

# Brief Security Databricks : rendu

## Partie 1 : Veille sur les Systèmes de Sécurité

### Livrable

- Un rapport qui synthétise vos recherches en décrivant chaque méthode de sécurité, ses avantages, ses limitations, et des cas d'utilisation.



## Mes notes

fonctionnalités clés de la sécurité du stockage Azure

- **Contrôle d'accès basé sur les rôles Azure (Azure RBAC)**

attribuer un rôle Azure : 3 éléments :

- principal de sécurité : objet qui représente un utilisateur, un groupe, un principal de service ou une identité gérée qui demande l'accès aux ressources Azure
- définition du rôle : ensemble d'autorisations (lecture, écriture, suppression)
- étendue : 4 niveaux : groupe d'administration, abonnement, groupe de ressource ou ressource  
⇒ attribution de rôles : associer une définition de rôle à un utilisateur, un groupe, un principal de service ou une identité gérée

- **Signature d'accès partagé (SAS)**

fournit un accès délégué aux ressources du compte de stockage > contrôle précis sur la manière dont un client peut accéder aux données

c'est un jeton ajouté à l'url d'une ressource de stockage

3 types de SAS :

- délégation d'utilisateur SAS > informations d'identification Microsoft Entra > Blob storage et Data Lake Storage
- service SAS > sécurisé avec la clé de compte de stockage > stockage Blob et Data Lake Storage
- compte SAS > sécurisé avec la clé de compte de stockage

- **Microsoft Entra ID**

fait partie de l'offre PaaS et fonctionne comme un service d'annuaire géré

accès à un ensemble de fonctionnalités qui ne sont pas disponibles nativement dans AD DS, telles que la prise en charge de l'authentification multifacteur, la protection de l'identité et la réinitialisation de mot de passe en libre-service

La principale force de Microsoft Entra ID réside dans la fourniture de services d'annuaire, le stockage et la publication des données des utilisateurs, des appareils et des applications, ainsi que la gestion de l'authentification et de l'autorisation des utilisateurs, des appareils et des applications

Microsoft Entra ID est avant tout une solution d'identité, conçue pour les applications basées sur Internet utilisant les communications HTTP (port 80) et HTTPS (port 443)

- **KeyVault**

c'est un coffre fort qui permet de

- gérer les clés (création et contrôle des clés de chiffrement)
- gérer les secrets (jetons, mots de passe, certificats, clés API...)

- gérer les certificats (provisionner, gérer, déployer des certificats SSL/TLS publics et privés)

de manière centralisée et simplifiée car en utilisant des URI (qui permettent aux applications de récupérer les versions d'un secret), aucun code n'est nécessaire pour protéger les infos secrètes stockées dans KeyVault

- **Storage Access Keys**

2 clés sont générées par compte de stockage qui peuvent être utilisées pour autoriser l'accès (soit via l'autorisation de clé partagée soit via les jetons SAS signés avec la clé partagée)

quand l'utilisation de Entra ID n'est pas possible, on peut utiliser des jetons SAS (de délégation utilisateur) avec une portée d'accès limitée

## **Partie 2 : Création et ingestion sécurisée sur le data lake**



## Mes notes

### Mécanismes de gestion des identités et des secrets

- **Principaux de service** : Ce sont des identités créées pour les applications afin qu'elles puissent accéder aux ressources Azure de manière sécurisée.
  - **Principal de service secondaire** : Utilisé pour accéder aux secrets dans Key Vault.
  - **Principal de service principal** : Utilisé pour accéder aux données dans Data Lake Storage Gen2.
- **Key Vault** : Un service pour stocker et gérer les secrets, clés et certificats de manière sécurisée.
- **Stratégies d'accès** : Elles définissent les permissions accordées aux principaux de service pour accéder aux ressources spécifiques.

