

4. Gestion des données urbaines dans les nuages informatiques

- Brève histoire des nuages informatiques
- Modèles de service et de déploiement
- Technologie clé : la virtualisation
- IaaS : les points de vue utilisateur et fournisseur
- PaaS : programmation et déploiement des applications
- **Stockage de données**
- Traitement de données
- Traitement de flux de données

Stockage de données dans le cloud



Transfert de données



Stockage temporaire ou persistant
de données

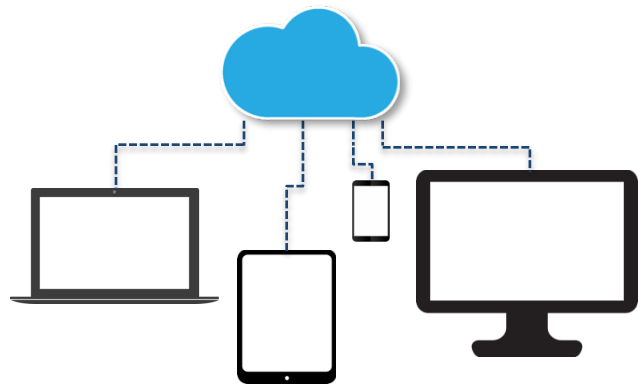
Stockage en libre service (SaaS)

Utilisateur final

- Synchronisation de répertoires de données accédées depuis plusieurs machines
- Copie de sauvegarde de données (backup)
- Partage de données (travail collaboratif, public)

Développeur d'applications

- API publique



Stockage en libre service (SaaS)

Offres

- Taille de l'espace de stockage gratuit
- Coût mensuel du stockage additionnel
- Taille limite des fichiers (2-100Go)
- Système d'exploitation supporté sur PC (OS X, Linux, Windows)
- Support des téléphones mobiles Android, IOS, Windows Phone (téléchargement, streaming)

