EsgynDB 安全技术白皮书 2.8.0

2021/04

目录

前言

关于本指南

本指南讨论了如何在 EsgynDB 中解决授权书认证方式，会计，数据保护问题。

目标读者

本指南的目标读者为 EsgynDB 系统管理员和用户。

相关文档

本指南为 EsgynDB 文档库的一部分，EsgynDB 文档库**包括但不限于**以下文档:

文档名称	说明
EsgynDB 安装部署指南	本文介绍安装 EsgynDB，包括安装前准备、安装 Hadoop 发行版、故障排除、配置、启用安全功能、提高安全性和卸载 EsgynDB 等。
EsgynDB 命令行工具指南	本指南介绍了如何在客户端工作站上使用 EsgynDB 命令界面（trafci）连接和查询 EsgynDB。trafci 使您可以交互地或从脚本文件运行 SQL 语句。
EsgynDB 管理指南	资料库（Repository）实例是一个数据仓库，用于从 EsgynDB 实例收集可管理数据。
易鲸捷加载转换指南	本指南介绍如何将数据加载转换到易鲸捷数据库。
易鲸捷 Designer 用户指南	本文介绍易鲸捷图形化数据库管理工具
易鲸捷迁移工具用户指南	本文介绍如何安装和使用易鲸捷迁移工具。
易鲸捷 DTM 技术白皮书	本文介绍 EsgynDB 技术架构，组件介绍，技术特点等。
EsgynDB 数据库规划文档	本文介绍节点数量规划、数据目录和安装部署目录规划、集群角色分配规划等。
EsgynDB 常见问题提排查与解决	本文介绍如何排查和解决 EsgynDB 的常见问题。
EsgynDB 灾难恢复手册	本文介绍 EsgynDB 灾难恢复设计原理，方案建议以及使用手册。
EsgynDB 备份恢复手册	本文介绍 EsgynDB 备份恢复设计原理，方案建议以及使用手册。
EsgynDB 数据库扩容指南	本文介绍 EsgynDB 如何更换节点，增加节点，删除节点等操作。
EsgynDB 客户端安装手册	本文介绍 EsgynDB JDBC， ODBC 以及 Trafci 驱动安装。
EsgynDB JDBC 程序员参考指南	本文介绍 EsgynDB JDBC 驱动连接设置，开发人员指南。

EsgynDB ODBC 程序员参考指南	本文介绍 EsgynDB ODBC 驱动连接设置，开发人员指南。
EsgynDB SPSQL 存储过程用户手册	本文介绍 EsgynDB SPSQL 存储过程的使用。
Esgyn DBManager 用户手册	本文介绍图形化数据库监控运维工具 DB Manager 的使用。
EsgynDB 数据库迁移指南	本文介绍如何将常见关系型数据库（Oracle、MySQL、SQL server 等）迁移至 EsgynDB。
EsgynDB LOB 大对象用户指南	本文介绍如何使用 EsgynDB 大对象。
EsgynDB SQL 用户手册	本文是 EsgynDB 的 SQL 使用手册。

1. 概览

EsgynDB 在 Hadoop 之上提供了一个运营型 SQL 引擎，该解决方案针对 Hadoop 大数据环境中的运营工作负载。包括的功能有：

- 全功能 ANSI SQL 语言支持
- 完全 ACID 支持读/写查询，包括对多行，表和语句的分布式事务保护
- 异构存储引擎访问，包括对数据存储的本地访问
- 增强了对客户端应用程序的高可用性支持
- 使用优化的查询内并行性支持大数据集
- 通过编译和运行时优化来提高 OLTP 工作负载的性能
- 安全访问数据集

本文档讨论了每个系统中必须存在的基本要素，并讨论了如何在 EsgynDB 中解决它们。这些要素包括：

- 授权书
- 认证方式
- 会计
- 数据保护

如果没有适当留意这些要素，您的环境将会不安全。通过将 EsgynDB 提供的功能与 EsgynDB 生态系统中的功能和产品结合使用，您可以组装一个应满足大多数安全需求的软件环境。

1.1 先决条件 (Prerequisites)

读者应该熟悉 EsgynDB 架构，以及 EsgynDB 代码如何与 HBase 和 Hadoop 基础架构中的其他产品交互。

读者应该熟悉 Kerberos，Active Directory（AD）和 LDAP 协议。并且对安全证书有基本的了解。

读者应该熟悉权限管理，ACL 和 Linux 文件权限。

读者应了解可用的 SIEM（安全信息和事件管理）解决方案，以及如何将其合并到 EsgynDB 生态系统中。

最后，读者应该熟悉数据保护机制，例如防火墙，数据加密，防止中间人攻击等。

1.2 术语 (Terminology)

1.2.1 安全系统的基本要素 (Basic elements of a secure system)

身份验证:

根据 WhatsIt 身份验证“提供了一种识别用户的方法，通常是在授予访问权限之前，让用户输入有效的用户名和有效的密码。许多公司提供单一登录解决方案，该解决方案将令牌传递到系统的各个过程和入口点。如果令牌通过召集，则操作可以继续。”

授权:

仅证明您的身份是不够的。您还需要获得特权或访问权限，才能在 EsgynDB 环境中执行操作。这是通过授权完成的。根据 WhatsIt 授权“是执行策略的过程：确定允许用户使用的活动，资源或服务的类型或质量。”基本上，授权定义了用户登录到 EsgynDB 后可以做什么。

会计:

安全系统还有其他一些要素。各国政府已启动安全措施，并制定了许多标准，在各自国家开展业务的公司必须遵循这些标准。这些属于问责制和会计领域。根据 WhatsIt，“问责制衡量用户在访问期间消耗的资源。这包括系统时间或用户在会话期间发送和/或接收的数据量。会计是通过记录会话统计信息和使用情况信息，用于授权控制，计费，趋势分析，资源利用和容量规划活动。”

数据保护:

在《设计用户界面软件的准则》中，作者将数据保护定义为：“数据保护试图确

保计算机处理数据的安全性，防止未经授权的访问，破坏性的用户操作以及计算机故障。数据保护必须解决两个普遍的问题。首先，必须保护数据免受未经授权的访问和篡改，这是数据安全问题；其次，必须保护数据免受授权系统用户的错误的侵害，实际上是为了保护用户免受自己犯的错误的影响。”

1.2.2 其他术语 (Other terminology)

身份提供者:

身份提供者存储可以在身份验证时使用的用户信息。这可能是客户配置的目录服务器(Open LDAP 或 Microsoft Active Directory), 平台上配置的目录服务器和/或底层 OS 身份验证 (读取/ etc / passwd）。

EsgynDB 扩展了数据库引擎，可由数据库自身作为身份提供者代替由外部目录服务器或操作系统提供身份。

组, 角色, 用户和所有权:

组

组将多个用户关联到一个实体。该组可用于分配特权和权限。有两种类型的组需要考虑，包括目录服务器组和 Linux 组。

EsgynDB 扩展了数据库引擎，可于提供数据库存储的组。

- 目录服务组（OpenLDAP 和 AD）

目录服务组定义了一组用户，并赋予他们对功能部件和功能的通用访问权限。企业安全管理员在目录服务中创建组，数据库安全管理员通过 REGISTER GROUP 命令注册这些组。 可以将特权和角色授予组。

- Linux 组

用于 EsgynDB 的平台是 Linux。存储过程使用的某些数据库对象（例如 UDF 库和 Java 文件）未存储在数据库中，而是存储在平台文件中。这些文件使用标准的 Linux 安全性进行保护。在 Linux 中，每个文件都分配了一组文件权限。文件权限

