Data Backup and Recovery for High Performance Computing

Data Backup and Recovery for HPC

High Performance Computing centers are broadly working to modernize their IT infrastructure to meet the challenges of tomorrow. Organizations across various industries are facing a pressing need to deliver high-performance, cost-effective, highly secure and robust backup and recovery solutions in their HPC infrastructures. This whitepaper discusses the considerations and advantages of using Bacula Enterprise as a central data backup and recovery system within a HPC environment, and how it can – and does – facilitate a specific, yet critical part in an HPC-using organization's enterprise-wide approach to digital modernization.

Version 1.0, March 2, 2021

Contents

1 Introduction

An ever-growing number of complex applications in a wide range of business-types and research areas is significantly increasing demand for HPC. Organizations across various verticals, such as government and defense, education, chemicals, health, manufacturing, energy and utilities, need to resolve complex calculations and problems.

HPC solutions can handle vast volumes of data with ease and can extensively support high performance data analysis. In addition, these solutions can deliver faster processing of data with a high degree of accuracy. These benefits offered by HPC solutions have further accelerated the adoption of these solutions across industry verticals. Further fueling this increased use are non-traditional HPC users, leveraging public cloud HPC solutions to solve machine learning and artificial intelligence challenges.

ITC centers of organizations using HPC face an ongoing challenge to adapt and improve their IT operations to meet these and other challenges of tomorrow. New and different approaches to security, efficiency and performance are needed – and are indeed currently being adopted – to achieve these improvements.

Bacula anticipates that technology and innovation improvements in the HCP space will increase, with special focus on areas such as Edge computing, Hybrid Cloud, or massive data sets, where Artificial Intelligence is now being used to train machine learning models. At the same time, computing capacity has increased to train larger and more complex models more quickly.

In parallel, new governance principles are being introduced into many areas of the sectors heavily using HPC (defense, government, higher education research, etc.), such as automation, adaptability, promotion of transparency and inherent accountability. In turn, a variety of new management styles are also seeing increased adoption. New projects that employ new methodologies are increasingly being supported by senior leaders within these organizations. This is contributing to a change in organizational culture, along with the development of new collaborative processes, technologies, and tools to automate the process and to apply consistent governance across a large organization.

So how does an organization find a backup and recovery strategy that fits into all the above needs? The sections below examine these needs and provides a solution.

2 New Management and Project Development Styles

Embracing new, modern methodologies such as Agility can, in turn, help large organizations with HPC to adopt new processes and technologies. Using the Defense sector as an example, these new processes have played a key part in introducing new processes such as a successful DevSecOps (DevSecOps is an augmentation of DevOps to allow for security practices to be integrated into the DevOps approach) structure. This has been done by implementing it in multiple, iterative phases. Perhaps beginning with some small tasks that are easy to automate, the project leader can then gradually build up the DevSecOps capability and adjust the processes to match. New approaches may facilitate a software system to start with

a Continuous Build pipeline, which only automates the build process after the developer commits code. Over time, it can then progress to Continuous Integration, Continuous Delivery, Continuous Deployment, Continuous Operation, and finally Continuous Monitoring, to achieve the full closed loop of DevSecOps.

Bacula understands that legacy backup and recovery solutions that are still being used today in defense and other organizations are often unlikely to have the flexibility needed to meet tomorrow's IT methodologies, requirements, technologies (e.g. new types of VM's, Containers and Clusters) and platforms, and are often in danger of becoming unfit for purpose. Bacula Enterprise directly addresses these issues of change in the IT environment (such as the DecSecOps example) and therefore is becoming an increasingly popular solution within these sectors.

This white paper examines some of the changing needs (present and future) of the HPC sector, and the reasons why Bacula is used extensively, especially in government, research and military organizations using either HPC and/or requiring backup of high volumes of data.

3 IT Environment Complexity

The IT environment of many organizations using HPC today continues to get more complex as data is moved between on-premise, Cloud, Edge, and off-site locations. In addition, developing technologies and applications, such as virtual machines, containers and big data repositories mean an ever-changing range of data and datatypes that need safeguarding. Not only must these organizations support multiple and different IT environments, but they also have to cope with the tremendous growth in the volume of data that they need to regularly manage and take care of. This introduces new backup and restore challenges in addition to a large range of other growing demands, such as security compliance requirements, RTO's, RPO's and ever-tightening budgets.