Sécurité

- Protection des fichiers par mot de passe
- Chiffrement des données



Stockage de données dans le cloud - infrastructure

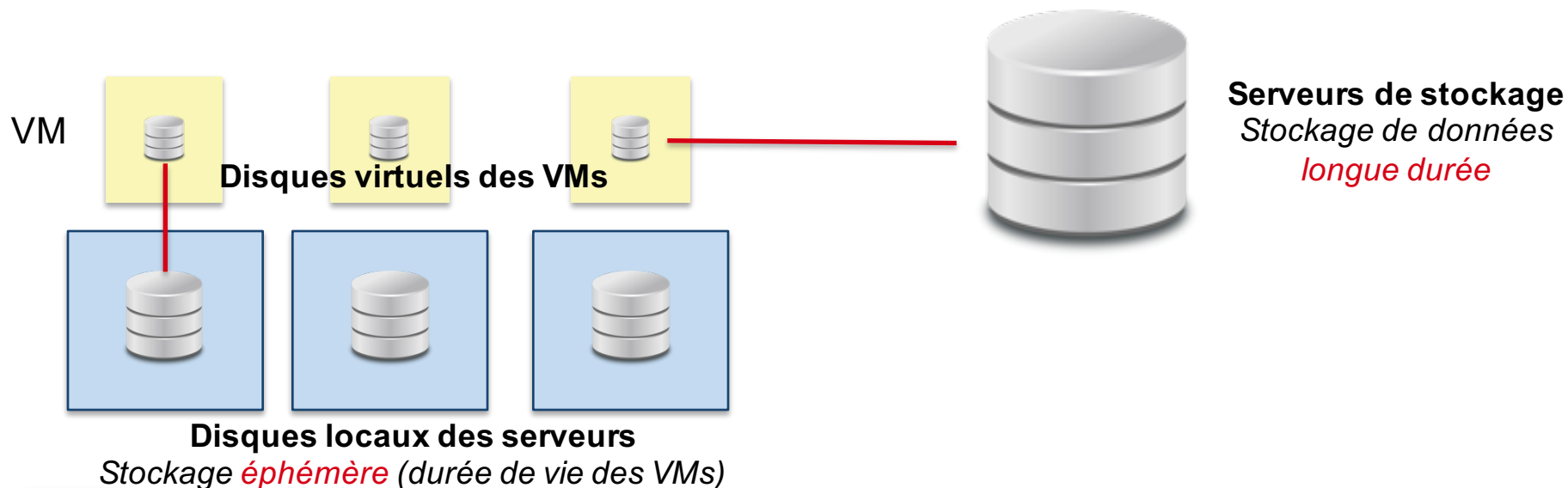
Deux types de données



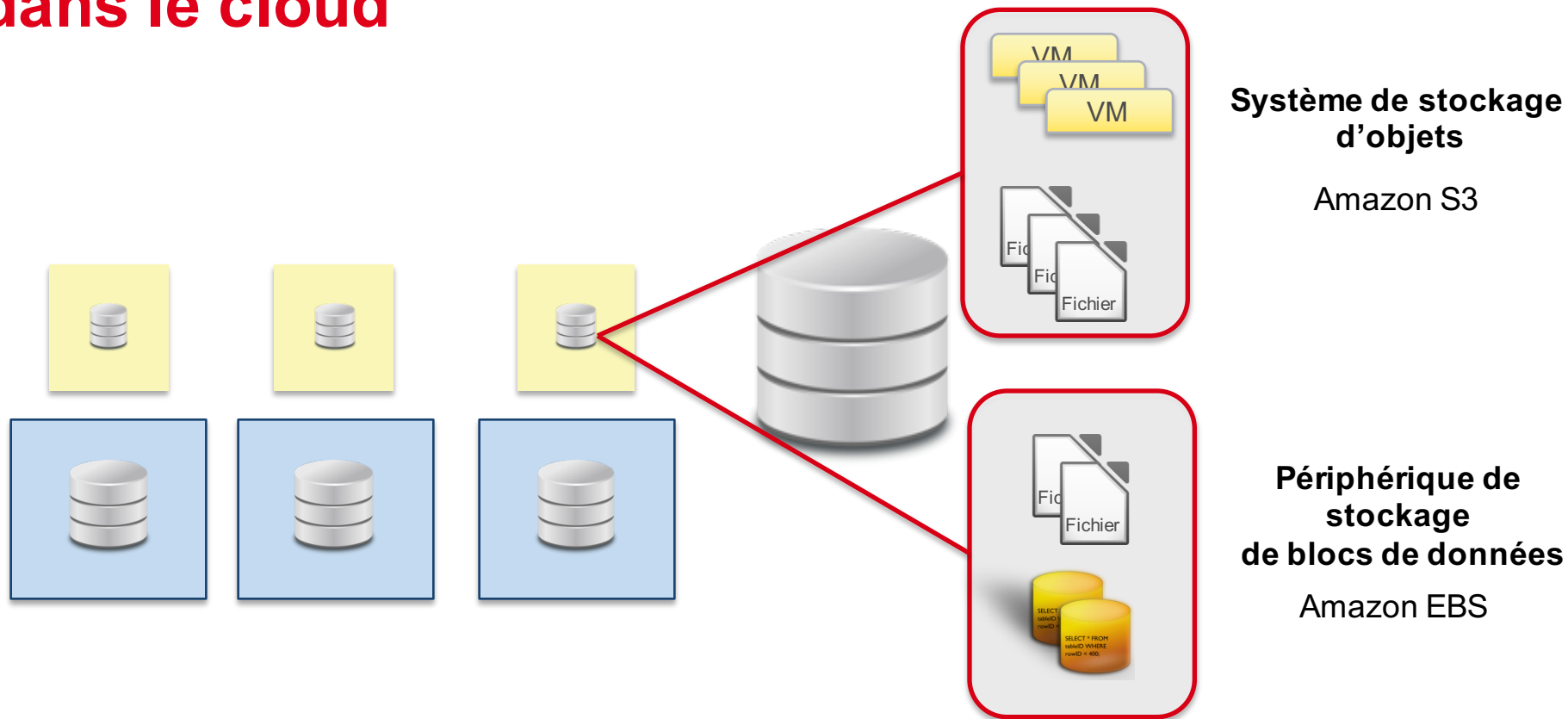
Images des machines virtuelles (VM)