描述了用户，组和其他所有人的读取，写入和执行权限。在 Linux 安全性上下文中引用组时，它是指 Linux 文件权限中的组属性。

- 数据库组

基于目录服务组的扩展，由 EsgynDB 数据库本身存放和管理组员关系，可脱离目录服务器工作。

角色

角色为用户或组提供隐式分配权限的灵活性，而不是单独分配权限。可以为用户或组授予一个或多个角色。可以将角色授予一个或多个用户或组。

权限被授予角色。将角色授予用户或用户组之一后，授予角色的权限将对用户可用。如果将新权限授予角色，则那些权限将对所有被授予角色的用户可用。从用户或用户组之一撤消角色后，授予该角色的权限将不再对用户可用。

可以将 SQL 对象的权限授予角色。将角色授予数据库用户或用户组之一后，该用户将基于授予该角色的权限继承权限。从数据库用户或用户组之一撤消角色后，角色权限也将被撤消，但用户保留基于其他授予的角色的所有直接授予的特权和权限。

ANSI SQL 标准在检查用户权限方面受到限制。根据该标准，授权验证期间只能使用分配给当前用户和当前角色的权限。因此，如果需要多个角色的特权来授权访问，则不可能实现。因此，ANSI 产生了一个称为“基于角色的访问控制”的标准，该标准描述了基于角色的访问控制的基础，允许客户定义在一个会话中默认情况下和交互情况下哪些角色与用户相关联。

用户

在 ANSI 中，数据库用户是未明确定义的。对于 EsgynDB，用户在目录服务中定义，并映射到数据库可识别的名称。企业安全管理员在目录服务中创建用户，数据库安全管理员将这些目录服务用户注册为数据库用户，并通过 REGISTER USER 命令为其分配数据库用户名。数据库用户的使用信息存储在 EsgynDB 元数据中。只有注册的数据库用户可以访问 EsgynDB 数据库。新版本的 EsgynDB 数

据，提供了本地用户认证技能，可以由数据库本身管理认证用户的信息，此功能上与现有的目录服务用户一致、可兼容，由用户决策使用那种。

EsgynDB 关于数据库用户的使用信息存储是通用的，只需要在数据库录入与 LDAP 记录相同的认证信息，俩种认证方式是可以互换的。

就本文档而言，当用户名指定为不合格时，它是指数据库用户名。

ANSI 具有当前用户的概念，并且对于会话，它是否指定当前用户是可选的，只要存在当前角色即可。对于 Trafodion，我们始终要求当前用户存在，并且当前用户在会话期间不能更改。

所有权

- 对象所有权:

在对象创建时，有关对象，对象所有者和其他特征的信息被写入 EsgynDB 数据库。某些对象（例如表）也会创建存储数据的物理对象。在元数据中，所有对象均由授权 ID 拥有，该 ID 是创建对象的用户名的数值形式。创建物理文件后，将为其分配标准 Linux 权限。 Linux 文件权限和其他访问控制列表（ACL）由 HDFS 管理。在 Linux 文件上定义的实际权限不必与 Trafodion 中定义的 ID 相同。生态系统中的不同层执行其自己的用户 ID 映射。

- 文件所有权:

有些数据库使用的文件存储在平台上。这些文件包括用于存储过程的.jar 文件，用于用户定义函数的库文件以及大对象（LOB）。 EsgynDB 还有几个目录，他们用于存储临时数据以进行排序，备份和还原以及批量加载。

数据库特权

模式，对象和列级别的特权由 SQL 引擎强制执行。

系统（组件）特权

系统特权是指在系统级别授予的特权。在 EsgynDB 中，这些称为组件特权。这些类似于数据库特权，但是这里的细微差别是与对象相关联的授予和吊销，而不是授予和吊销，它们实际上是与可以在 EsgynDB 中执行的操作相关联的组件上。

加密:

加密将敏感信息从纯文本转换为不可读。列 obj 上有多种加密数据的选项
物理磁盘或磁盘池。 数据通常有三种状态，包括静止数据，使用中的数据和传输中的数据。 有时，还有第四点-已处理数据。 这些术语的含义有所不同。 就本文件而言：

- 静止数据是不使用时存储的数据。
- 使用中的数据是操作系统内存和数据库管理系统中使用的数据。即，活动数据。通常还假设数据位于单个节点中时正在使用中。如果将其移动到另一个节点，则数据将进入“传输中”状态。
- 传输中的数据是通过网络和其他电子介质传输的数据库。
- 已处置的数据是不再存在的数据。

多租户

多租户允许多个客户或租户共享单个系统的资源。 EsgynDB 支持用于资源管理和数据隔离的多租户环境。

管理员和用户：

企业安全管理员

企业安全管理员是指管理整个企业信息安全的人员或办公室。企业安全管理员除其他外，添加/删除 LDAP 组。从这些组中创建，分配或删除 LDAP 用户；或在整个企业中使用 LDAP 时在企业的组之间移动 LDAP 用户。如果在整个企业中使用 Kerberos，他们也可以管理 Kerberos。

群集安全管理员

群集安全管理员是指在运行 EsgynDB 的 Linux 平台上管理信息安全的人员或办公室。这包括在平台上 LDAP 服务器，平台上 Kerberos 和 Linux 中配置和管理用户，以及管理群集的安全性。管理群集上的安全性除其他外，包括资源保护（例如入侵检测系统），数据保护（例如加密）和合规性验证（例如审核）。

实例安全管理员

实例安全管理员是指为一个或多个 EsgynDB 实例管理信息安全的人员或办公室。实例安全管理员将实例配置为可通过 LDAP 访问；与平台安全管理员合作，设置与实例关联的文件的权限，并管理实例的安全性。

数据库管理员

数据库管理员是指管理一个或多个 Trafodion 数据库的人员或办公室。每个 EsgynDB 实例包含一个数据库。数据库管理员负责管理数据库中的对象（例如表）；分配组件，架构，对象和列特权；和管理数据库中的操作。

数据库用户管理员

数据库用户管理员是指管理一个或多个 EsgynDB 数据库的数据库用户和角色的人员或办公室。数据库用户管理员负责将 LDAP（外部）用户映射到数据库用户（注册），创建角色，将角色管理职责分配给数据库用户以及将角色授予数据库用户。

外部用户名

外部用户名识别目录服务器中已知的用户。它用于建立数据库的身份。

数据库用户名

数据库用户名识别数据库已知的用户。它在数据库内部用于识别所有权和特权。
大数据世界中可用的安全功能概述

1.3 大数据世界中可用的安全功能概述

1.3.1 Hadoop

Hadoop 是一个不断发展的开源项目，起源于 Google。它支持跨旨在扩展到数千个系统的各种服务器上的分布式数据处理。 Hadoop 的初始版本确实考虑过安全性。如今，已添加了许多安全功能，有一些处于孵化阶段。以下文章尽管有些过时，但对 Hadoop 安全状态进行了很好的总结。

Hadoop 由 HDFS（Hadoop 分布式文件系统）驱动，HDFS 是跨多个节点的许多磁盘的存储层。为了安全起见，这意味着在 Hadoop 中运行的查询可能会从许多磁盘和许多节点读取和写入数据。为此，数据在节点之间以及同一节点上的进程之间移动。

Hadoop 和 HDFS 的安全性功能：

- 进行身份验证可以防止对 NameNode，DataNode，JobTracker 或 TaskTracker 的未经授权的访问，并可以防止未经授权的服务加入群集的 HDFS 或 MapReduce 服务。这是通过支持 Kerberos 票证及其自身的内部令牌验证过程来处理的。
- HDFS 通过支持单独的访问控制列表（ACL）提供授权检查。 HDFS 的 ACL 支持的一部分提供了组映射服务，该服务允许实施者使用 Linux 或 LDAP 组
- Hadoop 支持通过其透明加密功能和密钥管理（KMS）产品对静态数据进行加密。

