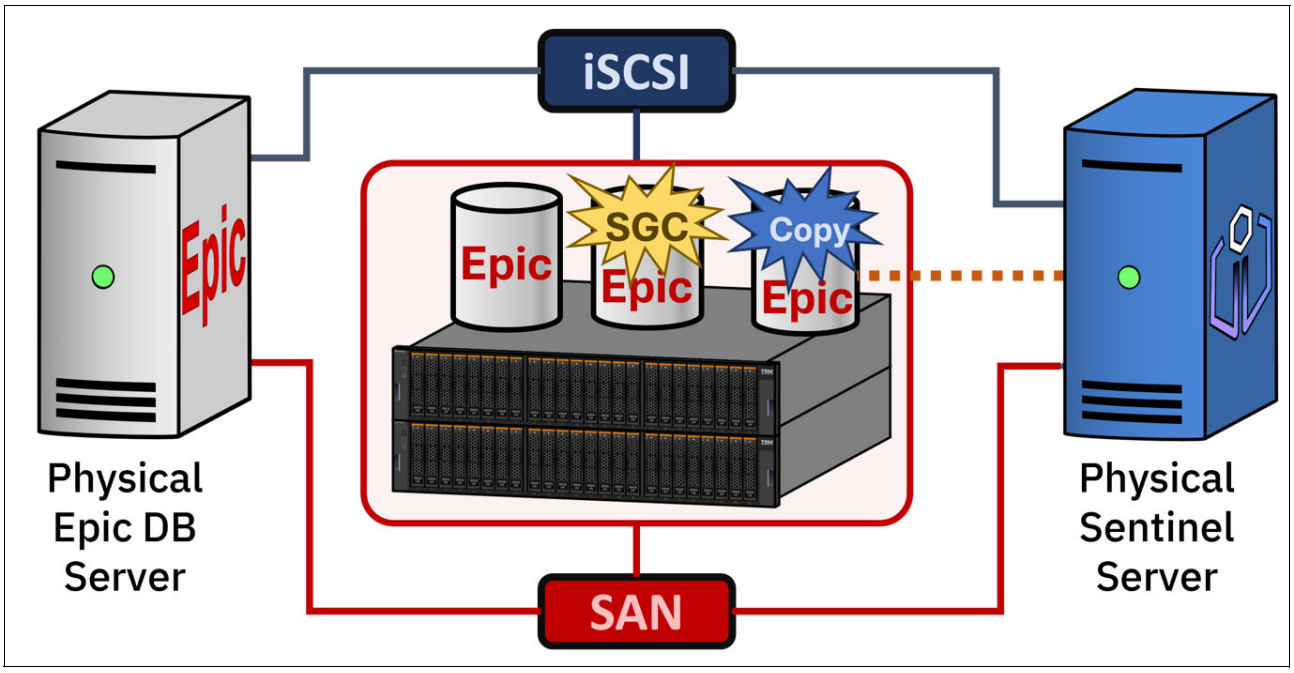


Figure 3-5 Physical Epic DB server and Physical Sentinel Server

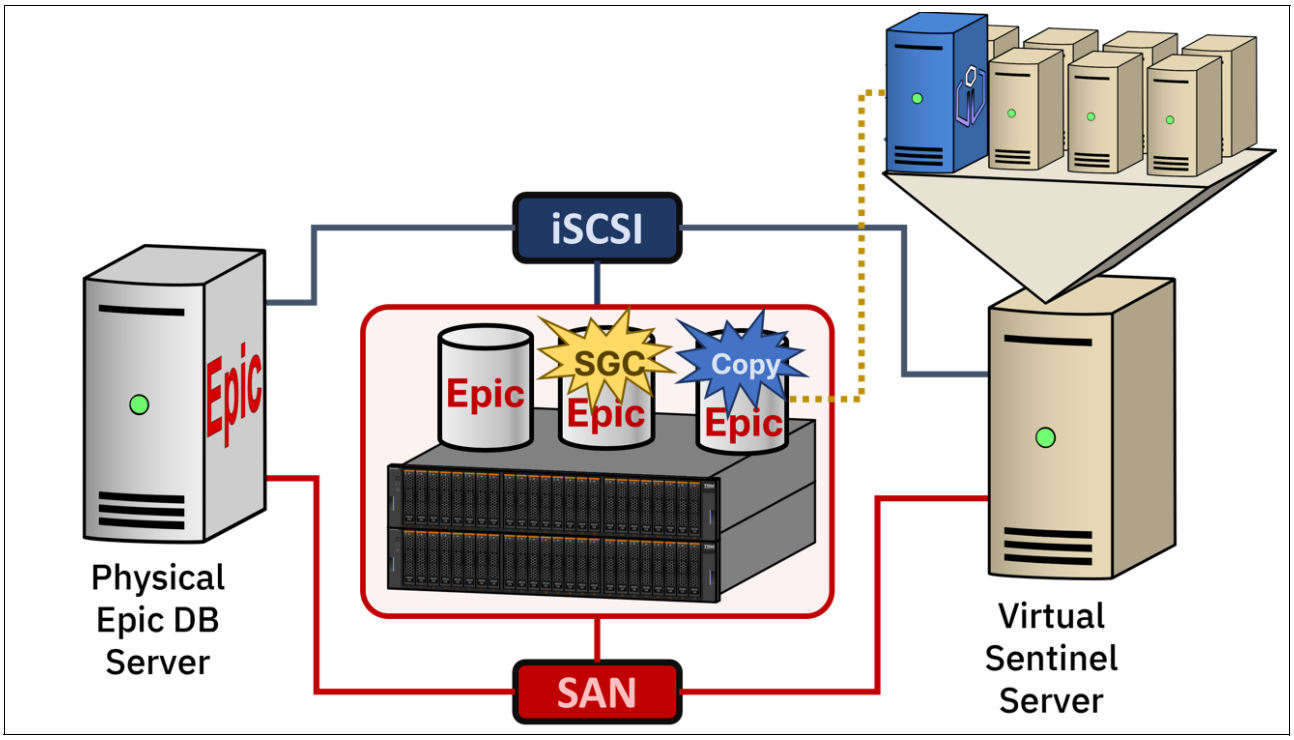


Figure 3-6 Physical Epic DB server and virtual Sentinel server

A backup job workflow will perform the following high-level tasks:

Note: There are many more tasks within this workflow, but we are not going into that level of detail in this chapter.

1. Quiesce the Epic database.
2. Take a Safeguarded Copy of the SAN or iSCSI volumes.
3. Unquiesce the Epic DB.
4. Create a copy of the Safeguarded copies.
5. Share the volumes to the Sentinel machine. If the volumes are served over iSCSI, the process will be identical when using a physical or virtual Sentinel machine. For volumes shared over the SAN, the process will differ slightly between physical or virtual Sentinel machines. If a physical Sentinel host is used, the copy volumes will simply be mapped to that host. If Sentinel is hosted in a VM, the copy volumes will be mapped to the appropriate ESXi host and then defined as pRDM volumes to the Sentinel VM.
6. Scan the copy of the Epic DB.
7. Reverse the process to dismount the volumes and destroy the copies of the Safeguarded Copy volumes.
8. Flag the recovery point created by this backup job as having passed or failed the scan and raise an alert if it did fail.

Note: There are configuration steps omitted from this chapter for brevity. For example, to map volumes to either a physical machine or an ESXi host, there must be host definitions within Storage Virtualize and the SAN zoned, or iSCSI configuration completed so volumes can be mounted to be scanned.

3.4 Setting up an IBM Storage Copy Data Management and IBM Storage Sentinel Environment to scan Epic databases

This section will describe the high-level steps needed to design and deploy a CDM and Sentinel environment to protect and scan your Epic databases. Many of the tasks are described in more detail in other chapters in this book and we will try to avoid too much redundancy.

1. Plan and implement your supported server and storage deployment, as outlined earlier in this chapter. This will need to include predefining Safeguarded Copy volume groups.
2. Plan and implement the security settings and user accounts needed for creating Safeguarded Copies, integrating with vSphere, logging into CDM and Sentinel, and so forth. This is the time to decide you will use local accounts or LDAP/AD, what the scope of authority will be for each account, and so forth.
3. Plan and implement your Sentinel farm at the scale needed to perform the desired scanning of your Epic databases (including any other workload that you plan to support with the Sentinel farm). You need to plan how you will distribute your scanning workloads across the farm, and how often you will be able to scan your application servers with this farm.
4. Deploy a new CDM virtual appliance, if needed. If using LDAP/AD for authentication, configure the security directory. To configure CDM to use LDAP/AD accounts, first register the LDAP server on the Provider Browser in the Configure Page (Figure 3-7 on page 40) and then import the LDAP group by going to the Access Control panel in the Configure Page, add a New User and Import the LDAP Group (Figure 3-8 on page 40).

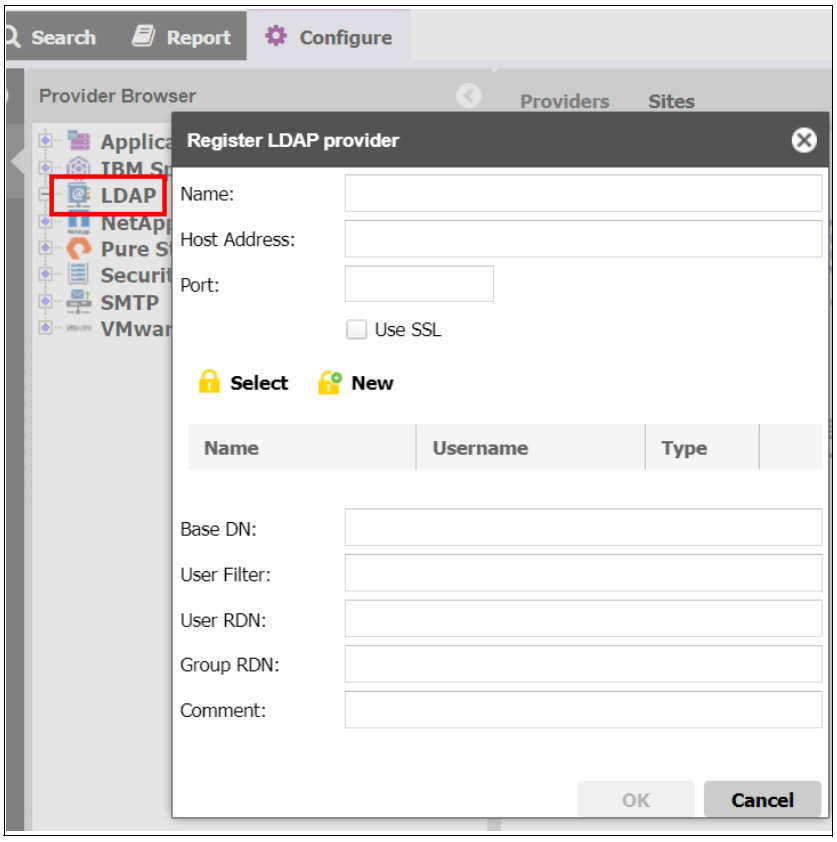


Figure 3-7 Register the LDAP server

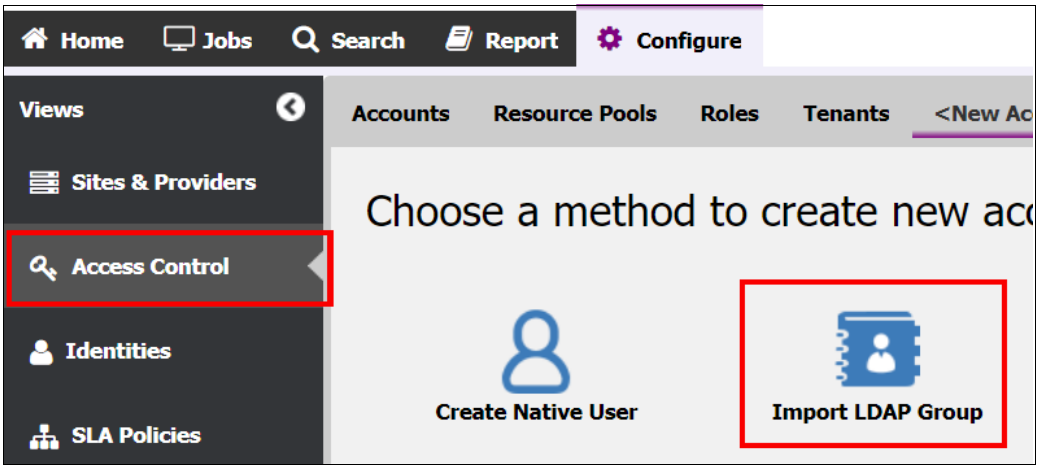


Figure 3-8 Import LDAP Group

5. Define a CDM Site that will contain your storage, vSphere, application server and Sentinel components. You configure Sites on the Sites and Providers section of the Configuration tab. See Figure 3-9 on page 41.

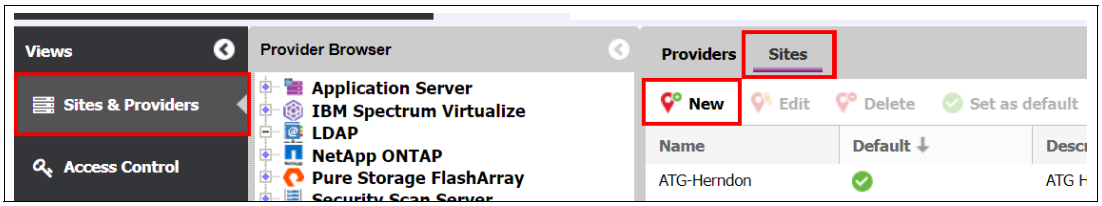


Figure 3-9 Configure Sites

6. Register your storage components, using the appropriate Storage Virtualize user accounts. This will automatically add the storage to the daily Storage Virtualize inventory job and kick off an inventory of the newly registered storage. You register the storage under the IBM Spectrum Virtualize node in the Provider Browser on the Configuration page. See Figure 3-10.

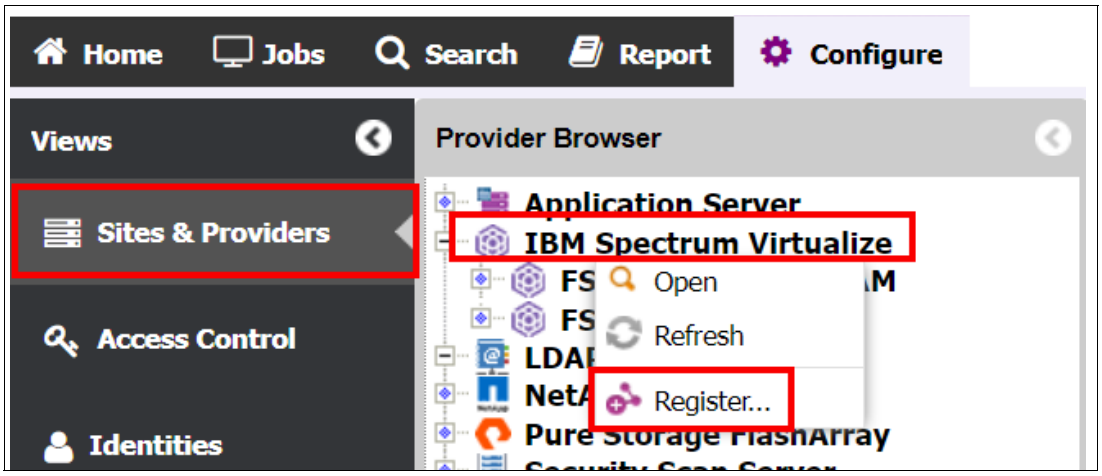


Figure 3-10 Register your storage components

7. If your application server or Sentinel servers are VMs, register your vCenter server(s) to CDM using the appropriate account or certificate. This will automatically add the vSphere environment to the daily vSphere inventory job and start an inventory of the newly registered components. You register your vCenter server(s) on the VMware node of the Provider Browser on the Configuration page. See Figure 3-11 on page 42.

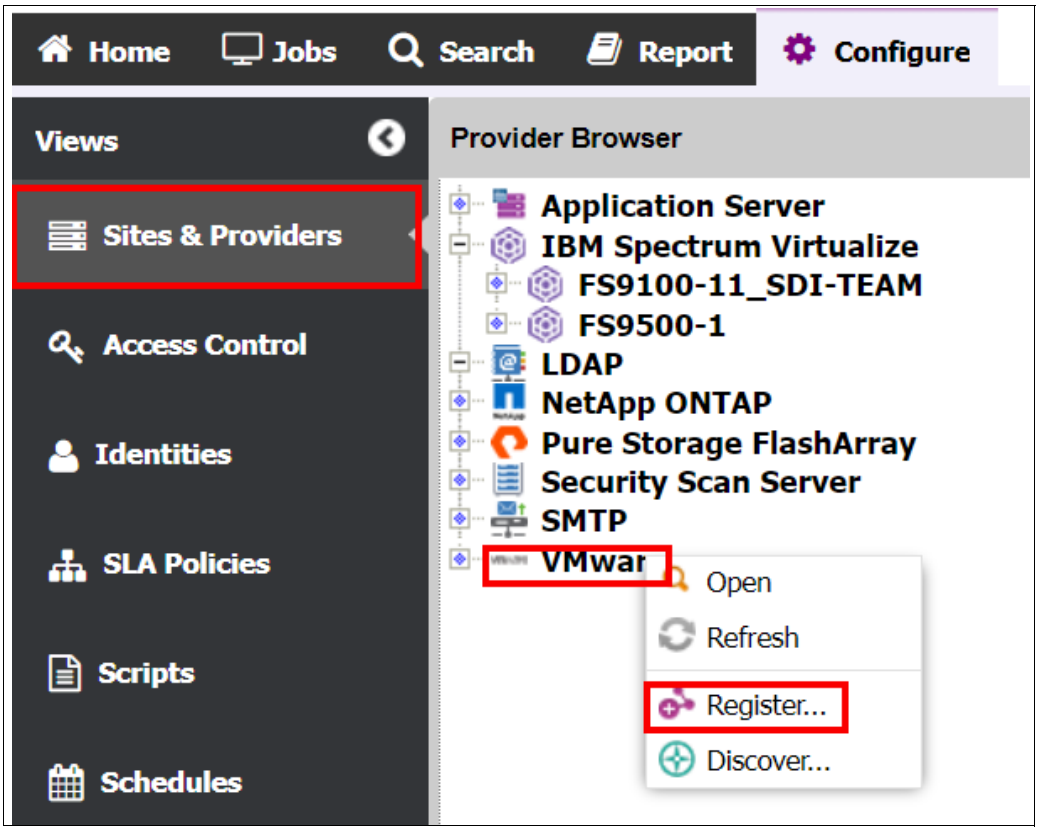


Figure 3-11 Register your vCenter server(s)

8. Register your Epic DB application servers. If your DB servers are VMs, wait until the vCenter inventory has been completed and you have validated the Epic VMs were found during the inventory by expanding the navigation tree in the Configuration tab where you registered the vCenter server. You should see the ESXi hosts and the VMs defined to vCenter in that tree, including the application servers you need to register. You register the Epic DB server in the Application Server node on the Provider Browser on the Configuration Page. See Figure 3-12 on page 43.