Données manipulées par les VMs



Services de stockage de données persistantes dans le cloud



Focus sur Amazon S3 et EBS

- **EBS**

- **Périphérique disque extensible** pour les machines virtuelles
- Réplication sur différents périphériques en option pour la fiabilité
- Différents supports : SSD, disque magnétique

- **S3**

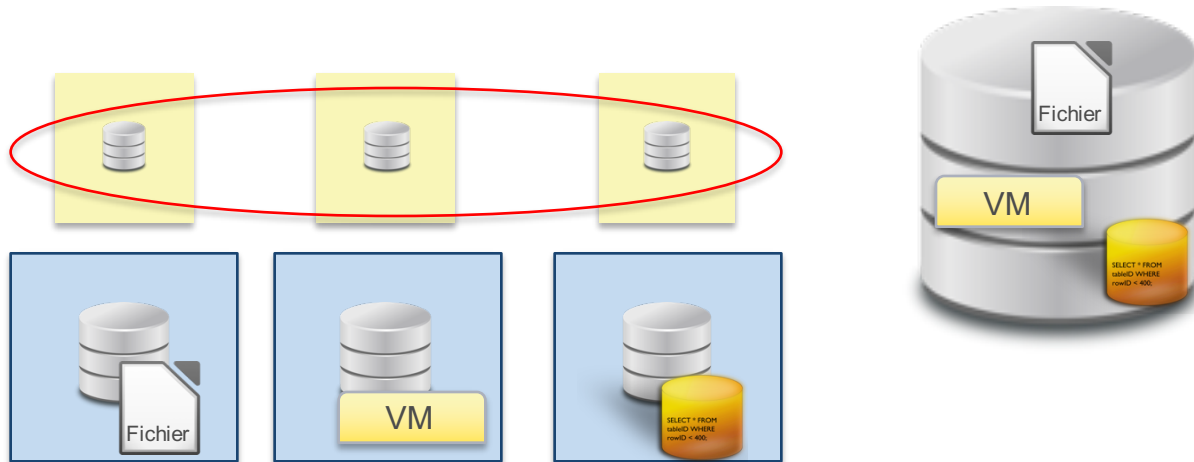
- **Stockage d'objets**
- Persistance et haute disponibilité
 - ✓ Stockage dans plusieurs zones de disponibilité
- Interface web

Déploiement de systèmes de fichiers dans les clusters virtuels

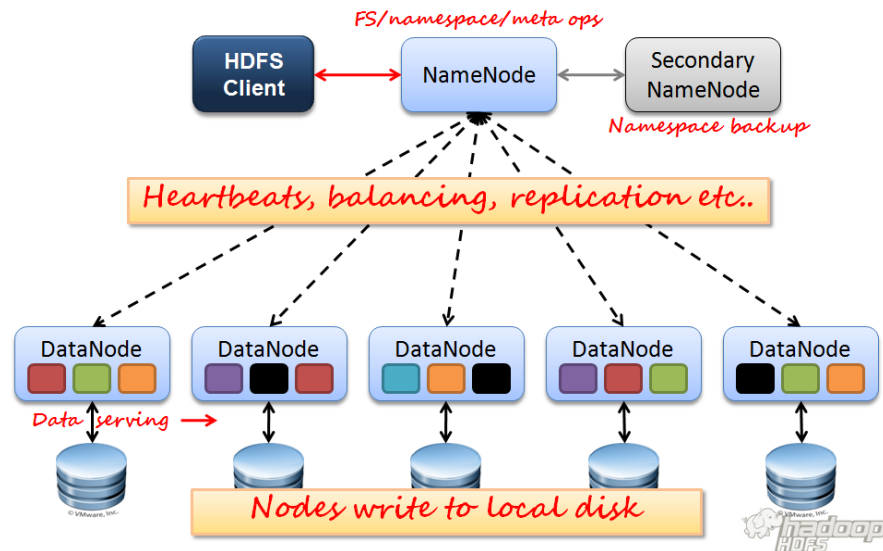
Système de fichier local (ext3, LFS ...)

