Hewlett Packard Enterprise

Technical white paper

Check if the document is available in the language of your choice.

SECURITY WITH HPE EZMERAL DATA FABRIC

CONTENTS

EXECUTIVE SUMMARY

Every day, businesses are seeing multiple high-profile cyberattacks across the globe, generating tremendous pressure to create and maintain a strong line of defense in their protection of data.

In most high profile attacks, a common thread is that business and consumer data was compromised—consumers' credit card number, social security number, banking information, and such—which highlights the importance of data protection.

The out-of-the-box protection scheme from HPE Ezmeral Data Fabric is designed to maintain a strong line of defense in the protection of data, which is the focus of this paper.

OVERVIEW

The HPE Ezmeral Data Fabric approach is to have data protection scheme built directly into the platform and enabled by default. HPE Ezmeral Data Fabric is designed with out-of-the-box security. The platform offers a robust and unmatched protection scheme for data. The security model of HPE Ezmeral Data Fabric is built directly into the platform, supporting its ability to apply security protection directly. The data comes into and out of the platform without requiring an external security management server or a particular security plug-in into each ecosystem component. The security semantics are applied automatically by design for data being retrieved or stored by any ecosystem, application, or users, out-of-the-box.

Our complete data protection includes:

- **Secure by default:** With HPE Ezmeral Data Fabric, your data platform is secure with default out-of-the box security capabilities. All network connections are encrypted with authentication enabled, and all data is stored encrypted.
- **Platform-based security:** A data platform with built-in security designed to apply security semantics automatically as data is being stored and retrieved from the platform. Solve all four pillars of security (authentication, authorization, auditing, and encryption), using platform-level capabilities that don't require external security tools or plugins. Such a solution is, therefore, complete and cannot be bypassed by components that have not been carefully altered to work with an external security tool.
- **Encryption:** Data is protected by encrypting all data being transmitted over the wire and encrypting all that is stored in HPE Ezmeral Data Fabric.
- **Data governance:** The DataOps governance framework within the HPE Ezmeral Data Fabric is built on an open architecture, allowing customers to extend and use the right technology to support their processes that match their use cases. The framework can track and manage the full data transformation process to achieve a complete data governance and data lineage monitoring solution.
- **Data lifecycle management:** This is the ability to fully control the placement of data on different nodes with varied performance characteristics as well as the ability to offload data for archival purposes.

SECURITY BY DESIGN

Security by design is the HPE Ezmeral Data Fabric approach to be free of vulnerabilities and impervious to attacks with built-in safeguards and adherence to best programming practices. While it is difficult to design software to be 100% vulnerability-free, HPE Ezmeral Data Fabric is focused on minimizing the more common mistakes in software engineering by embracing a wide range of industry security principles, focused on building security directly into the product. Some of the critical tenets built into our platform are:

Principle of defense in depth: One of the HPE Ezmeral Data Fabric hallmark security capabilities is focused on having security at each layer of the architecture, making it difficult and unlikely for an intruder to compromise your data. The core of this security is built directly into the platform, letting you apply security protection directly as data comes into and out of the platform without requiring an external security manager server or a particular security plug-in into each ecosystem component. In conjunction with the core platform, the ecosystem for HPE Ezmeral Data Fabric offers an additional layer of security, leveraging robust native security capabilities within HPE Ezmeral Data Fabric.

Establish secure defaults: HPE Ezmeral Data Fabric has embraced the secure by default principle with the behavior goal to be secure out-of-the-box with an option allowing the user to reduce the security if appropriate. It provides comfort, knowing the platform is secure without relying on complex, brittle scripts and complicated documentation steps, hoping nothing goes wrong along the way.

Keep security simple: The easy-to-use security capabilities bring a simple yet robust approach to security, which prevents users from making mistakes configuring security. This is accomplished by an innovative, alternative Kerberos design within HPE Ezmeral Data Fabric. Kerberos is well known in the industry as being very robust and very difficult for many users to understand, manage, and correctly configure. There is nothing simple about Kerberos. HPE Ezmeral Data Fabric implemented native security capabilities to leverage the robustness of Kerberos while simplifying it and making it easy to use. In particular, HPE Ezmeral Data Fabric self-manages the keys used to encrypt and sign tickets, instead of relying on Kerberos keytab files, which are very difficult for most users to maintain.

