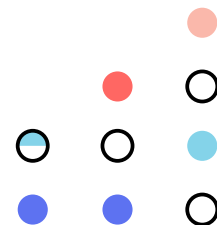


https://minet.net

8 mars 2023

Formation interne - MiNET

CEPH



Benoit
Nathan

Poti sommaire

01

Introduction

03

Architecture

02

Concepts de base

04

Configuration

https://minet.net/

01

Introduction

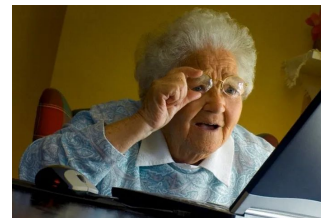
Ceph, c'est quoi ?
Pourquoi Ceph ?

https://minet.net/

Ceph



« Ceph est une solution de stockage distribué, sans point unique de défaillance, extensible jusqu'à l'exaoctet et hautement disponible. Les données sont répliquées pour permettre au système d'être tolérant aux pannes. »

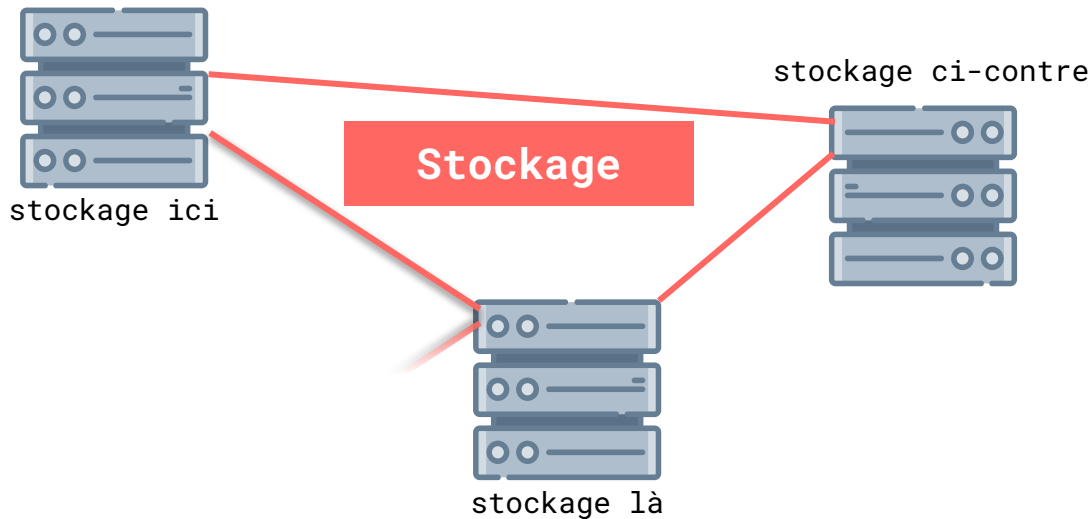




<https://minet.net/>



Stockage distribué



Le stockage est distribué à travers un ensemble de machines, appelées *nœuds*.





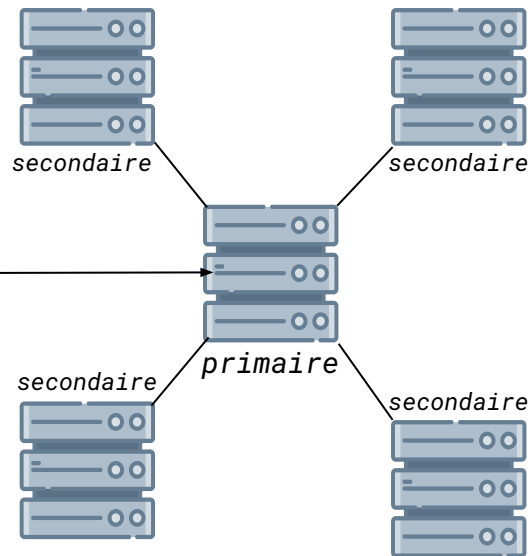
<https://minet.net/>



Point unique de défaillance



Élément d'un SI dont
une panne entraîne
l'arrêt complet du
système.

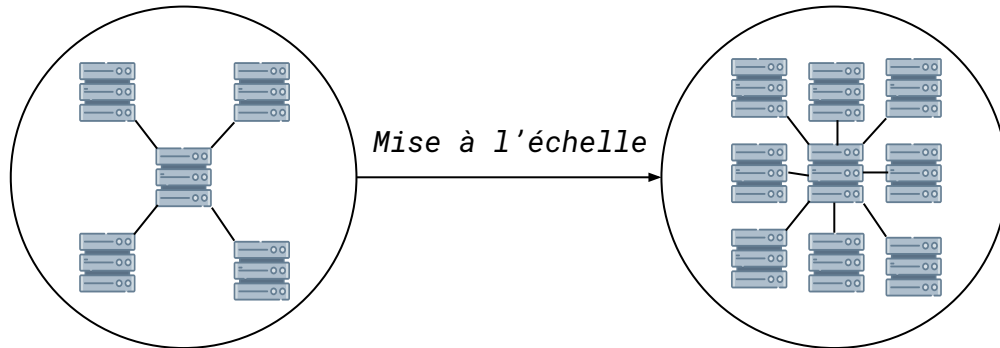




<https://minet.net/>



Extensibilité



Avant

Après

Besoin de **peu**
de stockage

Besoin de **beaucoup**
de stockage

Capacité d'un système à
maintenir ses
fonctionnalités et ses
performances face à un
changement d'ordre de
grandeur de la demande.





<https://minet.net/>



Autres systèmes similaires

Pool ZFS

- **ZFS** : Système de fichiers de très haute capacité
- **RAID 1+0** : Confère formation précédente

S'intègre avec Proxmox VE

Cluster IPFS

- **IPFS** : Système de fichiers interplanétaire : stockage décentralisé, immuable et versionné
- **Cluster** : Cluster de nœuds IPFS isolé

S'intègre pas avec Proxmox VE...



https://minet.net/

02

Concepts de base de Ceph

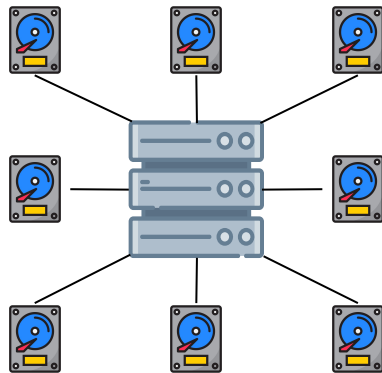
Pools, RADOS, RGW, MDS,
etc.



<https://minet.net/>



Pool de stockage



Collection de volumes
de stockage.