1.3.2 HBase

HBase 是 Apache 开源 NoSQL 数据库，NoSQL 以分布式方式在 Hadoop 之上运行，支持对大数据的实时读写访问。在 Hadoop 上运行的其他数据库专门用于快速执行数据检索。 HBase 可以充分执行数据检索，但也可以有效的进行更新。对于大量更新数据的 OLTP 系统来说，这是一个很好的解决方案。

EsgynDB 使用 HBase 存储自己的系统元数据，HBase 是 OLTP 应用程序的推荐使用存储引擎。

HBase 提供以下安全功能：

- 与安全 Hadoop 和 Kerberos 身份验证集成
- 允许安全管理员授予读取，写入，执行，创建和管理权限的 AccessController 协处理器或 ACL（访问控制列表）
- Per-cell ACLs (HBASE-7662)
- 通过 VisibilityController 协处理器（HBASE-7663）的可见性标签
- 透明加密（HBASE-7544）
- 支持用户管理员（超级用户）管理身份验证和授权。

通过 AccessController 协处理器启用安全 HBase。该协处理器在每个区域服务器或主服务器中运行并拦截操作。一旦操作被拦截，将进行检查以确保用户具有必要的特权，然后再继续操作。如果不是，则不执行该操作。如果特权检查有效，则操作继续。

此外，HBase 可与 Apache Ranger（孵化）一起使用。

1.3.3 Hive

Apache Hive 是一种开源产品，在 HDFS 文件之上提供了类似 SQL 的界面。将 HDFS 数据转换为一行后，即可使用标准 SQL 查询和处理。这对于查询大量数据很有用。

Hive 以两种格式管理数据：

- Record Columnar File（RCFile），这是默认格式。
- Optimized Row Columnar（ORC），这是一种针对行的列进行优化的格式。

Hive 根据使用方式支持基于安全性控件的多种不同变体。使用 Hive 作为存储引擎的产品（例如 EsgynDB，Impala 和 Spark）使用 HDFS 文件权限和配置属性来保护它的安全。

Hive 通过标准 SQL 授权撤销命令支持基于安全性的控件。当 Hive 用作 SQL 语言时，这是相关的。此支持使使用者可以控制通过用户和角色对对象和列的访问。

Hive 还具有默认（传统）模式，但是它不是很安全，因此不建议这样做。

此外，Hive 与 Apache Sentry 和 Apache Ranger（正在孵化）一起工作。

1.4 发行

1.4.1 Cloudera 发行

Cloudera 将安全性描述为四个支柱：

- 外围-通过身份验证和网络隔离来保护对群集的访问。
- 数据-通过使用加密，令牌化和数据屏蔽来防止未经授权的可见性。
- 可访问性-通过使用权限和其他授权技术定义用户和应用程序在系统上可以做什么和不能做什么。
- 可见性-通过审核和沿袭报告数据的来源，如何使用以及谁在使用。

可以通过 Cloudera Manager 获得某些支持，但是其中很大一部分是 Cloudera Navigator 的一部分。

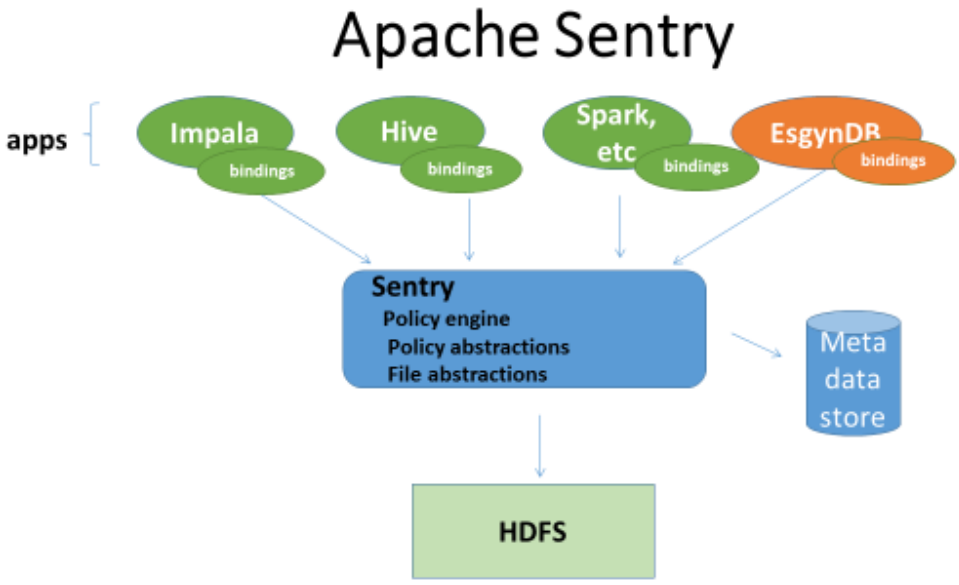
此外，Cloudera 提供了一种称为 Sentry 的产品，该产品增加了更多安全功能。

1.4.1.1 Cloudera 导航器

<添加导航器功能>

1.4.1.2 Apache Sentry

Apache Sentry 是为 Hadoop 生态系统提供 RBAC（基于角色的访问控制）功能的产品。 Apache Sentry 驻留在应用程序和 HDFS 之间。 Apache Sentry 当前与许多开源 SQL 查询框架集成在一起，包括 Apache Hive 和 Impala。



Apache Sentry 允许任何应用程序执行访问 HDFS 的特权请求。为此，Apache Sentry 创建并删除角色，授予和撤消角色的权限，以及授予和撤消组的角色。

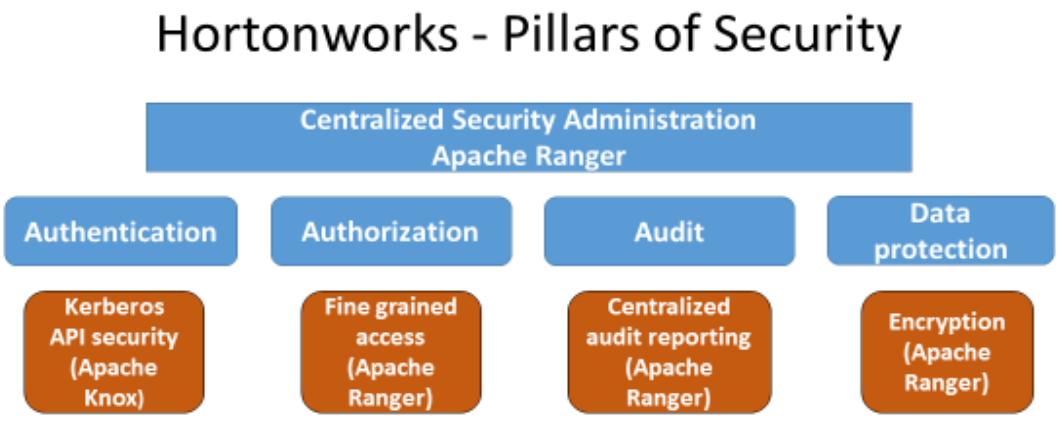
与 Apache Sentry 集成的应用程序调用 API‘s 来创建和删除角色，从角色添加和删除组，授予和撤消角色的权限，列出组的角色以及列出角色的权限。此外，该应用程序还负责管理包含用户列表的组，为对象分配角色，并根据从 Apache Sentry 返回的信息验证用户是否具有权限。

目前不支持 HBase，这是 Apache Sentry 的路线图

EsgynDB 支持由 Apache Sentry 管理的 Hive 数据。

1.4.2 Hortonworks

Hortonworks 通过身份验证，授权，审计和数据管理支柱来定义安全性。



Information extracted from Webinar "Enterprise Data Security Applied to Hadoop Feb 11, 2015 slide deck.