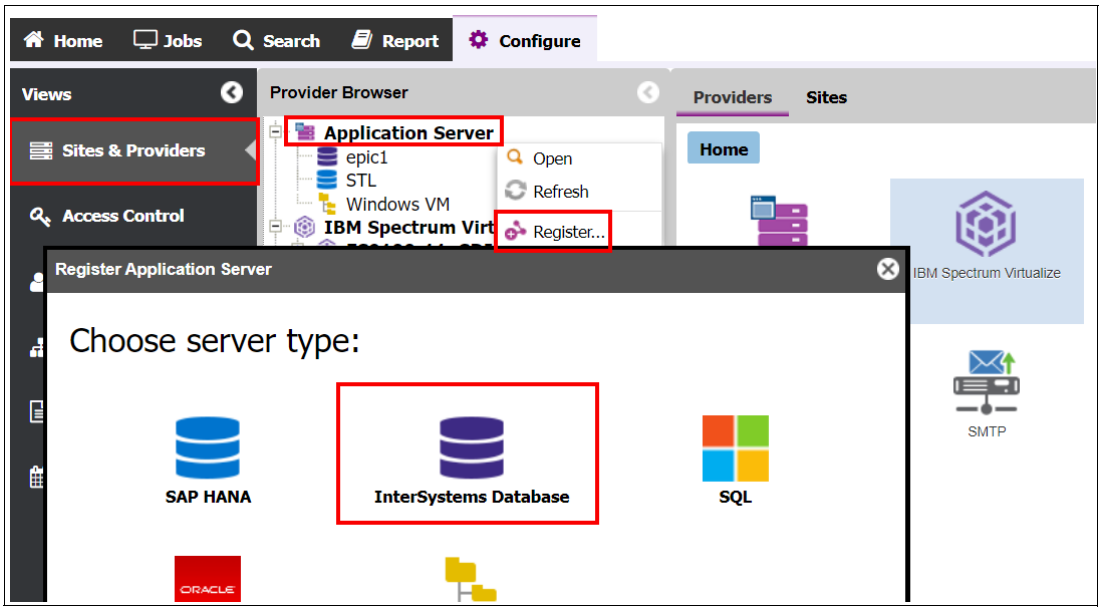


Figure 3-12 Register your Epic DB application servers

9. Register your Sentinel server(s) using the Security Scanner node in the Configuration tree. If your Sentinel servers are VMs, wait until the vCenter inventory has been completed and you have validated the Sentinel VMs were found during the inventory by expanding the navigation tree in the Configuration tab where you registered the vCenter server. You should see the ESXi hosts and the VMs defined to vCenter in that tree, including the Sentinel servers you need to register. You register the Sentinel server under the Security Scan Server node on the Provider Browser in the Configuration page. See Figure 3-13.

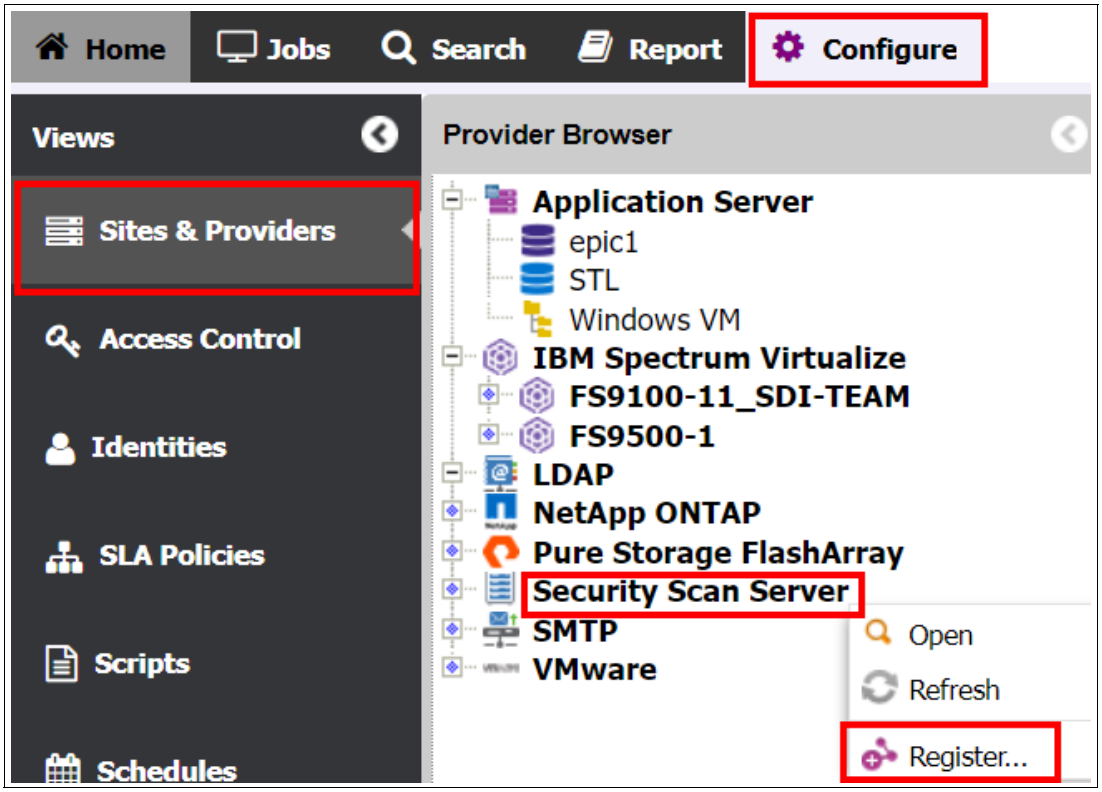


Figure 3-13 Register your Sentinel server(s)

10. Validate each of your components have registered correctly. Most importantly, view the results of the Epic inventory job to validate the Cache or IRIS instances were protected. You will see information on the application detected and the number of databases cataloged. Navigate to the Jobs page, expand the application servers and select Intersystems. Click on the **Default Inventory Job.** On the bottom panel, click **History** and then click on the job log hyperlink to open the job log page.

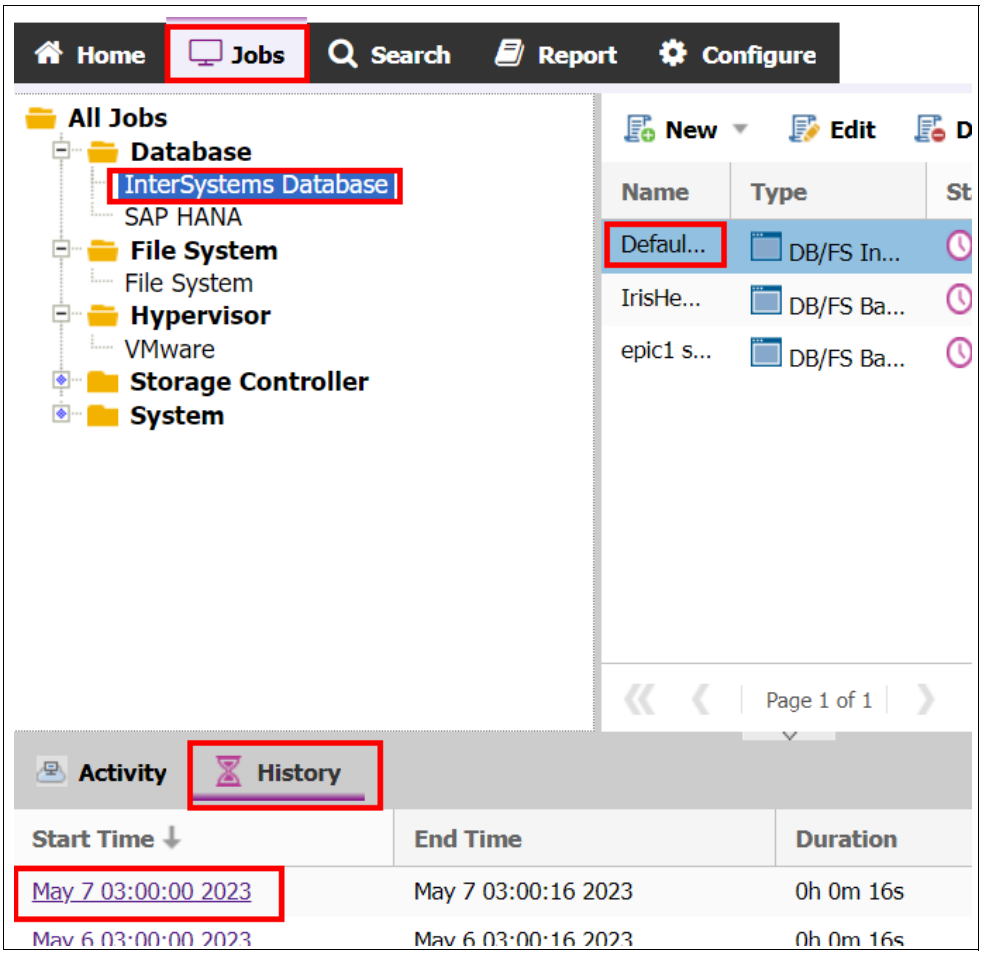


Figure 3-14 Click on the job log hyperlink to open the job log page

Figure 3-15 shows the job log page.

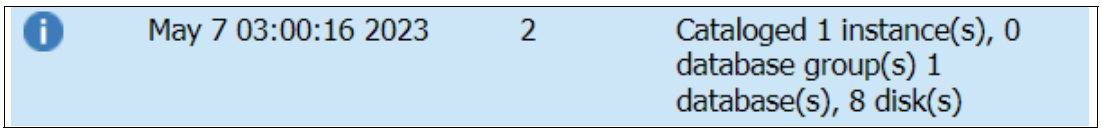


Figure 3-15 Job log page

11. Define an SLA for your Epic DB data protection. You will need to navigate to the SLA Policies section of the Configuration page, click **New** and then select **IBM Spectrum Virtualize**. See Figure 3-16 on page 45.

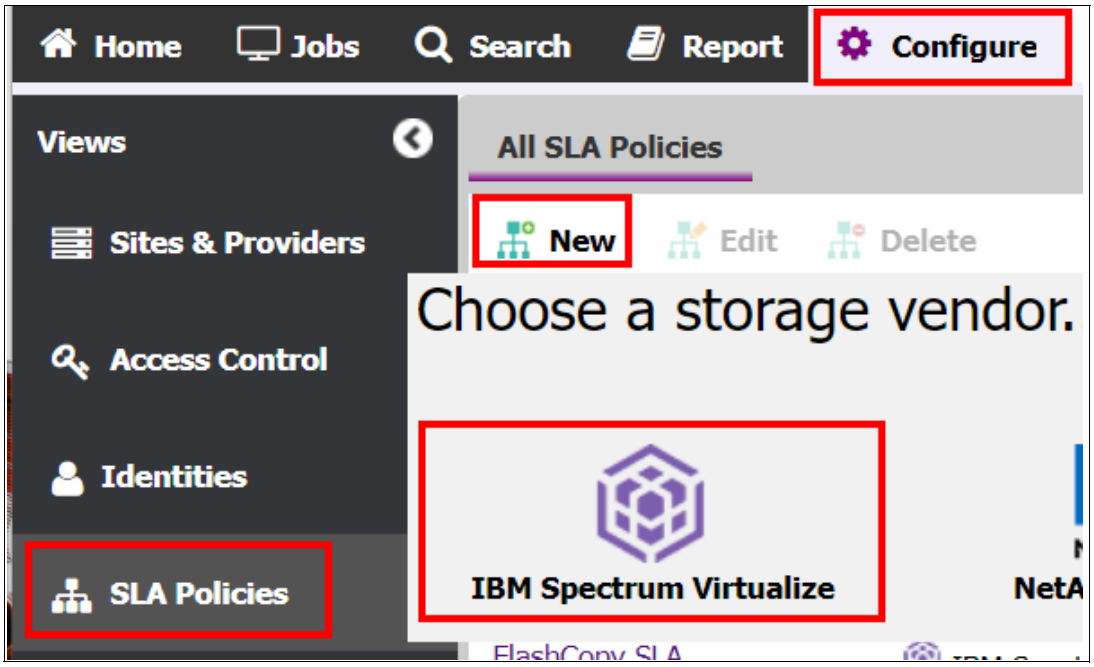


Figure 3-16 Define an SLA for your Epic DB data protection

a. Give the SLA a unique name (ideally one that identifies the type of protection the SLA manages). Add a meaningful comment.

b. Select the **Source** icon and enter the Frequency and Interval for this SLA.

c. Right-click on **Source** and add **Safeguarded Copy**. See Figure 3-17.

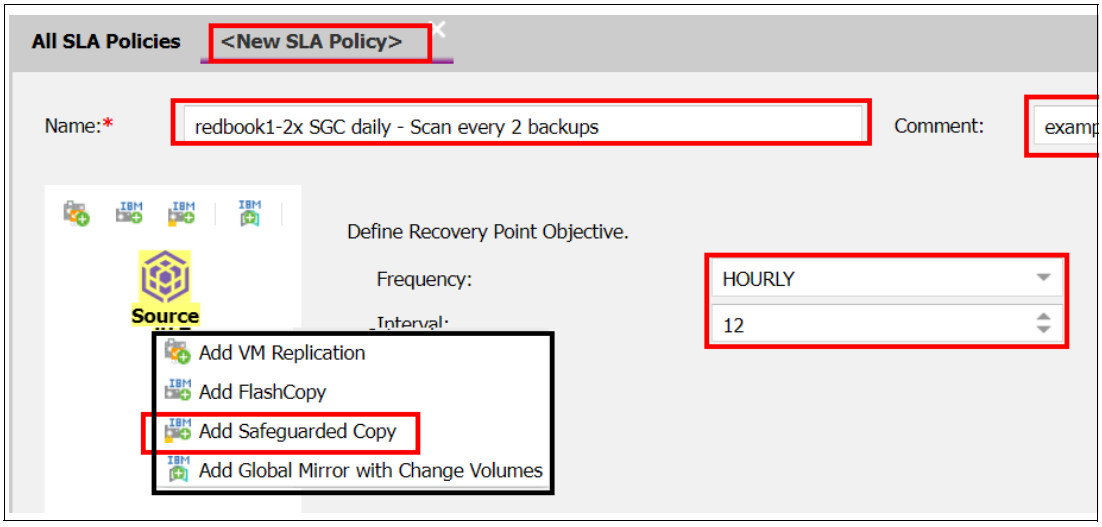


Figure 3-17 Add Safeguarded Copy

d. Select the **Safeguarded Copy Volume Group** to associate with this SLA.

e. Identify the number of days to retain the Safeguarded Copies.

f. Give this set of Safeguarded Copies a unique and meaningful name.

g. If desired, define a volume prefix for the Safeguarded Copies. It is recommended that you do use a unique and meaningful prefix to identify the CDM instance and SLA name, to help correlate Safeguarded Copies back to what created those copies.

h. Check the box to enable security scanning, identify how many backups between scans and select the Sentinel instance you have previously registered for this workload. See Figure 3-18

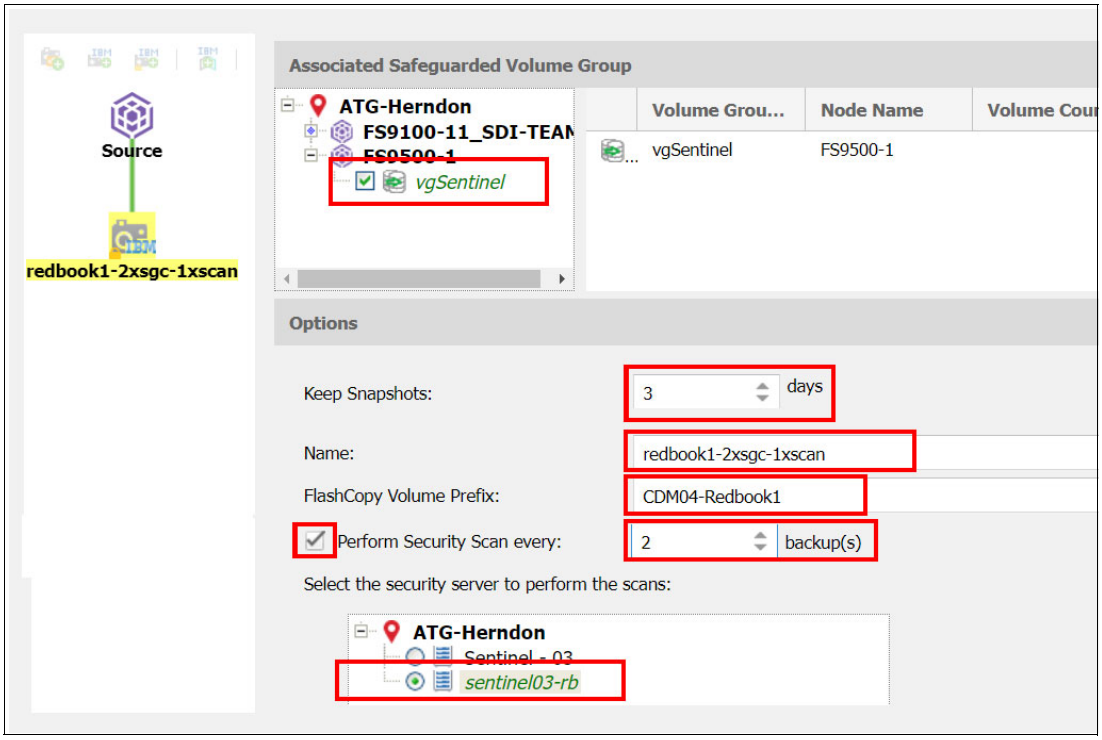


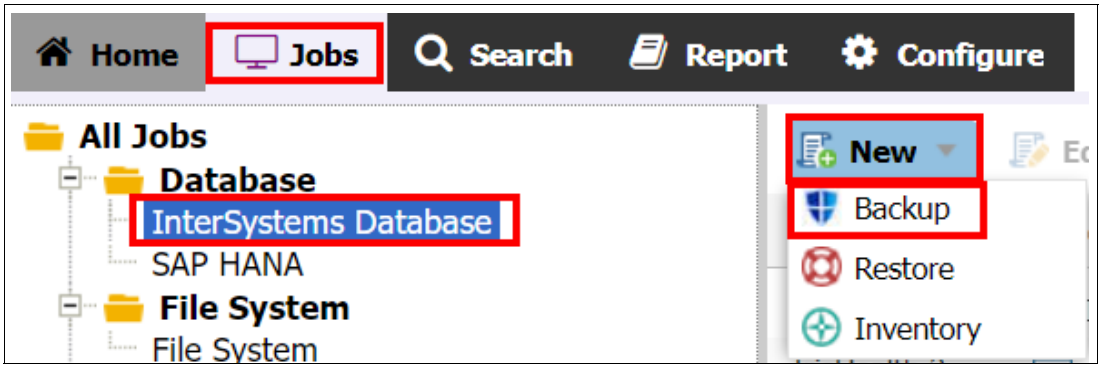
Figure 3-18 Completing the SLA configuration

i. Save your SLA.

12) Define a backup job for your Epic application server(s).

12. Navigate to the Jobs page and expand the Database node and click on Intersystems Database.

a. Click on **New** and add a backup job. See Figure 3-19 on page 47.



b. Give this job a meaningful name and comment.

c. Expand the node next to the site your Epic instance resides within and expand the Epic server's node. Check the box next to the instance or database, as you wish.

d. Select your SLA. See Figure 3-19.

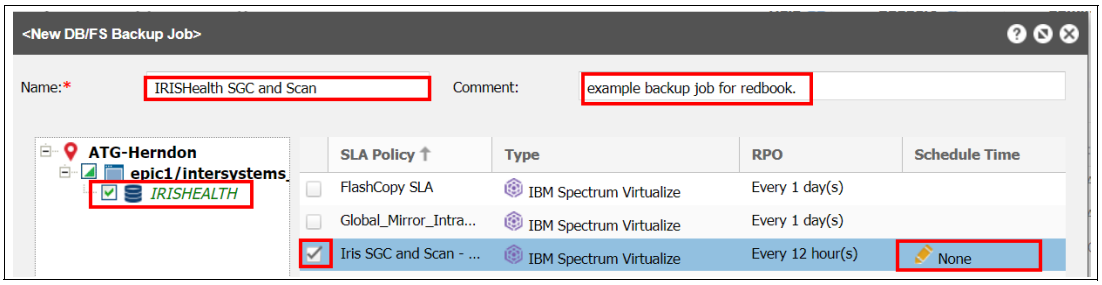


Figure 3-19 Select your SLA.

e. Click on the **Schedule Time** column so you can enable a schedule and set a start time. Keep in mind that how often the job will execute is defined in the SLA. For example, if you have defined the SLA to run every 12 hours and schedule the backup to start at midnight, it will run at midnight and at noon.

f. If you wish to define pre-job, post-job, pre-snapshot or post-snapshot scripts to run as part of this job, click the advanced button and identify the scripts you wish to execute.

13. Monitor your data protection as shown in Figure 3-20. You can manually start the backup job at any time by right-clicking on the job definition, but keep in mind that you may impact your production Epic instance if you run the job at time of high activity.

    Also, keep in mind the impact of running the security scan at that time. If you run the job manually or wait for its scheduled start time, please review the job log for the backup job. You should see activity indicating the instance was quiesced, a snapshot created, the instance was unquiesced and, if a security scan should be performed on this execution of the job, activity showing copies made of the Safeguarded copies and then assigning and mounting those copies to the security scanner. When the job shows as having finished, view the job log in the history panel. If the job has executed correctly, you should be able to go through the steps of defining a restore job and selecting a recovery point, as outlined in 3.5, “Performing a restore of an Epic Database Backup” on page 48.