Pool de stockage = 1 nœud

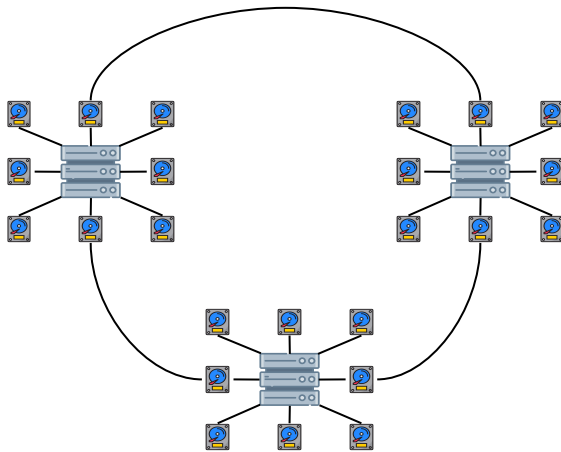




<https://minet.net/>



RADOS



*Reliable Autonomic
Distributed Object Store*

Moteur de stockage
objet fiable, autonome
et distribué sur lequel
se base Ceph.

Large collection de nœuds de
stockage dont les détails vous
seront fournis plus tard

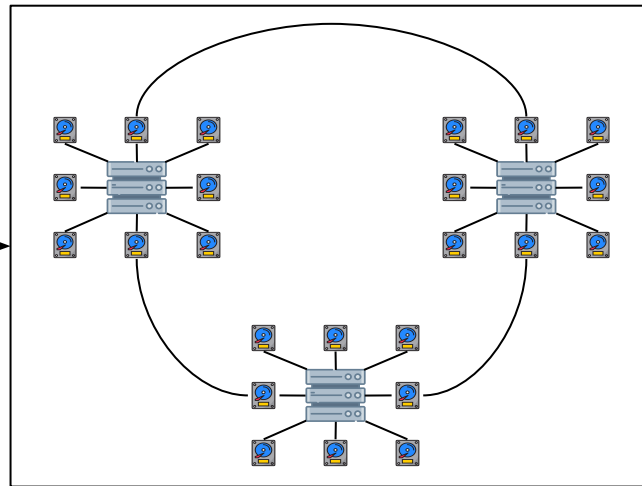
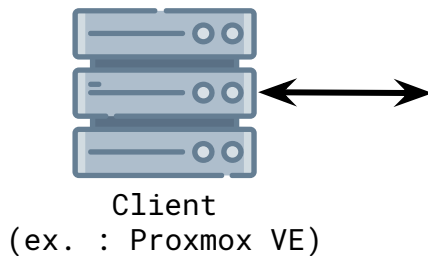




<https://minet.net/>



RGW



Cluster Ceph

REDOS Gateway
Ceph Object Gateway

Interface de stockage
objet pour fournir
une API REST aux
applications vers les
clusters Ceph.



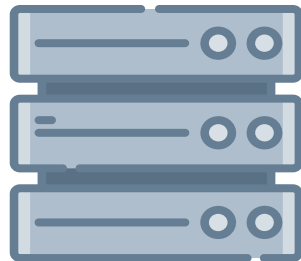


<https://minet.net/>

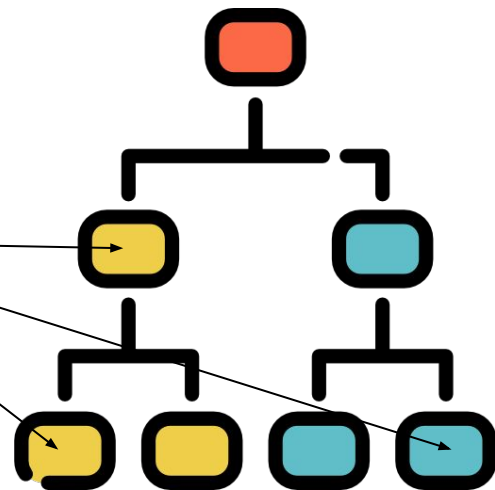


MDS

Serveur de
métadonnées



Gère l'espace de
noms du système de
fichiers de Ceph



CephFS

Metadata Server

Service gérant l'espace
de noms du système de
fichiers CephFS.

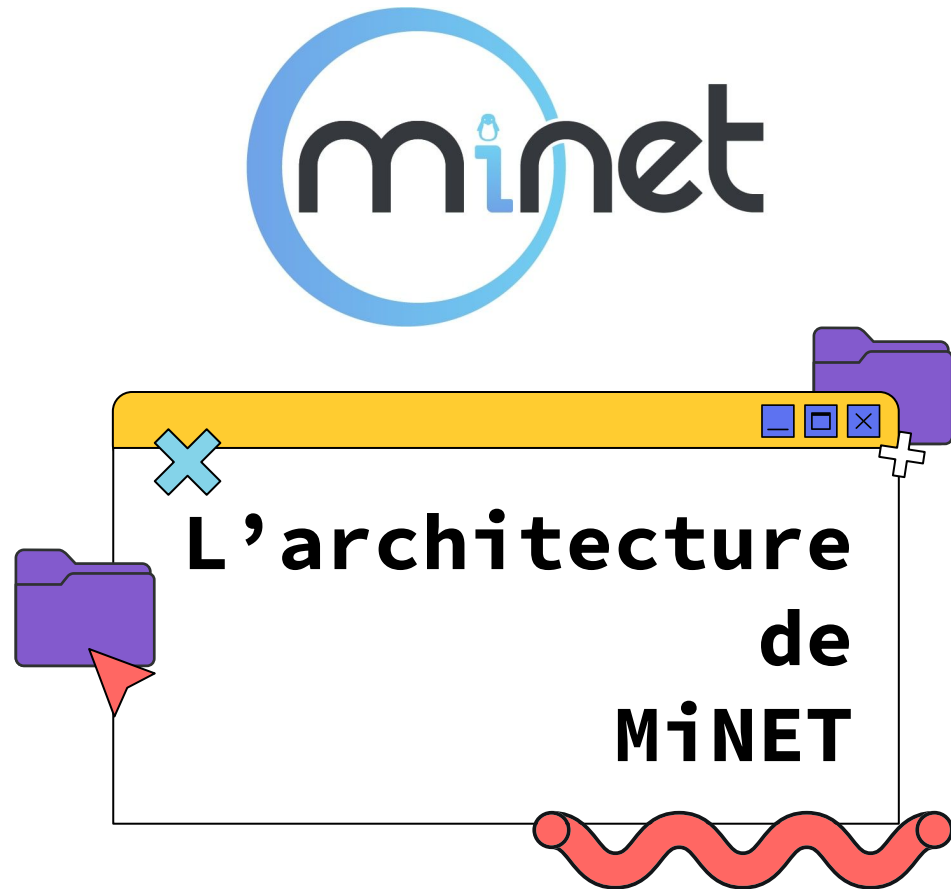
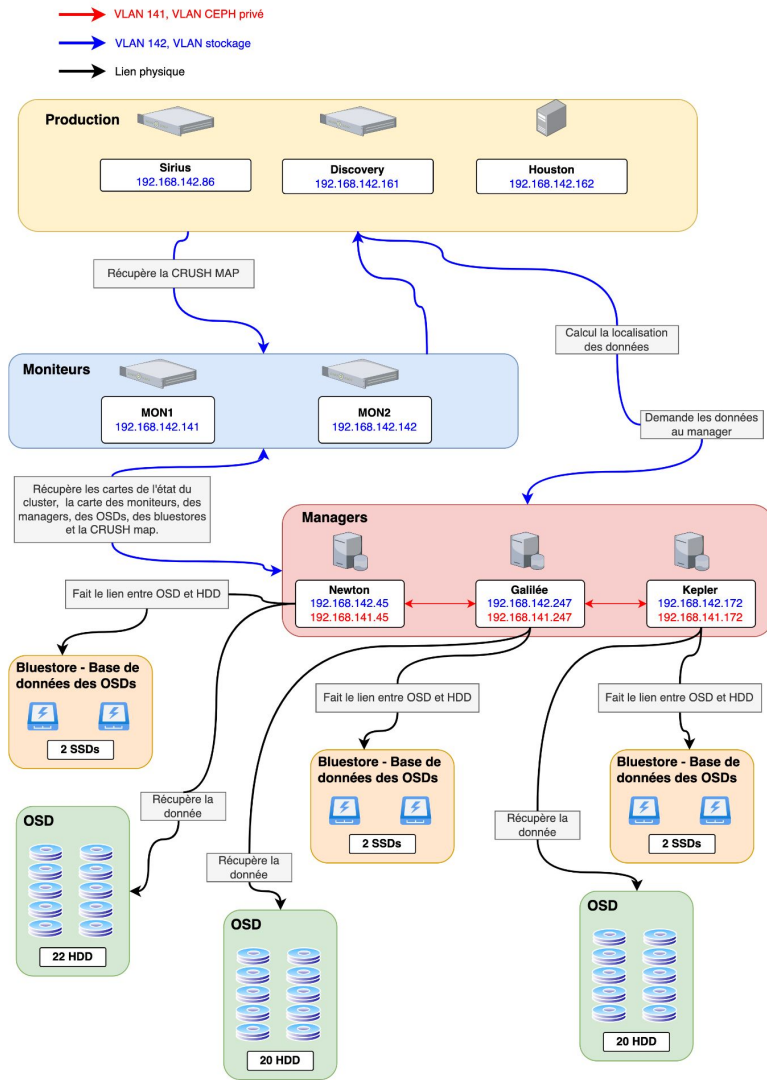


