# AFS Console

Jérôme Belleman

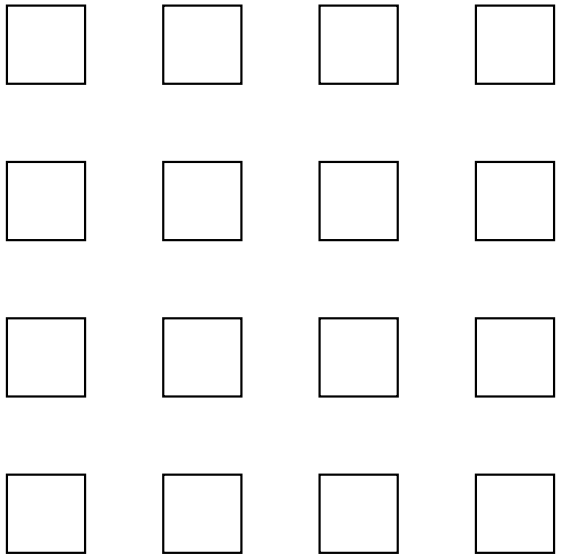
CERN

7 septembre 2010

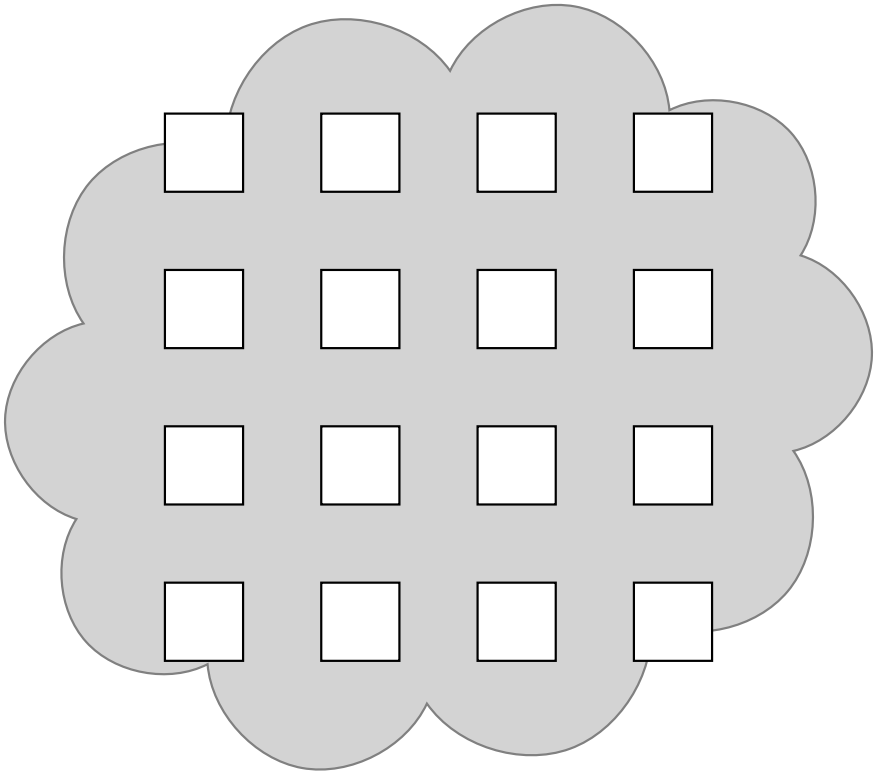
Section 1

# Présentation d'AFS

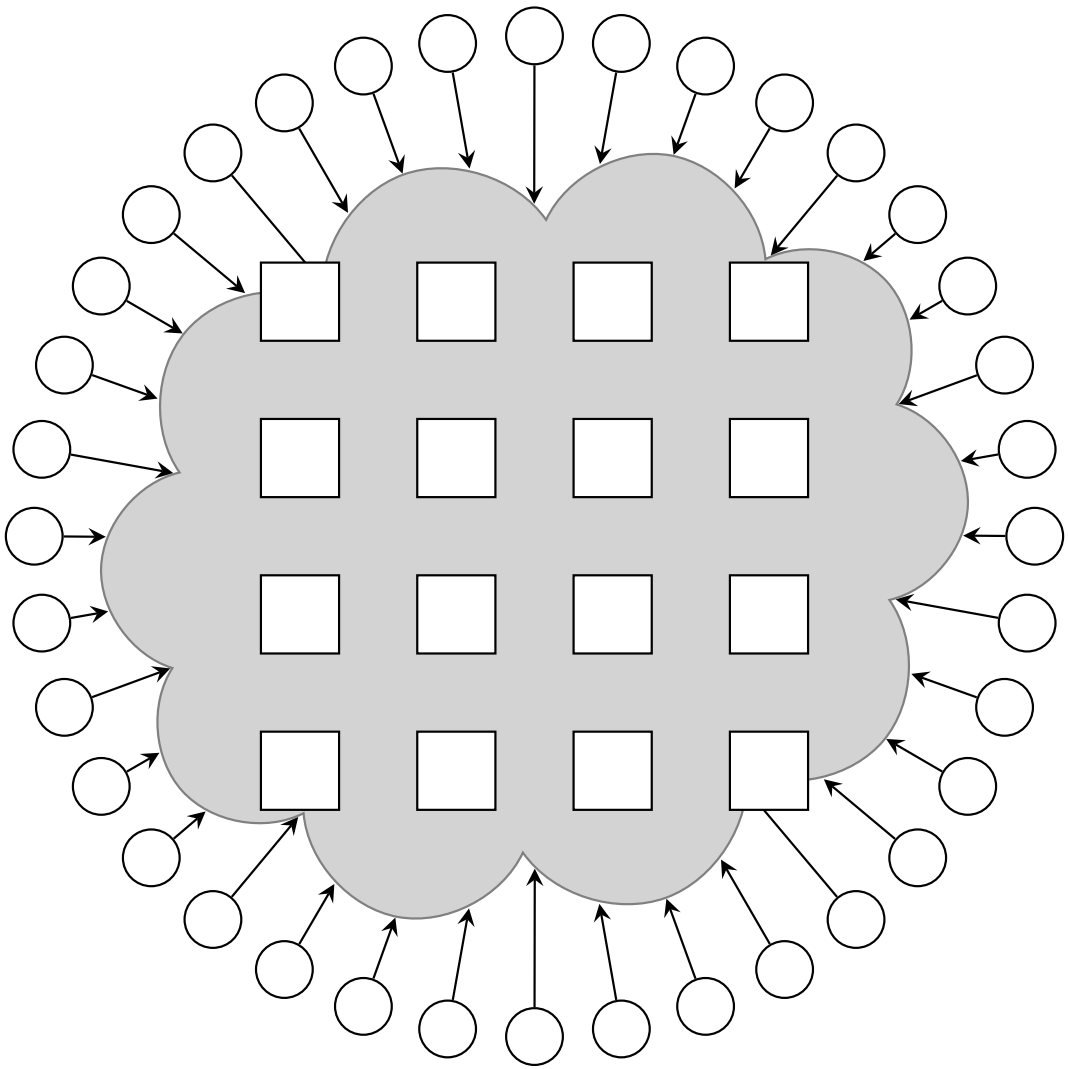
# Principe d'AFS à l'échelle d'un site



# Principe d'AFS à l'échelle d'un site



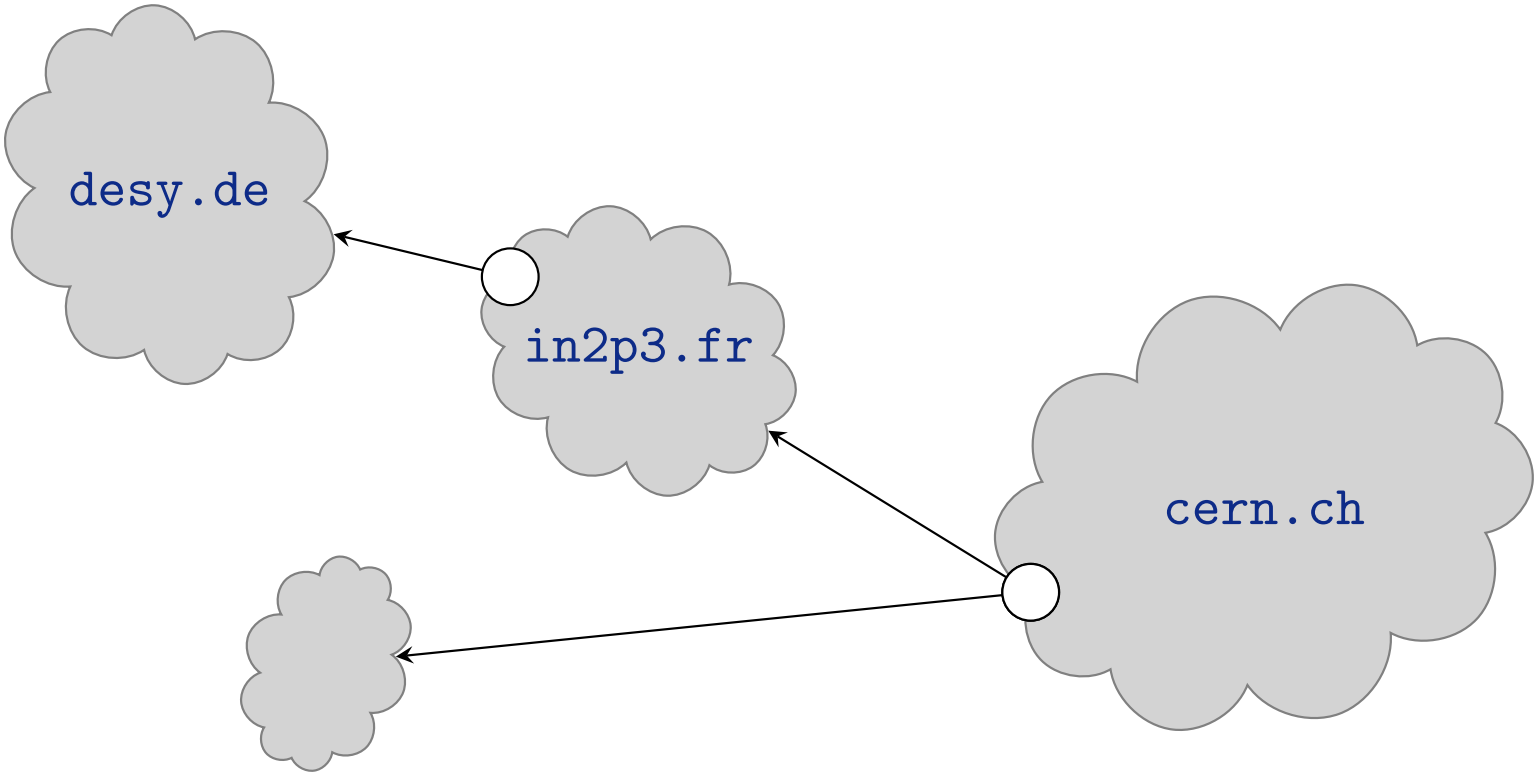
# Principe d'AFS à l'échelle d'un site



# Illusion de fichiers locaux

```
jay@pc42.cern.ch% cd /afs/cern.ch/user/jay
jay@pc42.cern.ch% ls
./              local/              .tcshrc
../             .login              texmf/
afsconsole/     moreafsstuff/       .vimrc
afsstuff/       .muttrc             www/
bin/            private/            .xbindkeysrc
.fvwm/          public/             .Xresources
.history        .reminders          .Xsession
.inputrc        stillmoreafsstuff/
```