### Étapes

1. Créez un compte de stockage en sélectionnant l'option **StorageV2** et en activant l'option **Enable hierarchical namespace**.
2. **Créer le principal de service Secondaire** :
  - Microsoft Entra ID : **Identité > Applications > Inscription des applications** et sélectionnez **Nouvelle inscription**  
Certificats & secrets > Nouveau secret client  
Notez l'**ID d'application (client)** et l'**ID du locataire**.  
keyvault-lmorel > secretkeyvaultmorel
  - Aller dans le Coffre de clés (Azure KeyVault) :
    - créer un coffre de clés en modifiant dans **Configuration de l'accès** : sélectionner "**Stratégie d'accès au coffre**" (détermine si un principal de sécurité donné, à savoir un utilisateur, une application ou un groupe d'utilisateurs, peut effectuer différentes opérations sur les clés, les secrets et les certificats)
    - mon coffre : kv-morel
    - **Stratégie d'accès** > créer > Sélectionnez le principal de service secondaire (keyvault-lmorel) et attribuez les permissions **Get** sur les secrets
3. **Créer le principal de service Principal** :
  - Microsoft Entra ID : **Identité > Applications > Inscription des applications** et sélectionnez **Nouvelle inscription**  
Certificats & secrets > Nouveau secret client  
Notez l'**ID d'application (client)** et l'**ID du locataire**.  
spdatalake-morel > secret-datalake-morel
  - **Coffre de clés (KeyVault)** : Objet > **Secrets > Générer/importer**

"Créer une clé secrète" : entrer les infos de la clé secrète du principal de service principal

**Configurer l'application pour utiliser le secret :**

Dans l'application, configurez l'authentification pour utiliser l'ID d'application et le secret du principal de service principal pour accéder aux ressources nécessaires.

- **Attribuer les rôles au principal de service principal :**
  - **dans le Data Lake Storage Gen2 :**
  - **Contrôle d'accès (IAM)** > ajouter une attribution de rôle > sélectionner le nom du rôle puis le membre (qui est en fait le nom du principal de service !)
    - Contributeur aux données Blob du stockage
    - (Délégation d'objet blob de stockage)



### Mes notes

Le rôle **Storage Blob Data Contributor** sur le compte de stockage permet de générer des User Delegation Keys et de créer des SAS Tokens.

Utiliser une User Delegation Key pour générer des SAS Tokens offre une sécurité accrue par rapport à l'utilisation des clés de compte de stockage. Cela permet de déléguer des permissions spécifiques sans exposer les clés de compte.

- **User Delegation Key** : clé générée à partir des identifiants d'un utilisateur ou d'un principal de service. Elle permet de signer des SAS (Shared Access Signatures) de manière sécurisée.
- **SAS Token (Shared Access Signature)** : jeton qui accorde un accès limité à des ressources Azure Storage, comme des fichiers ou des conteneurs, pour une durée et des permissions spécifiques.

## Journalisation



### Mes notes

Key Vault > Supervision > Paramètres de diagnostic > Ajouter

ensuite les logs sont visibles dans un dossier que l'on retrouve dans le conteneur sélectionné (ici storageegen2morel > stockage de données > conteneurs > insights-logs-auditevent





## Mes notes

- Expliquez les rôles des SP et la configuration.
- Justifiez les permissions attribuées.

le SP "secondaire" est protégé par un secret (variable d'environnement du script de connection) que l'on utilisera pour se connecter au coffre de clés

pour accéder au coffre fort, on définit une stratégie d'accès en "get" via le SP "secondaire"

pour le SP principal, on crée une clé secrète que l'on importe dans le coffre de clés dans le coffre de clés, on stocke le secret du SP "principal", que l'on récupère en se connectant grâce au "get" via le SP "secondaire"

ensuite, on génère une clé (User Delegation Key) grâce au rôle de "Contributor" donné au SP "principal" pour accéder au DataLake (via l'attribution de rôle dans le "Contrôle d'accès IAM") qui permet de créer un jeton (SAS-token) qui donne un accès limité à des ressources comme notre conteneur, pour une durée et des permissions déterminées