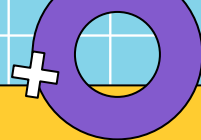
https://minet.net/

03

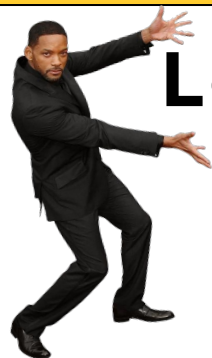
Architecture de Ceph

Moniteurs, Manageurs,
OSDs, c koi toussa ?





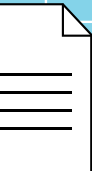
<https://minet.net/>



Les OSDs



Comment je stocke les données dans mes disques ?

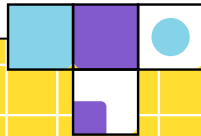
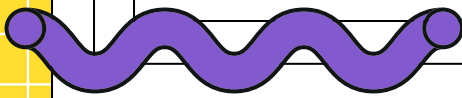
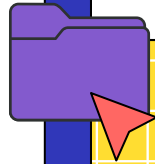




<https://minet.net/>

On délègue !

Les OSDs permettent de stocker et de récupérer les données de manière efficace, de répliquer les données pour la résilience et la disponibilité, et d'assurer une répartition équitable des données dans tout le cluster.





<https://minet.net/>



Les OSDs

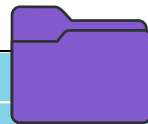
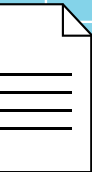
OSD=daemon

Stocke et accède aux données

OSD

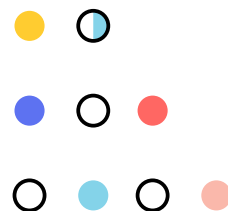
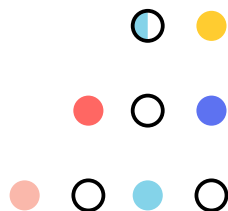
Il gèrent les métadonnées, comme les informations d'index et les emplacements de données

1 OSD gère 1 disque



https://minet.net/

On complique les choses ?