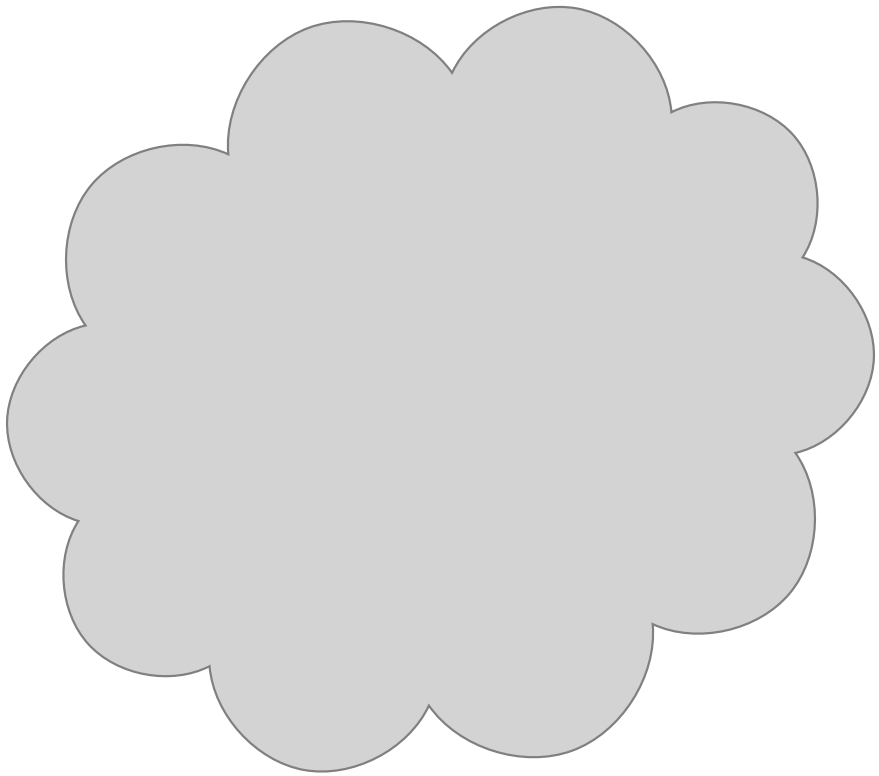
# Principe d'AFS sur le plan mondial



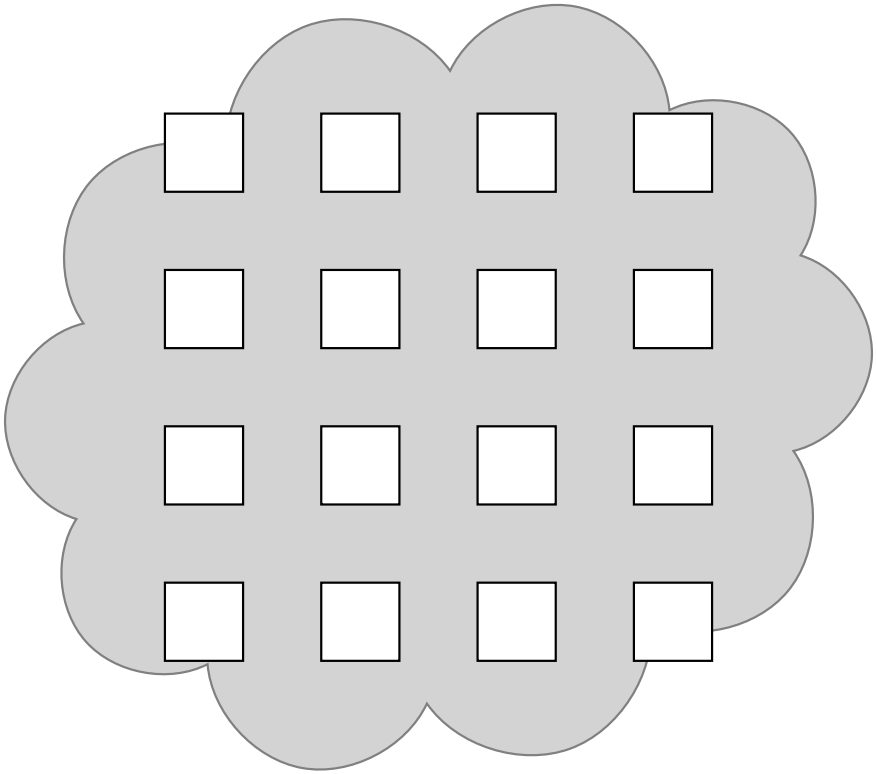
# "Le monde entier dans un sous-répertoire"

```
jay@pc42.cern.ch% ls /afs
biocenter.helsinki.fi/   in2p3.fr/
caspur.it/               itep.ru/
cern.ch/                 jpl.nasa.gov/
cs.pitt.edu/             lcp.nrl.navy.mil/
desy.de/                 math.cornell.edu/
doe.atomki.hu/           math.unifi.it/
epitech.net/             ncsa.uiuc.edu/
fnal.gov/                net.mit.edu/
freedaemon.com/          phy.bris.ac.uk/
grand.central.org/       physics.wisc.edu/
hep.caltech.edu/         physik.uni-freiburg.de/
hepix.org/               psi.ch/
hep.man.ac.uk/           slac.stanford.edu/
ics.muni.cz/             uni-mannheim.de/
ific.uv.es/              ...
```

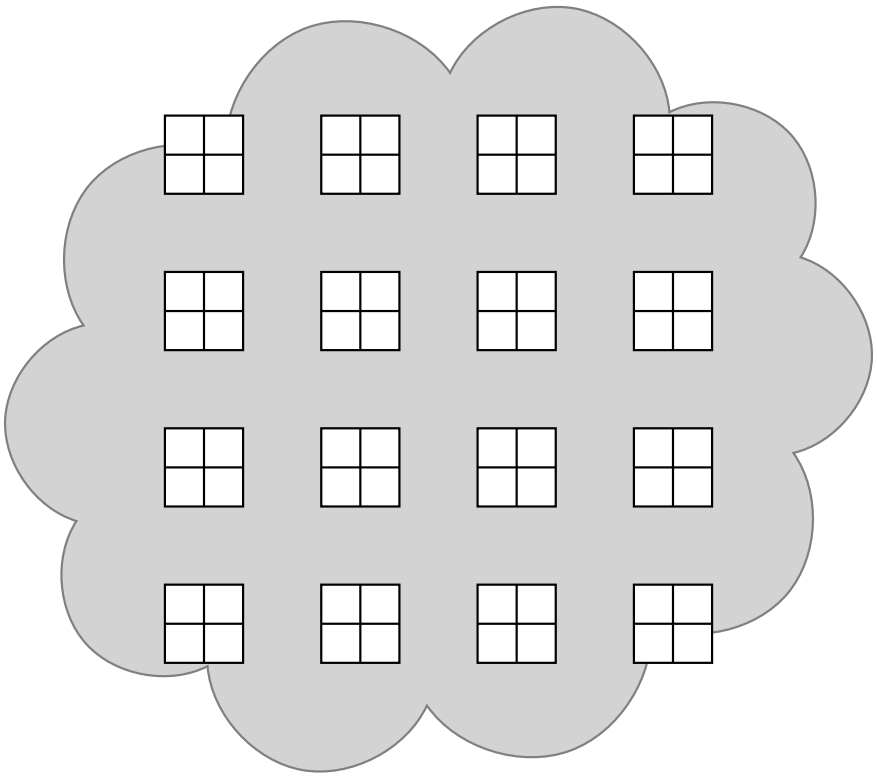
# Principe d'AFS à l'échelle d'un site



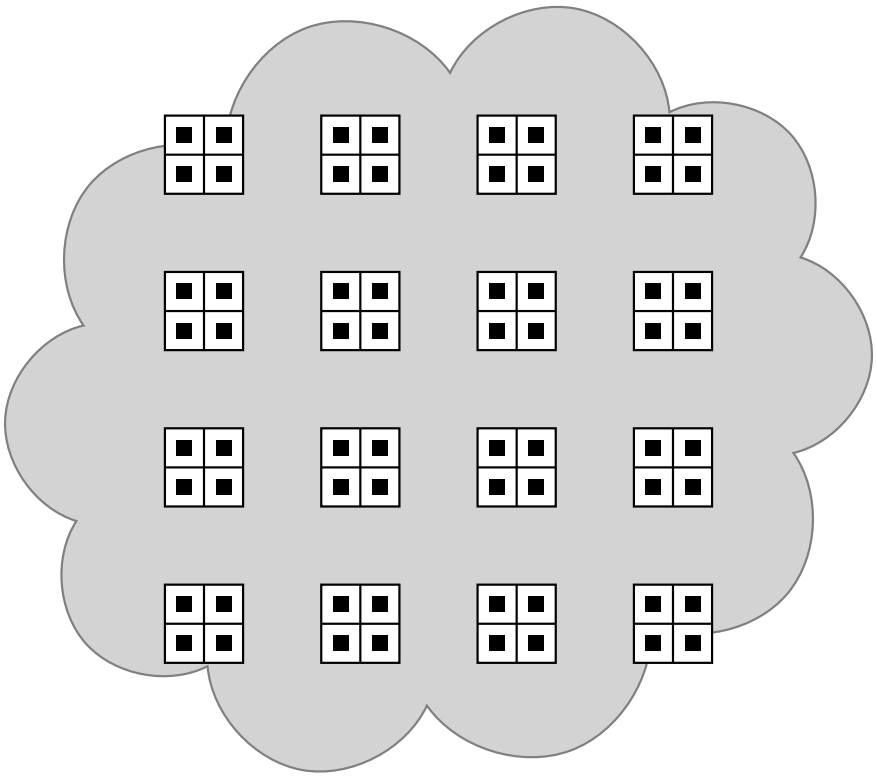
# Principe d'AFS à l'échelle d'un site



# Principe d'AFS à l'échelle d'un site



# Principe d'AFS à l'échelle d'un site



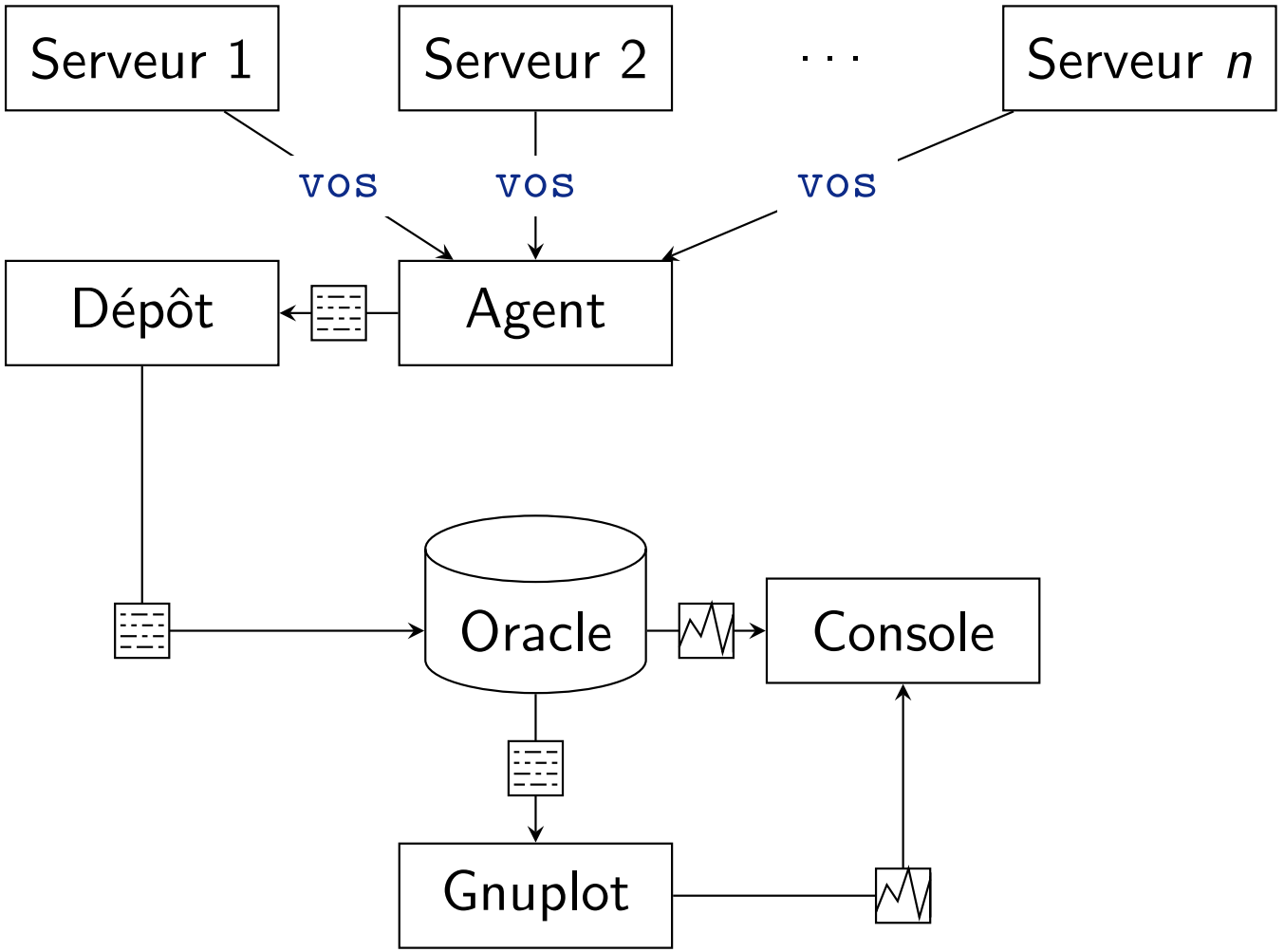
# AFS au CERN

- 40 serveurs
- 112 TB de capacité distribuée
- 25 000 utilisateurs
- 400 millions de fichiers
- 1,5 milliards d'accès par jour