**Journal Key Vault :** Fournissez les logs qui montrent les actions sur les secrets.

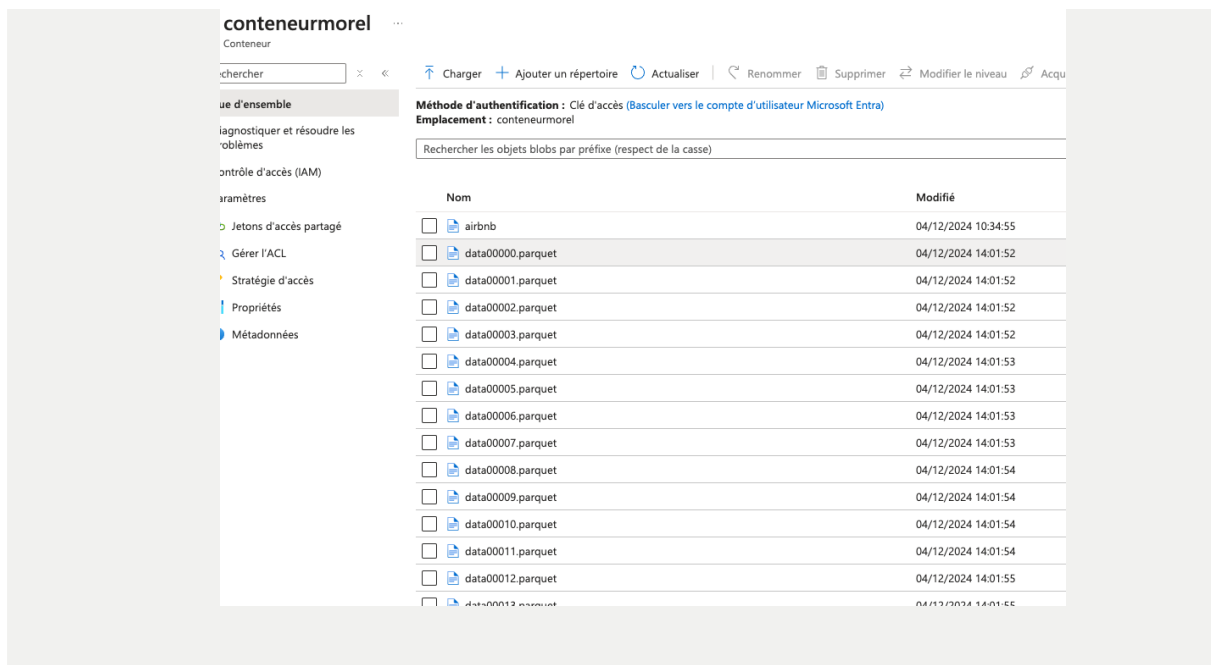
insights-logs-auditevent ...

Charger Ajouter un répertoire Actualiser Renommer Supprimer Modifier le niveau Acquérir le bail Résilier le bail

Méthode d'authentification : Clé d'accès (Basculer vers le compte d'utilisateur Microsoft Entra)  
Emplacement : insights-logs-auditevent / resourceid= / SUBSCRIPTIONS / 029B3537-0F24-400B-B624-6058A145FEF1 / RESOURCEGROUPS / RG\_LMOREL / PROVIDERS / MICROSOFT.KEYVAULT / VAULTS / KV-MOREL / y=2024 / m=12 / d=04

Rechercher les objets blobs par préfixe (respect de la casse) Afficher les objets supprimés

Nom	Modifié	Niveau d'accès	État de l'archive	Type d'objet blob	Taille	État
[...]						
h=09	04/12/2024 10:45:52					-
h=13	04/12/2024 14:06:43					-
h=14	04/12/2024 15:57:46					-
h=20	04/12/2024 21:32:39					-



## Partie 3 : Configuration d'Azure Databricks





## Mes notes

> créer un espace de travail Databricks (**databricks-lmorel**)

> créer un SP pour Databricks dans Entra ID et générer un secret

> puis stocker ce secret dans KeyVault (Objet>secret>générer/importer)

> Lancer l'espace de travail Databricks

Nouveau > Cluster : configurer le cluster

*un **cluster** est un ensemble de machines virtuelles (ou nœuds) qui travaillent ensemble pour exécuter des tâches de traitement de données*

> dans keyvault, créer une stratégie d'accès pour Databricks afin qu'il puisse accéder aux secrets ("get" et "list")

> puis ajouter "#secrets/createScope" à l'URL de l'espace de travail Databricks pour définir le scope

DNS : <https://kv-morel.vault.azure.net/>

ID de la ressource dans propriétés du coffre de clé /subscriptions/...

il faut créer un secret scope dans Databricks et le lier au Keyvault (pour que Databricks récupère les secrets directement depuis le Key Vault) : permet à Azure Databricks d'accéder aux secrets stockés dans Azure Key Vault de manière sécurisée

*Un **secret scope** dans Databricks est un conteneur pour stocker des secrets. Il permet à Databricks d'accéder aux secrets stockés dans Azure Key Vault sans avoir à les coder en dur dans vos notebooks ou scripts. Il agit comme un pont entre Databricks et Azure Key Vault.*