Separation of duties: HPE Ezmeral Data Fabric has been designed to support the administrators in administering the platform while not being allowed to see any user data.

Fail-safe security: HPE Ezmeral Data Fabric is designed not to allow unintended access due to an exception or failure in the system.

Detect intrusion: The auditing generates audit records for all data access and for any authorization denials. The audit data is sent to event store in HPE Ezmeral Data Fabric, allowing customers to send the security-relevant auditing data into the user's intrusion detection solution that is available in the enterprise.

Standardized integration: HPE Ezmeral Data Fabric uses Linux® account integration via Pluggable Authentication Modules (PAM), which is a standard functionality to authenticate users and access their user/group information. This is the same function any user employs when logging into a Linux box. As such, HPE Ezmeral Data Fabric works with any enterprise user registry that supports PAM with minimal configuration.

CORE SECURITY FEATURES

Authentication

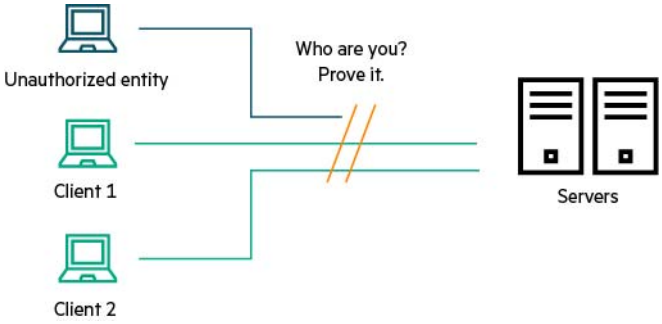

FIGURE 1. Authentication with HPE Ezmeral Data Fabric

User-based authentication: HPE Ezmeral Data Fabric leverages the Linux PAM, which give a broad range of registry support for authenticating a user to a cluster.

System-based authentication: HPE Ezmeral Data Fabric supports both a native security implementation of Kerberos and standard Kerberos.

Kerberos: This is a commonly used protocol for authenticating (identifying) users on a computer system, including Hadoop clusters. Kerberos is a ticket-based system in which the user first requests a ticket from the Kerberos server and the issued ticket is used as a trusted identifier to all services covered by that Kerberos server. The Kerberos integration with HPE Ezmeral Data Fabric lets you leverage your existing Kerberos infrastructure for authenticating users on your cluster.

HPE Ezmeral Data Fabric's Kerberos implementation: HPE Ezmeral Data Fabric provides a native authentication mechanism that operates equivalently to Kerberos but offers a much-simplified configuration, eliminating the need to manage the Kerberos keytab files, as the HPE Ezmeral Data Fabric system manages the keys automatically. Similar to Kerberos, the user will first request a ticket from the HPE Ezmeral Data Fabric CLDB server, using a user name and password that is validated, leveraging PAM interface, and then the ticket is used as a trusted identifier for all the services.

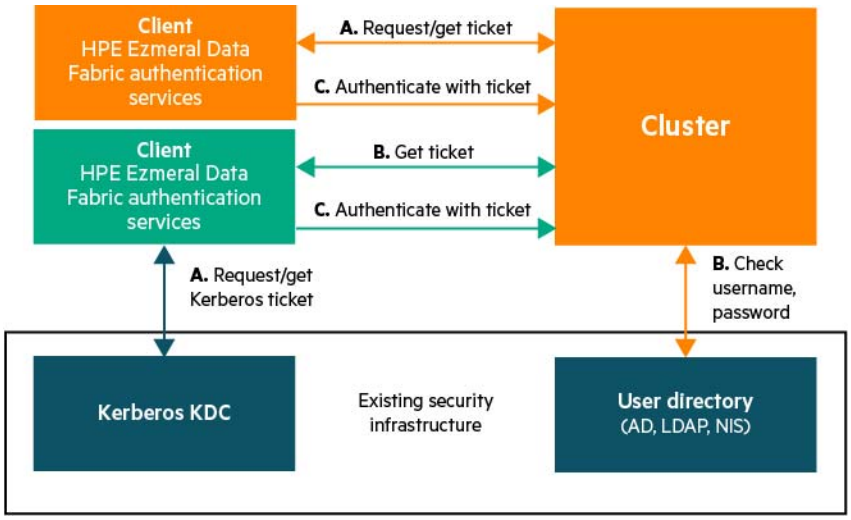

FIGURE 2. Authentication process with HPE Ezmeral Data Fabric

Login utility: The login utility supports user authentication with either username/password or Kerberos to generate a unique session token called a ticket. The following diagram outlines the process flow:

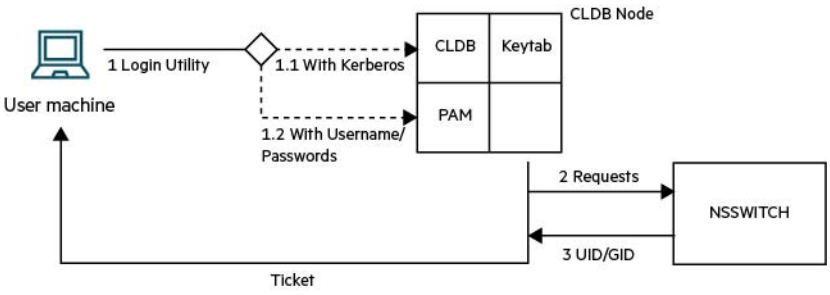

FIGURE 3. Authentication flow with HPE Ezmeral Data Fabric

Authentication flow

On clusters that use Kerberos for authentication, a ticket is implicitly obtained for a user that runs the command without first using the login utility. The implicit authentication flow for the login utility first checks for a valid ticket for the user and then uses that ticket if it exists. If a ticket does not exist, the login utility checks if Kerberos is enabled for the cluster, then checks for an existing valid Kerberos identity. When the login utility finds a valid Kerberos identity, it generates a ticket for that Kerberos identity.

When you explicitly generate a ticket, you have the option to authenticate with your user name and password or authenticate with Kerberos:

1. The user on the client machine invokes the login utility, which connects to a CLDB node in the cluster using HTTPS. The hostname for the CLDB node is specified in the mapr-clusters.conf file.
   a. When using user name/password authentication, the node authenticates using PAM modules with the Java Authentication and Authorization Service (JAAS). The JAAS configuration is specified in the doc.mapr.com/display/mapr/mapr.login.conf file. The system can use any registry that has a PAM module available.
   b. When using Kerberos to authenticate, the CLDB node verifies the Kerberos principal with the keytab file.
2. After authenticating, the CLDB node uses the standard UNIX® APIs getpwnam_r and getgrouplist, which are controlled by the /etc/nsswitch.conf file, to determine the user's user ID and group ID.
3. The CLDB node generates a ticket and returns it to the client machine, completing the login communication between the client and the CLDB.
4. After login, the client communicates with an HPE Ezmeral Data Fabric server. The server validates that the ticket is properly encrypted to verify that the ticket was issued by the cluster's CLDB.
5. The server also verifies that the ticket has not expired or been denied.

6. The server checks the ticket for the presence of a privileged identity, such as an HPE Ezmeral Data Fabric user. Privileged identities have impersonation functionality enabled.
7. The ticket's user and group information are used for authorization to the cluster, unless impersonation is in effect.

Authorization

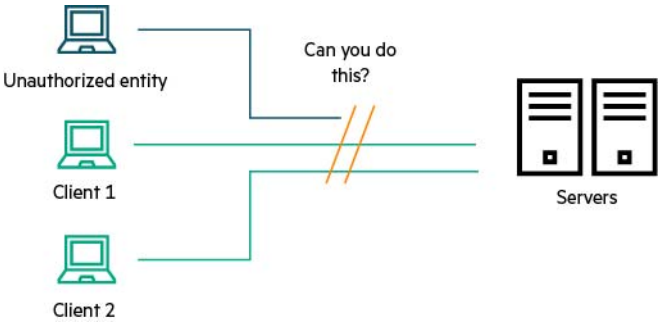

FIGURE 4. Authorization flow with HPE Ezmeral Data Fabric

HPE Ezmeral Data Fabric provides sophisticated authorization controls to ensure that users can perform only the activities for which they have permissions, such as data access, job submission, cluster administration, and such. An administrator can grant these permissions via the browser-based control system in the HPE Ezmeral Data Fabric management and monitoring interface or command line utilities.