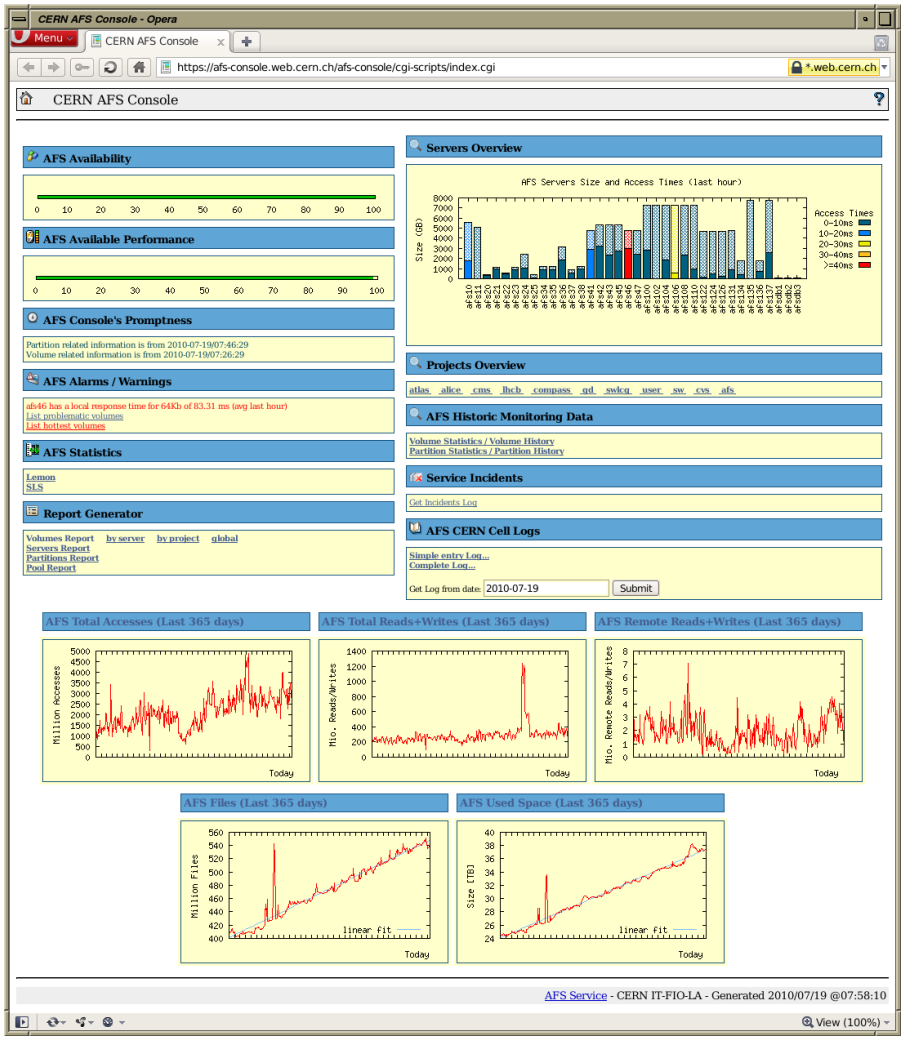
Section 2

# AFS Console

# Architecture d'AFS Console



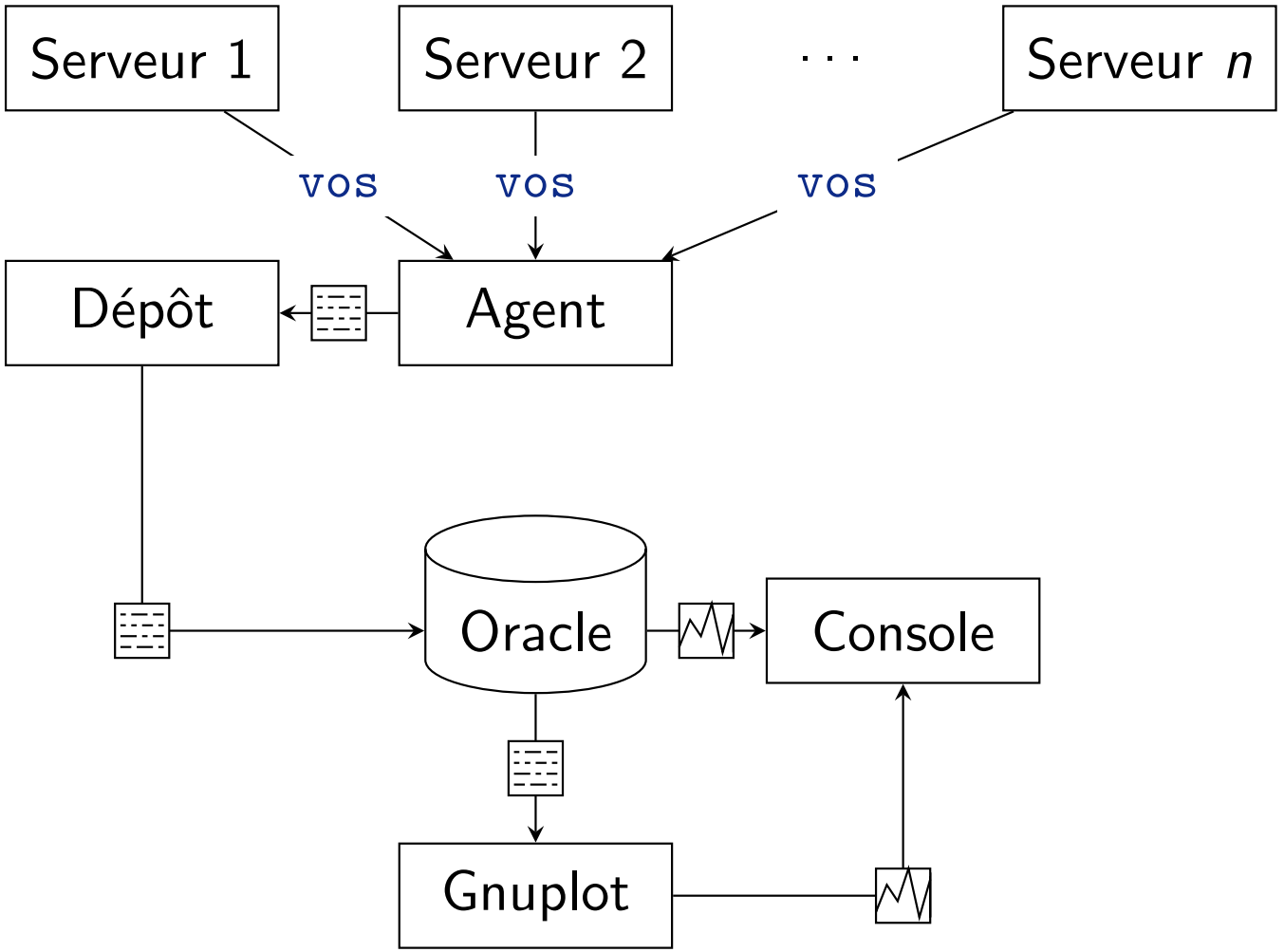
# Interface Web d'AFS Console



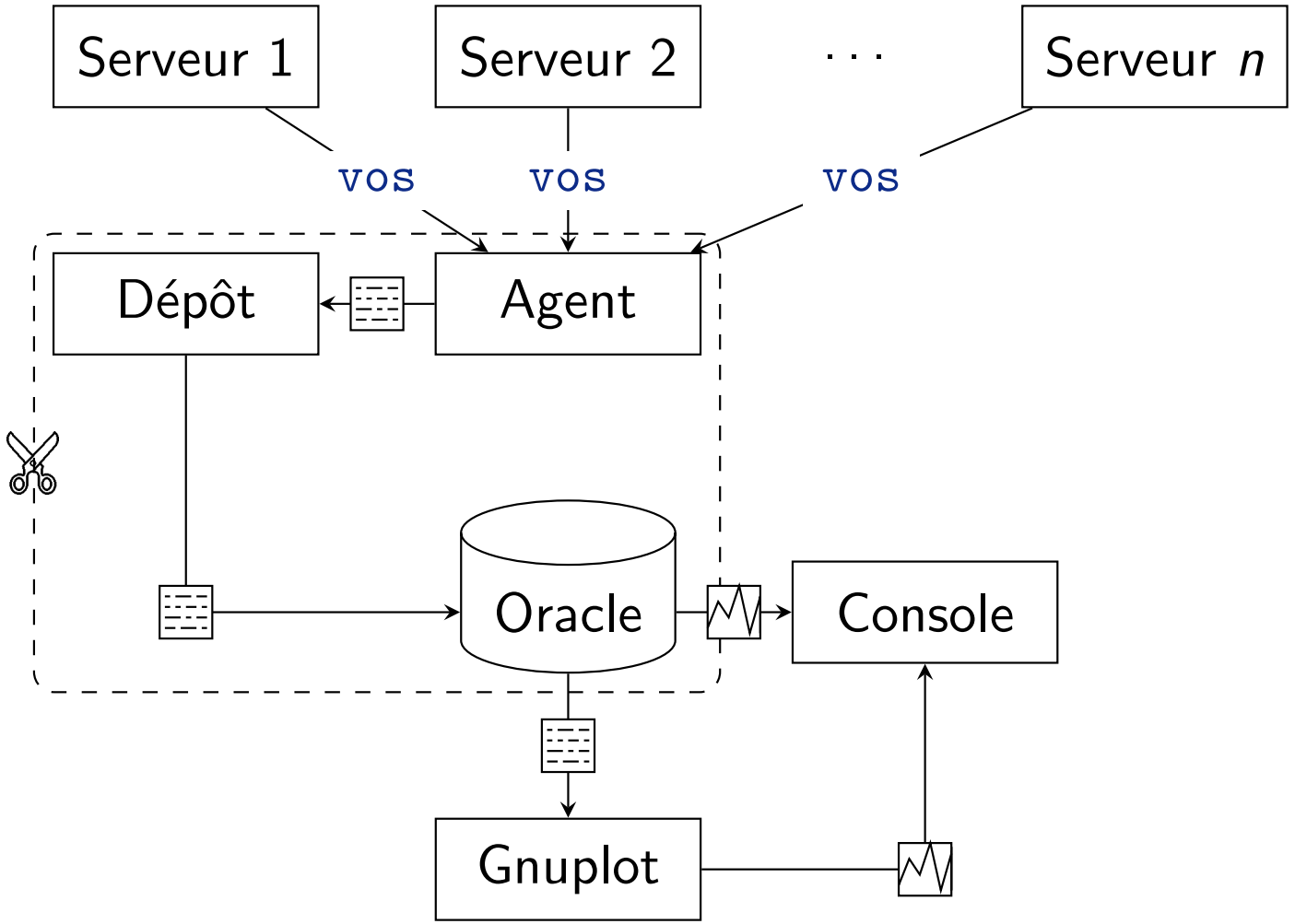
# Rendre AFS Console distribuable

- Indépendance vis-à-vis de Lemon
- Utilisation de n'importe quelle BD
- API générique
- Paquetage
- Performance de la base de données