Système de fichiers distribué (NFS, HDFS, GFS ...)

Système de fichiers parallèle (Ceph, GPFS ...)



Un système de stockage distribué: HDFS



Exploitation des disques locaux des serveurs de calcul pour le traitement de gros volumes de données

Impact de l'évolution des données sur la conception des bases de données dans le monde du cloud

- Volumes de données croissants
- Variété des données
- Vitesse
- Données semi- ou non structurées
- Données liées entre elles



Des bases de données relationnelles aux bases de données noSQL

- Bases de données **relationnelles** traditionnelles
 - Propriétés ACID
- Bases de données **non structurées - noSQL**
 - Deux des propriétés CAP (théorème de Brewer)
 - ✓ Passage à l'échelle
 - ✓ Performance
 - ✓ Disponibilité
 - ✓ Simplicité

Atomicity **C**onsistency
Isolation **D**urability

Consistency **A**vailability
Partition-tolerance

not only **SQL** : structure évoluant
dynamiquement (non relationnel)

Classification des bases de données noSQL

- Base de données clé/valeur



- Base de données orientée documents



- Base de données orientée colonnes (à enregistrement extensible)



- Base de données orientée graphes



Différentes catégories pour différents usages

Base de données clé/valeur

- Stockage d'informations de session
- Profil utilisateur ou préférences

Clé	Valeur
Clé	Valeur
Clé	Valeur

Simplicité

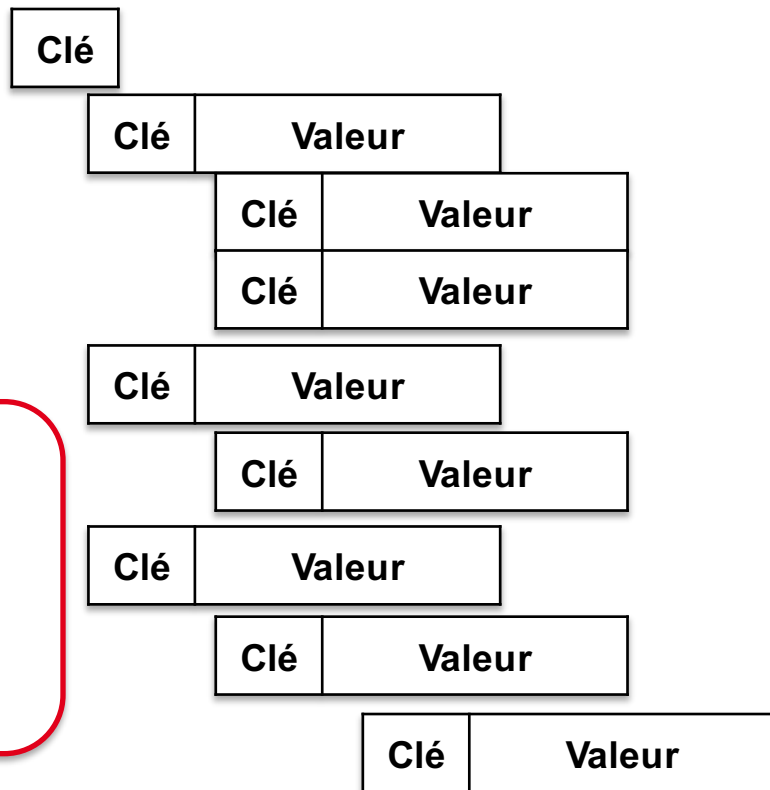
- ajout et suppression d'éléments
- recherche dans l'index

Base de données orientée document

- Application avec différents types d'objets et des recherches sur différents attributs
 - Analyse de données web en temps-réel
 - Nombre de pages vues
 - Nombre de visiteurs