UNIX file permissions: For files and directories in the HPE Ezmeral Data Fabric, you can leverage standard UNIX-style permissions to grant access to authorized users. Since HPE Ezmeral Data Fabric is a POSIX file system with full read/write capabilities, it can be accessed the same way that Linux file systems are accessed. This means existing file-based Linux applications can access files without any code changes or recompilation.

Access control expressions (ACEs): These are a powerful and flexible mechanism to grant permissions on structured and unstructured data. With ACEs, you get more flexibility than standard access control lists (ACLs). ACEs are Boolean expressions that allow "AND" and "OR" logic when defining permissions. The flexibility lets you specify fine-grained access control at the column and/or column-family level on the database in the HPE Ezmeral Data Fabric, an HBase binary and document database. Examples of ways you can grant permissions include:

- OR-based permissions found in standard ACLs
  - "Sales department" OR "marketing department"
- AND-based permissions
  - "VP level" AND "marketing department"
- Granular permissions
  - ("VP level" OR "director level") AND ("sales department" OR "marketing department") AND ("John Doe")

Files and directories: An ACE allows you to define access (allow list and deny list) to files and directories for a combination of users, groups, and roles. If ACEs are not set, POSIX mode bits for the file or directory will be used to grant or deny access to the file or directory.

Volumes: An ACE allows the whole volume to define allow lists (to grant access) and deny lists (to deny access) for files and tables within a volume.

Database in HPE Ezmeral Data Fabric: An ACE is used exclusively to set permissions for the tables, column families, and columns.

Event store in the HPE Ezmeral Data Fabric: A Kafka API-based pub/sub system, an ACE is used exclusively to set permissions for utilities.

ACLs: HPE Ezmeral Data Fabric supports ACLs to grant permissions for performing administrative tasks at both the cluster and the volume level. Examples of tasks include starting/stopping services, creating volumes, creating mirrors, and changing mirror properties. The ACLs also control which users and groups can perform specified tasks on specified job queues, including the ability to submit, terminate, or reprioritize jobs.

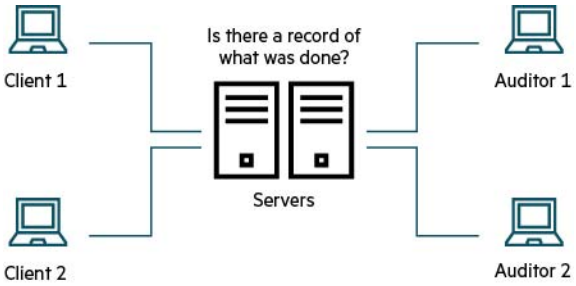

FIGURE 5. Auditing with HPE Ezmeral Data Fabric

Auditing

The auditing capabilities in HPE Ezmeral Data Fabric are critical for regulatory compliance as well as for understanding user behavior in the system. Regulations often require the ability to prove which user accessed which data, and logging user behavior helps in several situations, including identifying suspicious activities on sensitive data.

The HPE Ezmeral Data Fabric records access of data (files, directories, and table data) that are enabled for auditing, as well as operations on these objects and executions on the command line (CLI), it includes commands that modify the configurations of the cluster. Log entries are streamed in real time to the event store, written in JSON format, and can be analyzed with Apache Drill, your security information and event management (SIEM) solution, or other third-party tools. Log files are also retained for as long as you specify.

The HPE Ezmeral Data Fabric auditing consists of:

- All admin activities via CLI, REST, or the control system in HPE Ezmeral Data Fabric
- Authentications to the control system
- Operations on directories and files
- Operations on data base tables
- Operations on event store in HPE Ezmeral Data Fabric

In addition, each ecosystem offers operational auditing.