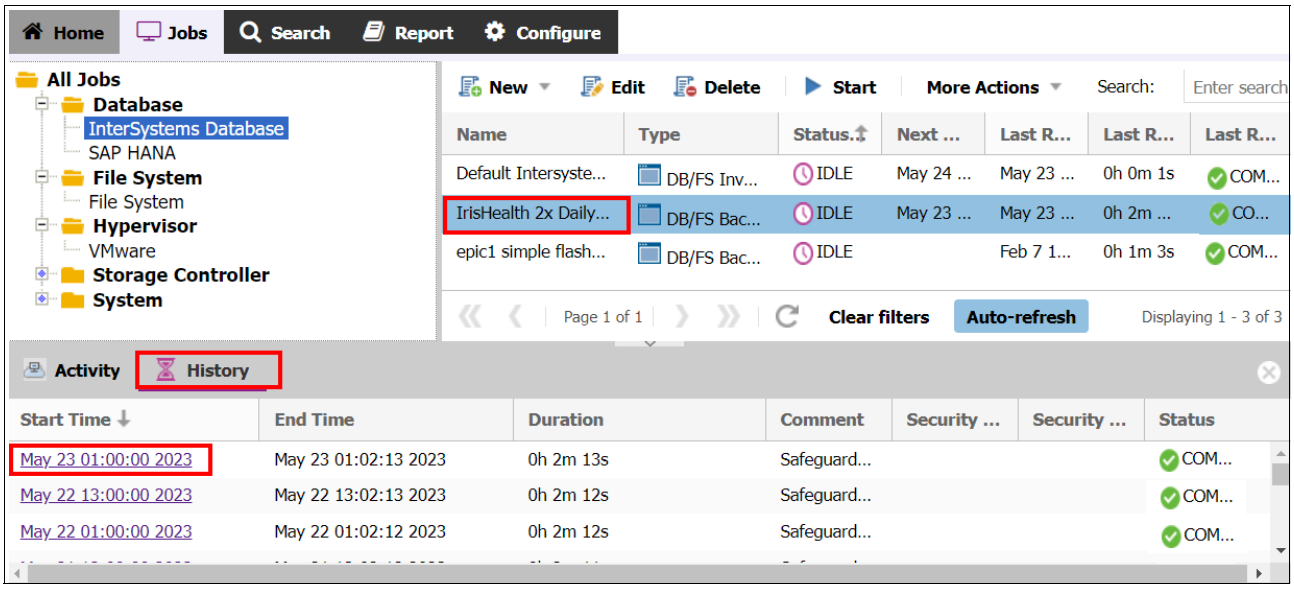


Figure 3-20 Monitor your data protection.

3.5 Performing a restore of an Epic Database Backup

Now that you have scheduled jobs creating Safeguarded Copies of the Epic databases and scanning them for malware corruption, you can perform a restore as needed.

Perform the following steps to define a Restore job.

1. On the Jobs page, expand the **Databases** node and select **Intersystems Database**. Click the **New** button and select **Restore**. See Figure 3-21.

Figure 3-21 Select Restore

2. Enter a meaningful name and comment, select to perform an Instant Disk Restore or Instant Database Restore. An Instant Disk Restore mounts the file systems to the target host but does not try to define or start the database. An Instant Database restore will define and start the database to an Epic database instance. For this example, we will define an instant database restore. See Figure 3-22.

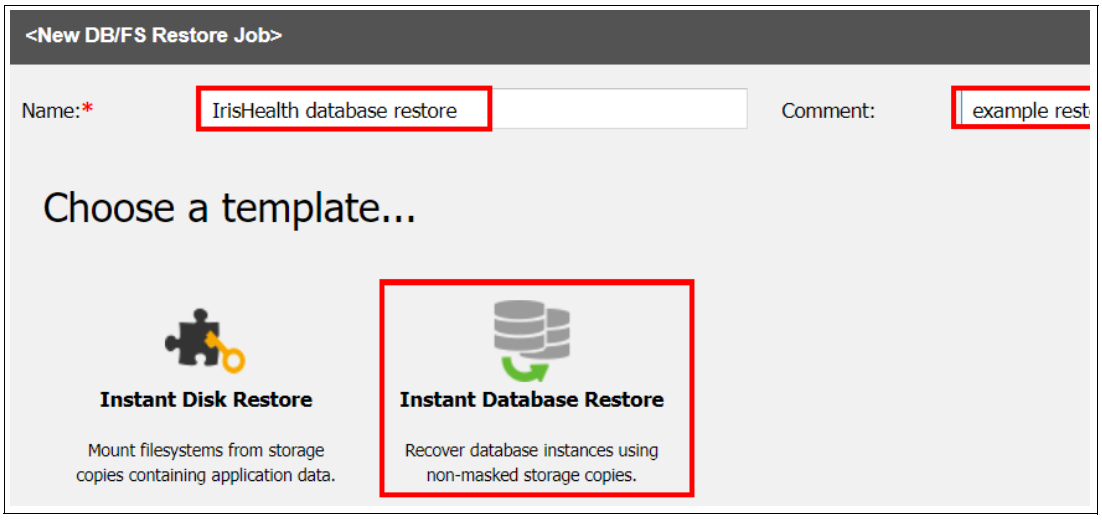


Figure 3-22 Instant Database Restore

3. The source icon will be automatically selected. Use the Application Browser tree to navigate and select the database to be restored. See Figure 3-23 on page 49.

Figure 3-23 Select the database to be restored

4. Click the **Copy** icon. Click the **Select Specific Version** or **Use Latest Successful Scan** button. Here you can click the **Use Latest** or **Use Latest Successful Scan** buttons to select those recovery points. If you wish to select a specific version, click the **Use Latest** text in the Version column. See Figure 3-24.

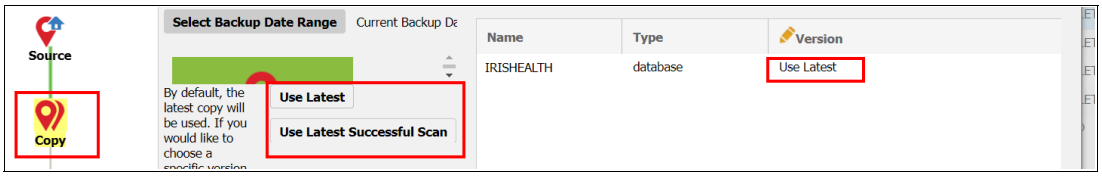


Figure 3-24 Click the Copy icon

5. You can then select a specific version. See Figure 3-25.

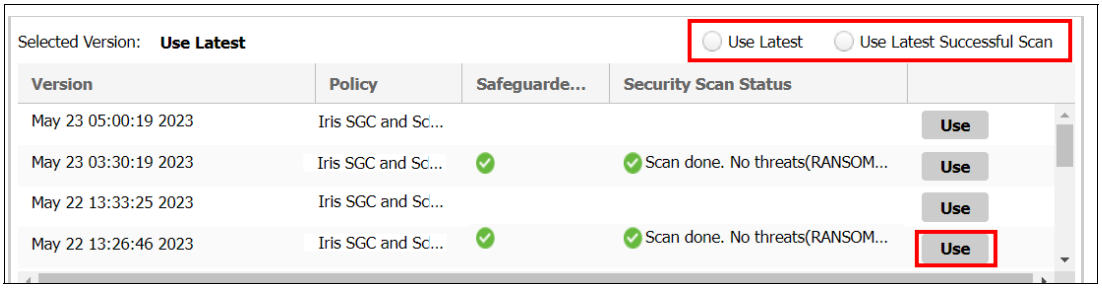


Figure 3-25 Select a specific version

1.5) Click the **Destination** icon. You can then select the database instance as the recovery target. Click **Create Job**. You can then start the job and monitor until completion.

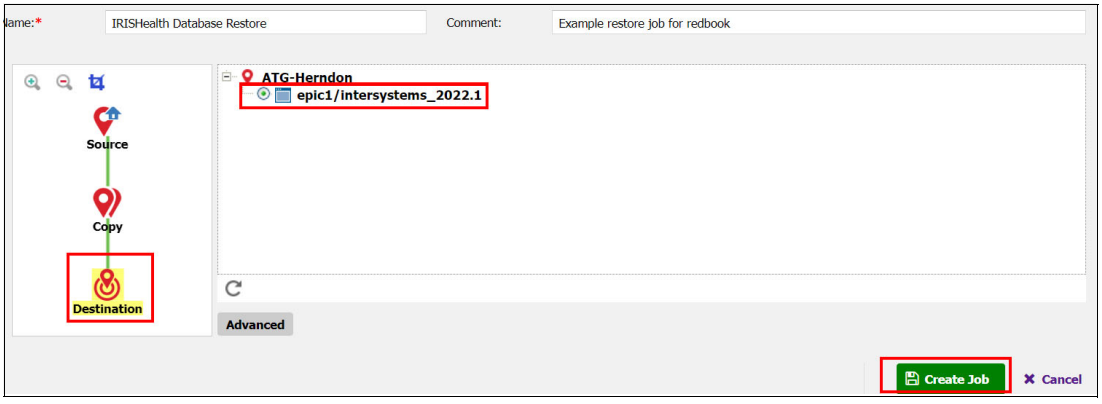


Figure 3-26 Create Job

When the job has completed mounting the volumes to the target, you will notice its status will change to Resource Active, as shown in Figure 3-27 on page 50.

Figure 3-27 Resource Active

6. At this point right-click on the job hyperlink in the Activity panel and select either **Cancel Restore** or **Make Permanent**. Cancel Restore will shutdown the DB, unmount the volumes and delete the temporary copies. Make Permanent will move those copies to a permanent status and complete the job.

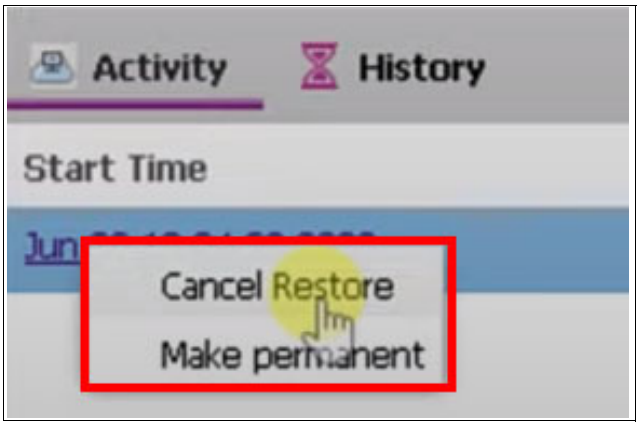


Figure 3-28 Cancel Restore and Make permanent options

4

Configuring IBM Storage Sentinel for SAP HANA

This chapter describes how to configure IBM Storage Sentinel to generate SAP HANA application consistent, storage-based backups and check these backups, proving that they do not contain malware. Immutable snapshots of SAP HANA databases can be scanned, and potential corruption identified.

If an attack occurs, IBM Storage Sentinel helps to identify the best Safeguarded Copy to use for restoring the SAP HANA data. It also automates the process to restore SAP HANA data to online volumes. Because a restore action uses the same snapshot technology as for the backup, it is much faster than the use of offline or cloud-based copies.

The main chapter topics are:

4.1 SAP HANA integration into IBM Storage Copy Data Management

SAP HANA (High-performance ANalytic Appliance) is a multi-model database that stores data in the server's main memory instead of keeping it on a disk. This results in data processing that is magnitudes faster compared to disk-based data systems, allowing for advanced, real-time analytics. Serving as a platform for enterprise resource planning (ERP) software and other business applications.

Depending on the needs of an enterprise, SAP HANA can be deployed on premises, in the cloud, or as a hybrid system, blending the privacy and control of an on-premises system with the lower cost, greater memory, and increased access of the cloud. Its ability to efficiently process enormous amounts of data makes it easily scalable to suit a growing business without sacrificing security or stability.

On the SAP HANA platform, developers can build their own tools and applications that integrate business logic, control logic, and the database layer with unprecedented performance.

IBM Storage Copy Data Management supports SAP HANA on application level and has integrated backup and restore features for the SAP HANA database. Therefore, IBM Storage Copy Data Management is capable to perform application consistent backups directly on the storage layer.

4.2 How SAP HANA creates persistence

SAP HANA is in some aspects different from other relational databases. While it is running, it stores all its data and its metadata in the server's main memory. Reads or updates of tables or columns are always done in memory, usually all the data is being read into memory when the database starts up. For completeness, SAP HANA can be configured in a way that parts of the data ("warm data") will be kept on disk. This feature is known as SAP HANA Native Storage Extension (NSE). As the data is kept in memory, SAP HANA needs to maintain data persistence across a database shutdown. The following paragraphs will show how this works.

In our sample configuration, the underlying storage for both data and log file systems is provisioned by IBM SAN Volume Controller or IBM FlashSystem volumes. To help with designing the storage layout of SAP HANA, IBM has released the *IBM System StorageTM Architecture and Configuration Guide for SAP HANA Tailored Datacenter Integration*, which can be downloaded at: https://www.ibm.com/support/pages/node/6355415.

4.2.1 SAP HANA volumes

The following lists HANA volumes:

- Data Area
  The data volumes are located at a dedicated XFS file system, mounted at */hana/data/<SID>*
- Log Area
  Equivalent to the data volumes, the log volumes reside on a different but dedicated XFS file system, *mounted at /hana/log/<SID>*

Transaction log disk operations

The transaction log ensures that all database transactions are logged in time order. Using this log, database administrators can easily step back to a defined point in time (known as roll back), or they can use the log to step forward to the most recent state of a database after restoring it from a backup. The transaction log is written continuously to disk and is always kept in sync with the actual database transactions. The log volume will be overwritten when it fills up. Therefore the log will be backed up by SAP HANA frequently. Log backups can be maintained by third party backup software like IBM Storage Protect for ERP. Using such a long term backup solution, even older database backups where no life transaction log does exist anymore can benefit from the log roll forward feature.

Transaction data disk operations

SAP HANA dumps its columns, tables and meta data periodically to the data area. These dumps are also known as *savepoints*. Savepoints are usually started every five minutes, but the frequency can be configured in SAP HANA. Whenever a savepoint is written, the former content will be overwritten. Savepoints will be read when the database starts up. In addition to savepoints, SAP HANA backup operators and administrators can also trigger database snapshots. A snapshot behaves like a regular backup. It contains a fixed time stamp and can be added to the database's backup catalog. Snapshots are written into the data area as regular files. Therefore, additional action is required to make a snapshot available to the backup catalog. The snapshot needs to be saved by either copying it to a safe location or by snapshotting the data area using IBM FlashCopy. After saving it, the snapshot will be committed to the database and SAP HANA will remove it from the data area afterwards.

4.3 IBM Storage Copy Data Management orchestration

In this section we discuss IBM Storage Copy Data Management orchestration.

4.3.1 SAP HANA data backup workflow

The complete SAP HANA data backup workflow will be orchestrated by IBM Storage Copy Data Management. This includes several steps, starting from the creation of the SAP HANA snapshot via SAP HANA SQL statements, freezing the XFS file system, taking a IBM FlashCopy of the underlying SAN volumes, and finally committing the SAP HANA snapshot to the database. The overall workflow is shown in Figure 4-1 on page 54.

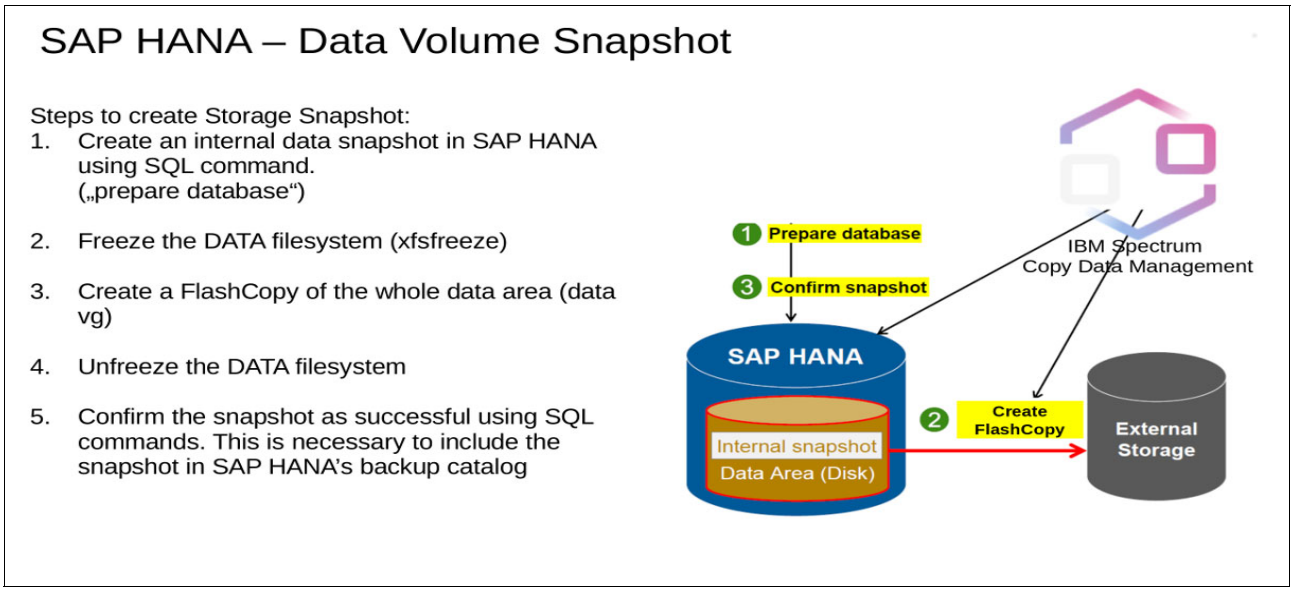


Figure 4-1 IBM Storage Copy Data Management - SAP HANA backup workflow

In a first step the registered components are scanned by IBM Storage Copy Data Management. The information collected by this scan is stored in the internal database of IBM Storage Copy Data Management. This allows IBM Storage Copy Data Management to perform backup and restore jobs fast, without the need to re-scan the systems before and after running a backup job.

To backup an SAP HANA database, IBM Storage Copy Data Management needs to know the complete data path from the SAP HANA data volumes down to the storage volumes where this data is stored. To configure the storage for SAP HANA properly, IBM recommends to follow the guidelines of the *IBM System Storage Architecture and Configuration Guide for SAP HANA Tailored Datacenter Integration*.

A schematic view of the data path is shown in Figure 4-2 on page 55.

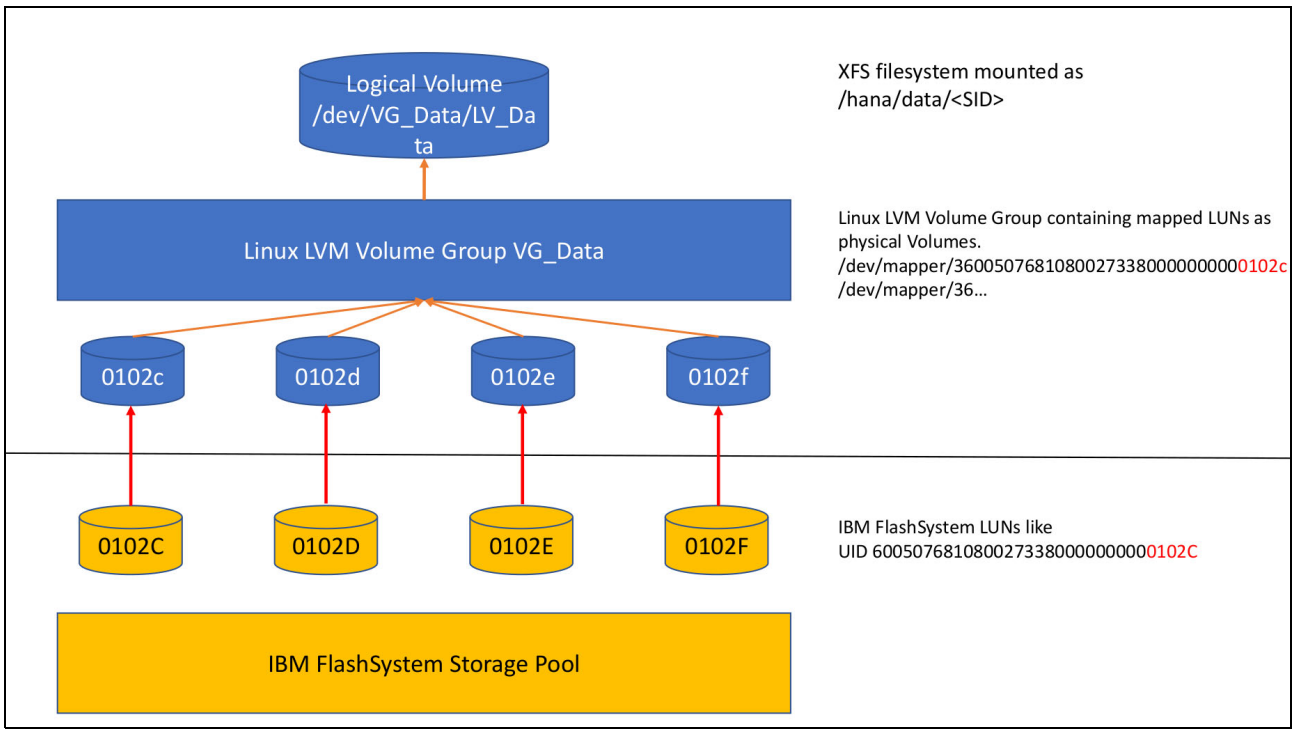


Figure 4-2 SAP HANA Data Volume: Data path in a SAN environment

In this example, IBM Storage Copy Data Management will identify the four Spectrum Virtualize volumes 0102C – 0102F as the LUNs holding the SAP data. These volumes will be backed up using the IBM Storage Virtualize FlashCopy feature.