> dans le compte de stockage, relier le SP au Datalake pour lui donner accès

le SP Databricks, via un rôle (IAM) peut accéder au Datalake

le secret du SP de Databricks est stocké dans le coffre

on crée une stratégie d'accès (access policy) dans KeyVault pour Databricks , et un scope (dans Databricks), afin que Databricks puisse accéder aux secrets ("get" et "list") de manière sécurisée ; le scope agit comme un pont entre Databricks et Azure Key Vault



## Mes notes

créer un notebook python qui permet d'accéder aux données du compte de stockage à l'aide d'OAuth 2.0 avec le principal de service

## Partie 4 : Monitoring et Alertes



## Mes notes

Azure Monitor > Journal d'activité

Insights > Espaces de travail Log Analytics >

Espace de travail Log Analytics

Créer une règle de collecte de données

Ressource : Créer un point de terminaison (endpoint-Imorel1)

Ressource : Créer un point de terminaison (endpoint-Imorel1)

## Livrables

- **Captures d'Écran** : Des dashboards, des alertes configurées, et des notifications reçues.
- **Rapport** : Décrivant le système de monitoring mis en place, les raisons des choix effectués, et comment le système répond aux besoins de l'entreprise.

Insights > Applications (insights-app-Imorel)

### Application Insights

Analysez les performances et l'utilisation d'une application web

✓ Validation passed

De base Étiquettes **Review + create**

#### SUMMARY

**Application Insights**  
par Microsoft

Subscription	Simplon - HDF - Data Engineer P1
Groupe de ressources	RG_LMOREL
Nom	insights-Imorel
Région	France Central
Espace de travail	log-analytics-Imorel [francecentral]

Supervision > alertes > créer une règle d'alertes (sur la ressource insights-app-Imorel)

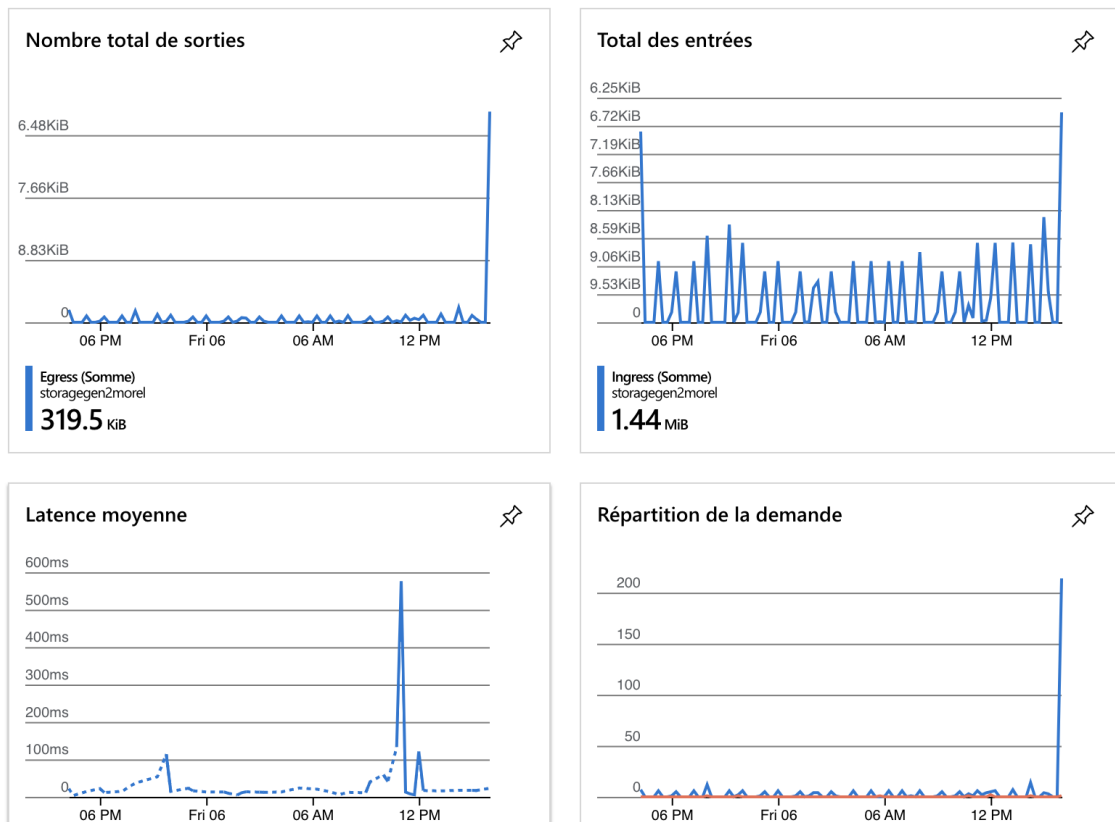
## Compte de stockage

### Activity Logs

pour le Datalake, les métriques pertinentes :