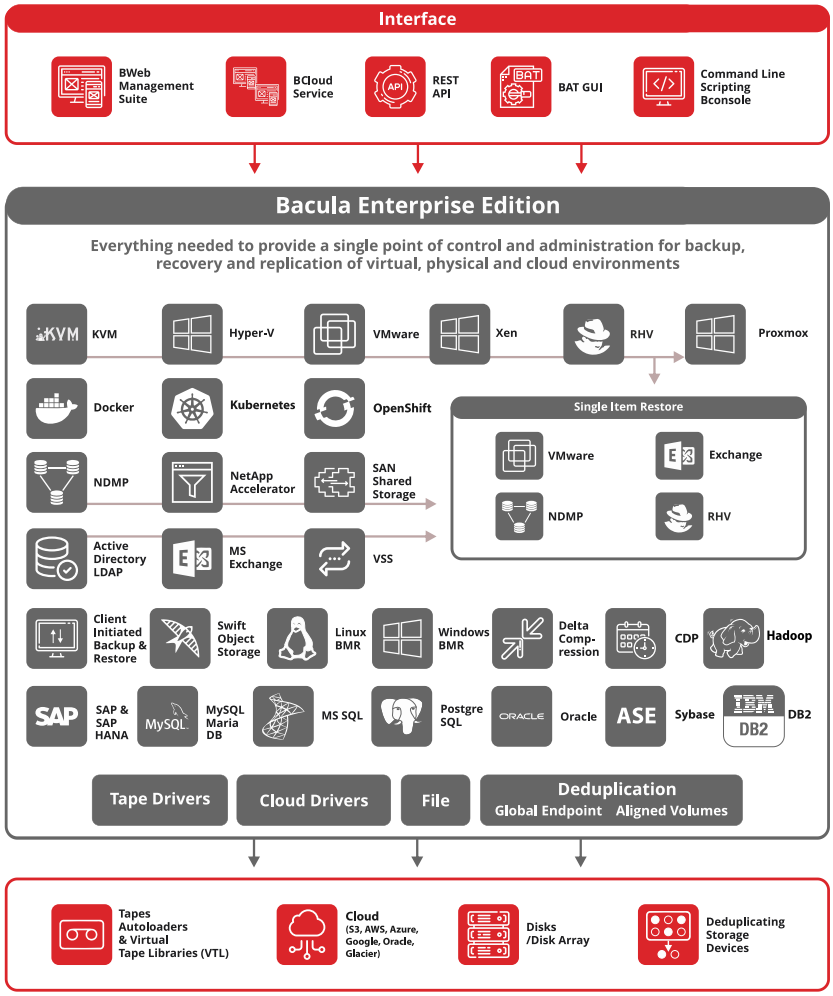
Bacula's response to these issues is to have an agile, modern and modular architecture that was designed using open principles for the new world of complexity and high data volume. It also treats IT security as the basic cornerstone of its entire functionality, where its product employs a purpose-built security foundation, which is integrated from end to end. In addition, Bacula recognized that the whole mind-set of the backup industry, which was built around metering data volume, was becoming unrealistic in a world where data volumes need to be free to grow.

Because Bacula recognizes that significant growth in an organization's data volume is practically inevitable, it utilizes a much lower-cost, fairer licensing model that is built around environments, rather than data volume. Bacula Enterprise raises the level of flexibility, automation and customization opportunities for all areas of a HPC user's IT infrastructure, far beyond that of its peers. At the same time, Bacula's security architecture is continuous and integrated into every part of its system.

4 Technical and Demanding IT Environments

Bacula Enterprise is enhanced by a constantly growing number of modules that delivers faster data recovery and minimal downtime to an IT infrastructure. These modules include PostgreSQL, MSSQL, MySQL, Oracle, SAP HANA, Sybase, Hadoop,

NDMP, NetApp, Delta, SAN Shared Storage, VMware, KVM, Hyper-V, Xen, Proxmox, Docker, Kubernetes, Bare Metal Recovery, VSS, Active Directory and of course high performance Deduplication. It also offers native hybrid cloud integration, via S3, S3-IA, Azure, Google Cloud, Oracle Cloud and Glacier interfaces. Despite integrating with such varied and large environments, Bacula automates security to protect the overall environment and data. Its tight access control and centralized authentication mechanisms are essential for the HPC IT environments of today and tomorrow. The diagram below gives a broad overview of some of the many technologies for which Bacula offers native integration.