When using FlashCopy, the storage system must ensure that all involved volumes are in a consistent state when the FlashCopy operation starts. IO-operations are not allowed during this time to ensure consistency of the copied data. Spectrum Virtualize uses consistency groups to guarantee time consistent flash copies.

With Safeguarded Copy, the write consistency is managed by the IBM Storage Virtualize volume group. However, currently IBM Storage Copy Data Management supports only “legacy” FlashCopy with Consistency Groups when it flashes the volumes. Figure 4-3 on page 56 illustrates this behavior. The FlashCopy operation takes just a few microseconds, so there is no measurable impact to the host IO performance.

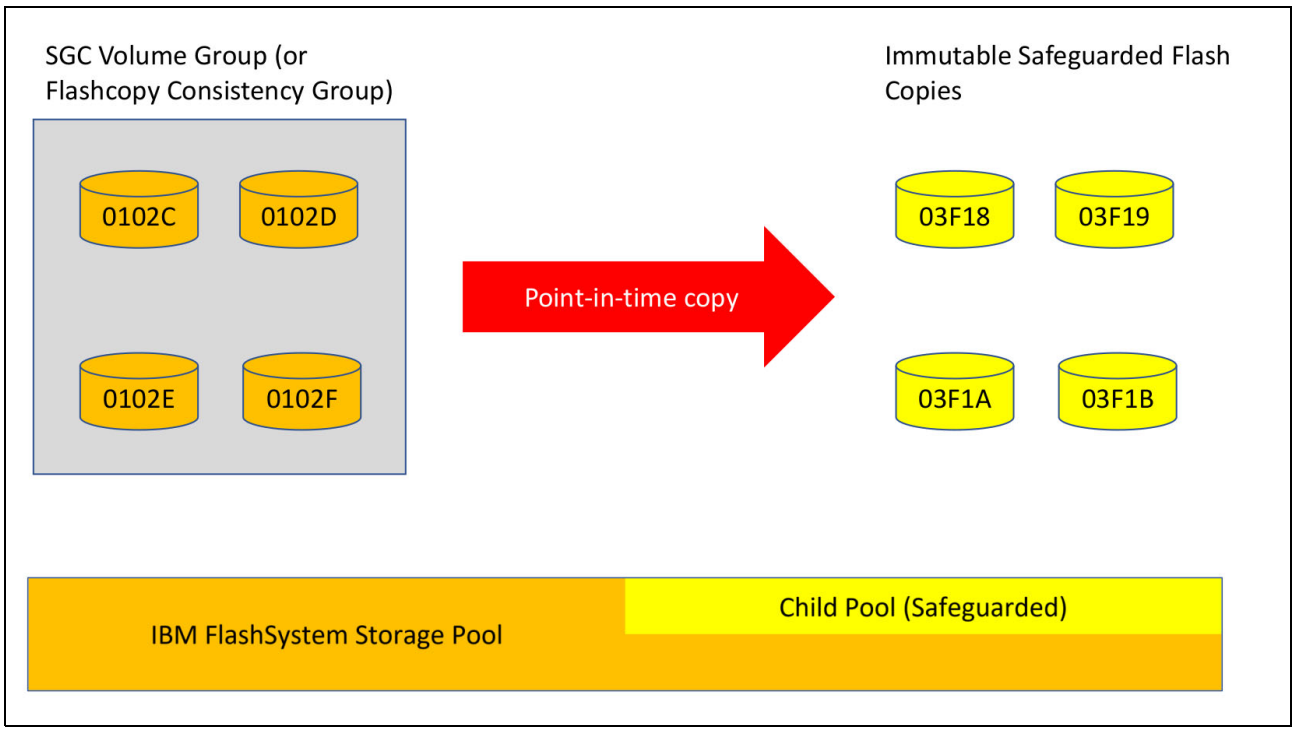


Figure 4-3 FlashCopy using Safeguarded volumes

The SAP HANA database needs to be prepared before creating FlashCopies of the data volumes. IBM Storage Copy Data Management does this by running a set of SQL statements, using an SAP HANA Database Backup Operator account. An SAP HANA snapshot, which is an online backup of the SAP HANA database, is written to the SAP HANA data volume. As soon as the snapshot has been created, I/O operations to the XFS file system are suspended, and the volumes are being flashed. When the flash copy operation is finished, I/O operations are resumed. As a last step, additional SQL statements are issued to inform the database that the snapshot has now been saved, a.k.a. “committed”. SAP HANA stores the backup information into its backup catalog and removes the snapshot file from the data area.

With IBM Storage Sentinel, the Scanning Engine becomes an active part of the SAP HANA backup workflow. In earlier versions of IBM Storage Copy Data Management, a SAP HANA backup job finished after it has completed the steps shown in Example 4-1 on page 58. Sentinel extends the workflow by mapping the Safeguarded copy backup to the Scanning Engine, so that it can be analyzed there.

Since Safeguarded copy volumes cannot be mapped to a host, IBM Storage Copy Data Management needs to perform some additional steps:

- Create additional FlashCopies where the Safeguarded Copy volumes are defined as source volumes.
- The additional FlashCopy target volumes are mountable and will be mounted to the Scanning Engine as a working copy.
- The scanning engine performs a scan, and reports the result back to IBM Storage Copy Data Management.
- Depending on the security scan results, the backup will be marked as either successful or failed by IBM Storage Copy Data Management
- Finally, the working copy will be dismounted and the FlashCopy target volumes will be removed.

4.3.2 SAP HANA restore workflow

IBM Storage Copy Data Management offers two different SAP HANA restore scenarios:

- Instant Disk Restore
  This is a *crash consistent* restore of the SAP HANA data area. No further action will be taken after the data area has been restored. Database Recovery will be handed over to the SAP HANA database administrator. If the database transaction log is available, he can perform a “roll forward” operation to the most recent state of the database, reducing the RPO.
- Instant Database Restore
  This is an *application consistent* restore of the SAP HANA database. After restoring the data area as with a Disk Restore, IBM Storage Copy Data Management will recover the database using the SAP HANA snapshot, which is now available as a regular file in the data area. Any existing transaction log will be cleared by the database restore.

These scenarios will fit into different use cases, and it is the use case which defines what kind of restore is required. For example, in case of a crash a Disk Restore should be chosen, to minimize the RPO and to give the SAP HANA administrators full control of the recovery process. There are other use cases for which a Database Restore makes more sense compared to a Disk Restore, for example restoring a test or development database to a previously defined point in time.

Note: IBM Storage Copy Data Management Instant Database Restore is an automated SAP HANA recovery which restores the database using the snapshot only. Any transaction log which might allow to roll forward to a more recent state will be deleted by this process without further confirmation. Use the Database Recovery feature carefully!

The behavior of a restore job, whether it is a Disk or Database Restore, can be configured in the advanced restore job settings.

- Mountpoint name
  A backup can be restored to either its original mountpoint or to a new one. A new mountpoint can be used for temporary restore operations, for example to check if the data is complete.
- Restore volume handling
  IBM Storage Copy Data Managements does not map the safeguarded copy volumes directly to the host. The default behavior is to create a set of additional recovery volumes and populate these by using he safeguarded copies as source volumes.

  The mappings are started with the “nocopy” option, which means that the restore volumes stay dependent on the backup volumes. If you need independent volumes, this can be changed by making the restore volumes *permanent*. IBM Storage Virtualize will copy the content in the background.

  Instead of creating a new set of recovery volumes, the existing SAP HANA production volumes, which are already mapped to the host can be overwritten with the content of the safeguarded copy volumes. This is called *reverse restore* and is the fastest way to restore data, and preferred for disaster recovery operations.
- Logical Volume Manager changes
  If the file system containing the SAP HANA data area is organized in an LVM structure, IBM Storage Copy Data Management can rename the restored volume group as well as the physical volume IDs. This allows coexistence of both the original data and a cloned set of this data on the same host.

Note: Not all features are available for all environments. For example, the *reverse restore* feature has restrictions in VMware environments when the data to protect is stored on VMware datastores.

4.3.3 SAP HANA requirements

This section explains the main requirements and configuration steps, that have to be done on the SAP HANA application server before registering the SAP HANA server in IBM Storage Copy Data Management 2.2.19. IBM Storage Copy Data Management installs an agent on the SAP HANA application server when the SAP HANA server is registered.

Prerequisites on the SAP HANA application server:

- The SAP HANA data, log and operating system (OS) file systems must be on separate volumes. Ensure, that the SAP HANA data and log volume group only contain volumes from a single and supported storage system.
- The SAP HANA client must be installed on the SAP HANA application server.
- Create a symbolic link to the SAP HANA client installation directory through the following command: **`ln -s /hana/shared/<SID>/hdbclient /opt/hana`**.
- Edit the sudo settings for the HANA administrative user (for example, *stladm*, where *STL* is the SID) to allow the user to run all commands without having to enter a password. Always use the **`visudo`** command to edit the /etc/sudoers file, since visudo validates the sudoers file when it is saved. Add the following lines, as shown in Example 4-1.

Example 4-1 visudo settings

```
stladm ALL=(ALL) NOPASSWD: ALL
Defaults:stladm !requiretty
```

- For SAP HANA 2.0 SPS 04 and SPS 05, the *hdbcli* module must be installed. The hdbcli module must be installed only after the complete installation of the SAP HANA client. Compile and install the binary hdbcli python module, as shown in Example 4-2.

Example 4-2 Installing the hdbcli module

```
[root@hana-host ~]# cd /hana/shared/<SID>/hdbclient
[root@hana-host ~]# tar vfxz hdbcli-<VERSION>.tar.gz
[root@hana-host ~]# cd hdbcli-<VERSION>
[root@hana-host ~]# python3 setup.py install
```

- If SAP HANA is installed on a virtual machine in VMware that is using VMFS datastores for SAP HANA data and log volumes, UUID must be enabled to allow IBM Storage Copy Data Management to discover the LUN IDs of the volumes. To enable it, power off the virtual machine, then select the SAP HANA virtual machine in the vSphere Client and click **Edit Settings**. Select **VM Options**, then the **Advanced** section. Select *E*dit **Configuration**, then find the *disk.EnableUUID* parameter. If set to FALSE, change the value to TRUE. If the parameter is not available, add it by clicking **Add Row**, set the value to TRUE, then power on the virtual machine.

For further configuration details like supported storage systems or OS versions, check the *IBM Storage Copy Data Management - All Requirements Doc*:
https://www.ibm.com/support/pages/node/6586522

4.4 IBM Storage Copy Data Management setup

As outlined in 4.3, “IBM Storage Copy Data Management orchestration” on page 53, IBM Storage Copy Data Management requires management access to all components which are involved in backup or restore jobs. These components are:

- IBM Storage FlashSystem or IBM SAN Volume Controller. Set up of IBM Storage FlashSystem is discussed in Chapter 2, “Orchestration and IBM Safeguarded Copy function” on page 17.
- SAP HANA database. Database access is required for several database queries, for running snapshots and for performing automated database recoveries.
- SAP HANA host. Host access is required for performing storage related operations, but also for OS related SAP HANA tasks.
- VMware vCenter (only required if SAP HANA runs in a virtualized environment). Storage attachment and mapping is managed by VMware, requires a vCenter user with appropriate permissions.
- On IBM PowerVM®, LPARs are treated as physical hosts.

4.4.1 Required user roles

When planning for IBM Storage Copy Data Management related users, a strategy of minimum required privilege should be mandated. For example, there is no need to add a vCenter admin user to IBM Storage Copy Data Management. Better create a dedicated user which has access to the involved resources only, and which also has no other capabilities than the minimum required. To learn more about VMware user permissions, refer to https://docs.vmware.com/en/VMware-vSphere/7.0/com.vmware.vsphere.security.doc/GUID-5372F580-5C23-4E9C-8A4E-EF1B4DD9033E.html.

The SAP HANA host user used by IBM Storage Copy Data Management is generally the OS database admin. This user is automatically created during the installation of SAP HANA. It follows a straight forward naming scheme, usually `<SID>adm`, where the SID is written in lower case letters, for example “stladm” if the SAP HANA SID is STL. This user does not own further permissions or capabilities in the operating system. To keep the user administration in IBM Storage Copy Data Management as simple as possible, the database admin will be used by IBM Storage Copy Data Management for both database management and OS administration. Therefore, the user’s privileges will be enhanced by configuring the **sudo** utility as described by section 4.3.3, “SAP HANA requirements” on page 58.

For administering the SAP HANA database itself, a dedicated DB user should be created. A database user needs to be defined inside the database, it has no dependencies to the operating system. SAP HANA grants privileges by assigning users to previously defined roles. It is - like for the OS user - best practise to follow a minimum required privilege strategy. For further information about SAP HANA backup users, refer to *SAP HANA Administration Guide for SAP HANA Platform- Authorizations Needed for Backup and Recovery*:

https://help.sap.com/docs/SAP_HANA_PLATFORM/6b94445c94ae495c83a19646e7c3fd56/c4b71703bb571014810ebb38dc59cf51.html?version=2.0.05.

Note: The predefined “DATABASE BACKUP OPERATOR” role is sufficient for most SAP HANA tasks triggered by IBM Storage Copy Data Management, except IBM Storage Copy Data Management’s database recovery restore job. For this particular use case, the “DATABASE RECOVERY OPERATOR” or “DATABASE ADMIN” role needs to be assigned.

4.4.2 Service Level Agreement (SLA) policies

The SLA policy defines how to protect a database or file system application. It uses specific features of the underlying Storage system; therefore, separate policies are predefined for different Storage systems. Since many Storage systems offer multiple backup features, additional policies may exist for a specific storage system. For IBM Storage Virtualize, four different types of SLA policies are available:

- VM Replication (VMware only feature)
- FlashCopy
- Safeguarded Copy
- Global Mirror with change volumes

Next to the backup type, the SLA policy also defines the Recovery Point Objective (RPO), the backup retention time, and the system which performs the backup. This grants the storage administrator full control of the backup process. For example, it defines not only the target storage system but also the target pool and other storage related settings like volume prefixes etc. It enables strong separation of duties: the application backup operator does not configure anything on the storage system but uses predefined SLAs as backup policies for the application.

Creating a Safeguarded Copy SLA policy is easy, if the involved objects are correctly defined, as described in Chapter 2, “Orchestration and IBM Safeguarded Copy function” on page 17.

IBM Storage Copy Data Management has an integrated wizard for creating SLA policies. It requires a unique name for the policy and the desired RPO target. Next, the FlashSystem volume group for SAP HANA needs to be selected. Finally, the retention rules and the security scanning server of IBM Sentinel need to be set, as shown in Figure 4-4 on page 60.

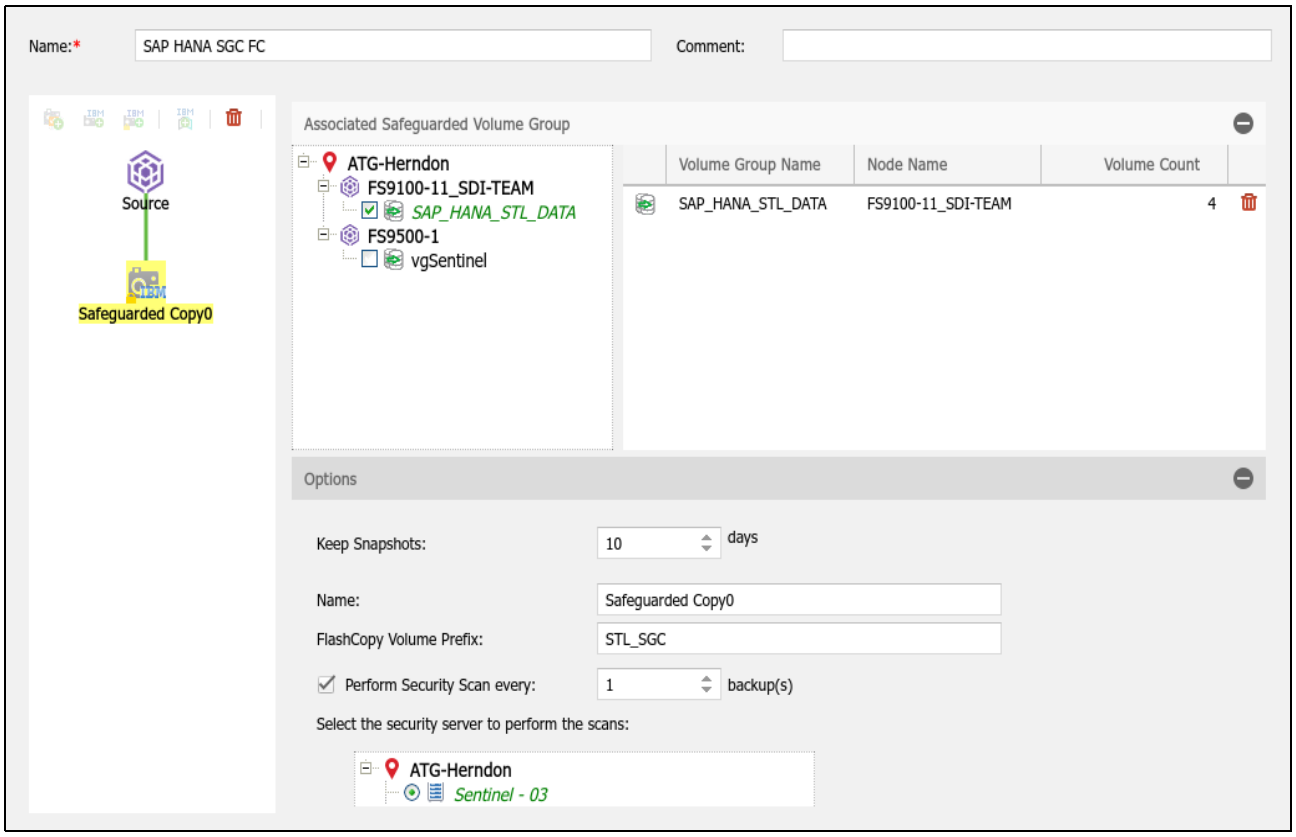


Figure 4-4 Define an SLA policy with Safeguarded Copy and security scan

There is one small but important change in the Safeguarded Copy SLA design compared to all former SLA definitions:

Traditionally, a SLA does not define backup sources, it is just a set of rules describing *how* to back up an application, but not *what* to back up. The latter is described by the job definition. One reason for this design is to achieve a proper separation of duty - storage administrators should do storage system related tasks, while application admins are responsible for backing up the application itself.

Safeguarded Copy depends on a volume group, which is a new entity in IBM Storage Virtualize. Volumes can be logically organized in such a volume group. The idea behind this new feature is to provide an easy way to assign policies, create snapshots etc. to a group of volumes. The volume group must exist prior the creation of the SLA - and it will be associated to that SLA. This implies that the SLA defines the backup source.