https://minet.net/

Les OSDs



Bluestore

Stocke les
métadonnées



1 OSD gère 1 disque

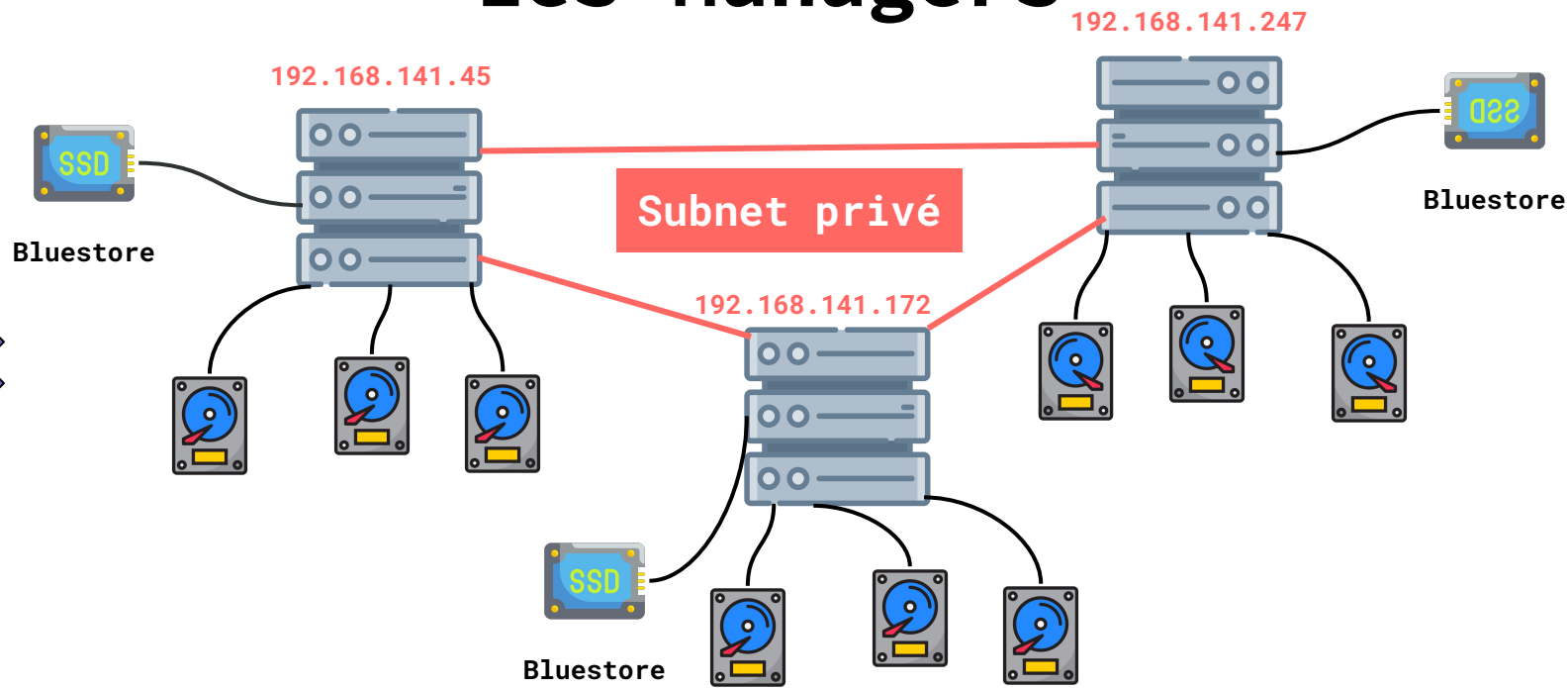


SSD

Il gèrent les
métadonnées, comme les
informations d'index et
les emplacements de
données

https://minet.net/

Les Managers





https://minet.net/

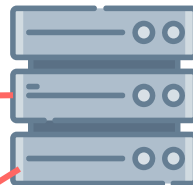


Les Managers

192.168.141.45



192.168.141.247



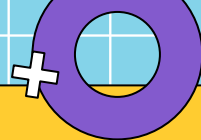
Subnet privé

192.168.141.172



Manager

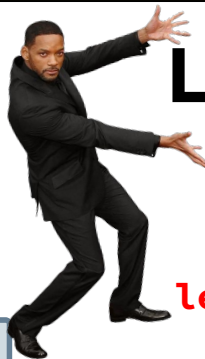
- Génère l'OSD map
- Gère les OSDs
- Rend transparent le cluster depuis tous les noeuds



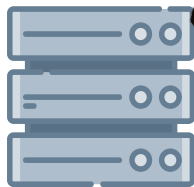
<https://minet.net/>



Les Moniteurs



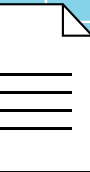
Comment je sais où sont
les données de la VM adh6 ?



Serveur de prod
Sirius



Serveur de stockage
Newton



https://minet.net/

Les Moniteurs

Comment je sais où sont
les données de la VM adh6 ?



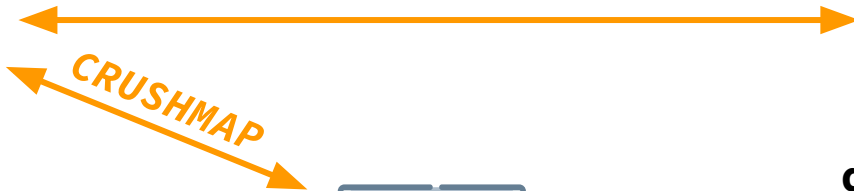
**Serveur
de prod**
Sirius



**Serveur
de stockage**
Newton



Moniteur



https://minet.net/

Les Moniteurs

Comment je sais où sont
les données de la VM adh6 ?



**Serveur
de prod**
Sirius



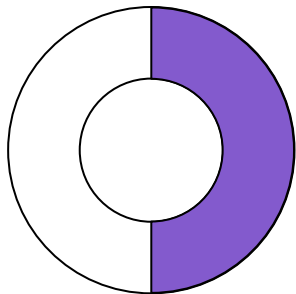
Moniteur



**Serveur
de stockage**
Newton

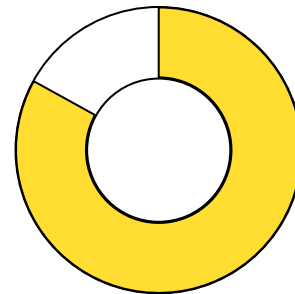
https://minet.net/

La réplication des données



100% autonome

Re-réplique les données
dès qu'un OSD est out



1 replicat/noeud

Jamais plus d'un réplicat
par noeud



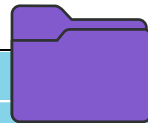
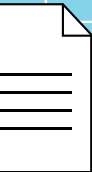
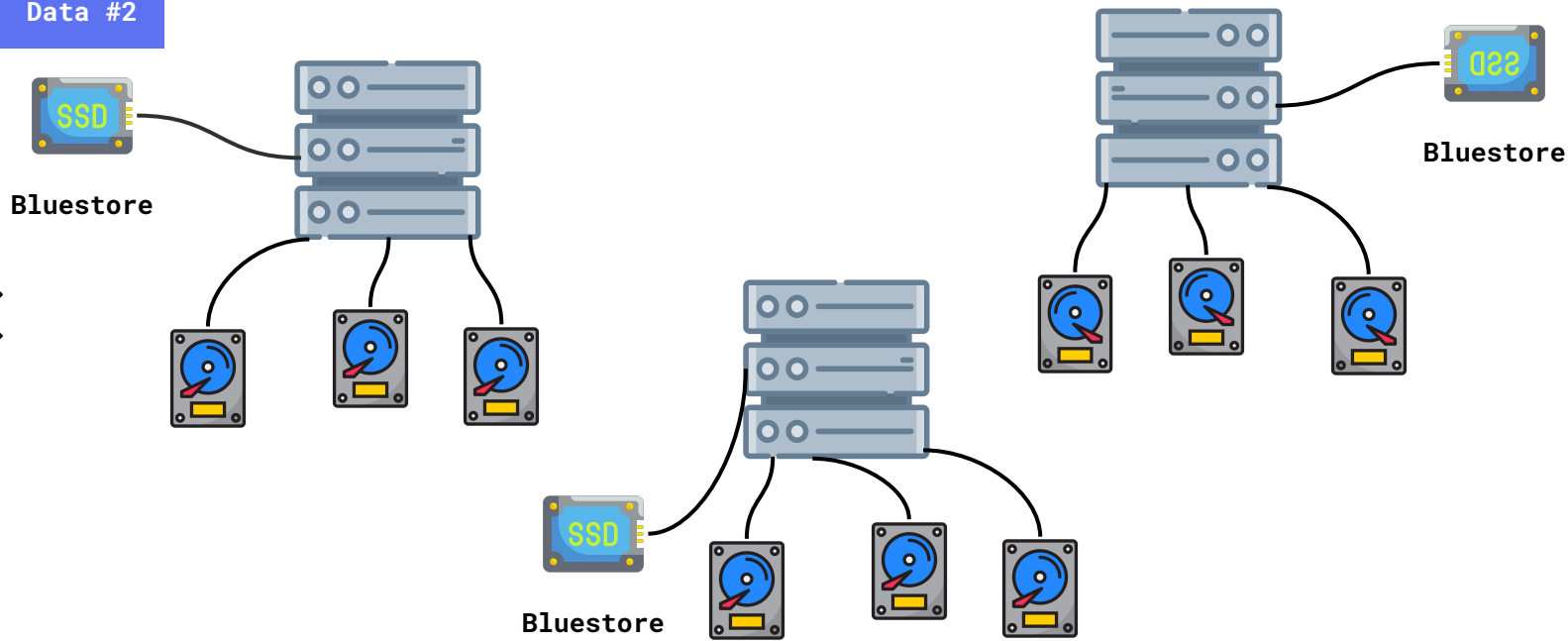
<https://minet.net/>