Some of the additional features available with Bacula Enterprise are:

- Centralized data control
- Highly configurable, especially for clusters, multiple OS's, disk, tape, virtual tape, robotic libraries and Cloud
- Scalable from a few machines to many thousands
- Simple onsite and off-site replication
- Bare Metal Recovery for both Linux and Windows platforms
- Deduplication at both the client and storage levels
- Integrated Snapshots and Virtual Full
- VM Performance Backup Suite for integration with many different types of hypervisors
- Natively integrated support for Docker (and its external volumes) and Kubernetes, including persistent data
- Continuous Data Protection
- Client behind NAT (for backing up remote devices)
- Especially broad compatibility with tape

5 Bacula: Support of Tape Libraries from All the World's Main Manufacturers

For HPC environments that have petabytes of data to store, tape can still be the best option for long term archival and meeting RPO's. It's also very effective regarding retention requirements and media preservation. Today's LTO-08 and LTO-9 tape drives have impressive specifications including transfer rates up to 400 MB/sec. native and storing up 18 TB of raw and up to 45 TB of compressed capacity. Bacula Enterprise is ideal for tape admins as we never license based on data volume and deliver unique features such as ACSLS support. Bacula supports tape libraries from all the world's top manufacturers and provides all tape library management operations. It also supports named user access to ACSLM, tape drive and volume locking in shared ACSLS environment, lock query and management, static tape drive location mapping and dynamic volume location mapping.

6 The Need to De-Risk Implementation

IT departments need to be able to integrate highly effective security measures with regards to the backup and restore system with minimal disruption to operations, while also making sure that innovative technologies like containers and microservices are completely covered. But to be able to get to that happy stage, initial integration of the backup system needs to be extremely low risk – with technology that presents a low barrier to success.

Bacula's implementation and configuration processes are extremely simple and straightforward for a competent systems administrator with Linux knowledge. The high degree of flexibility in the system allows Bacula to adapt and offer different options for interaction with various other systems. In addition, Bacula also offers the ability to create custom-made scripts at almost any level, to exactly fit the specific needs of a HPC environment.

As IT departments plan for the future, learning from previous mistakes has brought about an awareness that any new software (and provider) they may use must be one that offers demonstrable quality and security improvements over the traditional software lifecycle. In addition, project managers seek to "de-risk" planned projects wherever reasonably possible. These risk management considerations are no different where backup and recovery software implementation is concerned, which can be measured partly with these metrics:

- Mean-time to production: the average time it takes from when new software features are required until they are running in production.
- Average lead-time: how long it takes for a new requirement to be delivered and deployed.
- Deployment speed: how fast a new version of the application can be deployed into the production environment.
- Deployment frequency: how often a new release can be deployed into the production environment.
- Production failure rate: how often software fails during production.
- Compatibility with new mainstream IT technologies

Traditional backup and recovery vendors are usually not architected to cope effectively with today's and tomorrow's wide variety of systems and datatypes. Living with this shortcoming typically becomes untenable when a new requirement emerges and the IT department suddenly learns there will be custom work or high contracting costs, that there are real questions about the proposed solutions' effective security, or that it will not be supported until six to twelve months in the future.

Bacula takes a different approach. Its software de-risks implementations by using a modern, modular architecture and is designed to manage IT environments with especially high levels of data volume. In parallel with this, Bacula System's licensing plan is removed from data volume, making sure that HPC environments and research organizations can both grow and easily forecast future costs by removing data growth risks and concerns from their backup and recovery strategy. Bacula understands that data growth in HPC environments is to be expected, and therefore employs a sustainable (and fairer!) licensing model that is built around environments rather than data volume. This typically leads to significantly lower costs and risks for the end user. At the same time, Bacula's security architecture is continuous and integrated into every part of its system, radically reducing security risks in comparison to other solutions.

Bacula Enterprise provides unusual degrees of adaptability, automation and customizability to more successfully integrate into an organization's HPC infrastructure, in a way that the IT direction needs and requires – as opposed to other vendors that typically require that integration happens in a way their software dictates.

As an example of the importance of using a modern backup and recovery solution, applying traditional solutions in a governmental modern-day DevOps environment would severely risk slowing software delivery, exacerbate audit pain, and leave organizations with an incomplete view of compliance posture.