- 身份验证–通过 Kerberos 和 Apache Knox 管理
- 授权–通过 Apache Ranger 管理
- 审核–通过 Apache Ranger 管理
- 数据保护–通过本机加密和第三方进行管理

1.4.2.1 Apache Ranger（正在孵化）

Apache Ranger（孵化）是源自公司 XA Secure 的产品。 它为 Hadoop 生态系统提供基于角色的访问控制。 Apache Ranger（正在孵化中）位于现有产品之上，而不是与 Sentry 之类的 HDFS 集成。

支持的特权和权限基于基础产品（或存储库） 。例如，如果与 HBase 接口，则支持 HBase 提供的安全性。

1.4.2.2 Apache Knox

Apache Knox 设计用于外围安全性，并充当客户端和 Hadoop 环境之间的网关。 网关提供了可插入的 REST API，该 API 与 LDAP 和其他可能的身份提供者进行通信以对客户端进行身份验证。 有关更多详细信息。

2. EsgynDB 安全

各种各样的解决方案被各种实例吹捧。在 Esgyn 中，我们设计的安全性功能足够灵活，可以通过标
准集成接口与 Hadoop 生态系统中的工具（如 Kerberos，Sentry 和 Ranger）整合其他解决方案，还提
供具有增强功能的企业版。

在 EsgynDB 2.3 版中，我们支持:

- 强大的认证模型
- 兼容 SQL 的特权支持，包括
    - SQL 语句的特权粒度
    - 多个级别的特权支持，包括组件，架构，对象和列
    - 近乎实时的权限修改响应
- 集成的用户管理
- Kerberos 与 Hadoop 集成
- 与 Apache Sentry 和 Hive 表集成
- 通过以下方式审核运营:
    - 与谁在何时执行操作有关的存储库信息
    - 包含执行信息和故障详细信息的日志

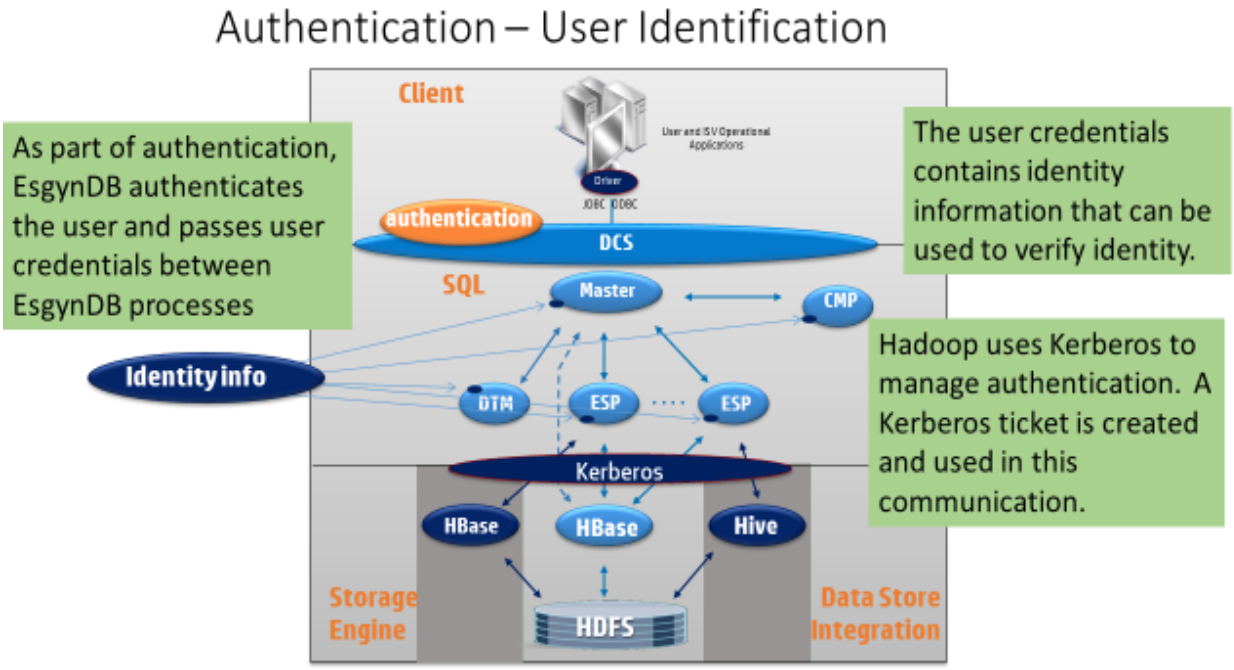
对于 EsgynDB 多租户版本，我们还支持

- 多个配置的目录服务器
- 综合组管理
- 模式隔离，支持多租户

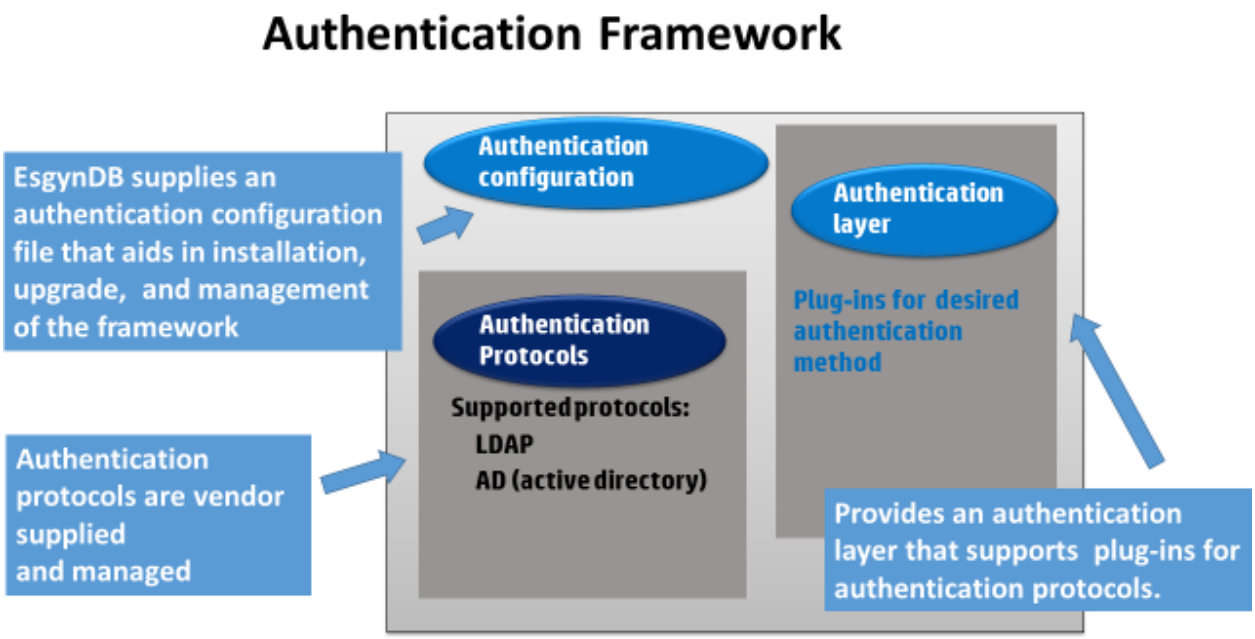
2.1 认证方式

在这种情况下，身份验证是验证连接到 EsgynDB 生态系统的人员的身份的过程。此

操作由身份验证框架管理。



身份验证框架通过验证用户名和加密密码来验证当前用户是否有权访问系统。安装 EsgynDB 时，身份存储配置信息将被设置。 对于多租户版本， EsgynDB 支持由目录服务器或数据库本身进行身份存储。