# Architecture d'AFS Console



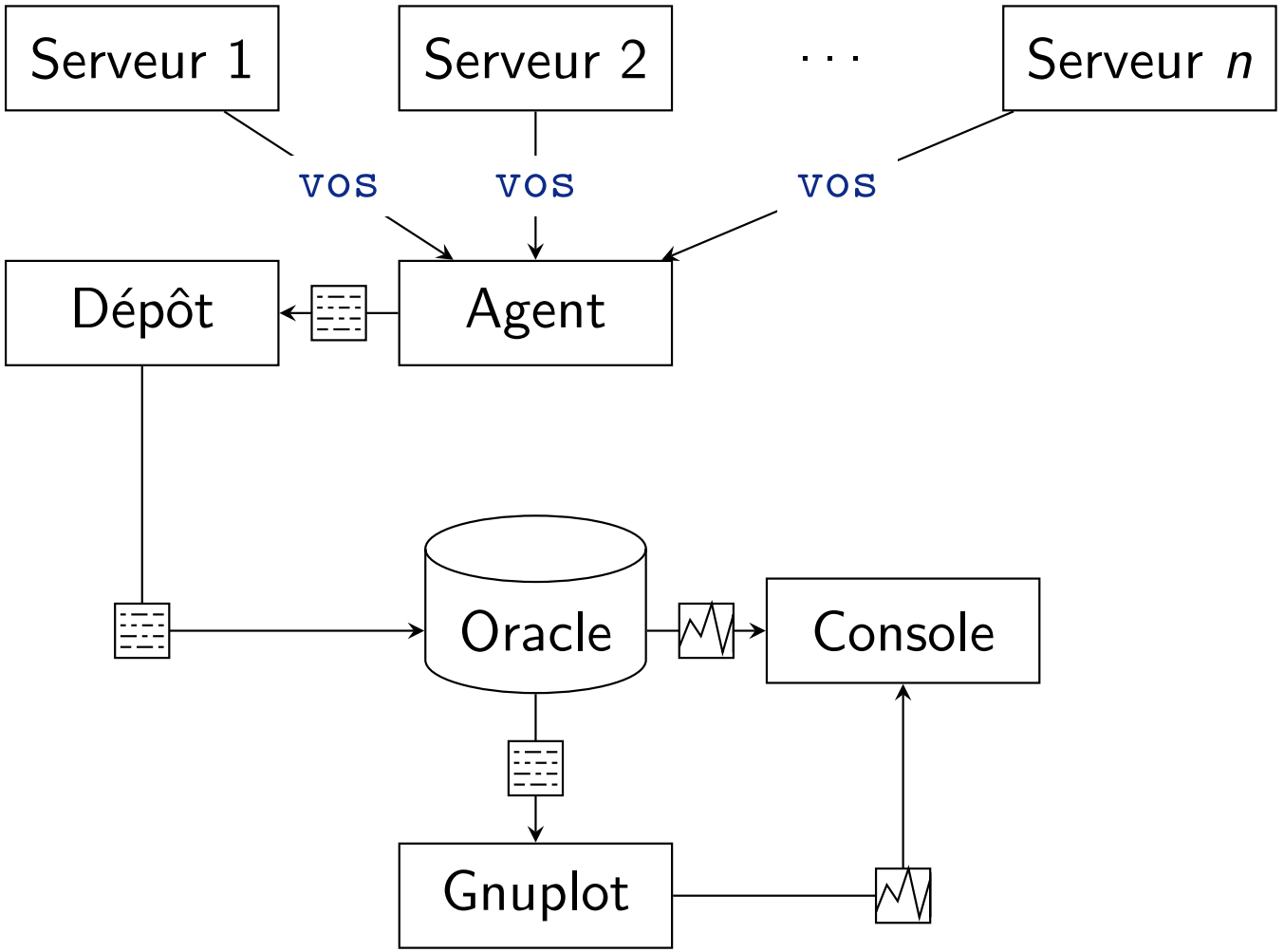
# Architecture d'AFS Console



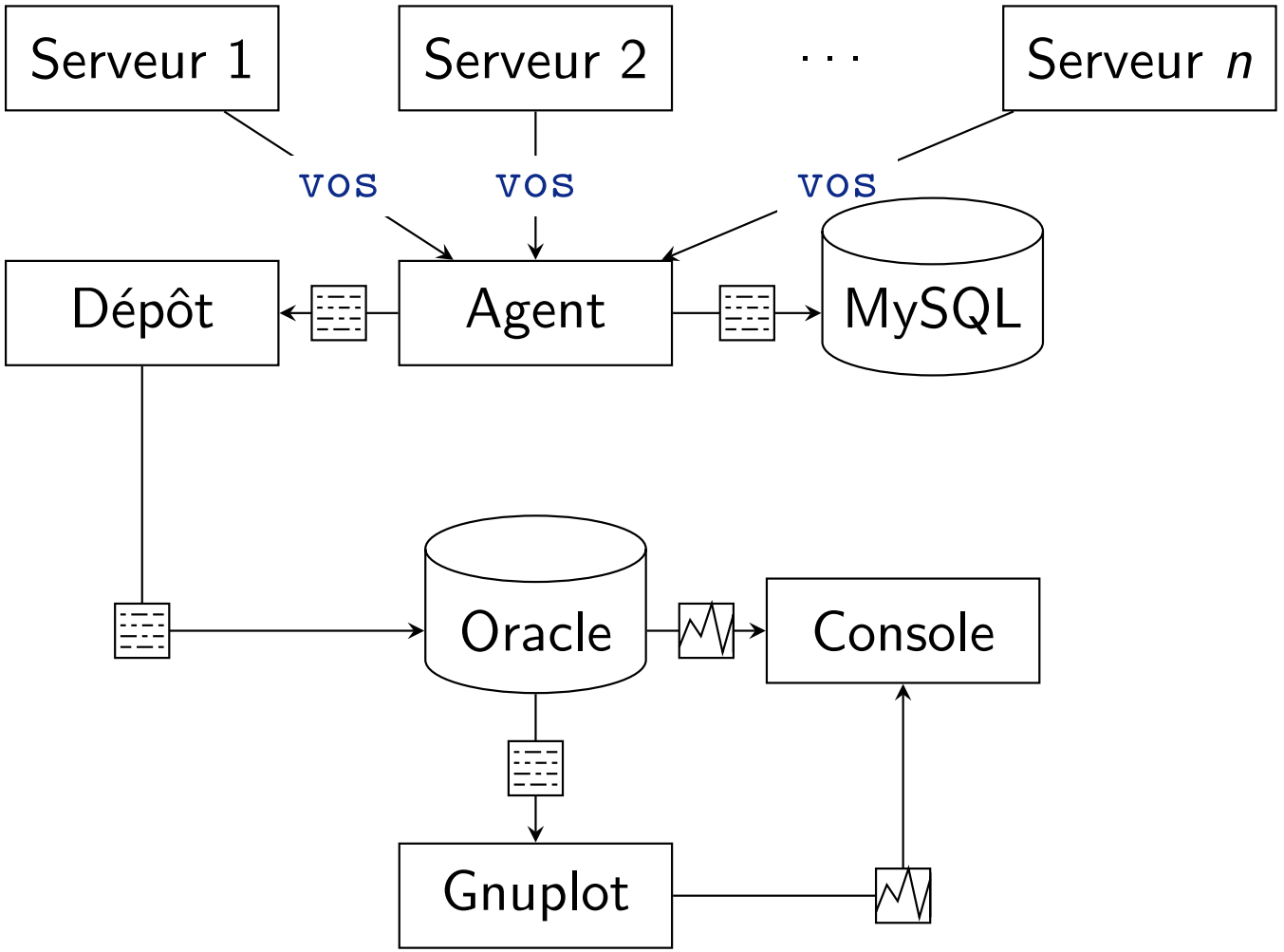
Section 3

# Capteurs

# Architecture d'AFS Console



# Architecture d'AFS Console



# Problématiques

- `fork()`
- Zombies
- Délais d'attente expirés
- Fraîcheur des échantillons

# Insertion de nouveaux échantillons

Toutes les 20 minutes :

1. Insère les nouveaux échantillons
2. Marque les anciens comme "anciens," *le cas échéant*

# Algorithme d'insertion

**Data**: Set $C$ of servers having reported, table $T$ of volume samples
**foreach** Server $S \subset C$ **do**
  **foreach** Volume $v \in S$ **do**
    $T \leftarrow T \cup \{v\}$
  **end**
  $t_{\max} \leftarrow \max\{r(\text{timestamp}), r \in T \wedge r(\text{server}) = S\}$
  **foreach** $r \in T$ **do**
    **if** $r(\text{server}) = S \wedge r(\text{recent}) = \text{True} \wedge r(\text{timestamp}) < t_{\max}$
    **then**
      $r(\text{recent}) \leftarrow \text{False}$
    **end**
  **end**
**end**

# Premier cas : tous les échantillons mis à jour

| Instant | Serveur | Volume | Récent ? |
|---|---|---|---|
| 1270521120 | afs42 | u.atlas | Oui |
| 1270521120 | afs23 | u.alice | Oui |
| 1270521120 | afs47 | u.lhcb | Oui |
| 1270521120 | afs47 | u.cms | Oui |
| 1270521120 | afs42 | u.totem | Oui |

# Premier cas : tous les échantillons mis à jour

| Instant | Serveur | Volume | Récent ? |
|---|---|---|---|
| 1270521120 | afs42 | u.atlas | Oui |
| 1270521120 | afs23 | u.alice | Oui |
| 1270521120 | afs47 | u.lhcb | Oui |
| 1270521120 | afs47 | u.cms | Oui |
| 1270521120 | afs42 | u.totem | Oui |
| 1270522320 | afs42 | u.atlas | Oui |
| 1270522320 | afs23 | u.alice | Oui |
| 1270522320 | afs47 | u.lhcb | Oui |
| 1270522320 | afs47 | u.cms | Oui |
| 1270522320 | afs42 | u.totem | Oui |

# Premier cas : tous les échantillons mis à jour

| Instant | Serveur | Volume | Récent ? |
|---|---|---|---|
| 1270521120 | afs42 | u.atlas | Non |
| 1270521120 | afs23 | u.alice | Non |
| 1270521120 | afs47 | u.lhcb | Non |
| 1270521120 | afs47 | u.cms | Non |
| 1270521120 | afs42 | u.totem | Non |
| 1270522320 | afs42 | u.atlas | Oui |
| 1270522320 | afs23 | u.alice | Oui |
| 1270522320 | afs47 | u.lhcb | Oui |
| 1270522320 | afs47 | u.cms | Oui |
| 1270522320 | afs42 | u.totem | Oui |

# Deuxième cas : serveur manquant au rapport

| Instant | Serveur | Volume | Récent ? |
|---|---|---|---|
| 1270521120 | afs42 | u.atlas | Oui |
| 1270521120 | afs23 | u.alice | Oui |
| 1270521120 | afs47 | u.lhcb | Oui |
| 1270521120 | afs47 | u.cms | Oui |
| 1270521120 | afs42 | u.totem | Oui |

# Deuxième cas : serveur manquant au rapport

| Instant | Serveur | Volume | Récent ? |
|---|---|---|---|
| 1270521120 | afs42 | u.atlas | Oui |
| 1270521120 | afs23 | u.alice | Oui |
| 1270521120 | afs47 | u.lhcb | Oui |
| 1270521120 | afs47 | u.cms | Oui |
| 1270521120 | afs42 | u.totem | Oui |
| 1270522320 | afs42 | u.atlas | Oui |
| 1270522320 | afs23 | u.alice | Oui |
| 1270522320 | afs42 | u.totem | Oui |

# Deuxième cas : serveur manquant au rapport

| Instant | Serveur | Volume | Récent ? |
|---|---|---|---|
| 1270521120 | afs42 | u.atlas | Oui |
| 1270521120 | afs23 | u.alice | Oui |
| 1270521120 | afs47 | u.lhcb | Oui |
| 1270521120 | afs47 | u.cms | Oui |
| 1270521120 | afs42 | u.totem | Oui |
| 1270522320 | afs42 | u.atlas | Oui |
| 1270522320 | afs23 | u.alice | Oui |
| 1270522320 | afs42 | u.totem | Oui |

# Deuxième cas : serveur manquant au rapport

| Instant | Serveur | Volume | Récent ? |
|---|---|---|---|
| 1270521120 | afs42 | u.atlas | Non |
| 1270521120 | afs23 | u.alice | Non |
| 1270521120 | afs47 | u.lhcb | Oui |
| 1270521120 | afs47 | u.cms | Oui |
| 1270521120 | afs42 | u.totem | Non |
| 1270522320 | afs42 | u.atlas | Oui |
| 1270522320 | afs23 | u.alice | Oui |
| 1270522320 | afs42 | u.totem | Oui |

# Troisième cas : volume déplacé/supprimé

| Instant | Serveur | Volume | Récent ? |
|---|---|---|---|
| 1270521120 | afs42 | u.atlas | Oui |
| 1270521120 | afs23 | u.alice | Oui |
| 1270521120 | afs47 | u.lhcb | Oui |
| 1270521120 | afs47 | u.cms | Oui |
| 1270521120 | afs42 | u.totem | Oui |

# Troisième cas : volume déplacé/supprimé

| Instant | Serveur | Volume | Récent ? |
|---|---|---|---|
| 1270521120 | afs42 | u.atlas | Oui |
| 1270521120 | afs23 | u.alice | Oui |
| 1270521120 | afs47 | u.lhcb | Oui |
| 1270521120 | afs47 | u.cms | Oui |
| 1270521120 | afs42 | u.totem | Oui |
| 1270522320 | afs42 | u.atlas | Oui |
| 1270522320 | afs23 | u.alice | Oui |
| 1270522320 | afs47 | u.cms | Oui |
| 1270522320 | afs42 | u.totem | Oui |

# Troisième cas : volume déplacé/supprimé

| Instant | Serveur | Volume | Récent ? |
|---|---|---|---|
| 1270521120 | afs42 | u.atlas | Oui |
| 1270521120 | afs23 | u.alice | Oui |
| 1270521120 | afs47 | u.lhcb | Oui |
| 1270521120 | afs47 | u.cms | Oui |
| 1270521120 | afs42 | u.totem | Oui |
| 1270522320 | afs42 | u.atlas | Oui |
| 1270522320 | afs23 | u.alice | Oui |
| 1270522320 | afs47 | u.cms | Oui |
| 1270522320 | afs42 | u.totem | Oui |

# Troisième cas : volume déplacé/supprimé

| Instant | Serveur | Volume | Récent ? |
|---|---|---|---|
| 1270521120 | afs42 | u.atlas | Non |
| 1270521120 | afs23 | u.alice | Non |
| 1270521120 | afs47 | u.lhcb | Non |
| 1270521120 | afs47 | u.cms | Non |
| 1270521120 | afs42 | u.totem | Non |
| 1270522320 | afs42 | u.atlas | Oui |
| 1270522320 | afs23 | u.alice | Oui |
| 1270522320 | afs47 | u.cms | Oui |
| 1270522320 | afs42 | u.totem | Oui |