With Bacula, the HPC user benefits from an especially secure, modern, scalable backup and data recovery solution that delivers hybrid cloud, virtual, container and physical machine compatibility all in one, single platform. Its modular design has helped HPC centers reduce risk by adopting more agile development practices and implement new, higher security measures.

Bacula's architectural flexibility is almost limitless in its application methods. A national defense organization using Bacula in the United Kingdom states that "*Because we have deployed Bacula using the method of an instance installed per deployed server, we are able to configure as part of a near fully automated build. This is a huge positive. It means far less work regarding pre-deployment configuration, as nearly all the work has already been completed during the server build stage. Another reason we especially appreciate Bacula is for its high data transfer rates and its compatibility with multiple storage media.*"

7 Meeting RPO's and RTO's

Frequently in large organizations, RTO and RPO (recovery time objective and recovery point objective) are two key metrics that must be considered in order to run IT in line with its requirements. The IT leader needs to couple backup and recovery requirements back to the organization's mission, and every utilized backup and recovery action should be aligned to the objectives. RPO's and RTO's are also needed to develop an appropriate disaster recovery plan that can maintain business continuity after an unexpected event.

Bacula's technology focuses on being able to achieve exceptionally fast recovery times, using a wide variety of different approaches that are relevant to the need in hand. For example, Bacula's MySQL recovery uses techniques that ensure consistency of the whole backup and can restore high transaction, high volume databases especially quickly. In an organization's high-transaction-rate MySQL databases, backup and restore recovery-time-objectives often become critically important even to the point that mission or business survival can be attached to them. There is often significant negative business impact if these systems are down too long. The longer the restore takes, the more a business's objective is compromised. Bacula's MySQL Percona module offers one of the fastest and easiest MySQL backup and recovery available in the industry today.

Another advantage is that Bacula's Percona backup software does not require the use of large temporary files. During restore, databases are restored to a temporary location, then made consistent using the "Prepare" option on the Percona tools, prior to actually modifying the live database. Rather than doing the "Prepare" work to make the database consistent at restore time, the "Prepare" can be done automatically by the module during the backup phase. For users of high volume, high transaction databases, this typically equates to a much faster data recovery than with any other vendor solution.

8 The Need for Especially High Levels of Security

While growth in HPC use continues, some concerns in connection to data security have somewhat affected adoption rates in some cases.

Widespread Ransomware and other malware incidents now require unusually high levels of security to be built-in to a backup and recovery system. For HPC, increased connectivity is a clear trend into the future, leading to a large number of connected devices which Bacula anticipates will result in greater risk, largely coming from cyberthreats that can exploit weaknesses in technology to compromise the integrity of networks, systems and data. Bacula recognizes that cybersecurity will be one of the foremost concerns for large organizations, exacerbated by their evolving needs as the nature and volume of their demands become increasingly complex and numerous. Echoing this evolution, retention regulations and policies are becoming clearer but more stringent, and these systems therefore need to be compliant.

HPC users are, correspondingly, seeing a need to make further improvements in their systems security, and IT managers are looking for effective ways to meet their organization's cybersecurity requirements, adhere to Security Technical Implementation Guides (STIGs) and Security Requirements Guides (SRGs), and industry best practices.