Data #1

Data #2

La réplication x2

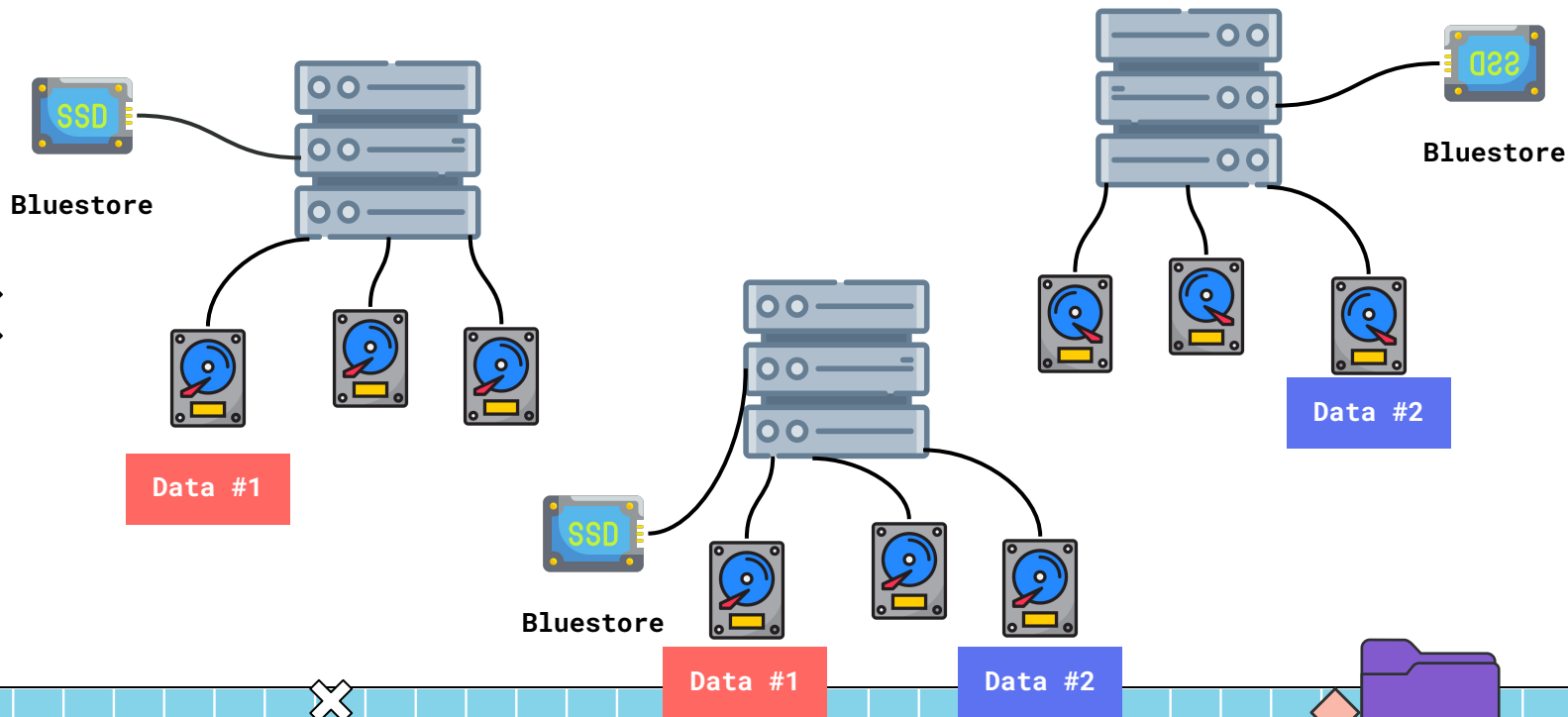




<https://minet.net/>



La réplication x2





Perdre 1 OSD

La



Perdre 1
serveur/3

n x2



Perdre 2
serveurs/3



Bluestore



Data #1



Bluestore

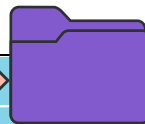


Data #1

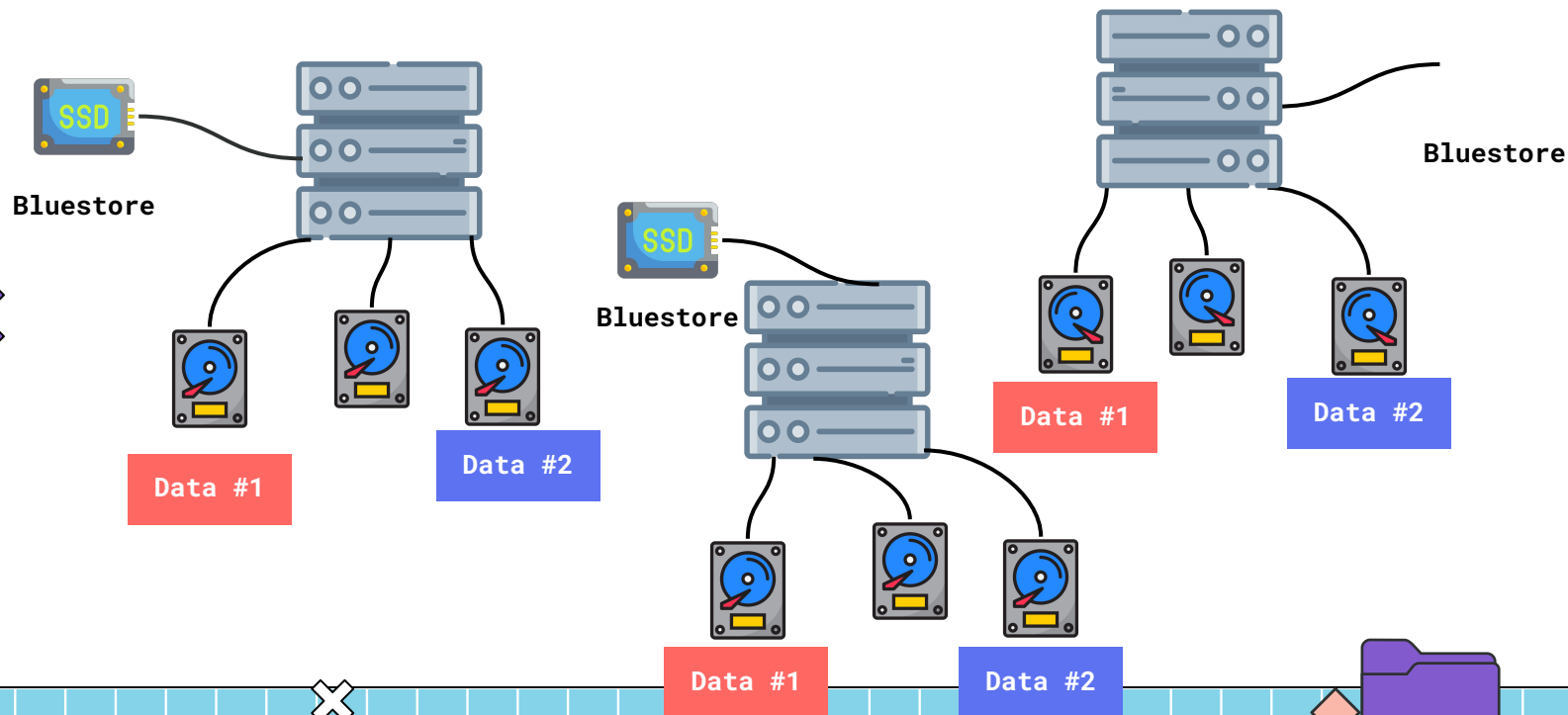


Data #2

Bluestore



La réplication x3





Perdre 1 OSD

La



Perdre 1
serveur/3

n x3



Perdre 2
serveurs/3



Bluestore



Data #1



Data #2