# Capteurs autonomes

- Scripts Perl indépendants
- Logique du capteur Lemon
- Patch du capteur Lemon

Section 4

# Base de données

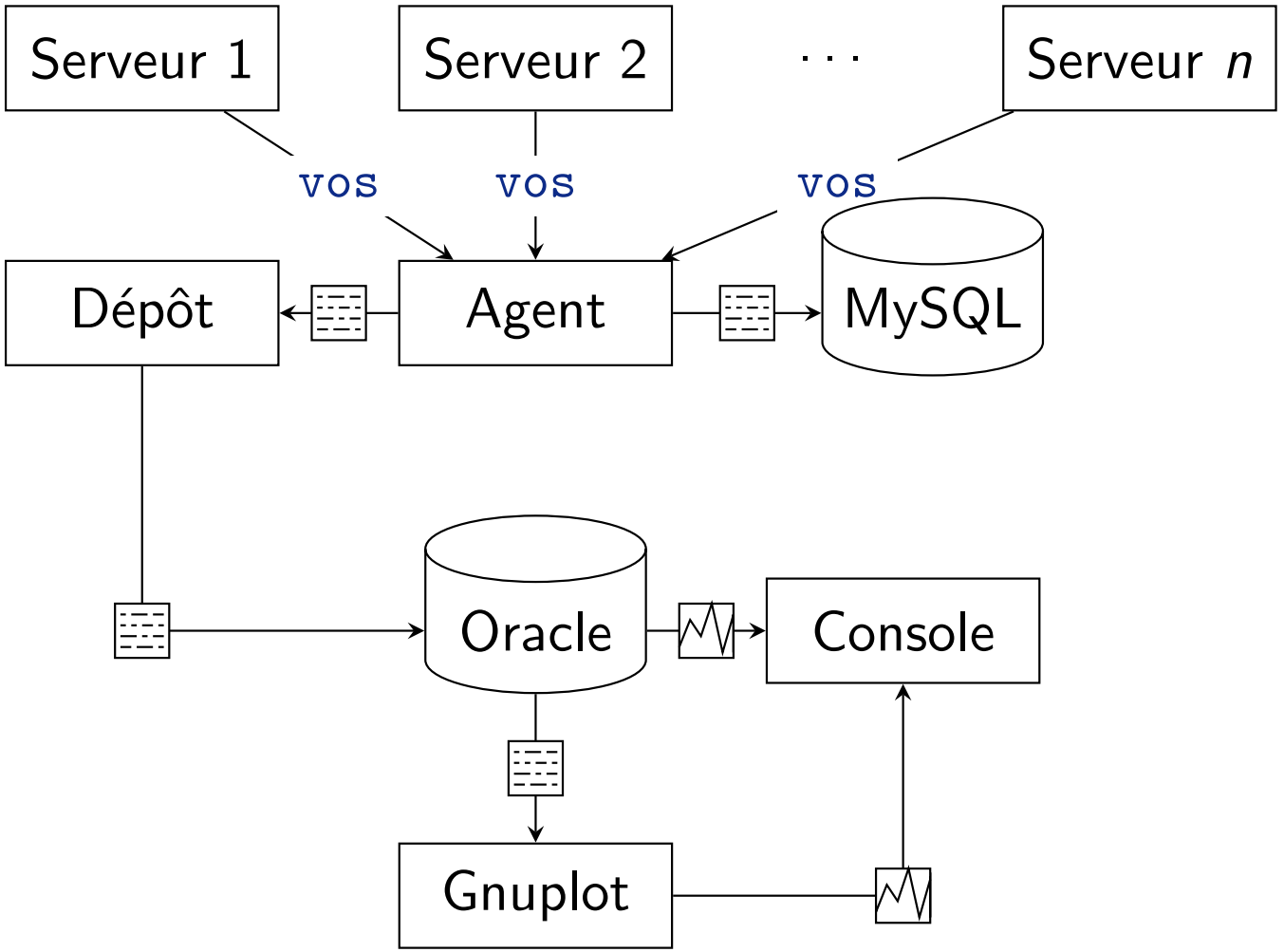
# Enjeux

- Indépendance vis-à-vis d'Oracle
- Plus grande liberté de développement → Performances accrues

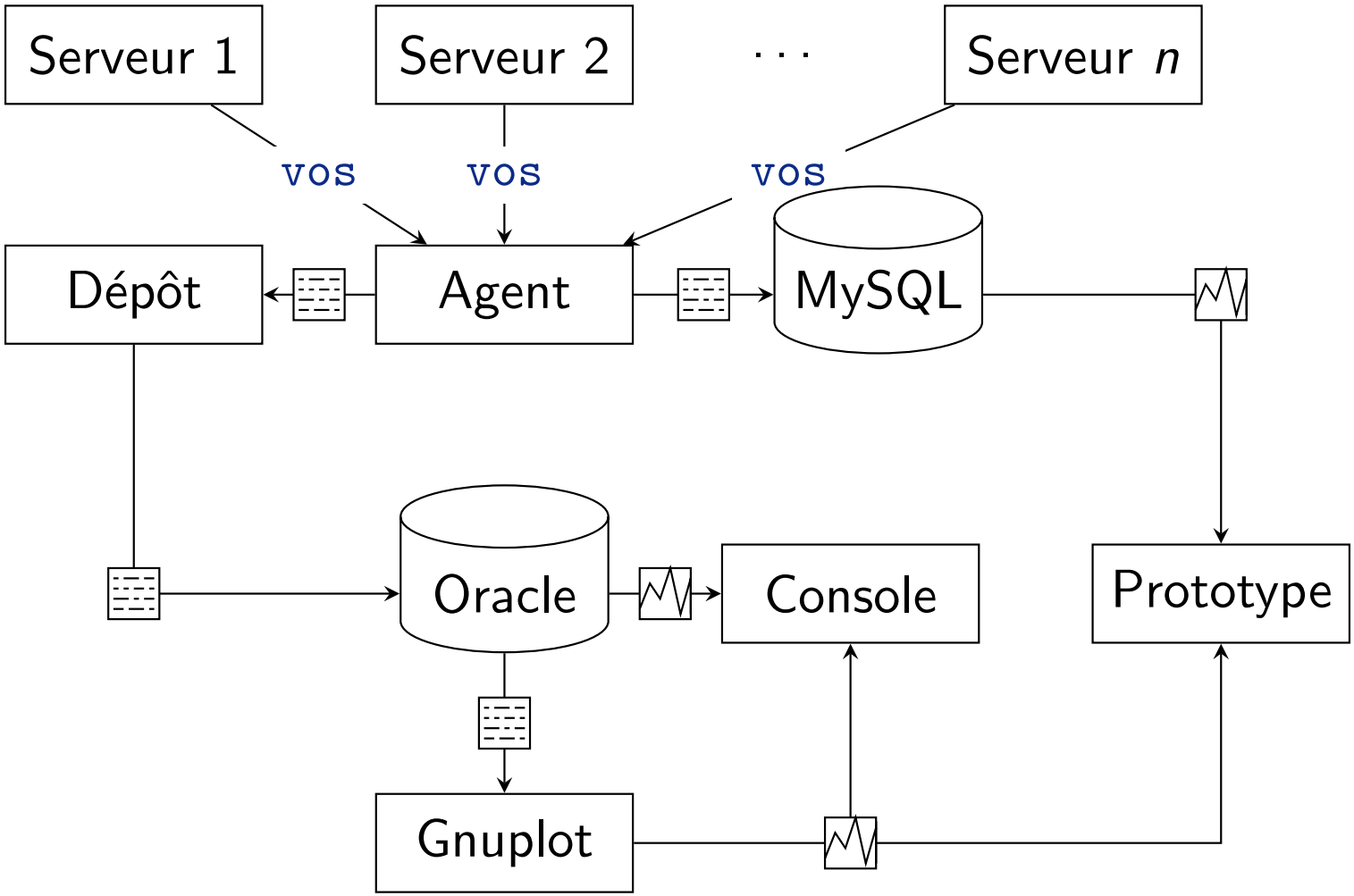
# Performance

- Vitesse en consultation
- Vitesse lors de l'ajout de nouveaux échantillons
- Vitesse de maintenance
- Cohabitation des 3 précédents besoins

# Architecture d'AFS Console



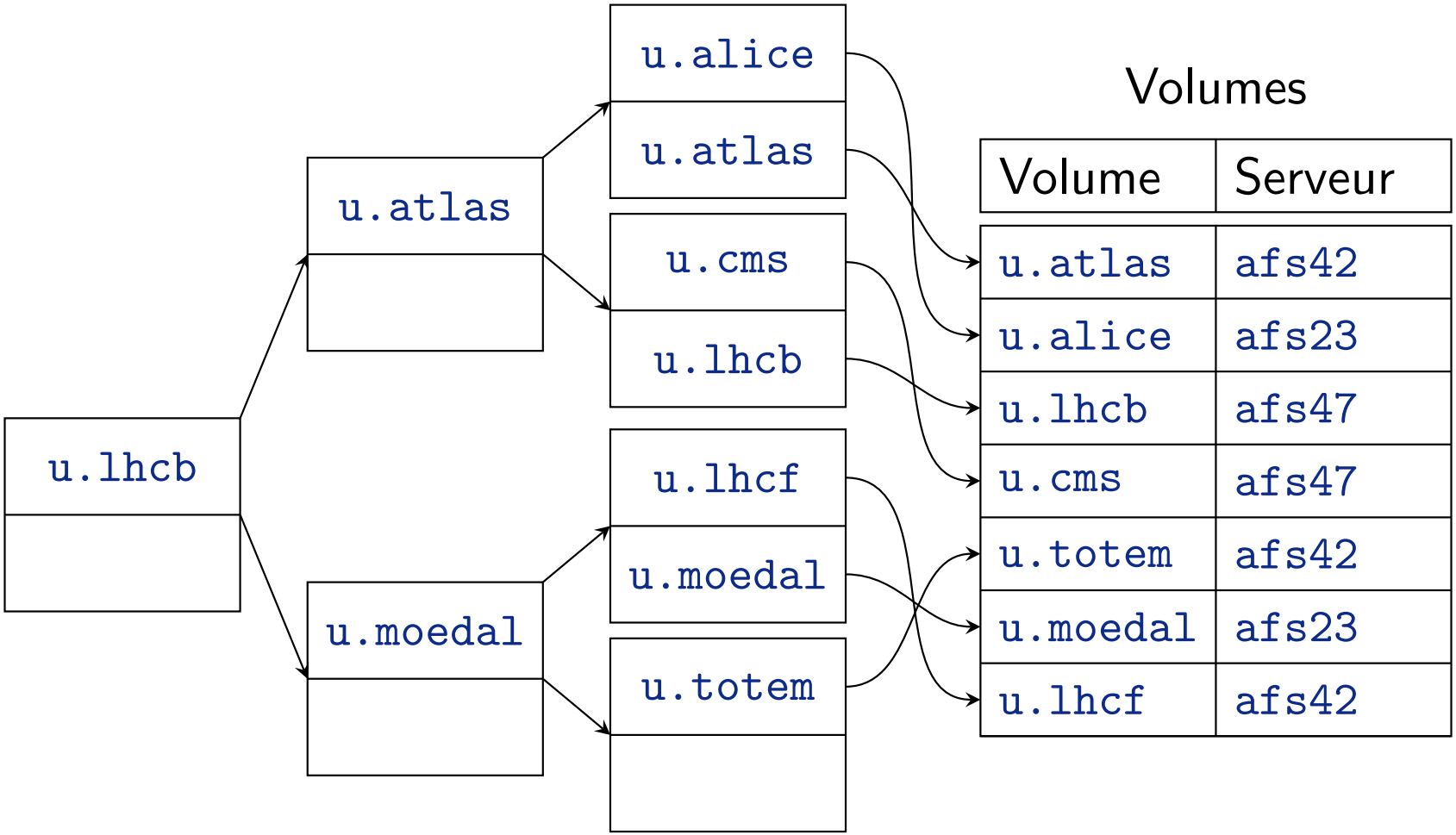
# Architecture d'AFS Console



# Pistes d'optimisations

- Optimisation des types
- Moteurs de stockage
- Configuration du serveur
- Contrôle sur la taille des tables
- Indexes

# Indexes


u.lhcb
u.atlas
u.moedal
u.alice
u.atlas
u.cms
u.lhcb
u.lhcf
u.moedal
u.totem
Volumes
Volume
Serveur
u.atlas
afs42
u.alice
afs23
u.lhcb
afs47
u.cms
afs47
u.totem
afs42
u.moedal
afs23
u.lhcf
afs42

# Indexage efficace

1. Identifier toutes les requêtes de l'application.
2. Définir un index pour chacune d'elles, en précisant éventuellement l'ordre des colonnes.
3. Pour les indexes où l'ordre des colonnes n'est pas important, arranger cette ordre afin d'essayer d'obtenir un index qu'on a déjà.
4. Supprimer les doublons.