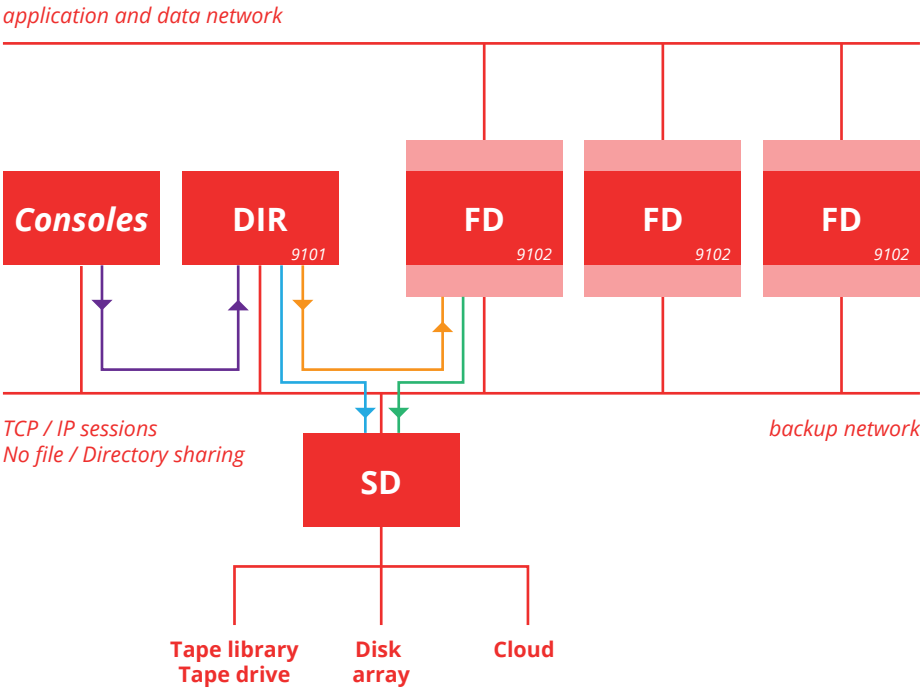
Bacula is unparalleled in the backup and recovery industry in providing for extremely high security levels. This ability spans specific elements regarding its architecture, features, usage approaches and customizability. Bacula has state of the art security built into each of its software layers. Some other features are:

- FIPS 140-2 compliant
- Verify the reliability of existing backed up data
- Detect Silent Data Corruption
- Data encryption cipher (AES 128, AES192, AES256 or blowfish) and the digest algorithm
- Automatic use of TLS for all network communications (can be turned off)
- Verification of files previously catalogued, permitting a Tripwire-like capability (system break-in detection)
- CRAM-MD5 password authentication between each component (daemon)
- Configurable TLS(SSL)communications encryption between each component
- Configurable Data (on Volume) encryption on a Client by Client basis
- Computation of MD5 or SHA1 signatures of the file data if requested
- Windows Encrypting File System (EFS)

One extremely important factor in Bacula's resilience to Malware attacks is its superior security architecture – The client is not aware of storage targets and has no credentials for accessing them. This high security architecture consists of some main secure-design elements:

- The client is not aware of storage targets and has no credentials for accessing them
- Storage and SD host are dedicated systems, strictly secured, only allowing Bacula-related traffic and admin access – nothing else.
- Bacula's "Director" (core management module), is a dedicated system with same restrictive access
- Bacula's Director initiates all activity and, in particular, hands out one-time access credentials to clients and Storage Deamons, which then only allow Bacula-related activity
- Bacula Enterprise provides no direct access from clients to storage; it is not in the protocol. Thus, even a compromised client cannot access any backup data, neither to read, to overwrite, to modify, or to delete it.

Bacula's superior security architecture



9 Ransomware

Backup and recovery is an IT department's last and most critical line of defense. Bacula anticipates the number of Ransomware attacks in 2020 and onwards to increase dramatically over previous years.

Where criminals extort money from companies is by encrypting their data so the company can't use it – and avoiding this situation is so much easier if the company under attack has effectively protected backups it can access and restore easily.

It may be initially difficult for an IT department to detect a Ransomware attack; first incursions can be, in practice, nearly undetectable. But once the malware starts encrypting, a good backup software should know, via detectable changes such as increase in File Change Rates, backup size increases, many file name changes, files vanishing and new ones appearing. All of this is tracked by Bacula and can easily be reported. As outlined in the section above, Bacula offers an especially robust architecture and feature set to protect against Ransomware, however, there are also some basic best practices that should be observed by any IT department:

- Use Different Credentials; User context for accessing the backup storage should be completely confidential.
- Access permissions are essential – don't rely on user accounts of file system access control, use dedicated services
- Make offline storage part of your strategy; This is one of the best defenses against propagation of ransomware
- Be sure to use the 3-2-1-1 Rule.
- Do not rely on different file systems to protect backup storage; more sophisticated types of Ransomware are coming

10 Bare Metal Recovery

Bare-metal recovery is a method where the backed up data is available in a form that allows an organization to restore a computer system from "bare metal", i.e. without any requirements as to previously installed software or operating system.

Typically, the backed up data-set will include the necessary operating system, applications and data components to rebuild or restore the backed up system - if necessary even to an entirely separate piece of hardware. This capability for a highly independent system recovery approach is strongly recommended for governmental research, defense, and related agencies.