Bluestore



Data #1



Data #2



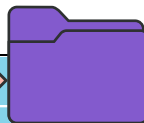
Data #1

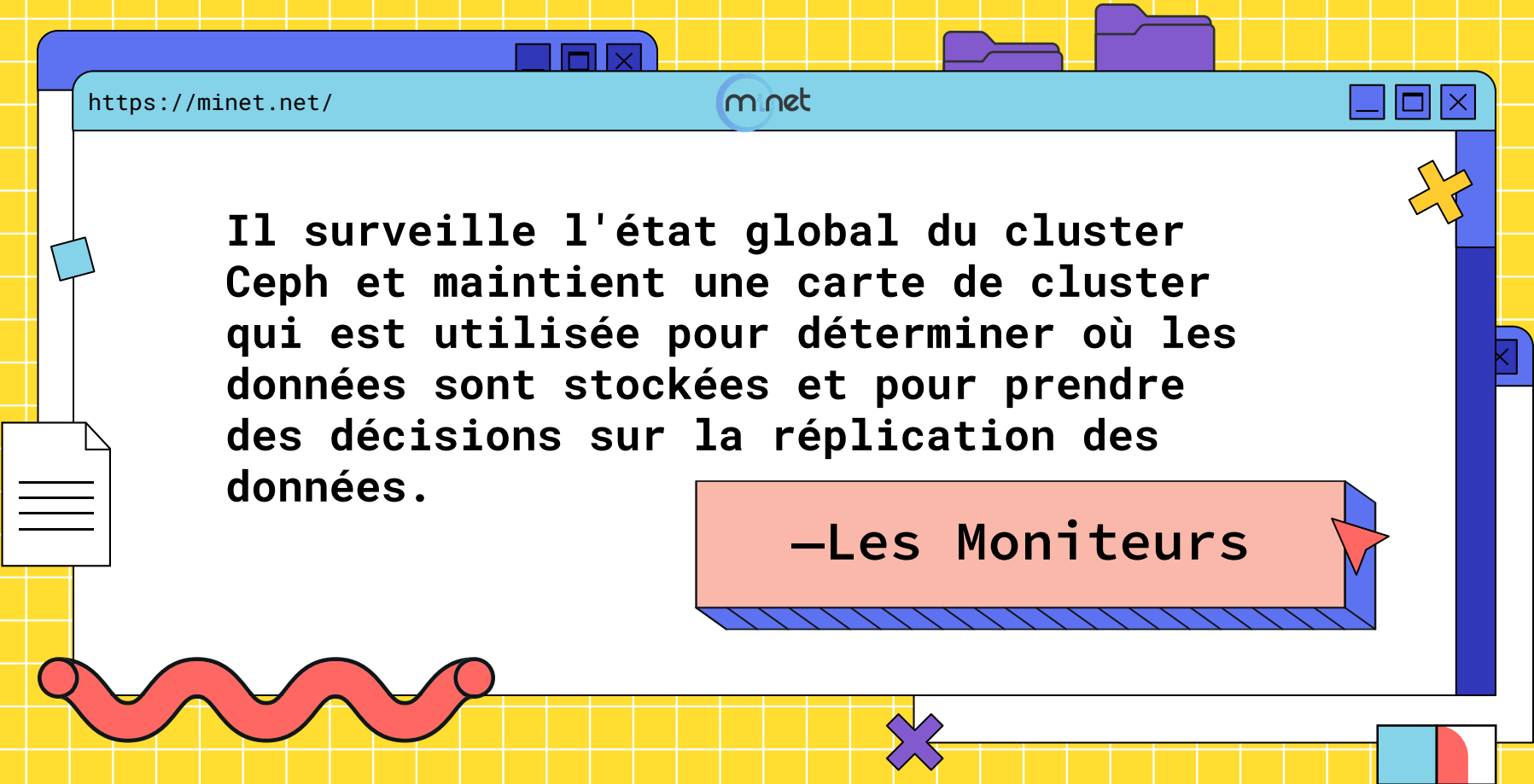


Data #2



Bluestore





https://minet.net/

m.net

Il surveille l'état global du cluster Ceph et maintient une carte de cluster qui est utilisée pour déterminer où les données sont stockées et pour prendre des décisions sur la réplication des données.

—Les Moniteurs

https://minet.net/

Les Moniteurs

Comment je sais où sont
les données de la VM adh6 ?