4.5 Perform SAP HANA backup and restore operations

This section describes the process how to run a backup and restore job for SAP HANA in an IBM Storage Sentinel environment. In a first step the registered components are scanned by IBM Storage Copy Data Management. The information gathered by this scan is stored in the internal database of IBM Storage Copy Data Management. After a backup has been done, the backup can be configured to be automatically scanned by the anomaly scanning engine.

4.5.1 Performing an SAP HANA backup job

After a backup job has been defined in IBM Storage Copy Data Management, it can be started either manually or by using the built-in scheduler. In this example we start the backup job manually.

1. Click the **Jobs** tab.
2. Select the SAP HANA backup job to run by clicking in the row containing the job name, as shown in Figure 4-5.
3. Click **Start**, or right-click the job name and select **Start**. A confirmation dialog box opens.
4. Click **Yes**. The job session runs.

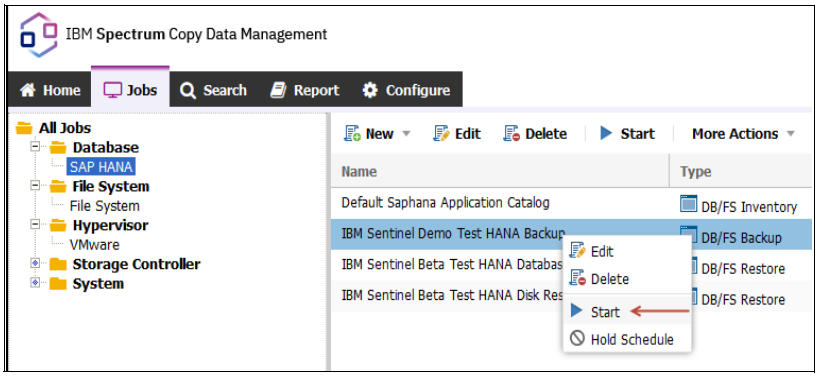


Figure 4-5 Start an SAP HANA backup job

When the backup is completed successfully, the backup job reflects the status as `COMPLETED`. If the backup has the Status `FAILED`, it could either be due to a setup error or the anomaly

scanning engine has detected an infected backup. To check for the cause of a failed backup, view the backup activity log.

In the lower left of the IBM Storage Copy Data Management GUI is the Activity pane. Click the job name to view the activity log for the specific job, including the job session's start date and time, duration, description, status, and associated messages, as shown in Figure 4-6. In this backup log example, the scanning engine has detected a possible malware threat in the backup data. For this reason, the backup job has the status `FAILED`.

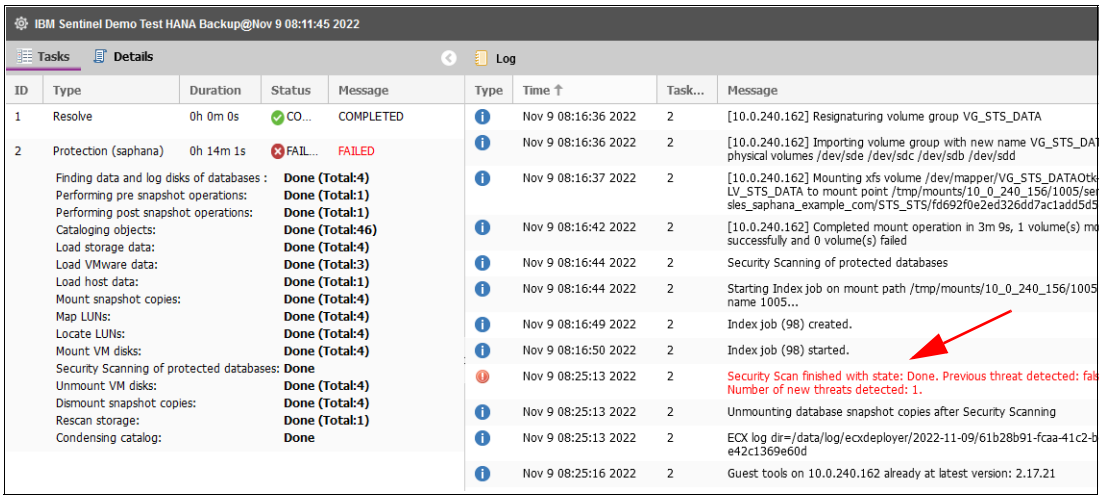


Figure 4-6 Backup job log with error message for infected backup

Backup jobs can be scheduled by the built-in IBM Storage Copy Data Management scheduler, or you can use standard job automation tools to start IBM Storage Copy Data Management backup jobs using REST-API calls.

For email notifications about backup and restore jobs, at least one SMTP server must be configured in IBM Storage Copy Data Management, as shown in Figure 4-7. The SMTP server must be added to IBM Storage Copy Data Management, before defining a backup job

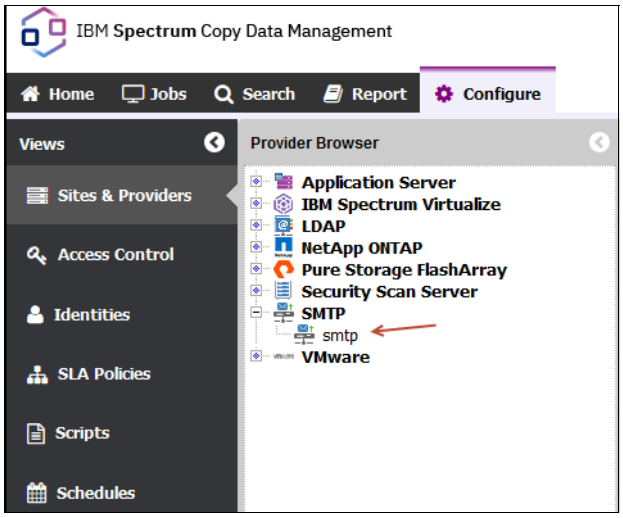
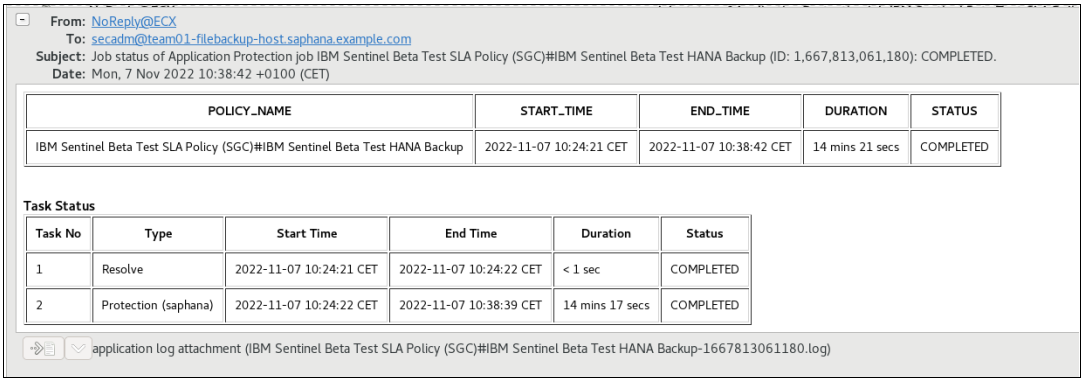


Figure 4-7 SMTP server registered in IBM Storage Copy Data Management.

IBM Storage Copy Data Management backup jobs can be configured to create status emails for every backup job. An example of an SAP HANA backup job status email is shown in Figure 15. This email also includes the job log.


From: NoReply@ECX
To: secadm@team01-filebackup-host.saphana.example.com
Subject: Job status of Application Protection job IBM Sentinel Beta Test SLA Policy (SGC)#IBM Sentinel Beta Test HANA Backup (ID: 1,667,813,061,180): COMPLETED.
Date: Mon, 7 Nov 2022 10:38:42 +0100 (CET)

POLICY_NAME	START_TIME	END_TIME	DURATION	STATUS
IBM Sentinel Beta Test SLA Policy (SGC)#IBM Sentinel Beta Test HANA Backup	2022-11-07 10:24:21 CET	2022-11-07 10:38:42 CET	14 mins 21 secs	COMPLETED

Task Status

Task No	Type	Start Time	End Time	Duration	Status
1	Resolve	2022-11-07 10:24:21 CET	2022-11-07 10:24:22 CET	< 1 sec	COMPLETED
2	Protection (saphana)	2022-11-07 10:24:22 CET	2022-11-07 10:38:39 CET	14 mins 17 secs	COMPLETED

application log attachment (IBM Sentinel Beta Test SLA Policy (SGC)#IBM Sentinel Beta Test HANA Backup-1667813061180.log)

Figure 4-8 Backup job e-mail send by IBM Storage Copy Data Management

4.5.2 SAP HANA restore job

An IBM Storage Copy Data Management database restore performs a complete recovery of the SAP HANA data volumes using the selected FlashCopy backup on the IBM FlashSystem. There are two types of restore (Instant disk or database restore), which are described in “SAP HANA restore workflow” on page 57. In an IBM Storage Sentinel environment it’s also ensured to only restore non-infected backups.

Performing an SAP HANA instant database restore job

An IBM Storage Copy Data Management instant database restore job performs a complete recovery of the HANA database using the selected FlashCopy backup on the IBM FlashSystem. When the database restore job has finished, the SAP HANA database is up and running and recovered to the point in time, where the backup was taken.

1. Login to the IBM Storage Copy Data Management web gui and click the **Jobs** tab. Expand the **Database** folder, then select **SAP HANA**. Click **New** and select **Restore**, as shown in Figure 4-9.

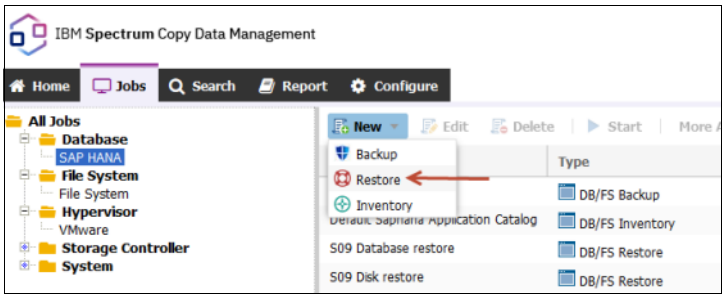


Figure 4-9 Create a HANA restore job

2. Enter a name for your job definition and a meaningful description. Select a template. Available options include *Instant Database Restore* and *Instant Disk Restore*. In this example we choose **Instant Database Restore**, as shown in Figure 4-10.

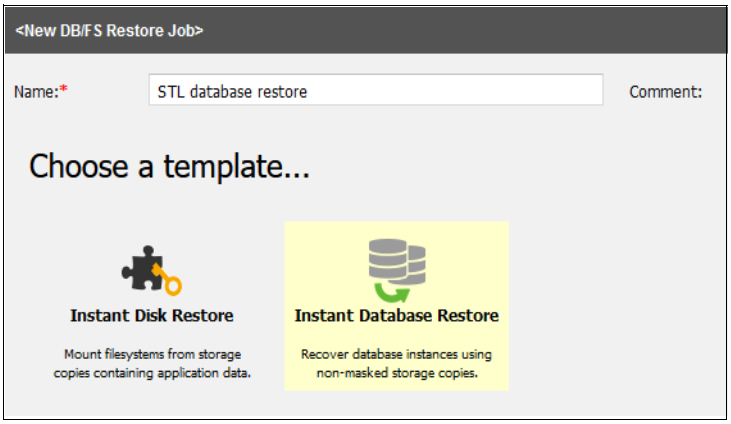


Figure 4-10 Choose the template “Instant Database Restore”

3. Click on the **Source** icon. From the drop-down menu select the Application browser to select a source site and the registered SAP HANA application server to view available database recovery points.

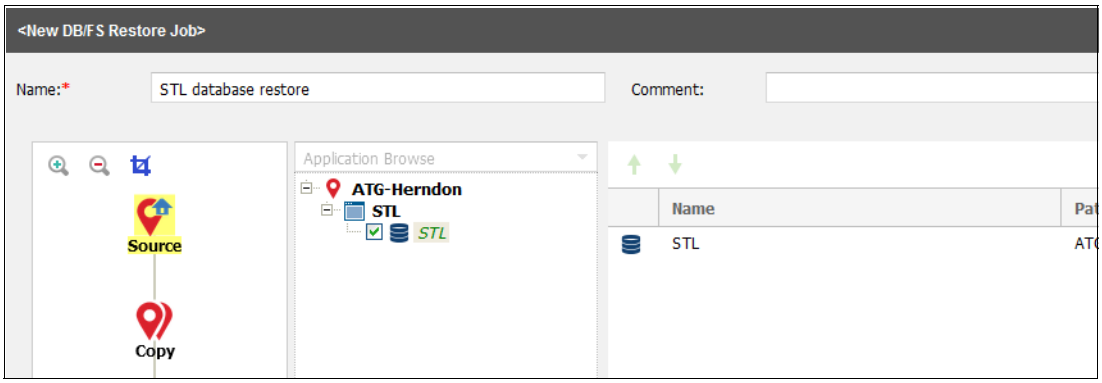


Figure 4-11 Choose the source (HANA application server) for the restore.

4. Click **Copy**. Sites containing copies of the selected data display. Select a site. By default, the latest copy of your data is used. To choose a specific version, select a site and click **Select Version**. Click the **Version** field to view specific copies and their associated job and completion time. See Figure 4-12.

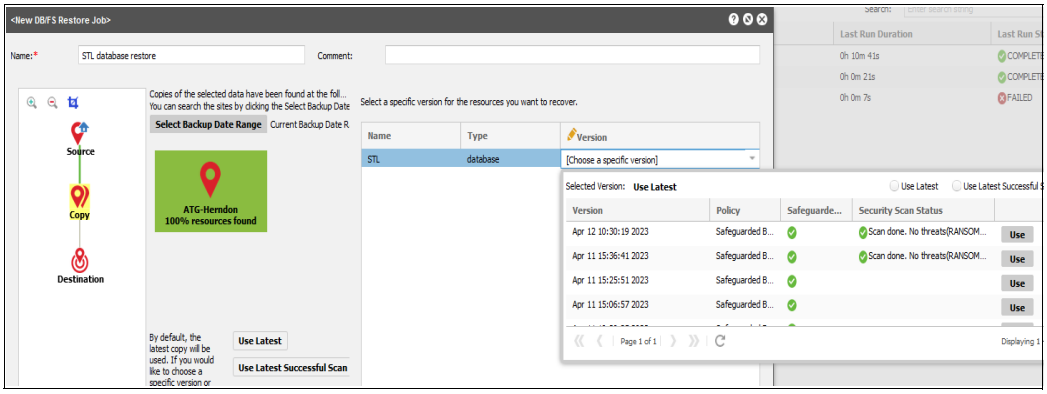


Figure 4-12 Choose a specific backup version

5. Click **Destination**. Select a source site and an associated destination. The database restore can be performed on the original HANA application server.
6. Before creating the restore job definition, click the **Advanced** button. Set the job definition options with the following values:
   - **Do not rename**
     Select this option to not rename mount points during recovery. IBM Storage Copy Data Management will mount them with the same path/name as the source.
   - **Revert: Enabled**
     SAP HANA databases can be restored using the snapshot revert feature. Supported on PureStorage and IBM Storage. This decreases the restore time for large databases.
   - **Protocol Priority: Fibre Channel**
     In this environment the protocol used for the attached volumes was fibre channel.
7. When the restore job information is correct, click **Create Job**. The job runs as defined by a schedule, or can be directly started from the Jobs tab.

Performing an SAP HANA instant disk restore job

An IBM Storage Copy Data Management instant disk restore job performs a complete recovery of the HANA database using the selected FlashCopy backup on the IBM FlashSystem.

The steps for creating an instant disk restore are similar to the instant database restore job described in “Performing an SAP HANA instant database restore job”. The only difference is to choose the **Instant Disk Restore** job template (shown in Figure 4-10 on page 64), when creating the instant disk restore job.

The job processing is similar to the database restore, but in opposite to an instant database restore, the instant disk restore requires a database recovery by the database admin using for example SAP HANA Cockpit or SAP HANA Studio. This restore process allows to do a roll forward of the database to a specific point in time using the SAP HANA tools.

4.6 Daily operations, best practices and maintenance

In this section we discuss daily operations, best practices and maintenance for using IBM Storage Sentinel for SAP HANA.

4.6.1 Adding capacity to the SAP HANA data area

Daily operations with the SAP HANA database often require to increase the data area by assigning additional volumes to the SAP HANA data volume group. During a backup, IBM Storage Copy Data Management always checks the volumes of the SAP HANA data volume group. New data volumes will be automatically discovered by IBM Storage Copy Data Management. The new volume must reside in the same storage system and has to be part of the same volume group as the existing data volumes, as shown in Figure 4-13.

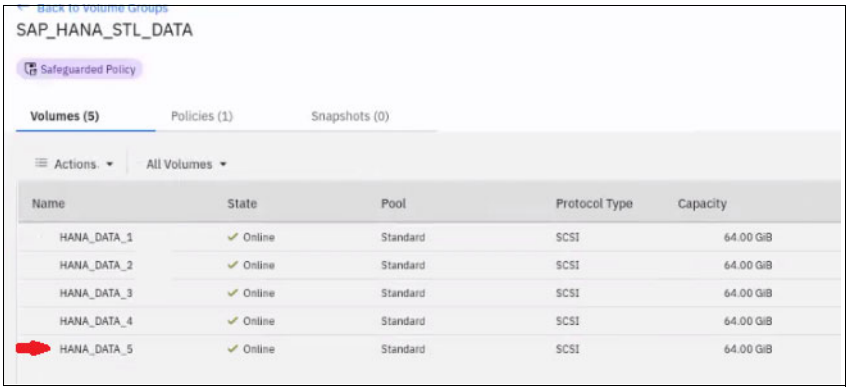


Figure 4-13 Creating a new volume for the HANA data volume group

In this example, a new volume *HANA_DATA_5* was created on the IBM FlashSystem to run in the HANA data volume group on the HANA server. The new volume was added online to the HANA data volume group *VG_STL_DATA*, as shown in Example 4-3.

Example 4-3 Adding the block device /dev/sdn to the HANA vg VG_STL_DATA

```
pvcreate --dataalignment 1M /dev/sdn
vgextend VG_STL_DATA /dev/sdn
lvextend -i1 -l +63832 /dev/VG_STL_DATA/LV_STL_DATA
xfs_growfs /hana/data/STL
```

To detect the changed volume configuration before starting the next HANA backup, a SAP HANA inventory job must be executed. Now the IBM Storage Copy Data Management HANA backup can run and will backup all volumes of the HANA data volume group.

Note: It is a best practice to check if the logical volume which should be extended is striped across multiple physical volumes *before* adding new volumes. To keep the LV extension striped, the number of volumes to add should be the same as the number of volumes used previously. For example, if the logical volume is striped across 4 physical volumes, another 4 volumes should be added. The extension can than be striped across those 4 new volumes, using the “-i 4” parameter in the **`lvextend`** command.

- Combining Safeguarded Copy backups with standard FlashCopy backups
  It’s not recommended to mix FlashCopy and Safeguarded copies for the same FlashSystem volumes. It can be configured by using two different SLA policies for a backup job. One policy is configured for Safeguarded Copy and the second policy for FlashCopy only. For each HANA backup one of the policies can be used.
- Backup of the IBM Storage Copy Data Management catalog.
  Since the catalog is a critical component of IBM Storage Copy Data Management it’s recommended to regularly perform a catalog backup. To manage your IBM Storage Copy Data Management catalog login to the Administrative Console (`https://<HOSTNAME>:8090/`). In the **Menu** select **Catalog Manager and Backup Catalog**.

5

Scanning engine and its technology

This chapter discusses technology and process of the malware scanning software of IBM Storage Sentinel. It describes the planning process, the different options for implementation, and additional considerations.

This chapter has the following sections:

5.1 Storage Sentinel architecture

The idea of scanning data is an important part of data validation. It is meant to ensure that data in safeguarded copy volumes are free of corruption or other changes often induced by a cyber-attack. It helps save time when it is needed most in situations such as:

- You restore data that is potentially compromised into a clean room environment and start analyzing if it is safe to use. This process will be done manually and requires skill and time if there is no automated and intelligent tooling.
- You restore data that has been compromised. While you might realize this quickly you still need to restore another version, or maybe many more, before you find a clean one.
- Your environment is under attack, and as a precaution you restore data that you believe to be clean, but has actually been compromised earlier, and the malware has been lying in wait for a specific time to actually render the data inaccessible.