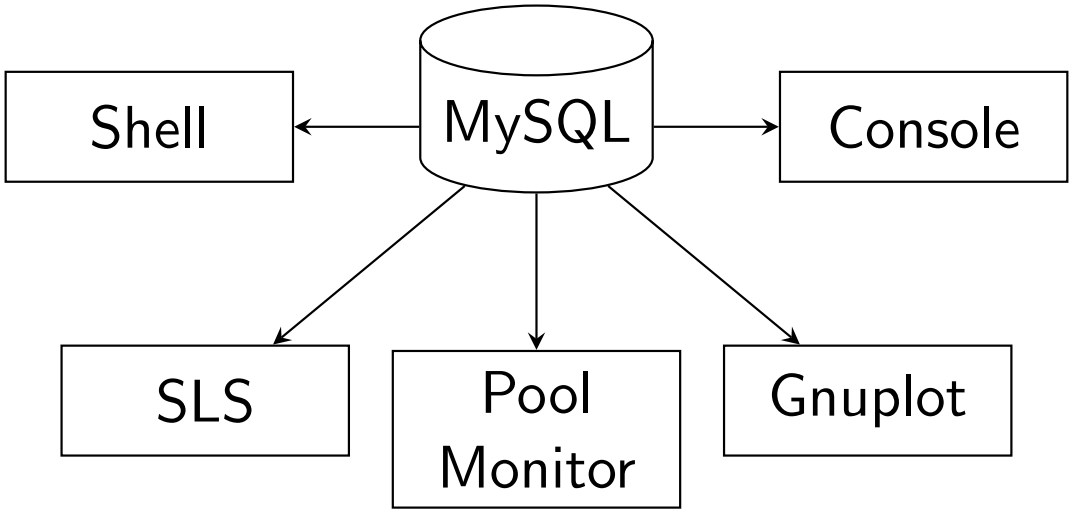
Section 5

# API

# Objectifs

- Interface générique à la BD
- Éviter que le code se répète
- Simplifier l'utilisation et le développement
- Encourager de nouveaux utilisateurs

# Utilisateurs

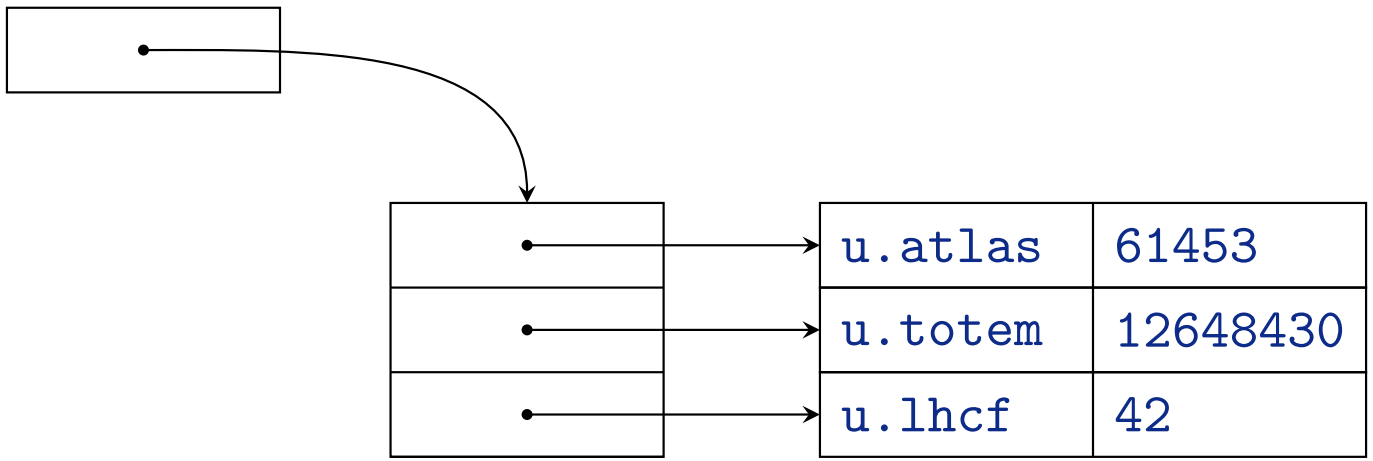


# Interface de bas niveau

```
sub volumes($cols, $srvprt, $vol, $proj,
            $timespan, $orderby, $rowc)
sub partitions($cols, $srvprt, $timespan,
               $orderby, $rowc)
```

```
my $vols = volumes('%volumename/max(size)',
                   'afs42/b', undef, undef,
                   '1275618720/1275705120',
                   undef, 3);
```

# Structure retournée



# Interface de haut niveau

```
sub volumes_lastupdate()
sub partitions_hottest($_srvprt, $_timespan,
                       $_orderby, $_rowc)
sub volumes_volxsltr($_ref)
sub partitions_plot($_srvprt, $_timespan)
...
```