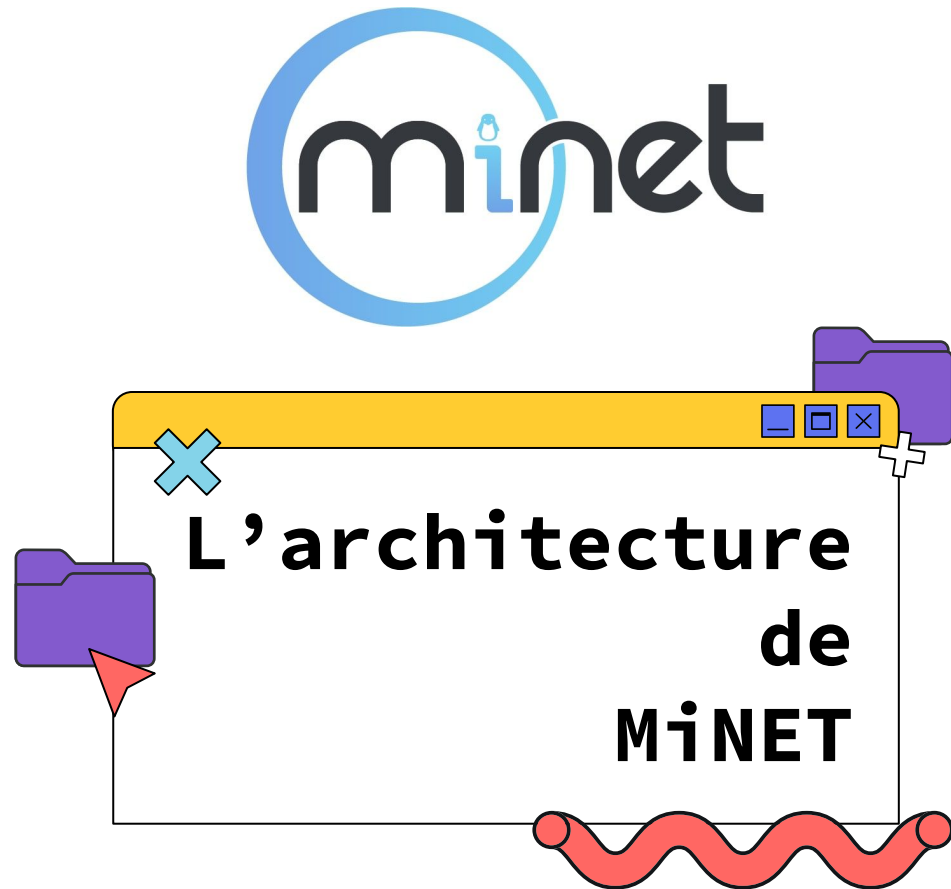
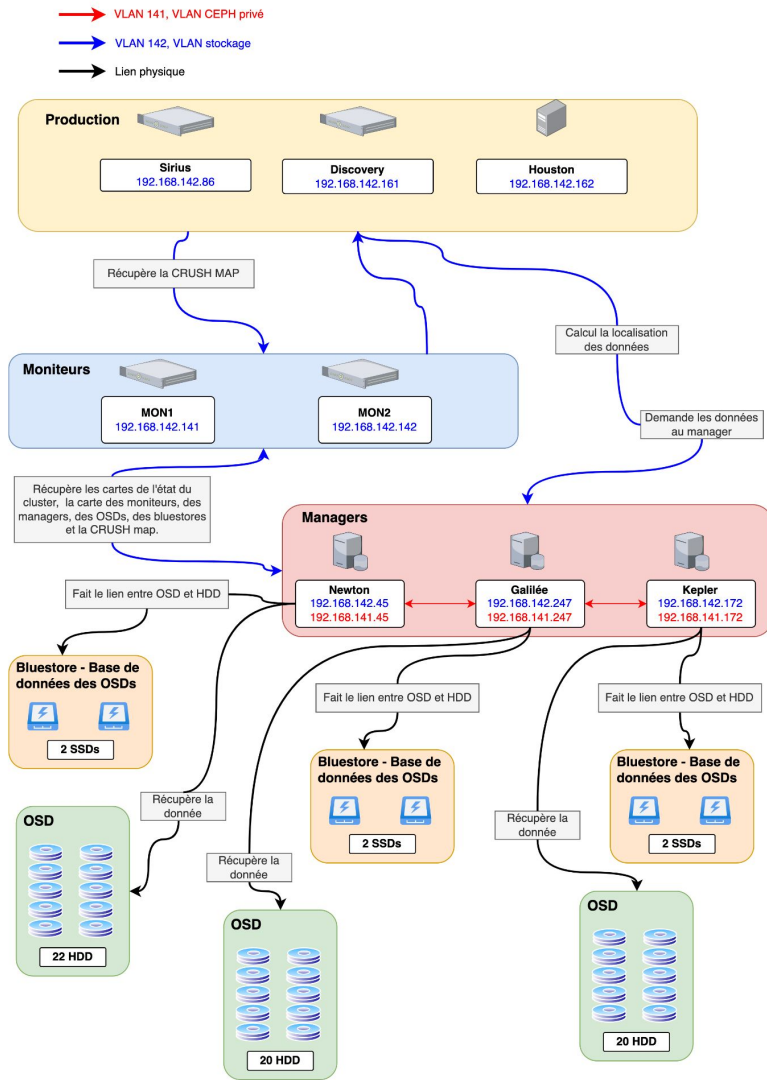
**Serveur
de prod**
Sirius