Figure 5-1 shows the overall idea of scanning the data. A safeguarded copy of production data is again copied into a recovery volume, which then is mounted to the scanning engine in a cleanroom environment. The scanning engine analyses the data found for anomalies or signs of malware behavior.

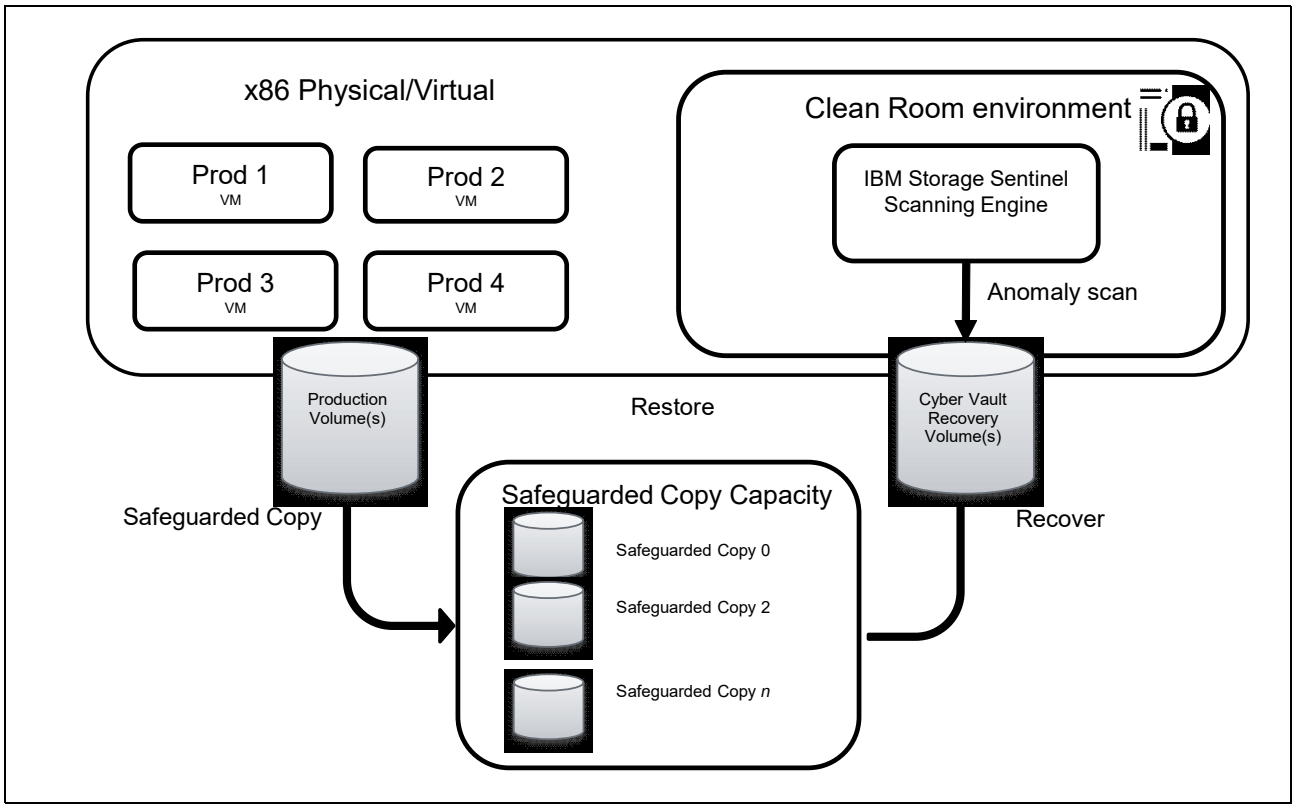


Figure 5-1 IBM Storage Sentinel Scanning workflow

5.2 Technology of the IBM Storage Sentinel scanning engine

In this section we discuss the technology of the IBM Storage Sentinel scanning engine.

5.3 The advantage of anomaly scanning versus signature scanning

Anti-virus software which has been inuse for many years scans file systems for known signatures of virus, or other malware software. Since ransomware developers constantly change their code to avoid detection, the signature databases need to quickly follow these changes, based on the discovery of new variants. What cannot be discovered by this type of scanning are so-called zero-day exploits, meaning new malicious software code that has never appeared before. The signature of such new code is not known yet by anti-virus software, until the vendors have understood its malware nature, and updated their signature databases and the anti-virus software accordingly. Current advanced anomaly scanning techniques, like the scanning engine of IBM Storage Sentinel, looks for the effects of ransomware upon the data, which do not change at the same speed as the ransomware signatures itself. Encrypting a file (or a portion of a file) or modifying and manipulating the metadata of a file are the main effects of triggered ransomware. The anomaly scan engine in IBM Storage Sentinel has been developed with the primary goal of discovering such data manipulations.

5.4 The scanning process

The scanning engine performs a full scan during the first time pass, and builds the metadata of what the data looks like now. Subsequent scans are incremental in comparing what the changed data looks like at that point to what it previously looked like. There are a number of checks that go into building the metadata and comparing the first versus subsequent scans.

The scanning engine uses Machine Learning technology to find anomalies which are considered typical signs of malware activity. It analyses more than 200 different data points of a file, and compares these with thousands of malware patterns, their effects, and previous data collected from tens of thousands of affected backups. This helps it to accurately understand if the files, or databases, have been compromised by ransomware. One of those data points is entropy, which is one of the tell-tale signs of encryption. The deep level of analysis allows it to identify the most subtle attacks where bad actors are using partial encryption to accomplish two tasks. The number one goal is to avoid detection by only changing a small portion of the file, which would not be caught by tools that are only basing their scans on metadata or exceeding thresholds. The second goal is to perform their attack as quickly as possible, and by only touching a portion of each file they compromise, they can move much quicker through a file system.

Unlike many standard anti-virus software products, with the scanning process the engine builds an index containing historical data about previous scans. The comparison between current and older scan data helps it understand if there are suspicious changes between subsequent scans, improving the accuracy and sensitivity of the scan results.

5.5 Scanning process for databases

In active database applications, corruption can go unnoticed until the database is taken offline. It will only fail as soon as the attempt to re-attach comes. Since the scanning engine

can detect corruption in a database's snapshot or backup, whether it was from ransomware or from any other issues, even most critical applications can be validated.

For databases, the scanning process differs from the one used for non-database files. Bearing in mind that the scanning engine is looking for signs of corruption due to a ransomware attack, in the database world, the scanning engine first identifies the type of a file, verifying that the headers and metadata match the known format of the database file. Second, it begins examining the structure of the file, making sure it is readable and has integrity. Lastly, it is scanned at the page level, verifying that the individual pages are intact and corruption free.

5.6 Machine Learning

The Machine Learning Model of the scanning engine is trained at the developer's labs. It is tuned and measured to meet a 95.5% recall rate with .002% false positive rate, using more than 200 analytics that evaluate changes in the file content and metadata from one observation to the next. These analytics are fed to machine learning algorithms in order to make probabilistic decisions of data integrity. The training consists of tens of millions of datasets including clean data, corrupted data, and suspect data, controlled detonation of live ransomware obtained by:

- Public subscription services
- Attack simulation based on academia and global research data
- Anonymized customer data sets

Once installed at a customer site, the machine learning engine is updated periodically, typically 3 - 4 times per year. Since it is not signature based but looks for the effects or malware to data, it does not require weekly or daily updates to operate successfully.

5.7 Scanning encrypted data

If encrypted data can be scanned depends on where the encryption has occurred:

- Data that are encrypted at the volume level cannot be read and analyzed by the scanning engine, but will fail to mount successfully and trigger an appropriate alert
- Data encrypted at file level: changes to file name / file type / file suffix can be discovered and analyzed if the change is suspected to be a malware effect (this type of corruption is more typical of a generic malware attack on a file system, and will render a DB inaccessible, which will trigger an appropriate alert).
- If a database is encrypted for security purposes, the scanning engine cannot read the data from each page but can still see the structure and understands the concept of user based application encryption. It can discover if the file contains corruption due to ransomware.

5.8 How to recognize and handle alerts

The first sign of an alert can be seen in the IBM Storage Copy Data Management GUI. In the job list you will see a failed scanning process with the message "*Threat detected*". An example is shown in Figure 5-2.

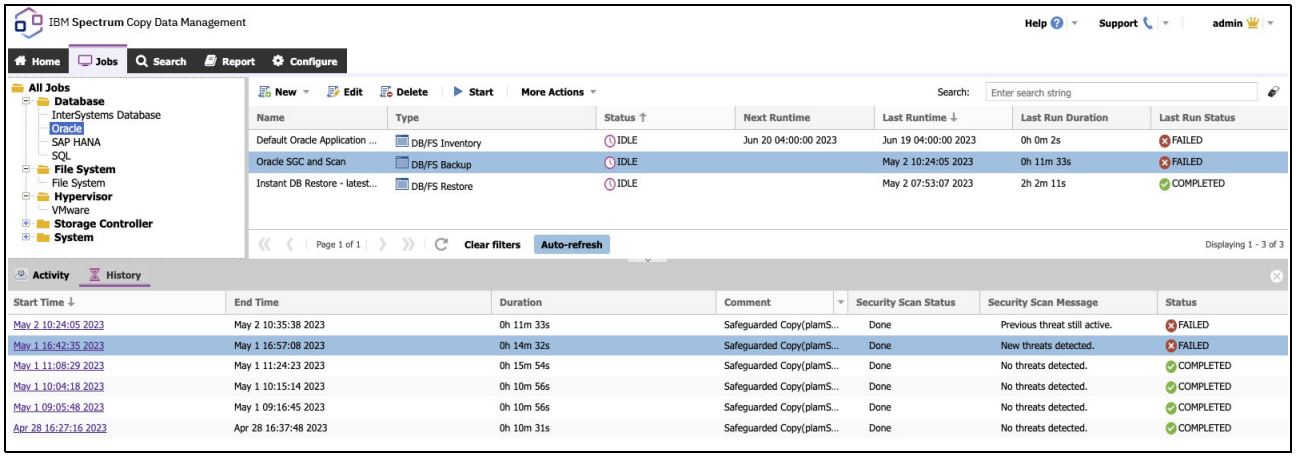


Figure 5-2 Threat detected message

If you analyze the job log you will see when exactly and on which volume the threat was detected, as shown in Figure 5-3 on page 71.

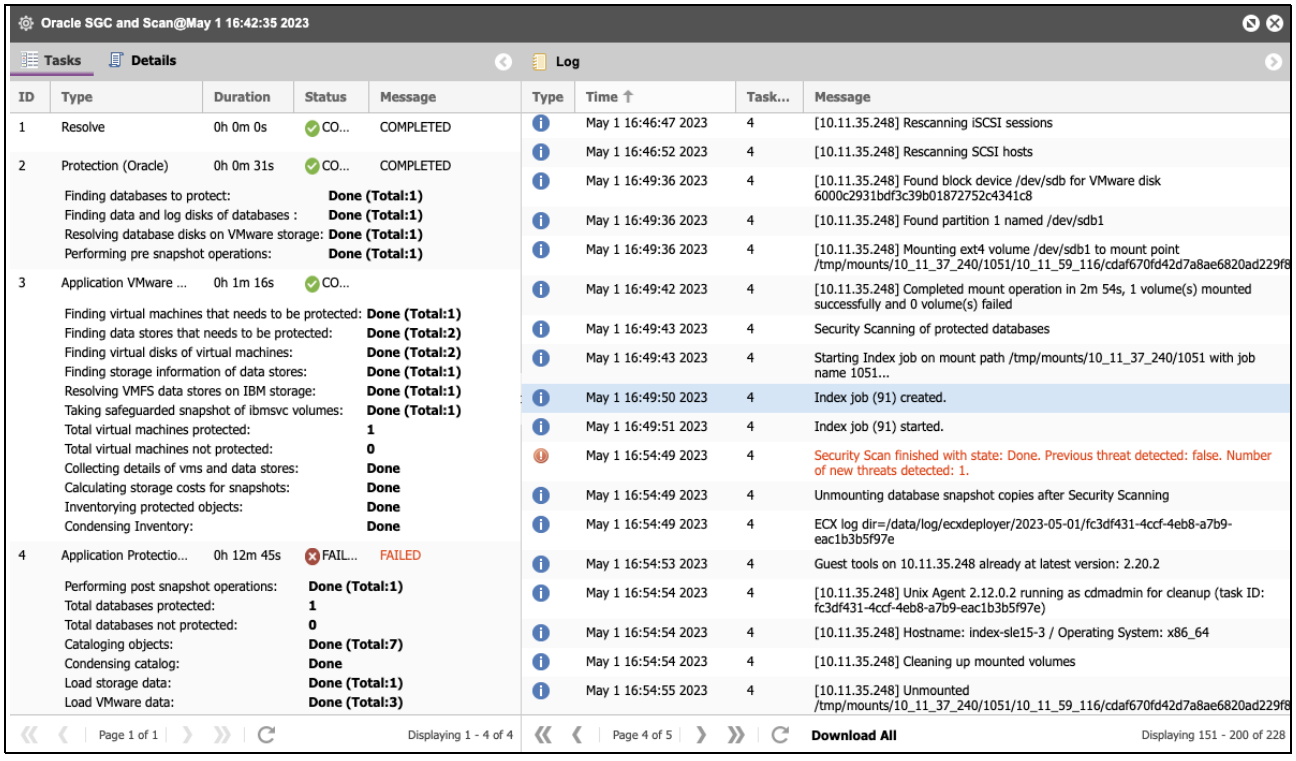


Figure 5-3 Job log showing details on the detected corruption

Finally, the IBM Storage Sentinel dashboard contains more details about the nature of the detected threat. See Figure 5-4.

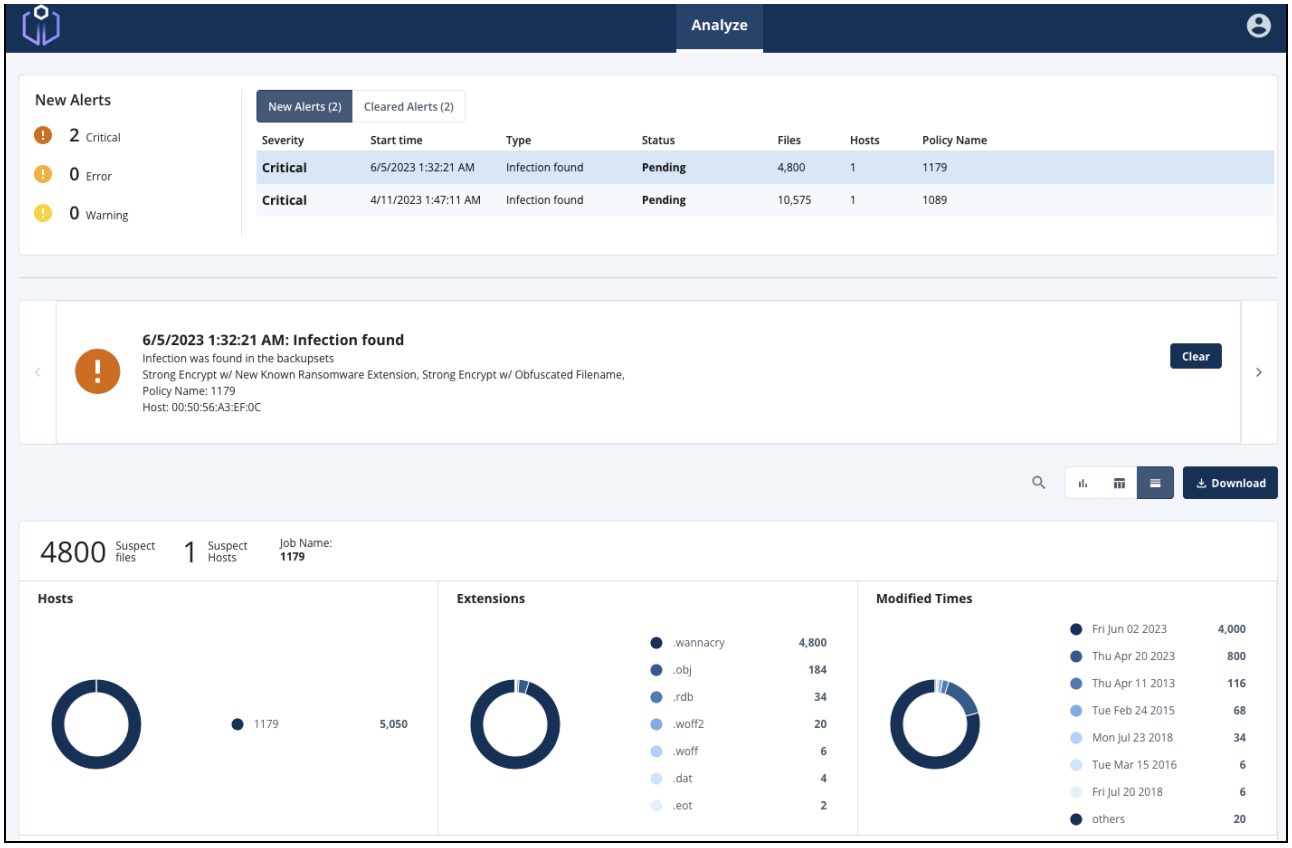


Figure 5-4 IBM Storage Sentinel Scanning Engine dashboard

5.8.1 After alert workflow

IBM Copy Data Management will raise the appropriate alert if a job fails due to Storage Sentinel detecting corruption. In addition, the scanning engine itself can be configured to trigger email notifications and SYSLOG output. The latter can be scripted into any solution that can query logs, such as the typical SIEM/SOAR tools available.

5.8.2 What to do when the scanning engine finds an issue

If a backup is flagged for potential corruption due to ransomware, IBM support needs to be contacted immediately. IBM support teams have been trained how to help their customers and will involve further levels of support up to the development labs. It is critical that a failed scan be analyzed by specialists to verify if an actual corruption has correctly been detected, or if the failed scan is actually a rare false positive.

5.8.3 How to deal with false positives

When an anomaly found by the scanning engine has been analyzed but is found to be an intended data change which is not caused by malicious activity, it is called a false positive. In such a case, the administrator should make a note of what files were involved. If the false positive is caused by a file type that the scanning engine server does not recognize, the information should be submitted to IBM support describing the file type for inclusion in future builds.

5.9 Scanning Engine planning considerations

While you can regard the scanning part of the solution mostly as a black box, there are some concepts and things you should understand:

5.9.1 Sizing considerations

Refer to the IBM Documentation for sizing recommendation of the scanning engine.

5.9.2 Scaling of scan workloads

Because the history of scans is one of the things that makes the Sentinel scanning so powerful, you need to be aware that when you change the scanner used for a specific instance of your application, the history is not transferred, and the new scan engine will create a new history. If you are running a single server, the best practice is to stagger the workloads throughout the window of processing. If you are running multiple servers, the scanning jobs should be balanced across the available scanning engines for optimum utilization of the processing window. Moving a job from one server to another will require a new initial scan and you will lose the scan history from the previous scans. Each scanning engine can run multiple jobs simultaneously. The scanning application is multi-threaded to improve performance.

The implementation of IBM Spectrum Sentinel creates one job per snapshot. Multiple snapshots will run as multiple jobs. If you have more jobs that a single server can handle, these should be split across two or more scan servers.

Storage Sentinel has a federated licensing scheme so licenses of all scanning engines can be managed from a central instance. See the IBM Documentation for details.

5.9.3 Virtual versus physical servers

Note that if the scanning engine is installed on a physical server it can only scan volumes used by physical application servers. If the scanning engine is installed on a virtual machine it is able to scan volumes used by both physical or virtual application servers.

When scanning virtual application servers, one may wish to limit the number of volumes that needed to have Safeguarded Copies created. For example, if the application uses volumes that are virtual disks on a VMFS datastore, you can create dedicated datastores for the VMs that contain an instance of the application. More commonly, customers can place the protected applications on dedicated disks served as physical raw device mapped (pRDM) volumes or using iSCSI. It should be noted that Storage Sentinel does not currently support vSphere vVols.