Encryption

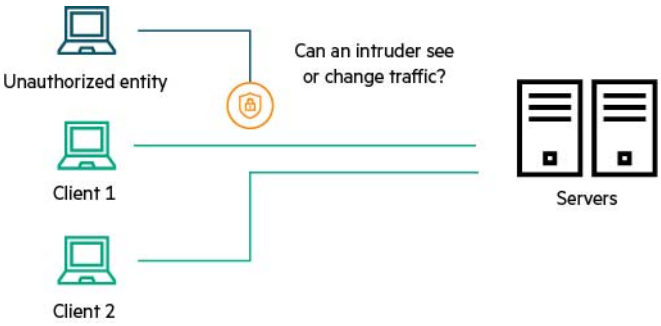

FIGURE 6. Encryption flow with HPE Ezmeral Data Fabric

HPE Ezmeral Data Fabric supports data encryption as an additional means of preventing unauthorized access to sensitive data. Encryption is used to avoid exposure to breaches, such as packet sniffing and theft of storage devices.

Over-the-wire encryption: To avoid data theft by packet sniffing, over-the-wire encryption is available between the HPE Ezmeral Data Fabric nodes, cluster, and ecosystem.

Encryption at rest: Encryption at rest not only prevents unauthorized users from accessing sensitive data, but it also protects against data theft via sector-level disk access. If data at rest encryption (DARE) is enabled, HPE Ezmeral Data Fabric automatically encrypts data at rest and manages the keys used to encrypt data seamlessly. There is no need for special utilities to encrypt or decrypt the data. New volumes are encrypted by default with the option to disable during volume create.

Field-level encryption: This enables securing specific sections of data residing in files. This capability logically behaves like access controls on a structured data set in a database management system. Some data elements in the files will remain open, while the secured data elements will be encrypted and can be decrypted by authorized users when used in conjunction with key management technologies. Some of our partners provide field-level encryption specializing in data security.

Format-preserving encryption (FPE) and masking: This is a mechanism for encrypting data, so that the format remains the same. This allows applications to access data that looks legitimate, instead of the typically garbled text that encryption outputs. This technique is particularly useful for analytical tasks that require readability in the encrypted data elements. Masking is similar in that it replaces sensitive data elements with an unidentifiable value, but it is not truly an encryption technique, so the original value cannot be returned from the masked value.

A significant benefit of these techniques is that the cost of securing a Big Data deployment is reduced. As secure data is migrated from a secure source into your platform, FPE or masking reduces the need for applying additional security controls on that data while it resides in your platform.

Both of these techniques are available from our partners specializing in data security.

HPE EZMERAL DATA FABRIC SECURITY PROTOCOLS

Protocol	Encryption	Authentication
MapR RPC	AES/GCM	ticket
Hadoop RPC and MapR-SASL	AES/GCM	ticket
Hadoop RPC and Kerberos	Kerberos	Kerberos ticket
Generic HTTP Handler	HTTPS using SSL/TLS	ticket, user name and password, or Kerberos Simple and Protected GSSAPI Negotiation Mechanism (SPNEGO)

HTTPS EXCLUDED CIPHERS

By default, the following weak TLS/SSL ciphers are excluded from the HPE Ezmeral Data Fabric HTTPS implementation:

- SSL_DHE_RSA_EXPORT_WITH_DES40_CBC_SHA
- SSL_RSA_EXPORT_WITH_DES40_CBC_SHA
- SSL_RSA_EXPORT_WITH_RC4_40_MD5

You can modify this list of excluded ciphers by editing the hadoop.ssl.exclude.cipher.suites property in the core-site.xml file. Restart the web servers that use the HTTPS protocol after changing the list of excluded ciphers. The following web servers use HTTPS:

- Control system
- JobTracker
- TaskTracker
- Node Manager
- ResourceManager
- HistoryServer
- CLDB
- HBase

CONCLUSION

Ensuring that business data is efficiently and securely managed begins with a data platform that is appropriately designed from the ground up. HPE Ezmeral Data Fabric delivers security out-of-the-box and offers ease-of-use security capabilities without compromising results in high data quality, integrity, and credibility for a better business outcome. HPE Ezmeral Data Fabric is designed with a robust and unmatched data protection scheme built directly into the platform. No external or open source security manager server with security plugins into each ecosystem component is required. The security semantics are applied automatically by design for data being retrieved or stored by any ecosystem, application, or users out of the box.

LEARN MORE AT

hpe.com/info/data-fabric

a50001588ENW, June 2020, Rev. 1

Hewlett Packard Enterprise