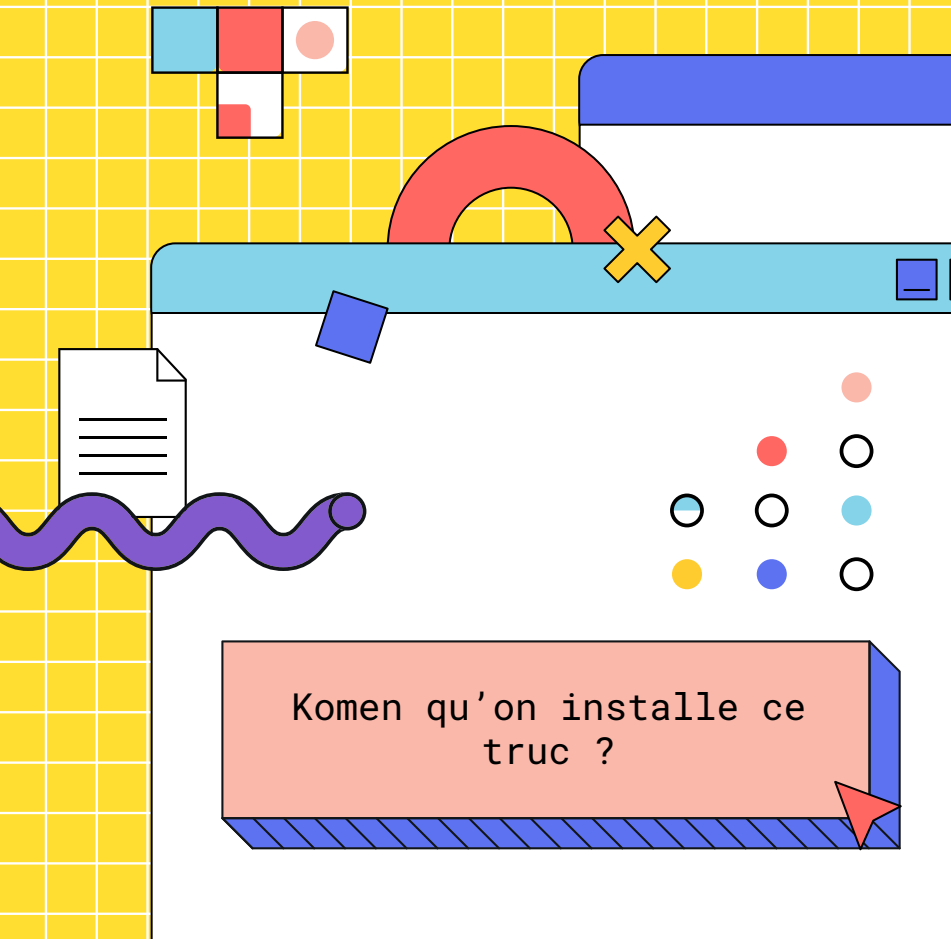


Moniteur



**Serveur
de stockage**
Newton



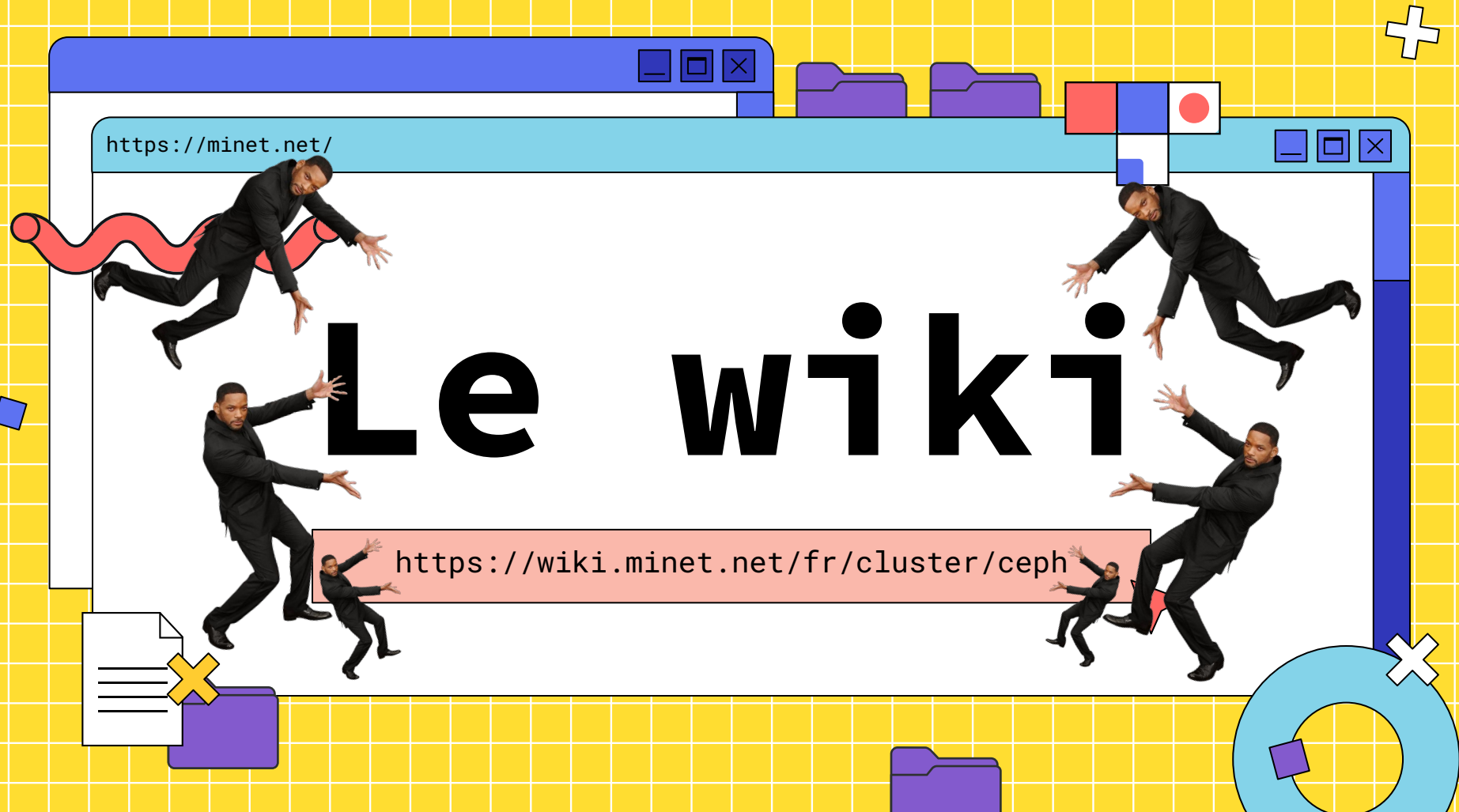


https://minet.net/

04

Configuration & Installation

Komen qu'on installe ce
truc ?



THANKS FOR YOUR ATTENTION



NO QUESTIONS PLEASE

<https://minet.net/>

Installation : C'est simple

```
1 apt install ceph ceph-mds
```

Déployer un cluster CEPH

**1-Déployer un
MON**

**2-Déployer un
Manager**

**3-Déployer les
OSDs**

**4-Se connecter
au cluster**





<https://minet.net/>



1-Déployer un MON (new cluster)

Installer
le 1er MON



Créer le
fichier de
configuration



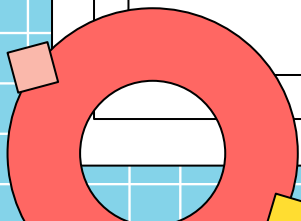
Créer le
keyring du
MON



Générer la
MON map



Lancer le daemon





<https://minet.net/>



1-Déployer un MON (old cluster)

Installer
le MON



Récupérer le
fichier de
configuration



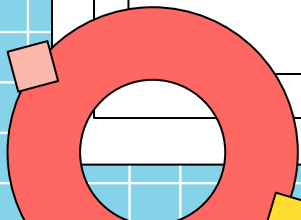
Créer le
keyring du
MON



Ajouter le MON
à la MON map



Lancer le daemon





<https://minet.net/>



2-Déployer un MDS

Installer
le MDS



Récupérer le
fichier de
configuration



Créer le
keyring du
MDS



Lancer le
daemon





<https://minet.net/>



3-Déployer les OSDs

Formater les
disques



Partitionner
les SSDs



Déployer les
OSDs



Lancer les
daemons



FAIRE UN RAID
ET MONITORER





https://minet.net/



4-Connecter proxmox à CEPH

Créer un
keyring au
client CEPH



Configurer le
'storage'
proxmox



Attention à bien
donner une IP dans le
VLAN public (142)



https://minet.net/

Monitoring



Le dashboard CEPH
est votre meilleur
ami.
Il permet de checker
la santé et gérer
votre cluster.

Dashboard

https://192.168.102.45:8888



Monitoring de tout
le cluster, avec
alerting et tendance
de CEPH

Grafana

https://grafana.minet.net



Monitoring de l'état
des SSDs pour
anticiper leur
changement

SSD

!!! À FAIRE !!!

Quelques commandes qui sauvent la vie

```
1 ceph -s # Get current CEPH health status
2
3 ceph osd tree # Architecture de tous les OSDs et leur état
4
5 rbd ls <pool> # Voir tous les objets RBD d'une pool
6
7 ceph auth get-or-create <client> # Récupérer/créer le keyring d'un clien
8
9 lsblk # Pour observer tous les disques et leur partition lvm
```

THANKS FOR YOUR ATTENTION



NO QUESTIONS PLEASE