Document : tout type d'objet sans pointeur

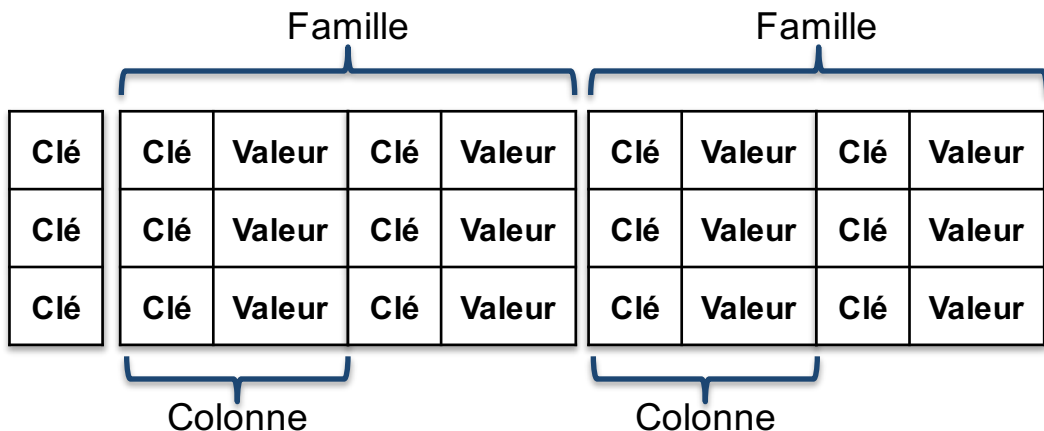
- Documents imbriqués, listes
- Index secondaires
- Lignes de taille variable
- Ajout d'attributs dynamiquement



Base de données orientée colonnes

- Même type d'utilisation que les bases de données orientées document
 - Débit plus élevé
 - Garanties de cohérence plus fortes
- Exemple
 - Site de commerce électronique

- Gros volumes de données scalaires
- Structuré en famille de colonnes
- Ajout de colonnes dynamiquement
- Partitionnement vertical et horizontal automatique

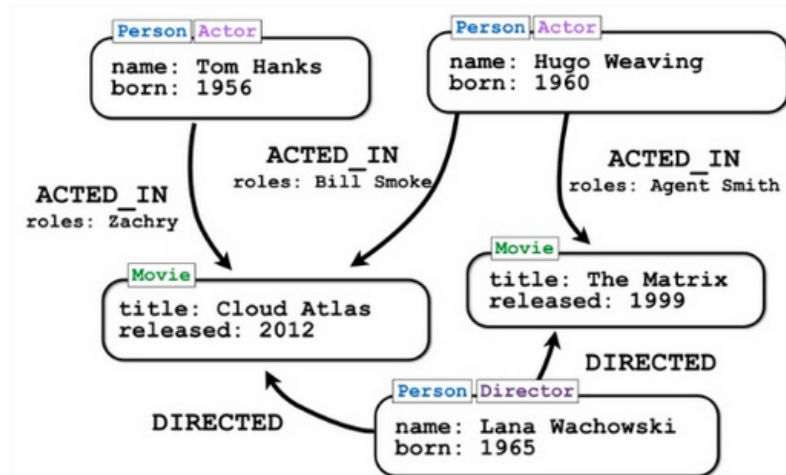


Base de données orientée graphe

- Réseaux sociaux
- Services de géolocalisation
- Moteurs de recommandation

3 notions :

- ✓ Nœuds
- ✓ Relations
- ✓ Propriétés



Déploiement de systèmes de fichiers et bases de données dans le cloud

- **IaaS** – Mise à disposition d'images de machines virtuelles préconfigurées
- **PaaS** – Gestion automatique du déploiement et de l'élasticité des systèmes de stockage



Système de fichiers : XtreamFS
Base de données relationnelle : MySQL
Base de données clé/valeur : Scalarix

Illustrations & photos : crédits

p. 2 : by netalloy, domain public, <http://www.freestockphotos.biz/stockphoto/15362>

p. 9 : © Hadoop HDFS

p. 10 : Domaine public