Storage Sentinel is deployed within a machine running x64 Linux (SLES 15). This means that the file systems on volumes on application servers running x64 Linux can be directly mapped and mounted to the scanning engine. However, Storage Sentinel also supports protected applications running on an IBM Power server and AIX operating system, which cannot be mounted directly to the Sentinel scanning engine. An AIX proxy machine on the Power platform will need to be deployed, so that Copy Data Management can mount the volumes to be scanned to the proxy, and share the file systems over NFS for the scanning engine to access. See Figure 5-5 on page 74.

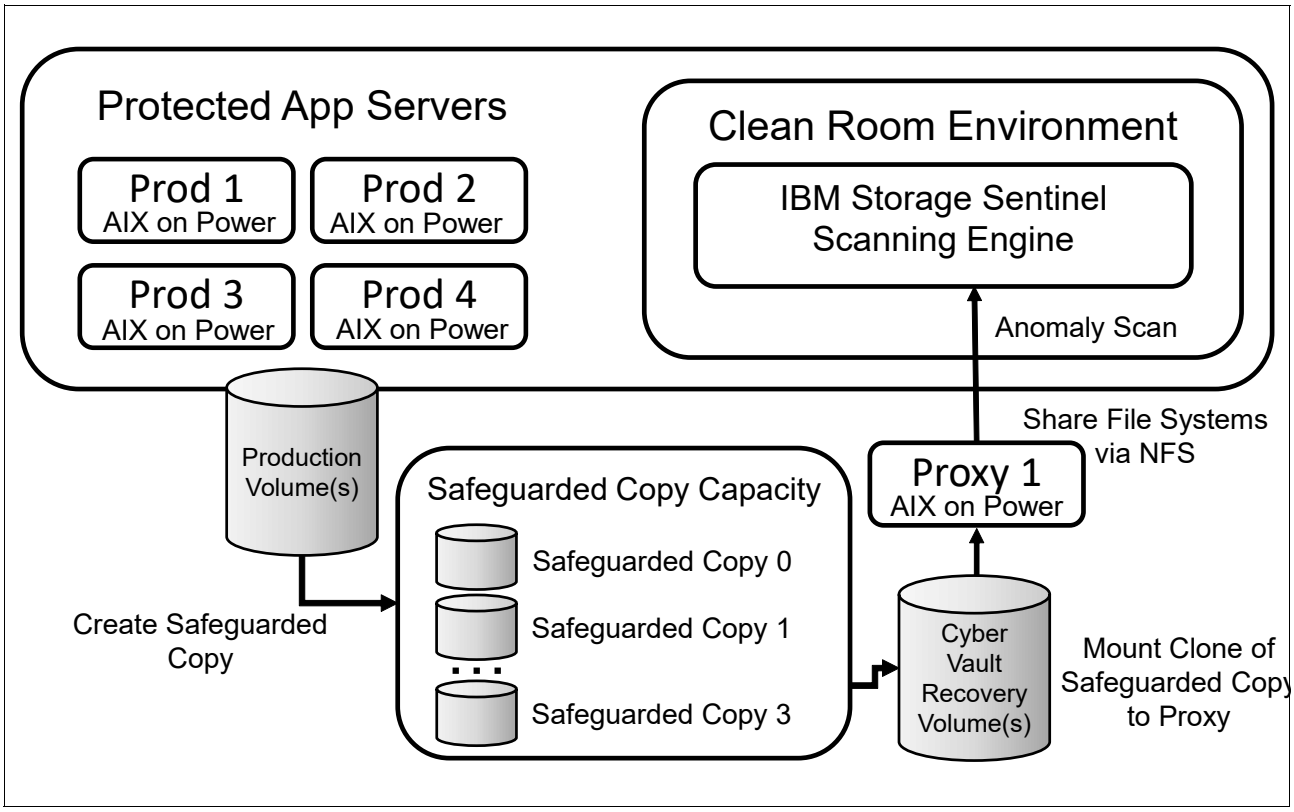


Figure 5-5 IBM Storage Sentinel configuration with an AIX proxy machine on the Power platform

5.10 Administration

In this section we discuss the administration of the scanning engine.

5.10.1 Monitoring the scanning engine

The basic day to day monitoring task is checking the index size and the amount of front-end data that the system scans, to make sure the user is not exceeding the licensed amount. In addition, the configuration should be backed up on a regular basis.

5.10.2 Backup and restore of the scanning engine components

Storage Sentinel includes utilities to allow backup software to take clean backups of the key application files. If you are running a federated Sentinel environment, back up the member instances first, then the manager. At recovery time, you will reverse this order and recover the manager first, then the members.

The default backup location Sentinel is `/opt/ie/backup`. If desired, you can specify a different location by modifying `/opt/ie/var/backup` location. The backup directory must reside in the same filesystem as `/opt/ie/var`.

For IBM Storage Protect (previously called Tivoli Storage Manager/TSM), follow these steps:

1. Add the following lines to `dsm.sys`:

```
PRESCHEDULECMD /opt/ie/bin/tsm_presched
```

```
POSTSCHEDULECMD /opt/ie/bin/tsm_postsched
```

Note: The files `/opt/ie/bin/tsm_presched` and `/opt/ie/bin/tsm_postsched` already exist on the Index Engines system.

On the Spectrum Protect server, do the following to complete the setup:

2. On the Storage Protect server, the copy group for the policy domain, policy set, and management class that the client belongs to and set Copy Serialization to Dynamic.
3. Define a scheduled job to back up the client directory /opt/ie/backup (or the alternate directory if you created one). Either an Incremental or Selective backup will work. There are multiple ways to specify what you wish to back up with Storage Protect. If you wish to have a dedicated backup schedule for protecting the Sentinel backup location, you can specify that location in the `objects` parameter in the schedule definition. You will need to include any subdirectories, so specify `Subdir Yes` in the `dsm.sys` stanza or include it in the schedule settings in the Options parameter. For example:

   ```
   UPDATE SCHEDULE examplePD BackupSentinel type=client action=incremental
   objects='/opt/ie/backup/*' options=-subdir=yes startdate=TODAY starttime=NOW
   dayofweek=any
   ```

   Keep in mind that if you are running a federated environment, you should back up the members before backing up the manager, so you will need at least 2 schedules.

   The Sentinel interface is used to recover from the current contents of the backup directory (Navigate to **Administration** → **System** → **Recovery** → **Recover From Backup**).

   If this is a pristine environment, you will need to install Spectrum Sentinel, restore the backup directory from the backup software, and then use the Sentinel GUI to rebuild the instance from the backup files. If you are running a federated Sentinel environment, restore the manager first, then the members.

5.10.3 Adding new applications

The registration of new applications and the scanning process for these are orchestrated in IBM Storage Copy Data Management. As you add workloads to be scanned, it is important to monitor your Sentinel license information so that you do not exceed your licensed capacity. The number of protected applications or the number of scan servers is not monitored or restricted by the license.

5.10.4 Adding new scanning engines

Additional scanning engines must be registered in the GUI of IBM Storage Copy Data Management. The scanning process for multiple applications should be spread across the scan servers for load balancing.

6

Overall Cyber Vault setup: Putting it all together

This chapter covers the IBM Cyber Vault architecture, a framework developed by IBM for highly automated protection of critical data. It discusses the planning process of IBM Cyber Vault, different options for implementation and additional considerations.

This chapter has the following sections:

6.1 Introduction to IBM Cyber Vault

The goal of the IBM Cyber Vault architecture is to establish a framework for highly automated protection of critical data. This ensures a very fast business recovery after different kinds of incidents. Immutable snapshots enhance the protection especially in the case of a successful cyber attack.

Many cyber attacks will corrupt, encrypt or even wipe application data which in most cases will cause affected applications to stop working. This architecture defines appropriate data protection measures along with environment monitoring, data validation, and recovery planning. The overall idea is that - at any time - a *golden copy* of critical data exists, which:

- Cannot be compromised even by a successful cyber attack.
- Is regularly tested for being free of any signs of corruption or even malware.
- Can be recovered quickly and safely.
- Can be used to recover critical business services as soon as possible after a cyber incident.

The IBM Cyber Vault architecture blueprint provides a detailed description of the overall concept. It can be downloaded at https://www.ibm.com/downloads/cas/ODKXBLR9.

In this context it is important that the set of business critical services and applications are identified and defined which should be protected and are first priority for recovery. These can be described as the *Minimum Viable Company (MVC)*. The application data that are required to recover the MVC are referred to as *primary workloads*; less critical data - that still are needed to recover the full company functionality - are the *secondary workloads*.

As a generic recommendation for improving resiliency against cyber attacks, primary and secondary workloads should be treated, as shown in Table 6-1. Note that these recommendations do not replace standard high availability and disaster recovery precautions.

Table 6-1 Recommendations for workload protection[1]

Property (feature / atttribute) recommended for resilience against cyber attacks	Primary Workloads (Minimum viable company recovery)	Secondary Workloads (Full company workload recovery)
Data retention	short term (days)	long term (days - weeks)
Data copies	immutable (on-array)	immutable if feasible, off-array
Recovery time target	minutes to hours	hours to days
Air gap	logical	locigal or physical

Figure 6-1 on page 79 shows an example of a complete architecture which is built upon the IBM Cyber Vault blueprint, covering primary and secondary workloads including data validation for both.

[1] *Ian Shave, Roger Kasten: The business Impact of Cyber Attacks, January 17, 2023, p. 7*

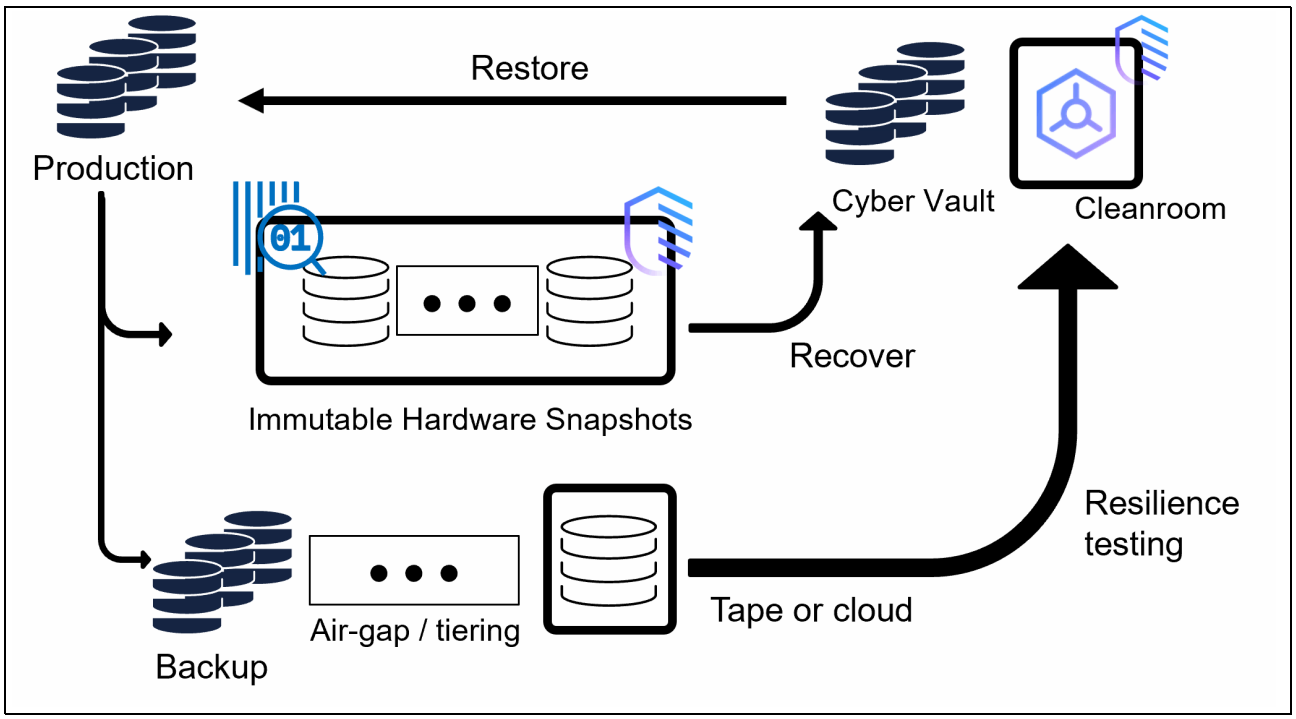


Figure 6-1 IBM Cyber Vault architecture example

6.1.1 The four steps to IBM Cyber Vault

In order to establish the IBM Cyber vault architecture the following four steps can be followed:

1. **Protect:** Protect critical data from cyber threats by periodically creating immutable data copies. These data copies can be established as crash consistent or as application consistent copies. Application consistent copies are more complex to establish since for example a database application needs to be quiesced during snapshot creation, but on the other hand enable a faster restart without the need to do a manual database recovery.
2. **Proactive monitoring:** IBM solution is QRadar® (`https://www.ibm.com/qradar`), but the architecture also allows integration with other Security information and event management (SIEM) tools, such as Splunk
3. **Data validation:** Data validation depends on the applications that need to be protected. The CV architecture blueprint mentions some examples. Alternatively (or in addition, if desired) the client can define their own testing method – it can be as simple as mounting a recovery volume to a VM with their application image and checking if it works as expected.
4. **Recovery:** The recovery processes again often are application specific and should be well defined, for example, target system for recovery, recovery process, level of automation, and so forth.

See Figure 6-2 on page 80.

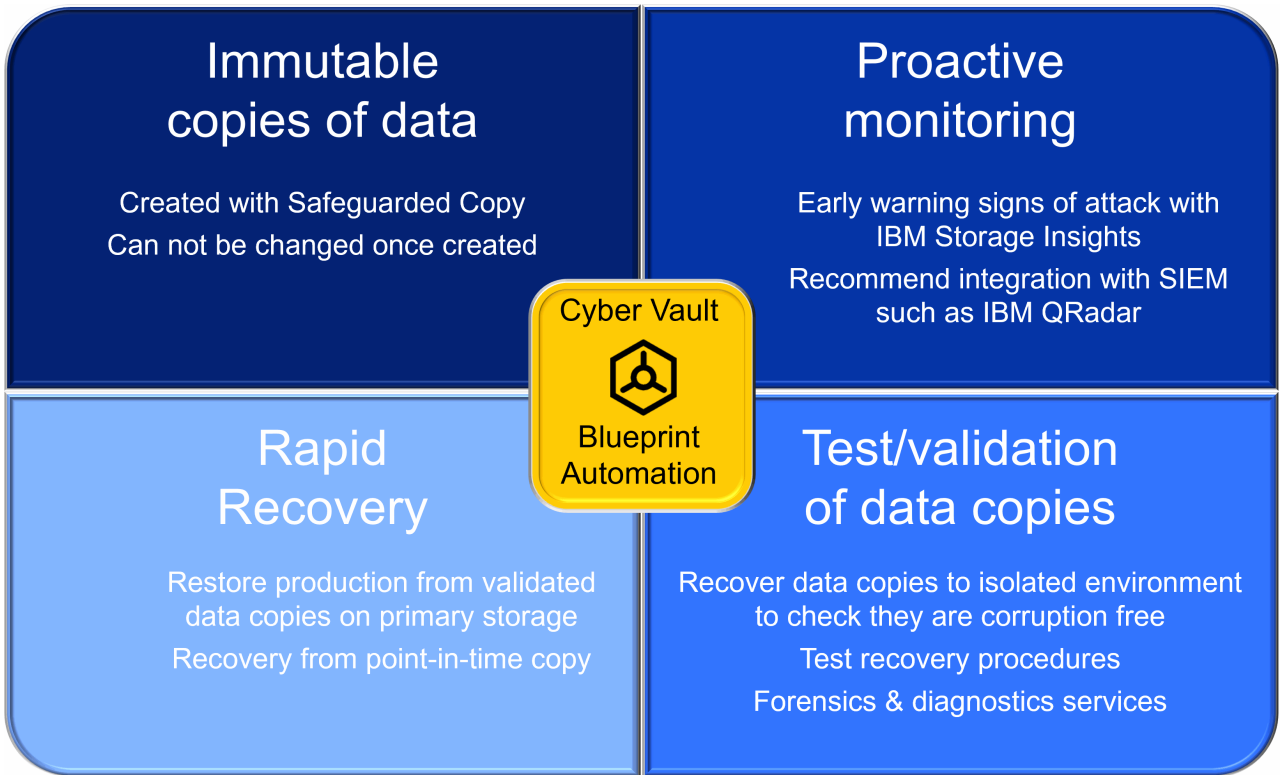


Figure 6-2 The four steps to IBM Cyber Vault

6.2 IBM Cyber Vault planning considerations

This section describes the required considerations for setting up a complete IBM Cyber Vault infrastructure to achieve an appropriate level of cyber resiliency regarding the business criticality of applications and data. After the parameters described in this section such as, RPO and RPO for the most critical business services, have been defined, the infrastructure components can be planned and sized for the expected capacities and workloads.

6.2.1 Definition of the Minimum Viable Company (MVC)

Before infrastructure planning takes place, the critical set of business services and applications must be defined, which will be protected by the IBM Cyber Vault. These should include all services and applications that are required to restart the business in an emergency mode. It allows the business to survive until the full set of services can be recovered. Statistics show that - without appropriate precautions for such attacks - even a partial business recovery may take a few weeks or even months. The idea is to shorten that time frame by identifying the most critical services, and prioritizing their recovery.

6.2.2 Establishing immutable copies of critical data

Since cyber attacks often aim to corrupt, encrypt, or even completely wipe critical data, it is essential to keep data copies that are immune to these threats. Standard backups are not safe, as the backup infrastructure is also a target for cyber criminals. Immutable copies, on the other hand, are immune to these threats because they cannot be altered or deleted. They are the most efficient way to achieve data resiliency and can reside within a storage array like

IBM FlashSystem, or in an air-gapped medium including tape. The frequency of taking immutable copies depends on the RPO (also see 6.2.3, “Crash consistency or application consistency?” on page 81), while their retention time should be appropriate to cover the process duration from the detection of an incident to the decision for recovery. Another consideration regarding retention time is the decreasing value of older data copies. If the business will not survive a longer outage anyway, there is no point in retaining data longer than the assumed survival limit. Does it still make sense to recover data older than, for example, one week? Or would this amount of data loss (or such a long outage period) jeopardize the business as a whole? For even longer retention times, an in-array immutable copy can be migrated to an air-gapped copy on lower cost media. Figure 6-3 on page 81 shows typical backup frequencies and retention times that clients use.

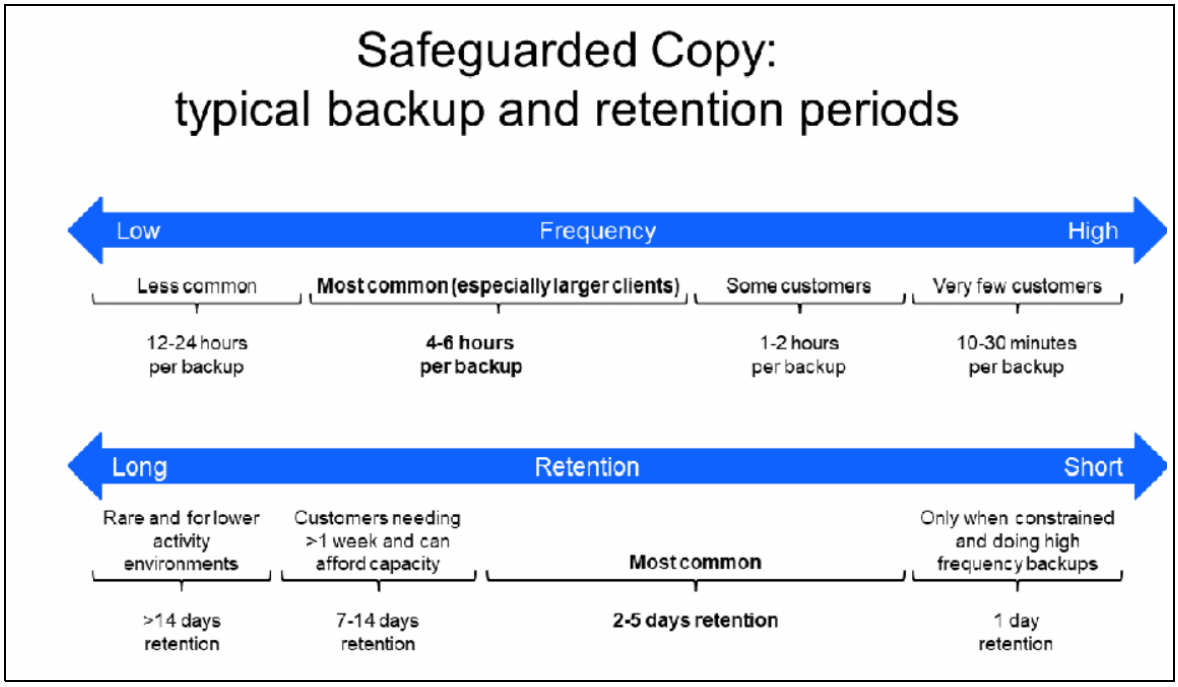


Figure 6-3 Safeguarded Copy: Typical backup and retention periods

6.2.3 Crash consistency or application consistency?

Just like any snapshot-based data copy, Safeguarded copies can be created in either crash-consistent or application-consistent mode.