Bacula Systems' bare metal recovery tool is available for both Linux and Windows Server, and enables a research organization to perform safe, reliable disaster recovery using Bacula Enterprise. With bare metal backup an organization can get its critical systems up and running again quickly and safely.

Bacula's bare metal backup software is available for Windows Server 2012, 2008, 2008R2, 2003 and for most flavors of Linux like Ubuntu, Debian, CentOS, Mint, Red Hat and SUSE.

11 Bare Metal Recovery as part of a Disaster Recovery strategy

Bare metal recovery can be accomplished in different ways. Many enterprises simply deploy a standard image, provision software, and then restore data and/or user preferences. In many cases, all data is already stored remotely and the system itself is largely unimportant. However, in many cases this is not a practical approach and

the ability to completely restore a machine to a point in time is a critical function of the disaster recovery implementation. The ability to restore a ransomware-encrypted computer to a recent point in time, including any user data stored locally, may be a necessary part of a layered defense. The same approach can be applied to virtualized systems, although there are usually preferable options available at the hypervisor level that Bacula can leverage for system-level recovery.

12 Disaster Recovery

Please request Bacula's separate white paper on Disaster Recovery - In this paper, you will find information on the following:

- Defining backup for Mission Continuity
- Obtaining Authorization and Commitment
- Defining Priorities
- Deciding on Technical Methodology
- Developing and Implementing the Plan
- Authorization
- Services and their Priorities
- Facility and Infrastructure Plan
- Plan Implementation

13 Different ways to interface with Bacula Enterprise

Bacula advises that as much flexibility as possible is made available to users of backup and recovery software. This can be invaluable in time-critical situations.

Using either GUI (BWeb Management Suite), BConsole, or a command line interface, Bacula Enterprise offers a wide range of different interfaces to configure, analyze, and manage backups and disaster recovery. BWeb is a multi-tenant backup management solution for Bacula Enterprise that saves time and resources.

It brings simplified backup and data recovery management. It also enables companies to exploit the large-enterprise capabilities of Bacula Enterprise, with a wide range of features for deploying, managing, reporting, optimization, and diagnostics. As a multi-tenant backup solution, it also allows government and other agencies to give a restricted permissions backup interface to sub-agencies with quota management and access management capabilities

14 Hybrid Cloud Technologies in HPC

While a majority of HPC work continues be in done on-premise, or in dedicated or private clouds, HPC workloads in public cloud are growing. New HPC-friendly services and options from large public cloud providers such as Microsoft Azure and Amazon Web Services are attracting traditional HPC users, who can now benefit from using public cloud to extend what they already do on-premise.

As part of Bacula's hybrid Cloud capabilities, it offers native cloud integration via S3, S3-IA, Azure, Google Cloud, Oracle Cloud and Glacier interfaces. For additional efficiency, restoring cloud data only requires the download of the specific data parts that are required for the restore job. In addition, Bacula's caching system provides an option to retrieve specific data that has been backed up to cloud, to be retrieved especially quickly from cache. The customizability and flexibility of Bacula's Cloud tools mirror the same flexibility and agility that Bacula's entire software platform itself delivers and comes together to present a great cost saving opportunity for any organization potentially needing to recover data from the Cloud.

15 Stand-alone Capabilities, and "Air-Gapping"

An "air-gapped" computer or network is one that has no network interfaces, either wired or wireless, connected to outside networks. Therefore, the organization's data is offline (disconnected) and cannot be accessed. To move data between the outside world and the air-gapped system, it is necessary to write data to a physical medium, and physically move it between computers. Many organizations build in Air-Gapping as part of their security strategy. Bacula Enterprise has especially strong characteristics in its ability to operate strictly as an independent, standalone entity, combined with the ability to write to a particularly large range of mediums. This facilitates and increases the options over time for an IT department to perform air gapping.

Bacula's software architecture is especially streamlined and lean, while its dependencies are easily satisfied. Bacula also readily lends itself to manual interaction. For example, in unusual or emergency and critical situations, Bacula is designed to be able to restore onsite in a variety of ways, including even just by delivering verbal instructions over a telephone if necessary. Therefore, it can be operated in a strictly localized fashion, with standalone interaction, without needing networked communication.

"Standalone" capability is also important in terms of business trust. Bacula Systems relies on trust to work with its customers and in no way attempts to intrude into an organization's privacy. For example, Bacula Systems does not do internal audits of its customers and respects the privacy of its users.