## Metrics

pour le Datalake, les métriques pertinentes :



## Insights

Dans Monitor, insights, compte de stockage

Abonnement	Transactions	Chronologie des transactions	Latence E2E	Latence du serve...	ClientOtherError...
✓ Simplon - HDF - Data Engineer	322				
storagegen2morel	322		25,94 ms	24,36 ms	6

Il est possible d'épingler des rapports, des graphiques afin de les retrouver facilement dans la vue d'ensemble (insights du compte de stockage)













pour configurer les notifications, les actions, créer un Groupe d'actions et choisir le mode d'alerte (SMS, email...)

## Monitor

vue d'ensemble de la performance et de la santé des ressources

> paramètres > paramètres de diagnostic, on voit l'ensemble des paramètres de diagnostic créés sur les ressources

dans le journal d'activité : on voit toutes les activités intervenues depuis un temps donné, à l'échelle souhaitée, par exemple sur l'ensemble d'un groupe de ressources

Groupe d'administration : <b>Aucun</b> Abonnement : <b>Simplon - HDF - Data Engineer P1</b> Gravité de l'événement : <b>Tout</b> Intervalle de temps : <b>Dernières 24 heures</b>						
Groupe de ressources : <b>RG_LMOREL</b> ✕  Ajouter un filtre						
46 éléments.						
Nom de l'opération	Statut	Heure	Horodatage	Abonnement	Événement lancé par	
>  Delete resource diagnostic setting	Échec	il y a 3 minutes	Fri Dec 06 2...	<a href="#">Simplon - HDF - Data Engineer P1</a>	Imorel.ext@simplonforma...	
>  Create or update action group	Réussite	il y a 7 minutes	Fri Dec 06 2...	<a href="#">Simplon - HDF - Data Engineer P1</a>	Imorel.ext@simplonforma...	
>  Create or update resource diagnostic setting	Réussite	il y a 8 minutes	Fri Dec 06 2...	<a href="#">Simplon - HDF - Data Engineer P1</a>	Imorel.ext@simplonforma...	
>  Delete Deployment	Échec	il y a 12 minutes	Fri Dec 06 2...	<a href="#">Simplon - HDF - Data Engineer P1</a>	Imorel.ext@simplonforma...	
>  Delete Smart Detector alert rule	Échec	il y a 16 minutes	Fri Dec 06 2...	<a href="#">Simplon - HDF - Data Engineer P1</a>	Imorel.ext@simplonforma...	
>  Delete Smart Detector alert rule	Échec	il y a 16 minutes	Fri Dec 06 2...	<a href="#">Simplon - HDF - Data Engineer P1</a>	Imorel.ext@simplonforma...	
>  Create Smart Detector alert rule	Réussite	il y a 17 minutes	Fri Dec 06 2...	<a href="#">Simplon - HDF - Data Engineer P1</a>	Imorel.ext@simplonforma...	
>  Delete Connection	Échec	il y a 32 minutes	Fri Dec 06 2...	<a href="#">Simplon - HDF - Data Engineer P1</a>	Imorel.ext@simplonforma...	
>  Add management locks	Réussite	il y a 2 heures	Fri Dec 06 2...	<a href="#">Simplon - HDF - Data Engineer P1</a>	jvangansberg.ext@simplo...	
>  List Storage Account Keys	Réussite	il y a 3 heures	Fri Dec 06 2...	<a href="#">Simplon - HDF - Data Engineer P1</a>	Imorel.ext@simplonforma...	
>  List changes of a single resource	Réussite	il y a 4 heures	Fri Dec 06 2...	<a href="#">Simplon - HDF - Data Engineer P1</a>	Imorel.ext@simplonforma...	

quand on crée une règle d'alertes, on associe un groupe d'actions (envoi email par ex)

exemple de règle d'alerte : sur le keyvault, sur toutes les opérations administratives

## Logs Analytics (espace de travail)

explorer les données de journalisation pour des analyses plus détaillées