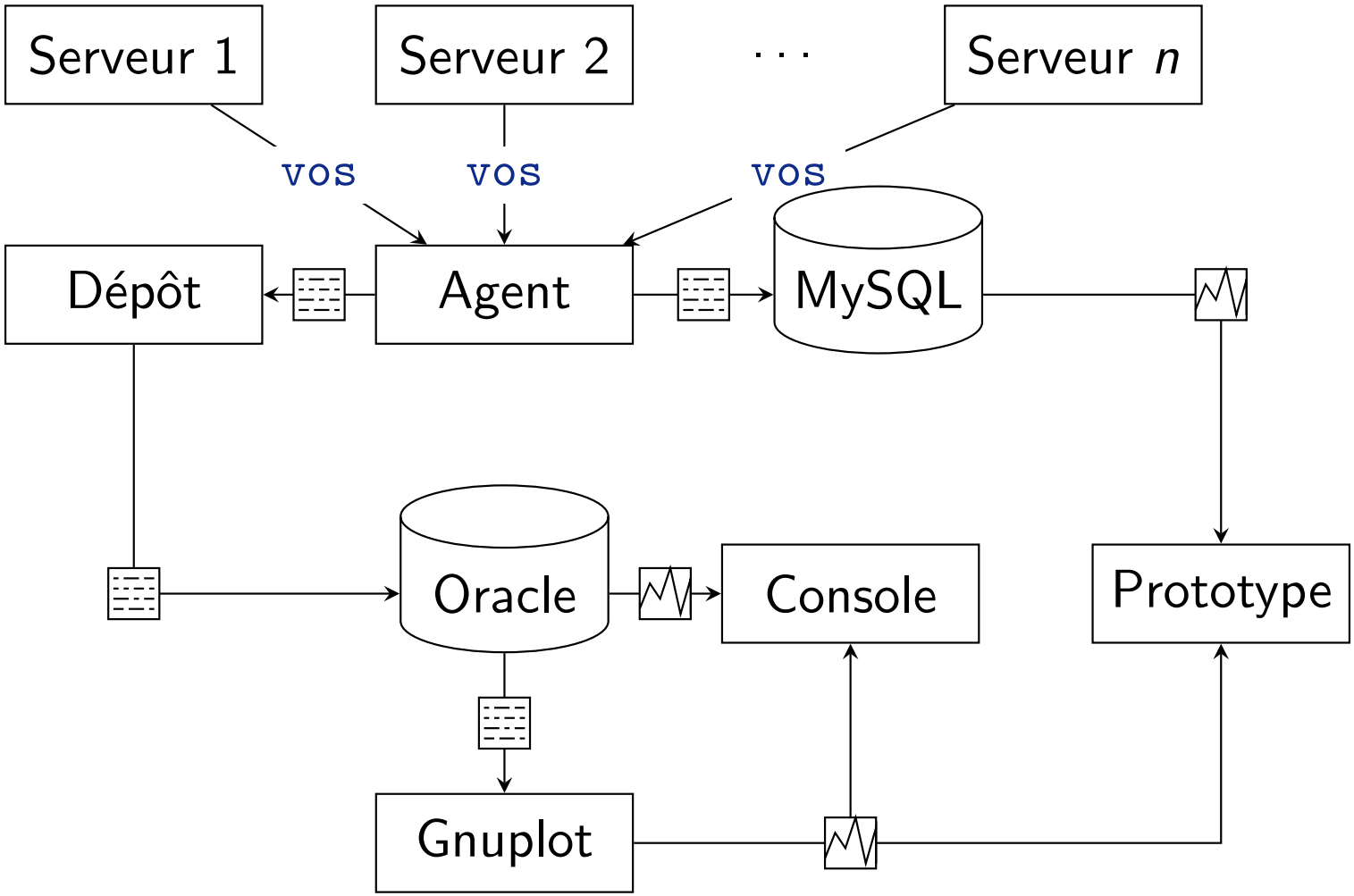
Section 6

# Production

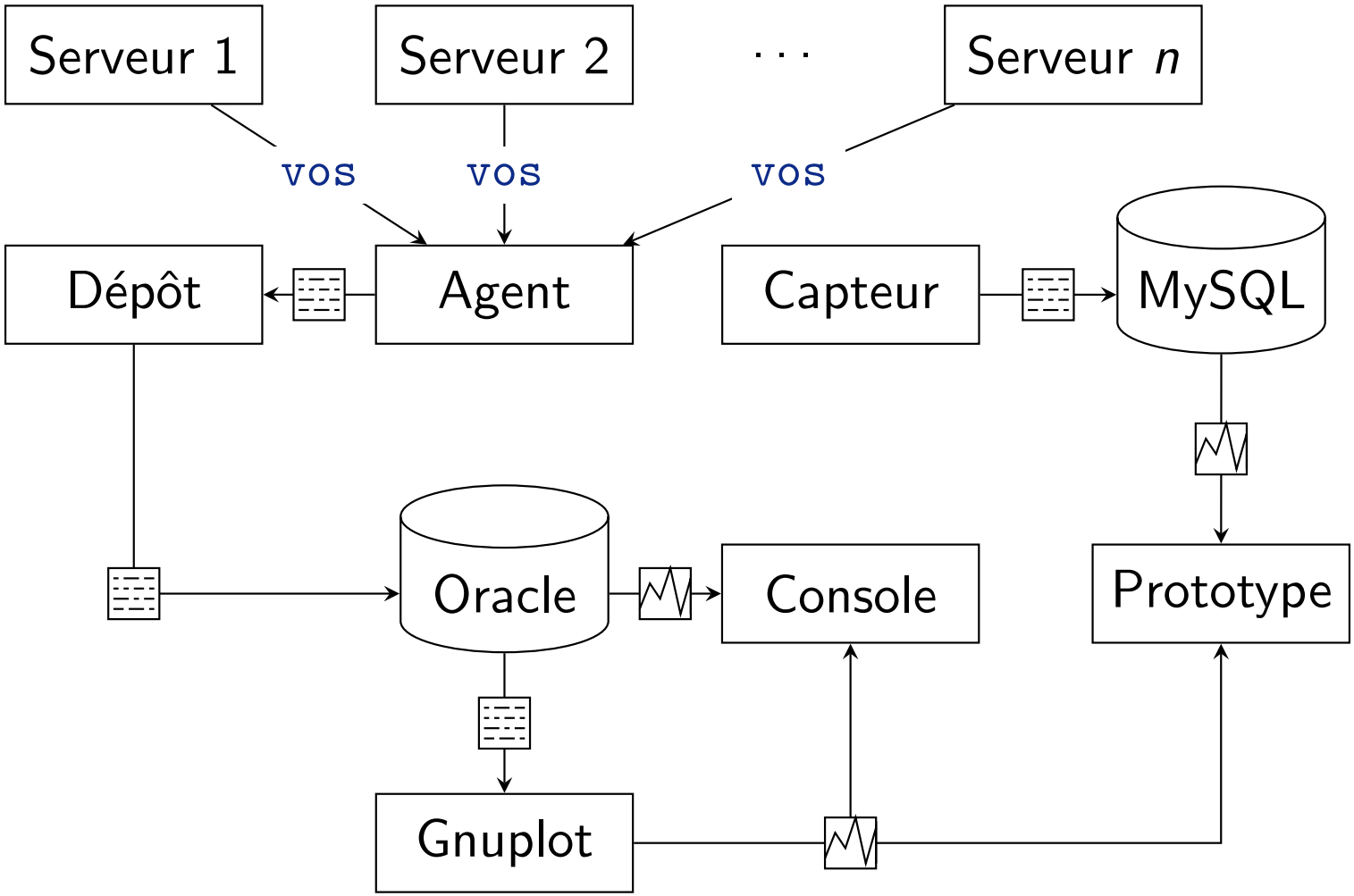
# Mise en production

- Définir un plan de mise en production
- Publier le plan de mise en production

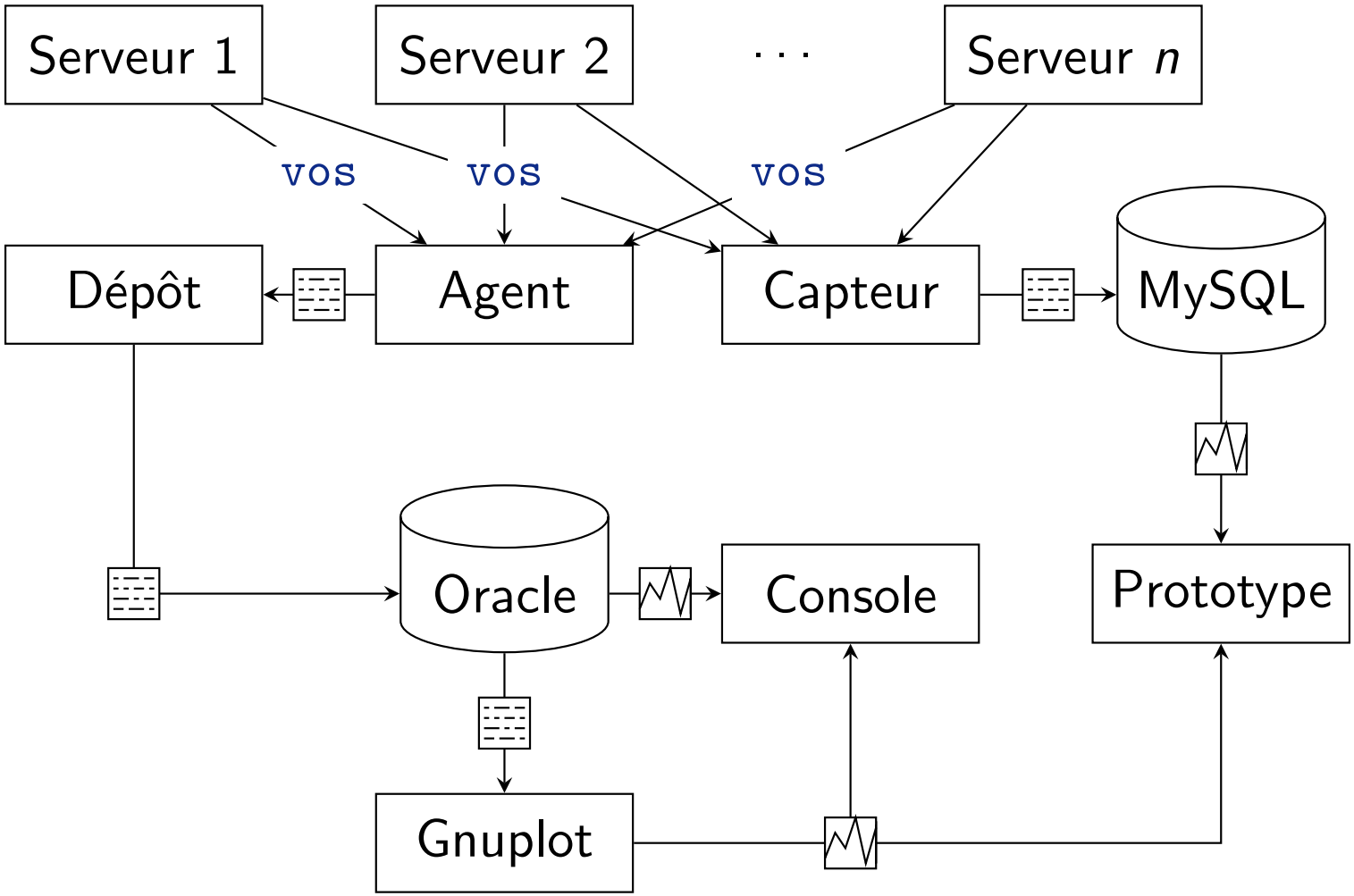
# Architecture d'AFS Console



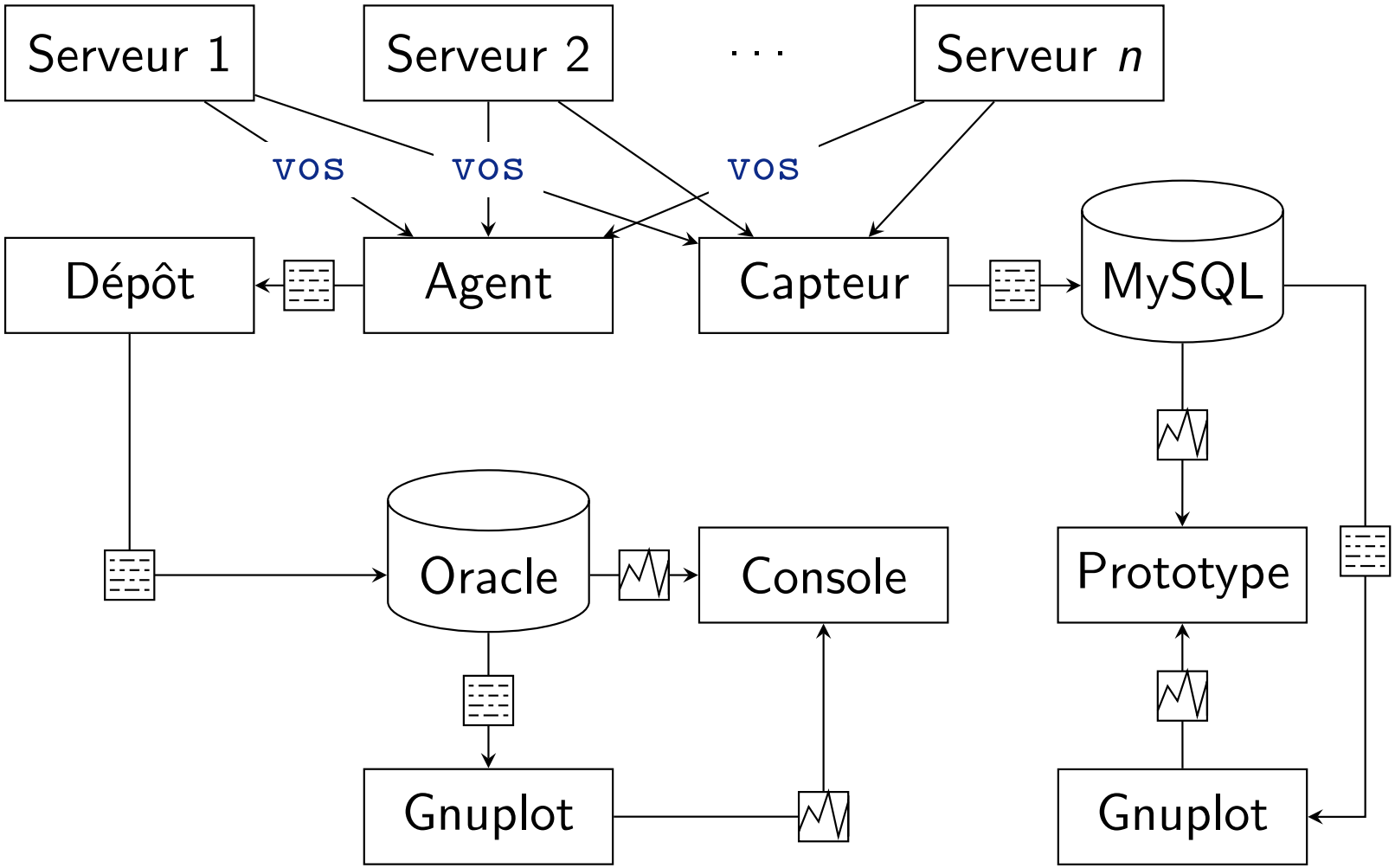
# Architecture d'AFS Console



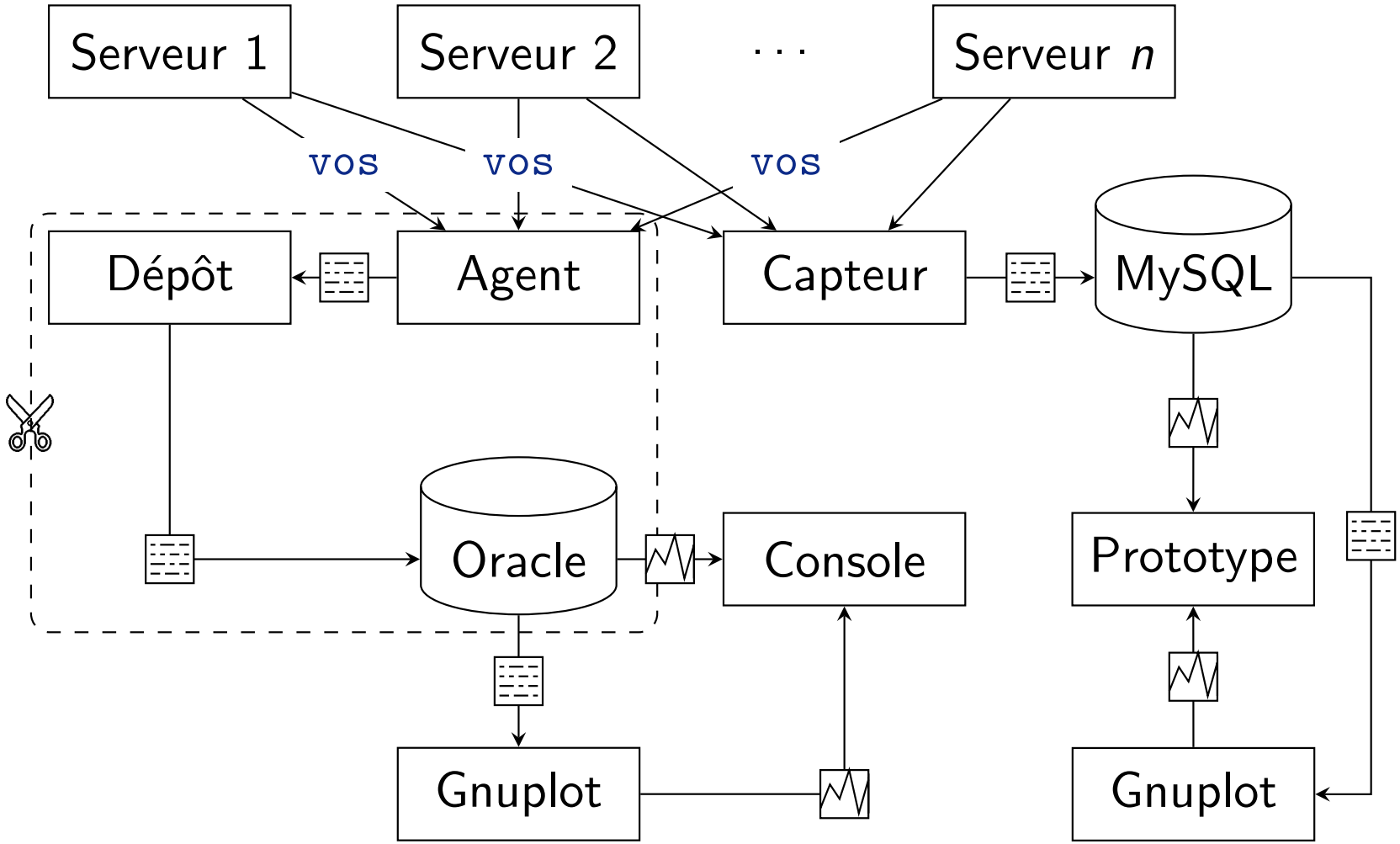
# Architecture d'AFS Console



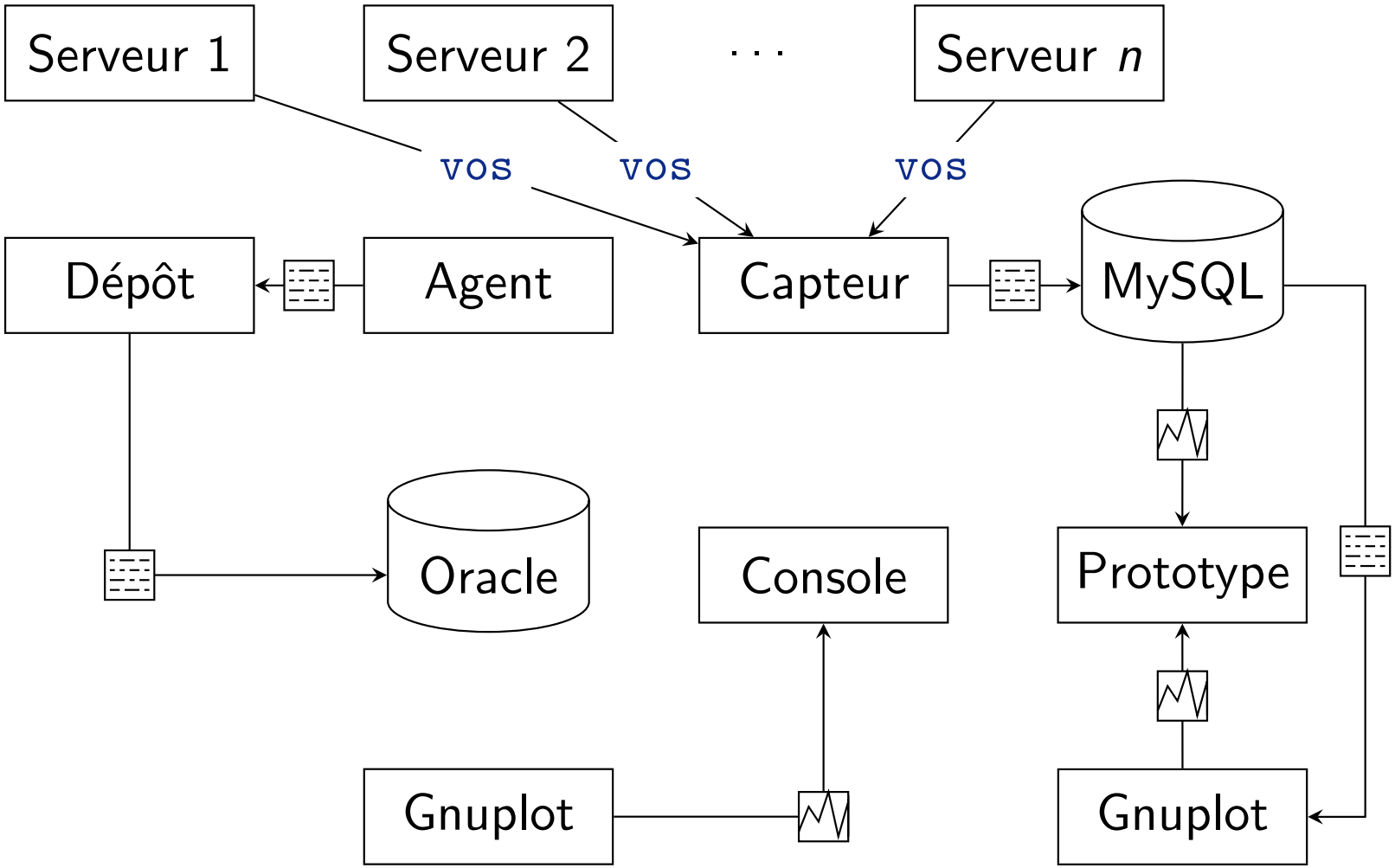
# Architecture d'AFS Console



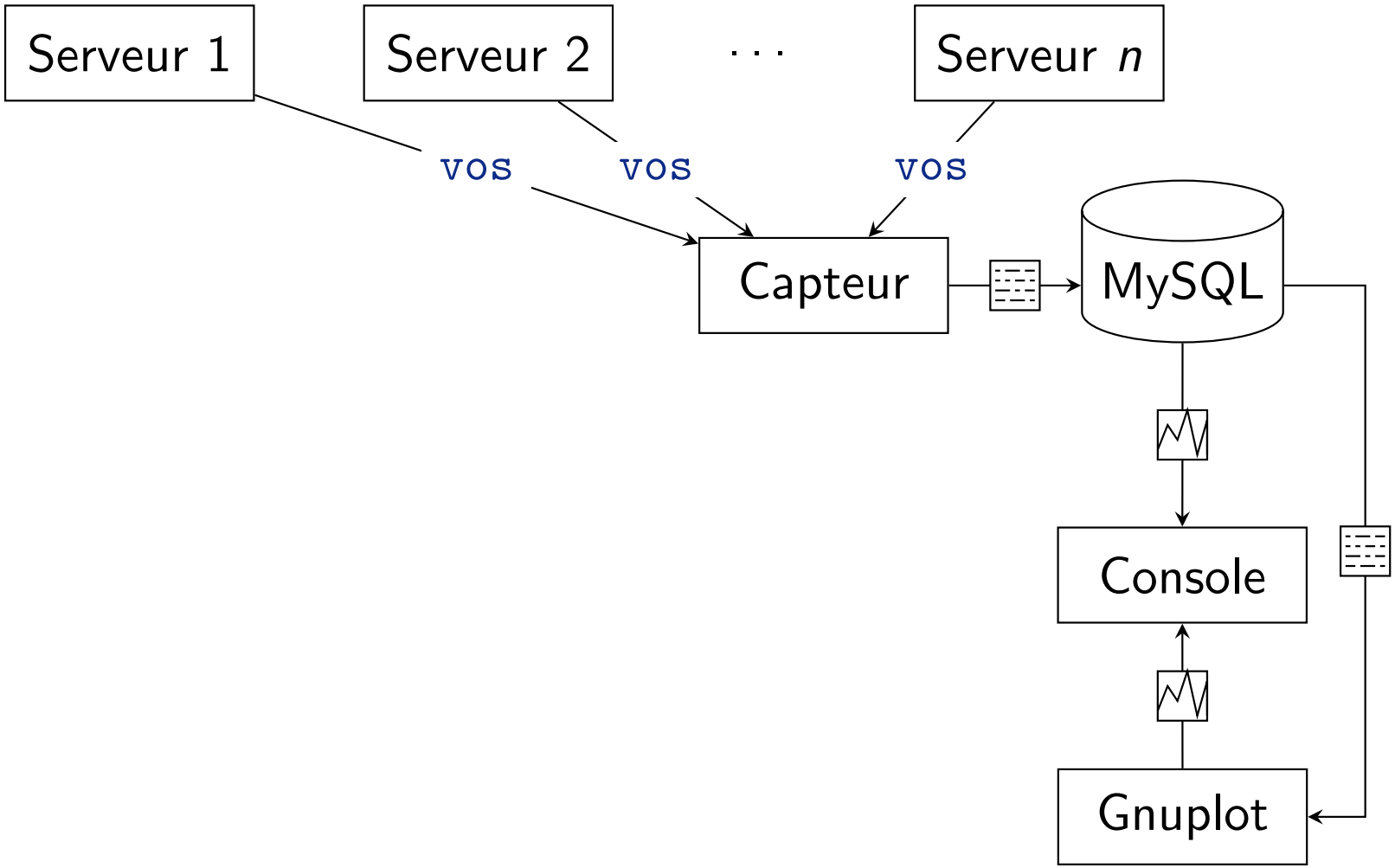
# Architecture d'AFS Console



# Architecture d'AFS Console



# Architecture d'AFS Console



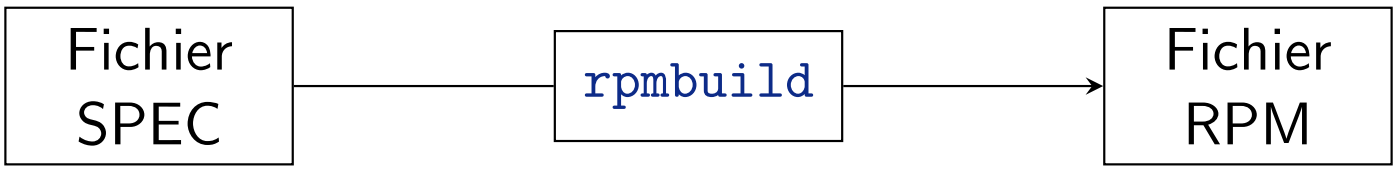
Section 7

# Distribution

# Besoins

- Distribution au CERN
- Distribution dans le monde
- Configuration semi-automatique

# RPM



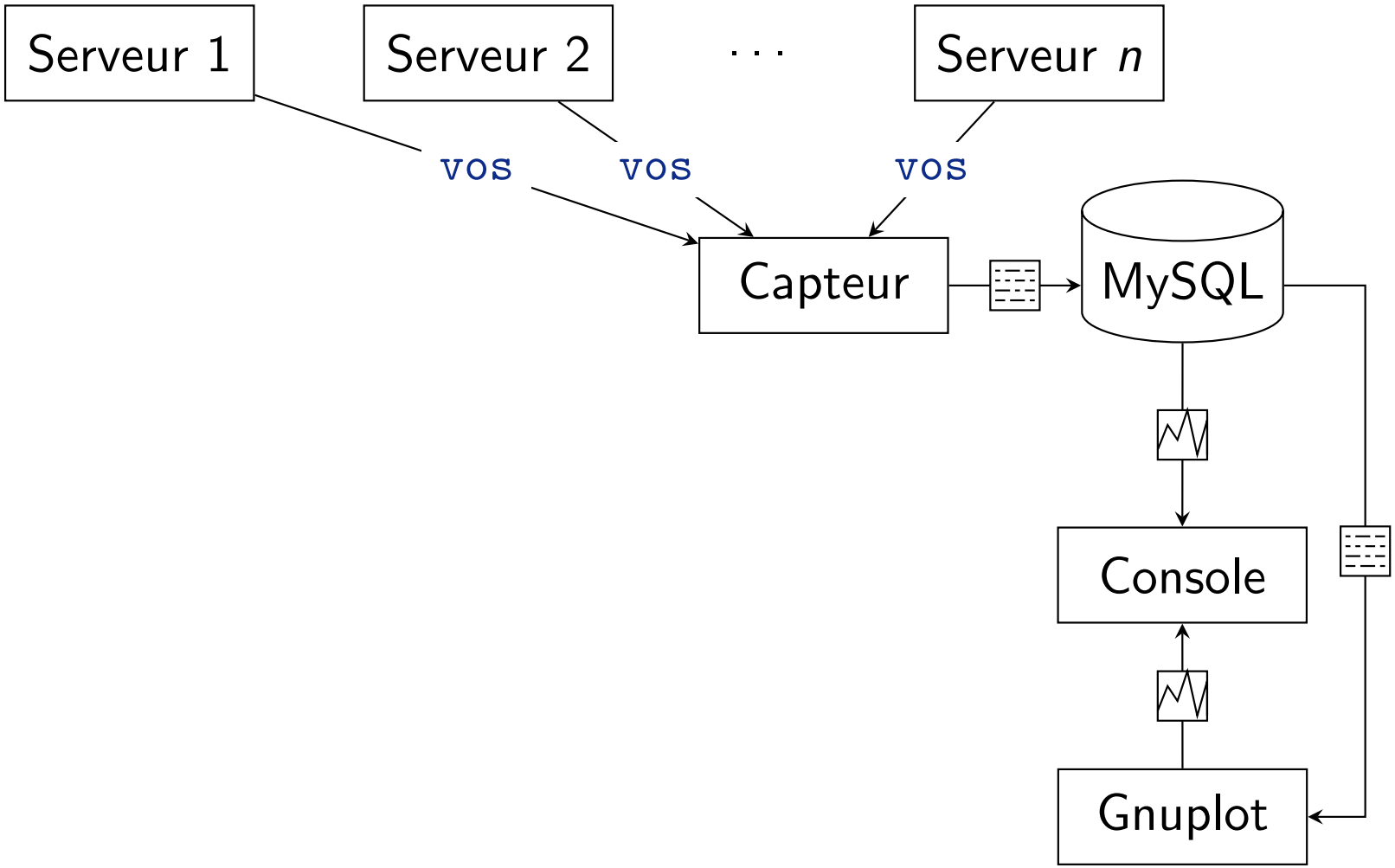
# Configuration post-installation

- Configuration de la base de données
- Compilation des bibliothèques
- Accès aux serveurs AFS du site
- Paramétrage des logs
- ...

Section 8

# Bilan

# Architecture d'AFS Console



# Rendre AFS Console distribuable

- Indépendance vis-à-vis de Lemon
- Utilisation de n'importe quelle BD
- API générique
- Paquetage
- Performance de la base de données

# Rendre AFS Console distribuable

- ☑ Indépendance vis-à-vis de Lemon
- ☐ Utilisation de n'importe quelle BD
- ☐ API générique
- ☐ Paquetage