16 Container Technology and Kubernetes in HPC

Backing up containers and using Docker, and/or Kubernetes to orchestrate and manage the containers is currently gaining popularity in many governmental environments. HPC has much to potentially gain from using containers, such as its qualities of a lightweight, standalone, executable package of software that includes

everything needed to run an application except the OS: code, runtime, system tools, system libraries and settings. By adopting this technology, organizations stand to gain from its portability between clouds or between clouds and on-premise servers. This helps alleviate Cloud Service Provider (CSP) lock-in, and closely aligns with Bacula's Hybrid Cloud methodology (see "Hybrid Cloud Technologies in HPC", above). However, a container's main quality that may help IT departments most significantly is its ability to start much faster than a Virtual Machine.

As container technology is increasingly deployed in a production environment, its associated data is more likely to be needed to be backed up. In addition, Kubernetes clusters that are used in a production environment are more likely to generate persistent data, where effective backup and recovery is advised – and likely a necessity.

Bacula is, at the time of writing, the only full enterprise solution in the world that backs up and recovers both Kubernetes cluster configuration and persistent data, as well as assisting in other ways to mitigate vulnerabilities found in Docker images, Kubernetes configurations, application deployments, and CentOS Linux Kubernetes hosts. Bacula's main Kubernetes capabilities are:

- Kubernetes cluster resources configuration backup
- Restore a single Kubernetes configuration resource
- Backup and recovery persistent data
- Restore a Kubernetes resource configuration to local directory
- Backup and restore of persistent volumes
- Restore a Kubernetes persistent volumes data to local directory

17 Avoiding Vendor Lock-In

Vendor lock-in can be difficult to avoid. However, there are some ways to mitigate against lock-in, including using technologies that are as "open" as possible, whenever possible. Container technologies present a good example of this. One of the potential benefits of using containers is the portability they enable. Since the application in the container is isolated from the environment it is stored in, organizations can move the container to other locations knowing that its applications will work in the same way without modification. In effect, this can help mitigate the worry of supplier lock-in for many IT departments.

Bacula complements this approach by providing backup and recovery at the container level. It provides advanced functionality backup and recovery for both containers and Kubernetes clusters.

In addition, large parts of Bacula's code are - or are based on - open source code. It also largely avoids using proprietary standards across its architecture. Bacula's modularity, adherence to open standards, flexibility and open source background helps to mitigate vendor lock-in significantly.

18 Scalability

A backup and recovery system must have the capability to scale to any type of operational requirement, including:

- Business systems
- Command and Control systems
- Embedded and Weapon systems
- Intelligence analysis systems
- Autonomous systems
- Assisted human operations

While this would appear to be a good assumption, there are only a limited number of backup and restore software solutions that truly fit this requirement. An absence of compatibility in certain technologies, such as databases, or storages and network types means that in some cases, IT departments get into the situation where it is necessary to use different, multiple backup systems in order to be able to span their entire IT environment. Multiple backup software from various vendors is not a solution that Bacula recommends. Instead, the approach should be to find a single solution that covers the entire IT environment, all from one platform.

Some HPC users' program application systems have geographically distributed operational regions across the country or even overseas. It is important that a backup and recovery system can be effectively used in such a model, including offering interfaces and user interfaces that are flexible and can therefore fit these different forms of "scale". Scalability also requires that the backup and recovery solution offers excellent stability and performance at scale. Bacula scales to many thousands of servers, used in mission critical situations, and offers a wide variety of user interfaces (both command line and GUI) that can be configured to provide localized control, or overall control.

19 Conclusion

Bacula Enterprise is designed to facilitate positive change within a HPC IT infrastructure. Its especially broad compatibility helps remove barriers, while its modularity and flexibility help improve agility and speeds new capabilities into the field. In an environment where new policies, processes and culture change are planned – or even in process – Bacula's flexibility and resilience enable IT leaders to future-proof the backup and recovery aspect of their strategy while at the same time exploit the overall lower significantly risk that Bacula's architecture represents for new deployments.

Bacula's approach allows organizations using HPC to protect more environments, with more security, much faster and with lower risk than they ever have before.

For More Information

For more information on Bacula Enterprise Edition, or any part of the broad Bacula Systems services portfolio, visit www.baculasystems.com.