认证过程

- 构建服务器时选择认证类型，群集的安全管理员需要配置了一台或多台 OpenLDAP 或 AD 目录服务器；或者直接使用数据库本身作为用户及组存储。
- 如果使用目录服务器作为身份存储，安全管理员需要在 LDAP 服务器上预先注册需要的用户名和组。
- 无论使用目录服务器还是数据库本身作为身份存储，均需要在 EsgynDB 数据库注册对应的数据库用户和组。
- 客户端应用程序向用户请求登录详细信息（例如用户名和密码）。
- 用户名和密码将与主机名一起传递到相应的驱动程序，以建立与 EsgynDB 的连接。
- 驱动程序获取主机证书。
- 使用客户端安全性库提供的 API 和先前获得的证书对密码进行加密。
- 用户名和加密密码通过网络发送到 EsgynDB。
- EsgynDB 进程作为连接协议的一部分被分配。
- EsgynDB 进程使用存储在 EsgynDB 证书库中的私钥来解密密码，并将解密后的密码和用户名发送到身份验证库。
- 认证库验证用户当前在数据库中注册。然后，根据认证类型，身份验证库将用户名和密码提供给已配置的目录服务器或在本地数据库进行认证。如果使用 LDAP 认证，将确保通信稳定，发生通信错误，则认证库将对所有已配置的服务器重试，直到列表用尽或接受/拒绝用户名/密码为止。身份验

证库还允许配置多个 AD 或 LDAP 服务器。连接逻辑确定在验证用户凭据时要联系的目录服务器。

- 如果用户名/密码被接受，则 EsgynDB 会将当前用户作为自身及其所有子级的有效用户。
- 如果用户多次输入错误密码，用户将被锁定拒绝登录。根据认证方式不同，需要安全管理员在 LDAP/AD 侧或本地数据库内执行解锁操作。
- 如果用户距离上次密码变更时间超过设定的密码生命周期，将会被锁定，但提供额外几次(可设定)宽限机会允许用户登录以改变密码。
- 多租户版本支持用户组绑定，登录时将检查登录用户隶属的组是否为待登录租户的绑定组，只有被绑定组的成员才可以登录。

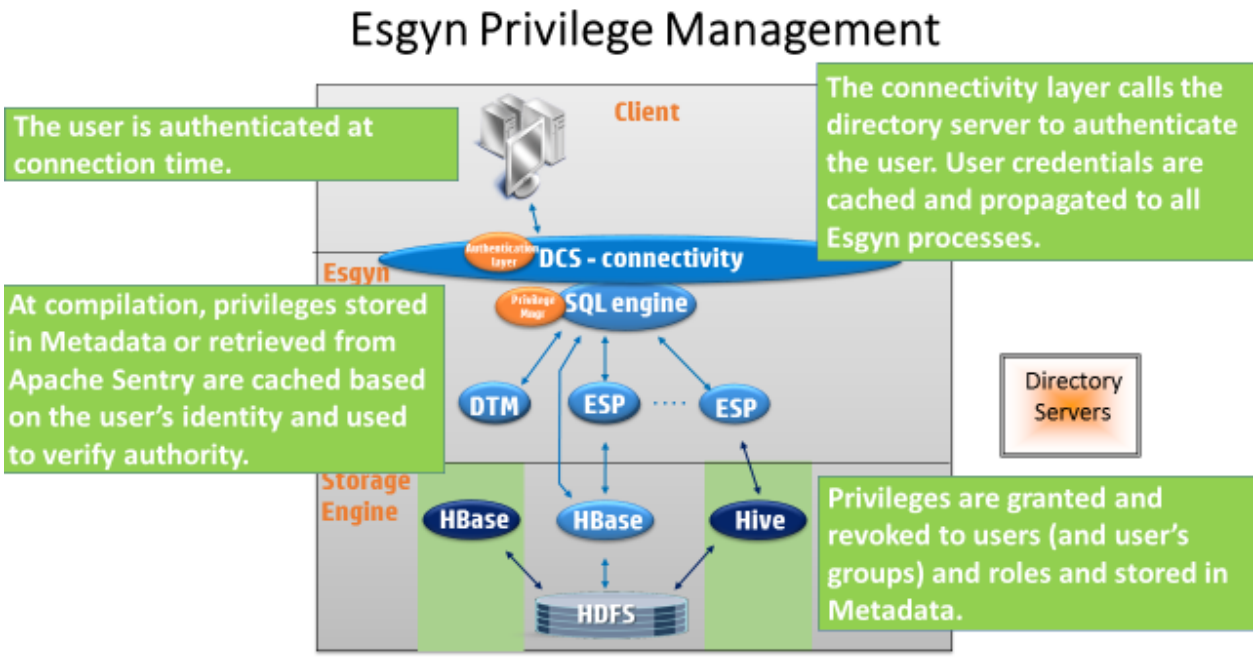
使用本地认证和 LDAP 认证的区别

认证类型	优点	缺点
本地认证	● 不需要额外配置 LDAP 服务器 ● 不需要维护 LDAP 服务器 ● 不存在通信成本	灵活性不如 LDAP 认证
LDAP 认证	● 认证流程灵活 ● 支持模糊认证(例如支持动态组、条件过滤)	配置复杂、需要额外维护服务器 数据库无法实时感知用户信息在 LDAP 侧发生变动

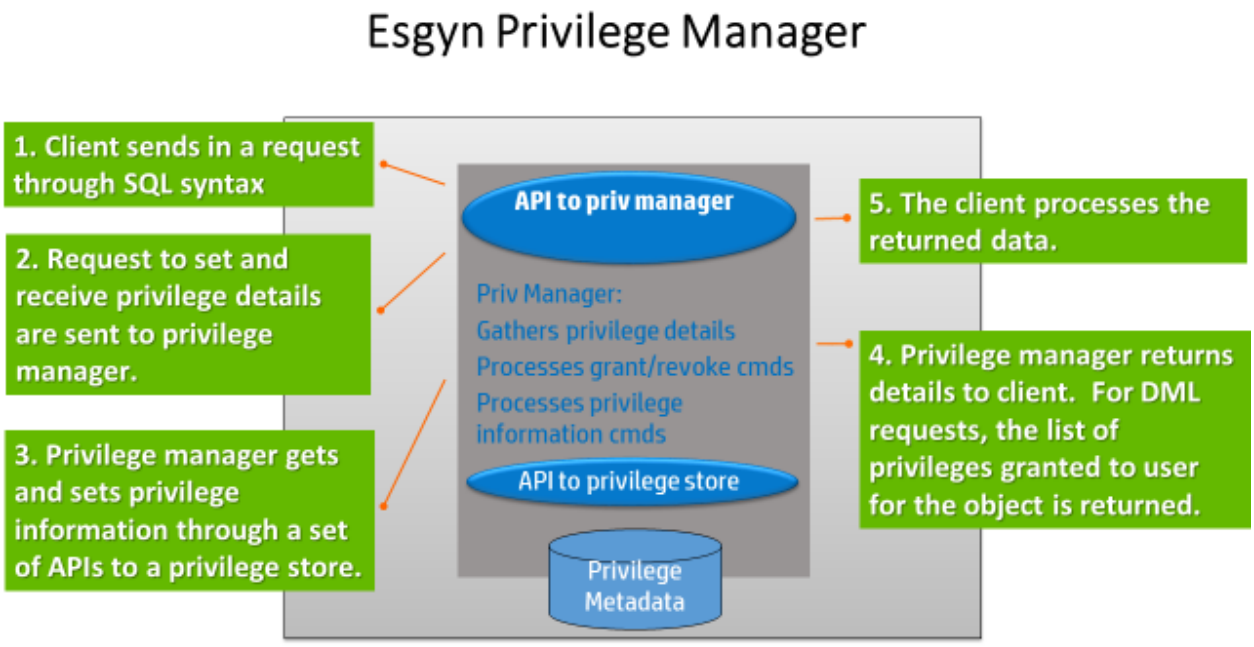
2.2 授权

授权或人员/实体可以执行所请求的 SQL 操作的验证由 EsgynDB Privilege Manager 组件管理。 权限管理器负责验证连接的用户是否具有执行 SQL 请求所需的权限:

2. EsgynDB 安全



权限管理器是 EsgynDB 中的一层，提供安全服务，包括为用户检索权限，授予和撤
消权限以及显示权限。 当前，它使用 EsgynDB 作为元数据存储。 EsgynDB 可以选择将 Apache Sentry 的元数据存储用于本机 Hive 表。



在对用户进行身份验证并将其凭据存储在 EsgynDB 进程中之后，用户可以自

由执行查询。在查询编译期间，将从元数据存储中检索当前用户，当前用户的组，当前用户（及其组）角色（包括特殊的 PUBLIC 角色）的权限。在系统（组件），模式，对象和列级别检索权限，然后将它们组合成一个简单列表。将根据用户拥有的权限检查执行查询所需的权限，并且请求通过还是失败。

数据库安全管理员在对象的生存期内授予和撤销权限。撤销权限后，权限管理器会向所有其他 EsgynDB 进程发送一条消息，通知他们更改。下次同一用户执行与权限更改有关的查询时，将重新编译该查询以获取新的更改。

权限提供对特定对象或组件执行操作的权限。可以通过多种方式向用户，组或角色授予或撤消特权：

创建对象或组件时，隐式特权将授予对象或组件的所有者。所有者保留对象生命周期的隐式权限:

- 创建对象或组件时，隐式权限将授予对象或组件的所有者。 所有者保留对象整个生命周期的隐式权限。
- 可以向用户，组或角色授予或撤消显式权限。显式权限可以由数据库用户管理员，对象所有者，或已通过 WITH GRANT OPTION 授予权限的用户授予或撤销。
- 授予用户或组的权限可以有各种来源。权限可以直接授予用户，用户所属的组，也可以通过角色继承。例如，用户从两个不同的角色获得对表 T1 的 SELECT 权限。如果从用户撤消了其中一个角色，则用户仍可以通过授予其余角色的 SELECT 权限从 T1 中进行选择。
- 因此，被授予角色的用户将获得该角色的所有权限。撤消任何此类权限的唯一方法是撤消用户的角色（或撤消角色的权限）。

可以授予和撤销以下权限：

- 组件，例如 SQL_OPERATIONS
- 模式，对象和对象列

GRANT [COMPONENT] PRIVILEGES 语句将权限分配给用户或组。REVOKE [COMPONENT] PRIVILEGES 语句从用户或组中删除权限。有关授予和撤消权限的详细信息，请参见[14]，以及当前支持的权限列表。

有关组件权限的说明。组件权限是系统级权限。我们支持三个组件，分别称为 DBMGR，SQL_OPERATIONS 和 WMS。有关可用组件权限的最新列表，请参

见[14]。

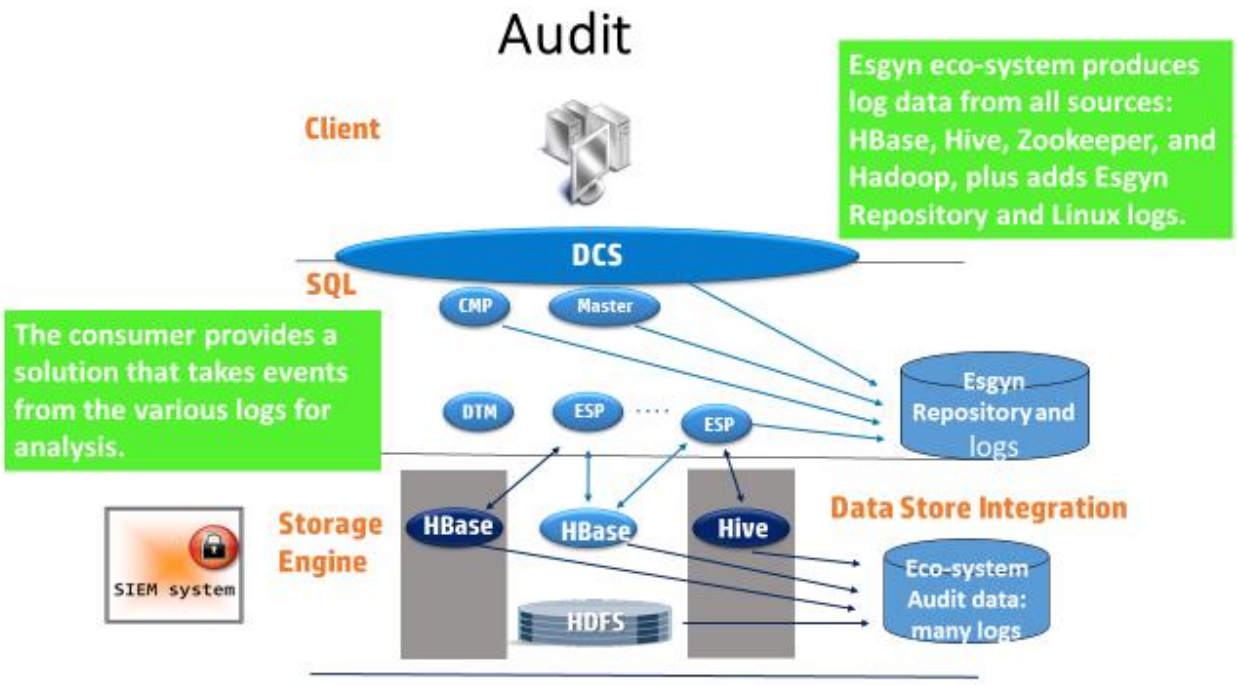
2.3 审计

每个人都害怕收到一封表明您的个人信息已被泄露的信。也就是说，一些可以识别您身份的数据（例如您的社会保险号）已丢失，被盗或放错了地方。因此，保护所有敏感信息很重要。但是，如果泄漏了敏感信息，则确定泄漏的发生方式以防止其他问题并确定可以获取哪些（或多少）信息非常重要。可能需要采取预防措施来修复泄漏造成的损坏。尽早发现问题并随后进行法证分析的一种方法是记录活动。借助发生的所有事件的详细信息和适当工具的使用，可以确定泄漏造成的破坏程度，并可以减少将来的安全泄漏。有效的日志记录是记录这些活动的一种方式。

有效的日志记录包括三个区域，包括日志生成，日志存储和日志监视。日志生成确定需要记录的内容并创建适当的事件。日志存储管理事件的存储位置，事件保存多长时间等。日志监视使用户可以查看实时活动和取证分析中发生了什么事件。有关日志的一个不错的总结，请参见[9]。

从历史上看，IBM 和 HP 等公司会将日志记录功能作为其产品的一部分提供。在后来的几年中，开源公司开始发展。审核日志记录更多变为附加功能。产品将以不同的位置和不同的格式记录活动。在过去的十年左右的时间里，日志记录已成为一项要求，要求公司在许多不同的来源中提供审核信息。许多政府制定一系列公司必须遵守的要求。为了解决这个问题，出现了一个全新的行业，称为 SIEM（安全信息和事件管理），以提供解决方案。