Crash-consistent backups are easier to create because the orchestration does not require knowledge about the application. However, the recovery process may take longer due to potential inconsistencies within the backup data. After a crash-consistent backup has been restored, these inconsistencies need to be repaired on the application level before the service can be restarted.

Application-consistent backups are more complex to create because the file system needs to be in a consistent state before the snapshot is taken. Many applications use their own methods to establish data consistency, such as setting themselves in backup mode or quiescing the application. One example of a method is halting all I/O processes and flushing the file system buffers, so that there are no I/O operations left pending to the file system. After the snapshot has been taken, the application can be resumed to normal operation. Since the snapshot now contains a consistent data set as seen from the application perspective, the recovery process is often quicker.

6.2.4 How to do proactive monitoring

Just like any other part of the infrastructure, also the Cyber Vault infrastructure should be monitored for any kind of security related incident. For integration with a *Security Information and Event Management (SIEM)* tool, in most cases message logs are forwarded to a SYSLOG server. Advanced SIEM tools, such as IBM QRadar, will correlate events from multiple sources and understand the relation between these which can help detect suspicious activities very early to prevent the actual attack.

A specific monitoring use case that should be implemented for IBM Cyber Vault is to control if the immutable copies are successfully created, according to the defined schedule. SIEM tools have the capability to check if, for example, a message sequence proving the success of this process occurs regularly, and raise an incident if these messages are not logged after a specified time interval.

Some SIEM tools are able to execute actions towards the Cyber Vault. One example is that - in case an early warning sign of a cyber breach is detected - an immediate immutable copy is triggered. If this copy has not yet been affected by the attack, it can help reduce the RPO. In case it has already been corrupted, it can serve as a source for forensic analysis which is very close to the time of the cyber attack.

6.2.5 RPO, RTO and data validation

For the application data in scope of the MVC, appropriate Recovery Point Objective (RPO) and Recovery Time Objective (RTO) targets must be defined. These targets may be derived from existing RPO and RTO targets for disaster recovery, but should be reviewed in light of the possibility of a cyber attack. The RPO target translates to the frequency of immutable copies taken, while the RTO corresponds to the application recovery process.

In addition, it is important to assess when and how often a validation process of the data copy should be initiated. IBM recommends running data validation as a periodic process that includes all copies taken. This ensures that the last good data copy is always known, so that the recovery process can be started immediately in the event of an incident. This strategy provides the optimum protection level along with the shortest expected recovery time.

Depending on the amount of data, it may not be feasible to validate every single data copy. For example, if the validation process takes longer than the time between two data copies, then it may not be possible to validate all of the data copies. In the event of an incident, a decision must be made about which data copy to use for recovery. Should the last validated copy be used, even if a more current but unvalidated copy is available? Or should some additional time be spent to validate the most current data copy? This decision involves a trade-off between quicker recovery and minimizing the amount of data loss.

To minimize the time spent on data validation for multiple applications or volume groups, you can either use more resources (memory or processing power) for the scanning engine, or you can parallelize the scanning process by using multiple scanning engines.

6.2.6 Recovery planning

The recovery of business services after a cyber attack follows similar principles like Disaster Recovery processes. The key difference is that - before a service can be recovered - it must be ensured that:

- The target infrastructure for recovery is not penetrated by an intruder or infected with malware. Depending on the expected RTO it may be required to hold available a specific infrastructure setup that can be activated upon short notice.
- The source data has been validated to not be corrupted by the attack, and free of any signs of malware or other types of infection.

If IBM Storage Sentinel is used, t, the available data copies can be easily checked for scanning results, and the latest, validated copy (the "golden copy") can be selected for recovery. For other data validation strategies, an appropriate selection process must be established.

For quick recovery, these processes should be automated. Automation helps to avoid human errors, but it also needs to be regularly tested to uncover infrastructure or software changes that could cause the automation process to fail. Additionally, a human decision checkpoint should be introduced, such as when selecting the best data copy for recovery.

6.2.7 Further considerations

The following section discusses further considerations for Cyber Vault implemenation.

Secondary workloads and tape technology

Data from secondary workloads should be treated similarly to primary data, while taking into account the differences described in Table 6-1 on page 78. The full picture of an IBM Cyber Vault architecture including primary and secondary data is shown in Figure 6-1 on page 79.

Physical tape should be considered for storing copies of secondary data. It offers a physical air gap, has options for immutability, and is the lowest-cost storage medium available. It can also be used to store copies of primary data for longer retention times if required. However, recovery times will be somewhat longer compared to a snapshot-based recovery.

Tape is a proven secure storage medium, as data on tape can only be deleted by overwriting tape cartridges. This process takes much longer than deleting data from a flash-based storage system. It also requires tape drives to be used. Since this activity would likely deviate from normal tape workload behavior, it can be detected by monitoring tape drive and cartridge usage with a SIEM tool.

Preserving suspect data for law enforcement or other organizations

There are many scenarios where an organization needs to preserve data that is believed to have been corrupted as part of a cyber attack. This preserved data can be used for post-mortem analysis of the attack or as evidence in a legal prosecution. The legal or contractual requirements for preserving this data vary by location, the type of organization, and contracts that may exist between the organization and its partners. An organization's policies on this matter should be established by its legal counsel and information security offices. It is especially important to confer with law enforcement and comply with rules governing the creation and management of this sort of evidence, such as chain of custody and securing the data against unwanted access.

It is often difficult to know exactly what data needs to be preserved after a cyber attack, but creating a clone of the suspect volumes with no expiration date is a good first step. An organization may also wish to create snapshots of the OS and application volumes for the server, and any SIEM/SOAR logs that may encompass the time frame of the attack. If centralized monitoring of storage activity has been implemented for the attacked host, retaining the reports for that same time frame could also be valuable.

In general, anything that could potentially be of value should be saved before the data expires. It is easy enough to delete later if any part of the data is found to be of no value.

It is recommended that an organization establish policies on what data is to be retained and in what format, and a procedure built to meet that policy with as little effort as possible (ideally, an automated procedure). Trying to figure this out after the fact is going to take valuable effort that is best spent recovering data and returning the key applications to service.

7

Supported patterns

This chapter describes supported patterns of IBM Storage Safeguarded Copy feature on Storage Virtualize based storage.

Safeguarded Copy can be deployed in a range of different topologies:

- “Safeguarded Copy on a single system” on page 86
- “Safeguarded Copy on Metro/Global Mirror relationship” on page 86
- “Safeguarded Copy in a HyperSwap environment” on page 87

Note: Sentinel anomaly scan server should be located at the location where safeguarded copies are located.

7.1 Safeguarded Copy on a single system

In single system environments, there are three main components:

- The production or source volume, which refers to the data being copied.
- The safeguarded backup capacity, also known as the child pool, which is the storage location for the safeguarded copies and is inaccessible to the host.
- The recovery volume, which is where the safeguarded copies are restored to for access.

Safeguarded Copy enables a user to create up to 32,000 immutable copies of a source/production environment. These copies are immutable and cannot be directly accessed by any server.

To access the copies, they must be recovered to a set of recovery volumes. This provides instant access to the data and can be accessed by a recovery system for various purposes, including data validation, forensic analysis, or to restore the data back to the production environment.

Safeguarded Copy is not a direct replacement for FlashCopy and both can be used as part of a Cyber Resilience solution. See Figure 7-1.

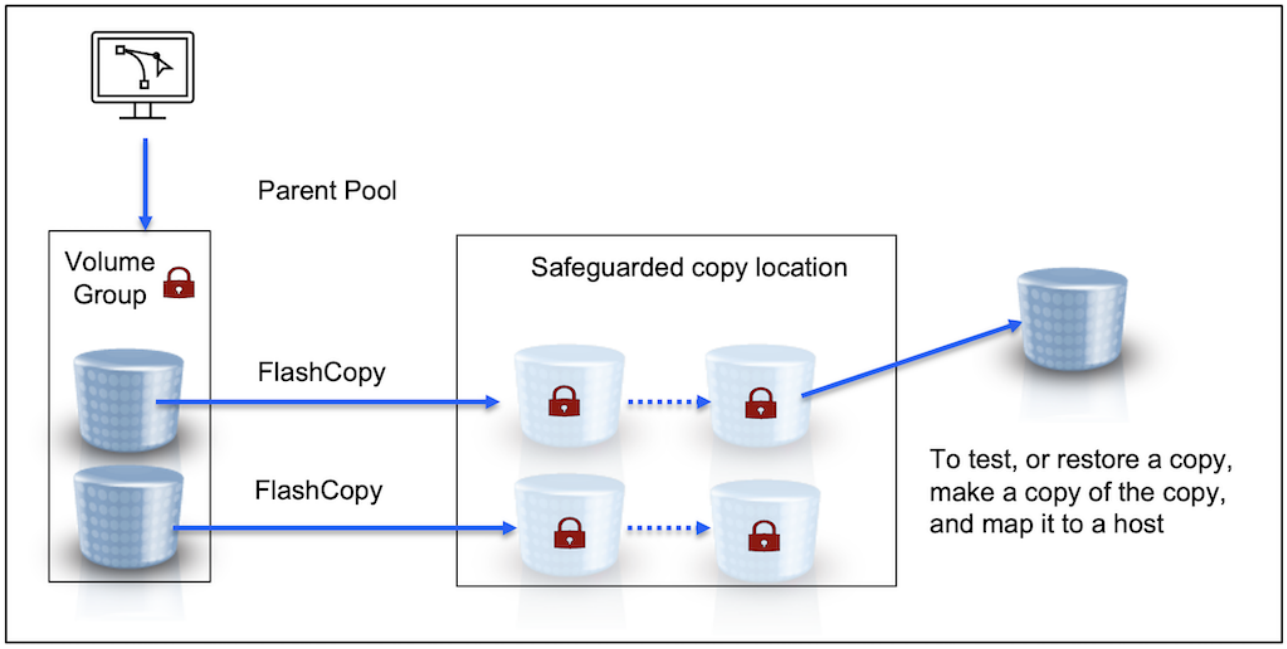


Figure 7-1 Safeguarded Copy on a single system

7.2 Safeguarded Copy on Metro/Global Mirror relationship

In Metro/Global mirror relationship the choice of doing safeguarded copies in one site or both sites is entirely customer preference. Obviously, there are capacity implications to planning for, recovery times and process all have to be considered, implemented and tested.

Note: Safeguarded Copies are taken from the primary and/or the secondary but you cannot create Remote Copy relationships on Safeguarded Copy point in time copies themselves.

Safeguarded Copy leverages FlashCopy, so it has a system/cluster boundary for taking and restoring safeguarded copies (SGCs). If the copies are only taken at the secondary site, you cannot use FlashCopy to simply flash back the SGCs to the source volume because it is in a different cluster. You must replicate the data back to the primary site.

In taking safeguarded copies at the secondary site, there is no need to pause or suspend MM, GM/GMCV. A crash-consistent point-in-time copy will be taken. Some clients may prefer or have a requirement for an application-consistent copy. To do this, they often use a script or cron job to pause the database, call CSMCLI to take a SGC, and then resume database read/write activity. They may also use an external tool like Storage CDM (or others) in conjunction with CSM. See Figure 7-2.

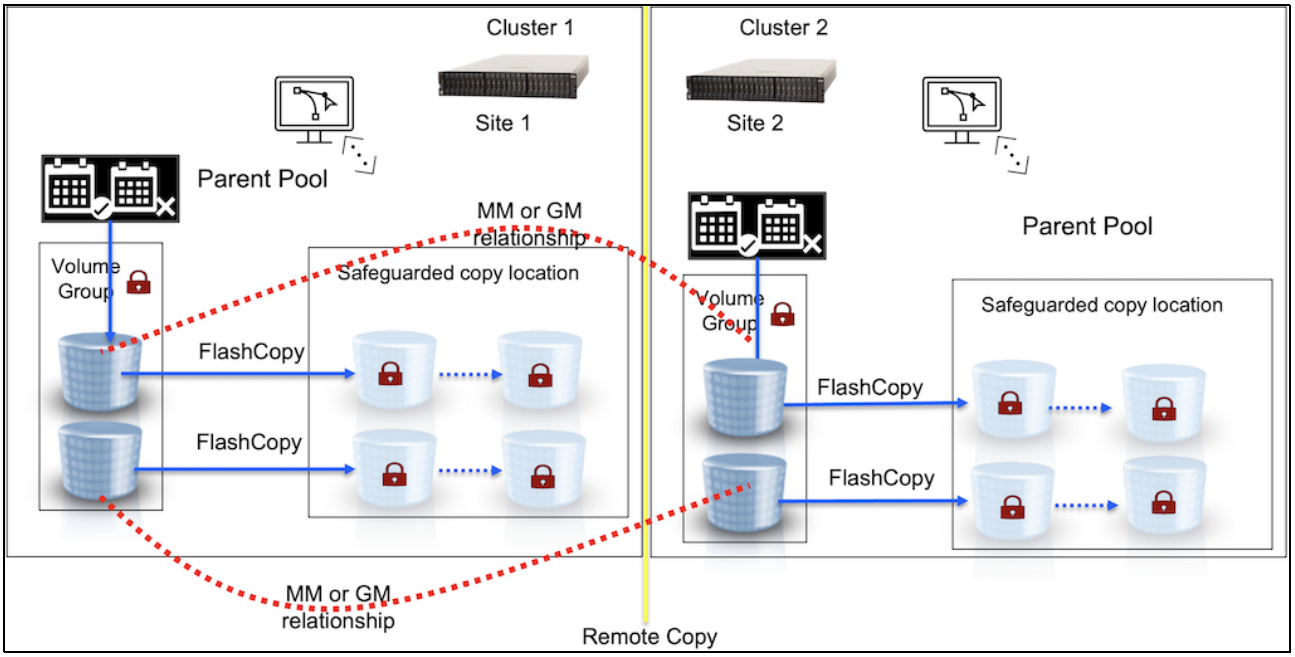


Figure 7-2 Safeguarded copy on Metro/Global Mirror relationship

7.3 Safeguarded Copy in a HyperSwap environment

Likewise in a HyperSwap® environment, you can choose whether to make the safeguarded copies at the primary site, secondary site or both. See Figure 7-3 on page 88.

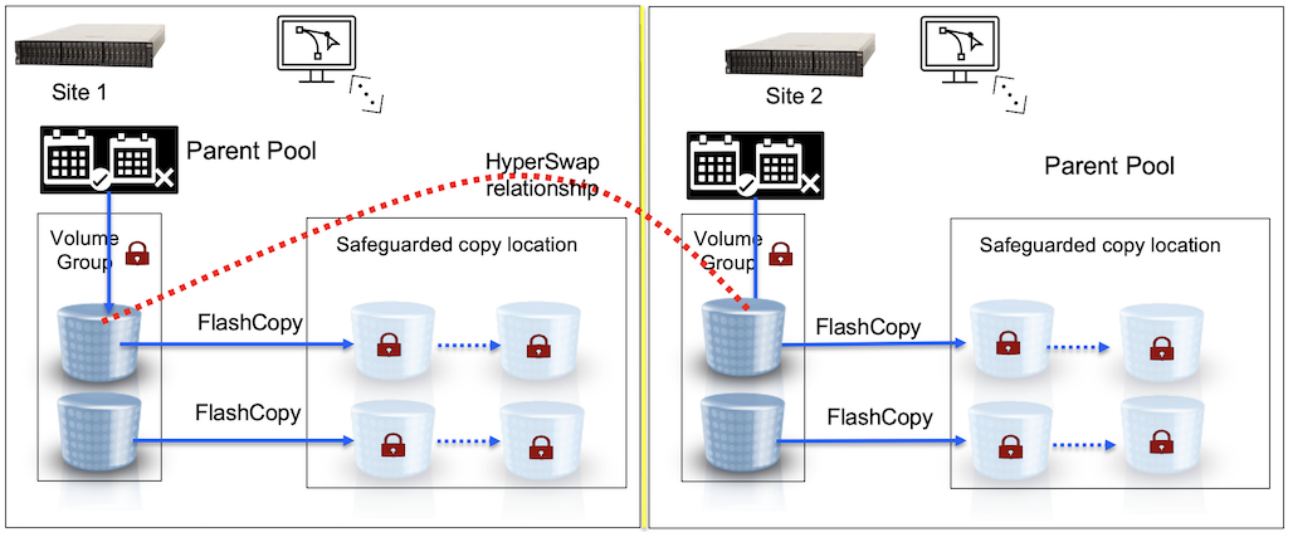


Figure 7-3 Safeguarded Copy in a HyperSwap environment

When considering data restoration in a HyperSwap environment, there are a few different factors to consider.

The most efficient method is to recover the data to the single-site copy, regardless of whether it is located on the primary or secondary site. Once the data has been recovered and a new single-site copy has been created, it should be mounted to the host and the data should be verified. If necessary, a secondary copy can be created in a HyperSwap relationship.

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this book.

IBM Redbooks

The following IBM Redbooks publications provide additional information about the topic in this document. Note that some publications referenced in this list might be available in softcopy only.

- *Implementation Guide for IBM Storage FlashSystem and IBM SAN Volume Controller: Updated for IBM Storage Virtualize Version 8.6*, SG24-8542
- *Performance and Best Practices Guide for IBM Storage FlashSystem and IBM SAN Volume Controller: Updated for IBM Storage Virtualize Version 8.6,* SG24-8543
- *IBM FlashSystem Safeguarded Copy Implementation Guide,* REDP-5654

You can search for, view, download or order these documents and other Redbooks, Redpapers, Web Docs, draft and additional materials, at the following website:

ibm.com/redbooks

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

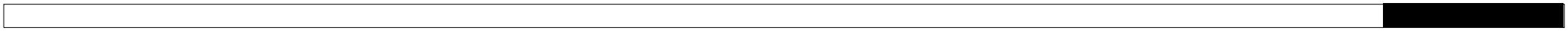
To determine the spine width of a book, you divide the paper PPI into the number of pages in the book. An example is a 250 page book using Plainfield opaque 50# smooth which has a PPI of 526. Divided 250 by 526 which equals a spine width of .4752". In this case, you would use the .5" spine. Now select the Spine width for the book and hide the others: **Special>Conditional Text>Show/Hide>SpineSize(-->Hide:)>Set** . Move the changed Conditional text settings to all files in your book by opening the book file with the spine.fm still open and **File>Import>Formats** the Conditional Text Settings (ONLY!) to the book files.

(1.5" spine)
1.5"<-> 1.998"
789 <->1051 pages

(1.0" spine)
0.875"<->1.498"
460 <-> 788 pages

(0.5" spine)
0.475"<->0.873"
250 <-> 459 pages

(0.2"spine)
0.17"<->0.473"
90<->249 pages



(0.1"spine)
0.1"<->0.169"
53<->89 pages

To determine the spine width of a book, you divide the paper PPI into the number of pages in the book. An example is a 250 page book using Plainfield opaque 50# smooth which has a PPI of 526. Divided 250 by 526 which equals a spine width of .4752". In this case, you would use the .5” spine. Now select the Spine width for the book and hide the others: **Special>Conditional Text>Show/Hide>SpineSize(-->Hide:)>Set** . Move the changed Conditional text settings to all files in your book by opening the book file with the spine.fm still open and **File>Import>Formats** the Conditional Text Settings (ONLY!) to the book files.

(2.5” spine)
2.5”<->nnn.n”
1315<-> nnnn pages

(2.0” spine)
2.0” <-> 2.498”
1052 <-> 1314 pages

SG24-8541-00

ISBN DocISBN

Printed in U.S.A.