- ☐ Performance de la base de données

# Rendre AFS Console distribuable

- ☑ Indépendance vis-à-vis de Lemon
- ☑ Utilisation de n'importe quelle BD
- ☐ API générique
- ☐ Paquetage
- ☐ Performance de la base de données

# Rendre AFS Console distribuable

- ☑ Indépendance vis-à-vis de Lemon
- ☑ Utilisation de n'importe quelle BD
- ☑ API générique
- ☐ Paquetage
- ☐ Performance de la base de données

# Rendre AFS Console distribuable

- ☑ Indépendance vis-à-vis de Lemon
- ☑ Utilisation de n'importe quelle BD
- ☑ API générique
- ☑ Paquetage
- ☐ Performance de la base de données

# Rendre AFS Console distribuable

- ☑ Indépendance vis-à-vis de Lemon
- ☑ Utilisation de n'importe quelle BD
- ☑ API générique
- ☑ Paquetage
- ☑ Performance de la base de données

# Perspectives

- Support d'autres SGBD
- Essor de l'API
- Nouvelle interface Web

# Connaissances acquises

- AFS
- Perl
- MySQL : configuration, optimisation, UDF
- Architecture d'une API
- Concepts de monitoring

# Expérience personnelle

- Relations professionnelles
- Organisation du travail
- Cours et présentations
- Candidature pour un poste
- Richesse internationale

# Références


Robert D. Schneider
*MySQL Database Design and Tuning*
MySQL Press, 2005


Baron Schwartz, Peter Zaitsev, Vadim Tkachenko, Jeremy D. Zawodny, Arjen Lentz & Derek J. Balling
*High Performance MySQL*
O'Reilly & Associates, 2008

# Questions?