SIEM（安全信息和事件管理）解决方案提供日志存储和监视功能。他们从整个系统的各种日志中收集信息，将事件集中到一个公共位置，并为客户提供报告功能。对于 EsgynDB，我们建议您使用当今可用的众多 SEIM 解决方案之一。



2.4 数据保护

为了保持数据合规性，EsgynDB 应该在安全的平台上运行-特别是，只有具有适当权限的人员才能访问 EsgynDB 生态系统中存储的数据。

有几个风险领域：

1) 外部威胁-通过网络接口访问 EsgynDB

我们的信念是，通过采用标准的网络安全措施，例如防火墙和 EsgynDB 身份验证，可以将外部访问带来的威胁风险降至最低

2) 群集级别访问-未经授权访问 EsgynDB 组件

人们担心，有人可以通过欺骗或其他方法访问文件和进程。 通过正确地使用文件权限，限制对文件和进程所在的群集的访问，并使用诸如 sudo（或 PowerBroker 之类的产品）之类的工具来限制有权访问群集的人员执行的操作，可以最大程度地减少未经授权的访问并限制欺骗。

3) 应用程序/ SQL 命令访问-使用标准 SQL 接口未经授权访问数据。

首先要考虑的是用户访问权限超出了他们的访问权限。通过使用数据库授权检查和加密数据，可以最大程度地减少应用程序和 SQL 命令的访问。

EsgynDB 运行所在的环境的数据保护可以分为三层：集群，实例和数据库。集群包含一组节点，EsgynDB 生态系统将在其中运行。群集托管一个操作系统，该操作系统为安全性提供了另一个考虑因素。

2.4.1 外部威胁和防火墙

防火墙提供了第一级安全性，该防火墙限制了对连接到群集外部网络的所有节点的访问。这些节 点通常仅限于头节点以及故障转移，登录和 I / O 节点。

如果您尚不支持防火墙，则应安装一个。

可根据要求提供操作 EsgynDB 所需的端口列表。

2.4.2 集群级访问

群集提供对安装过程的管理以及群集中包含的实例的配置。每个 EsgynDB 实例都提供对其包含的数据库的配置和管理。

Esygn 需要一个称为 trafodion 的 Linux ID 和一个相应的组 trafodion。此用户和组是在 EsgynDB 安装期间创建的。 EsgynDB 进程以 trafodion 的形式运行，HDFS 和 Linux 中文件的文件许可权由 trafodion 拥有。

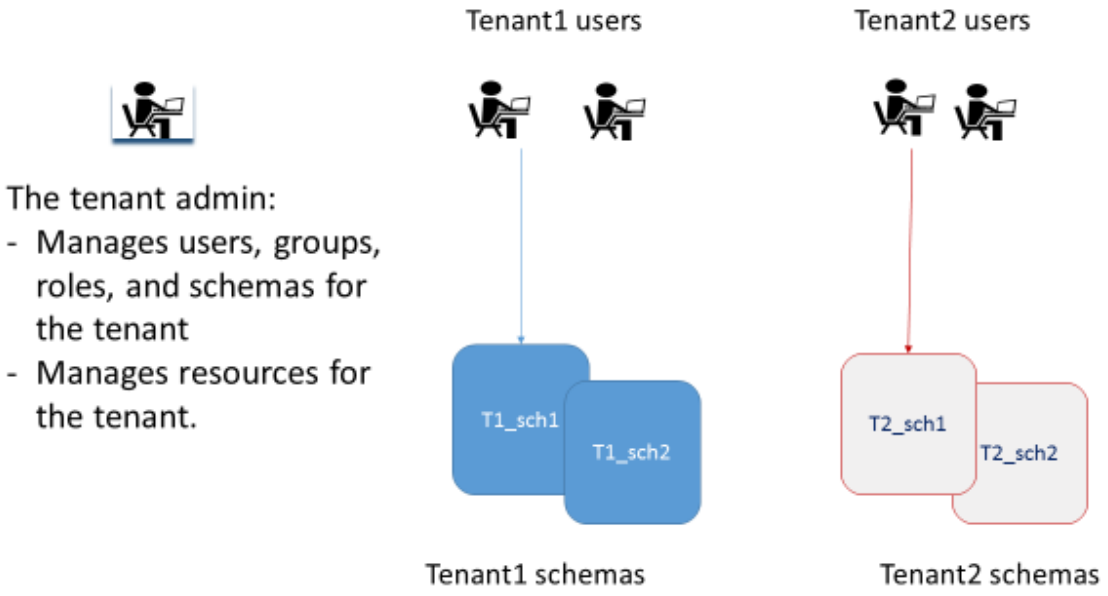
连接到群集时，请注意将其作为 Trafodion ID。这是一个匿名用户，以该用户身份运行的任何人都应记录其操作。 EsgynDB 通过 DB Manager 和 REST 界面提供了许多操作，因此由于 trafodion ID 的限制，需要直接在 Linux 上运行的运行操作受到限制。安装，升级。

2.5 多租户

EsgynDB 系统可以分为称为租户的多个组件。租户共享基本资源，例如 CPU 和内存。创建租户时，将为其分配一组资源。资源分配由 Linux 中的 cgroup 功能监视，该功能是限制和说明进程集合的各种系统属性的机制。除了资源分配外，还为租户分配了数据库资源（例如，租户管理员）并分配了一个或多个模式。对于多租户版本，我们支持租户之间的数据隔离。即，分配给一个租户的用户不能查看与另一租户关联的数据。

在 SQL 中，租户是一个包含以下内容的容器:

- 租户管理员-具有管理租户所需特权的角色
- 租户架构–由租户管理员管理的架构
- 租户用户组–与租户关联的 AD 或 LDAP 组
- 架构，用户，组和角色–由租户管理员管理
- 租户的资源分配使用了计算资源。



2.6 加密

当数据在不同状态之间流动时，可以使用多种方法来保护数据。传统上，数据是通过硬件或软件算法静态加密的。当它从静止转变为使用中的数据时，数据将被解密。从使用状态变为静止状态时，将对其进行加密。如果用于解密数据的密钥被错误的人解锁，则可能会损害数据。

使用中的数据通常被解密，并在软件级别用于执行操作。冒充受信任的用户可以通过欺骗软件来破坏数据。

传输中的数据是指数据跨网络边界流动的时间。通过在中间攻击中使用人来拦截数据，可以破坏数据。

当不再需要数据时，需要对其进行正确处理。如果信息仍以某种格式存在于数字设备上，则可能会破坏数据。

EsgynDB 利用 Hadoop 和其他 Hadoop 提供程序提供的加密解决方案，请参阅[13]。

- HBase 提供对透明加密和 HBase 级别的支持。
- HDFS 在 HDFS 级别提供对透明加密的支持。 HDFS 还提供了密钥管理解决方案（KMS) 和加密区域。

2.6.1 EsgynDB 加密功能

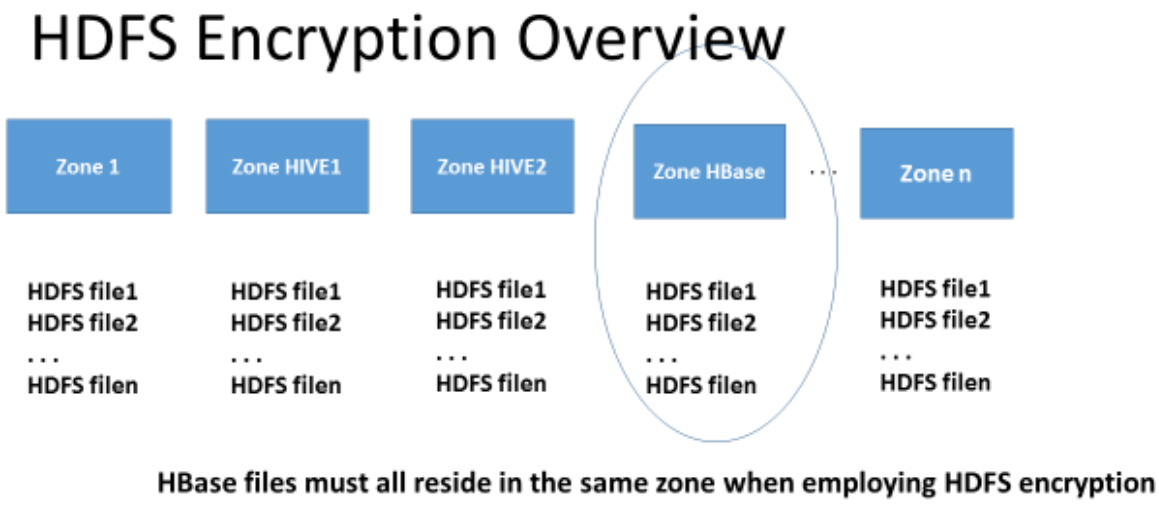
Esgyn 提供了两个内置函数，可让您使用 AES（高级加密标准）对字符串进行加密和解密。

- AES_ENCRYPT
- AES_DECRYPT

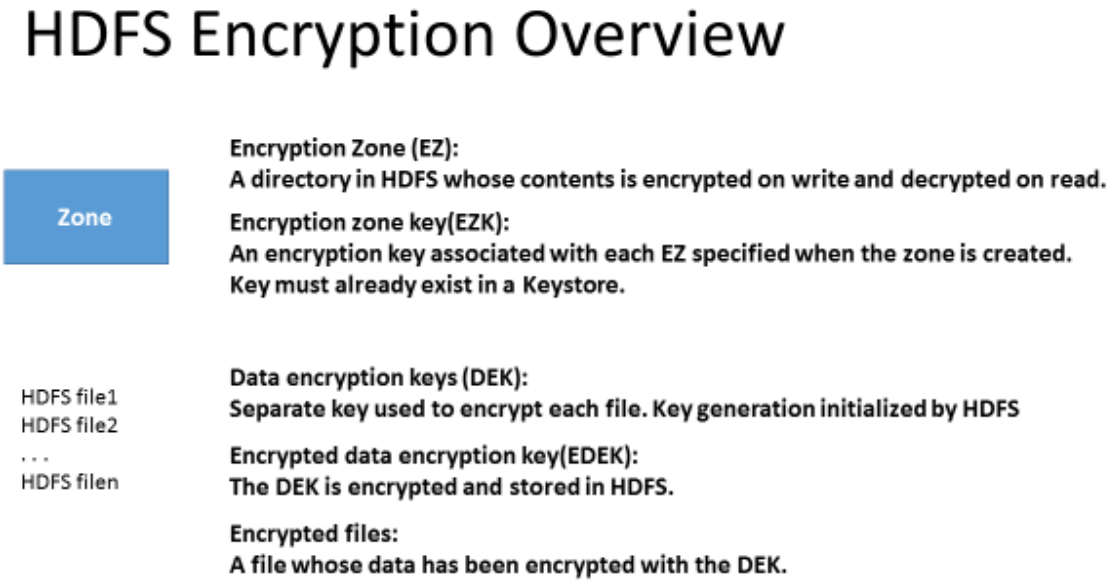
2.6.2 HDFS 透明加密

HDFS 加密作为 Hadoop 的一部分发布。使用此功能时，当 HDFS 客户端将数据发送到 HDFS 时，数据将被加密；当 HDFS 将数据发送回客户端时，数据将被解密。

HDFS 加密具有区域的概念，您可以将数据分为几个区域。每个区域都有其自己的主密钥，并且区域中的每个文件都用其自己的密钥加密。对于多租户应用程序，可以将租户的一个数据放入区域 1，租户的两个数据放入区域 2，等等。但是，存在一个局限性，即所有 HBase 数据必须驻留在一个区域中。 EsgynDB 使用 HBase 作为其表的存储引擎。



每个 HDFS 区域都有一个单独的加密密钥（EK），这些密钥存储在密钥存储区中。 将文件添加到区域时，将生成数据密钥（DEK），并使用此密钥对数据进行加密。 DEK 的管理由 HDFS 执行。 EK 的管理由客户执行。 EK 和 DEK 都可以加密。



2.6.3 HBase 透明加密

HBase 加密类似于 HDFS 中的单区域加密。客户在他们选择的密钥存储区中创建一个主密钥。创建表时，将指定 ENCRYPTION 选项，并且数据将由 HBase 加密。默认情况下，HBase 基于主密钥为每个表（列族）生成一个单独的数据加密密钥，并使用它来加密数据。 HBase 还有另一个称为 ENCRYPTION_KEY 的选项，该选项使客户可以指定自己的数据加密密钥。 HBase 使用用户指定的加密密钥，并使用主密钥对其进行加密。结果密钥用于加密数据。

EsgynDB 使用 HBase 创建和管理其表。它目前仅支持 ENCRYPTION 选项。

3. 数据安全前景和承诺

安全是一个不断发展的主题。 EsgynDB 正在与客户，法规遵从标准，Cloudera 和 Hortonworks 之类的发行版，第三方解决方案合作，并根据其自身的丰富经验提供一种提供市场上最佳安全解决方案的策略。

本节讨论 EsgynDB 中正在进行的一些激动人心的工作。

3.1 认证方式

如今，EsgynDB 支持 AD 和 LDAP 作为身份存储。它正在为其认证框架提供其他身份存储。我们正在寻找当今市场上可用的本机 Linux，Kerberos 和其他单点登录解决方案。

EsgynDB 的本地认证仅支持为数据库本身进行认证，暂时不能为其它组件提供服务。

3.2 授权书

Cloudera 和 Hortonworks 都在开发 Apache 产品，以提供 EsgynDB 今天提供的许多功能，但是它们不仅可以处理一种产品，还可以在生态系统中的所有产品上工作。 Apache Ranger 处于 EsgynDB 的集成路线图上。如[9]中所述，Apache Ranger（正在孵化）是一个身份验证层，位于 EsgynDB 之类的现有产品之上。EsgynDB 希望增强其 Privilege Manager 界面，使其成为 Ranger 下的存储库并利用其所有功能。除了与现有发行版集成之外，EsgynDB 还计划继续改善其特权支持，并正在研究：

- 行级权限
- 与 HBase 的单元级安全性集成在一起，
- 支持分层角色
- MAC（Mandatory Access Control)）功能

EsgynDB 正在通过支持基于特殊安全标签的 MAC（强制性访问控制）来改善安全授权。此设计也称为 LBAC（label based access control）。有两类需要带有标签：访问请求的主题或用户，以及需要额外级别的特权检查的目标或对象。安全管理员负责在与主题和目标关联的这些标签上设置特权。除传统特权外，还会检查这些标签上的特权。

EsgynDB 还在运行政府机构要求的安全兼容性测试，以证明我们的安全功能正常运行。

3.3 加密

EsgynDB 正在使用 AES 256 加密添加架构和表级别的应用程序加密支持。

3.4 数据编辑，掩蔽和沿袭

重要的是要知道您的数据在哪里，被谁接触，该数据，谁可以接触该数据，以及在需要将数据发送给其他用户时对数据进行混淆处理。 EsgynDB 当前正在通过管理审核日志并与现有解决方案进行接口以提供必要的功能来完成此任务。