



Cloud Computing

– Cours 5 –

Chapitre 5 : Le stockage des données dans le Cloud

Dr. MENNOUR R.

Faculté des nouvelles technologies

rostom.mennour@univ-constantine2.dz

Etudiants concernés

Faculté/Institut	Département	Niveau	Spécialité
Nouvelles technologies	IFA	Master 1	SDIA

Section 1 :

Importance du stockage

L'importance du stockage

Quantité de données

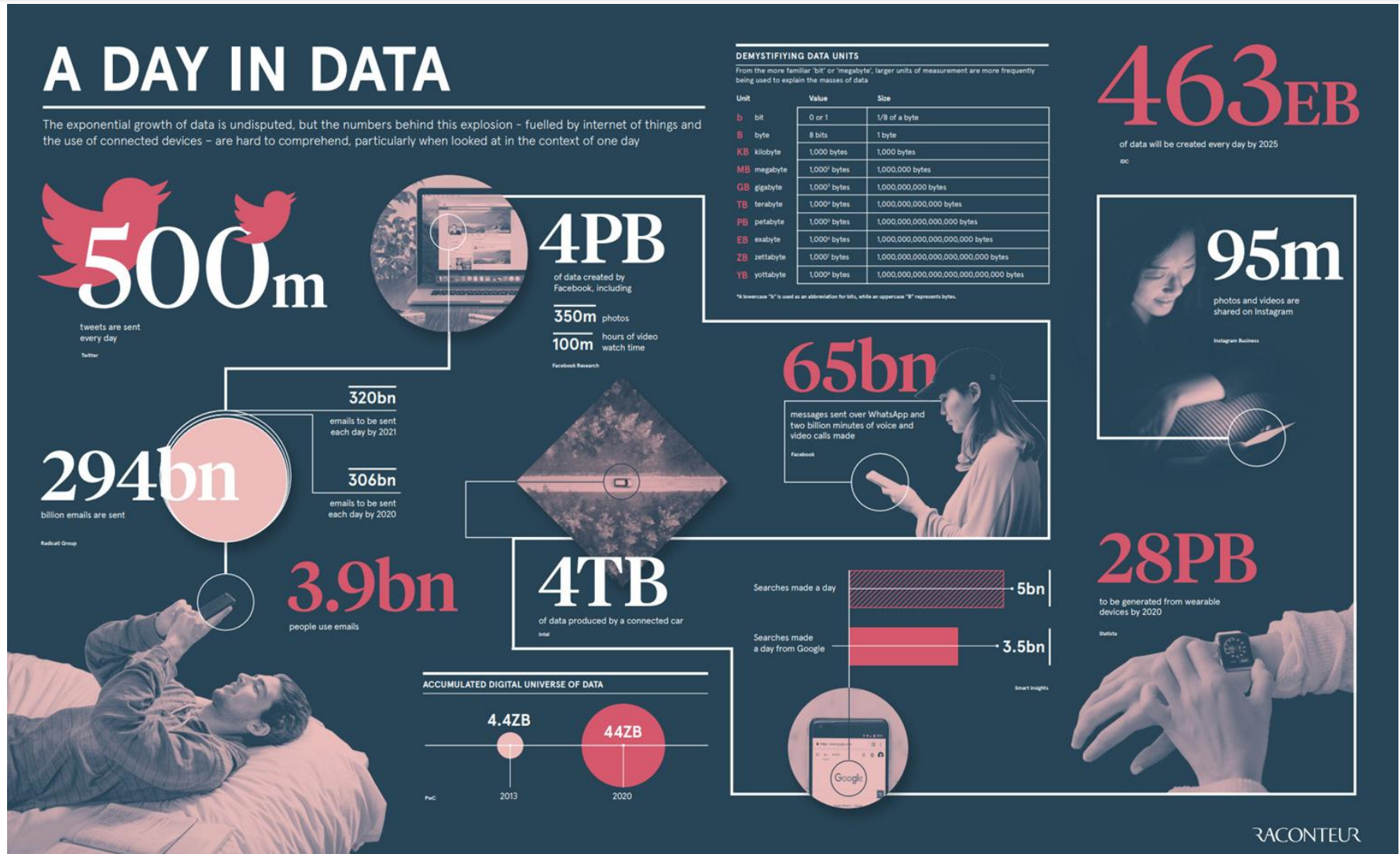
- en 2008, Google traitait environ 20 pétaoctets de données par jour (Google ne publie pas d'informations sur la quantité de données qu'il traite aujourd'hui).
- selon une estimation de Cisco, l'utilisation collective d'Internet dans le monde a atteint un zettaoctet en 2016.
- Selon la société de veille commerciale IDC, la «Global Datasphere» en 2018 a atteint 18 zettaoctets. Il s'agit du total de toutes les données créées, capturées ou répliquées. (Cependant, toutes ces données ne sont pas stockées et conservées.)
- La grande majorité des données mondiales a été créée au cours des dernières années et cette étonnante croissance des données ne montre aucun signe de ralentissement. En fait, IDC prédit que les données mondiales passeront à 175 zettaoctets en 2025.

Qu'est ce que ça veut dire

- Si vous deviez stocker 175 zettaoctets sur des DVD, votre pile de DVD serait assez longue pour faire le tour de la Terre 222 fois.
- Si vous tentiez de télécharger 175 zettaoctets à la vitesse de connexion Internet actuelle moyenne, le téléchargement vous prendrait 1,8 milliard d'années. Même si vous avez fait appel à toutes les personnes dans le monde pour aider au téléchargement, cela prendrait encore 81 jours.

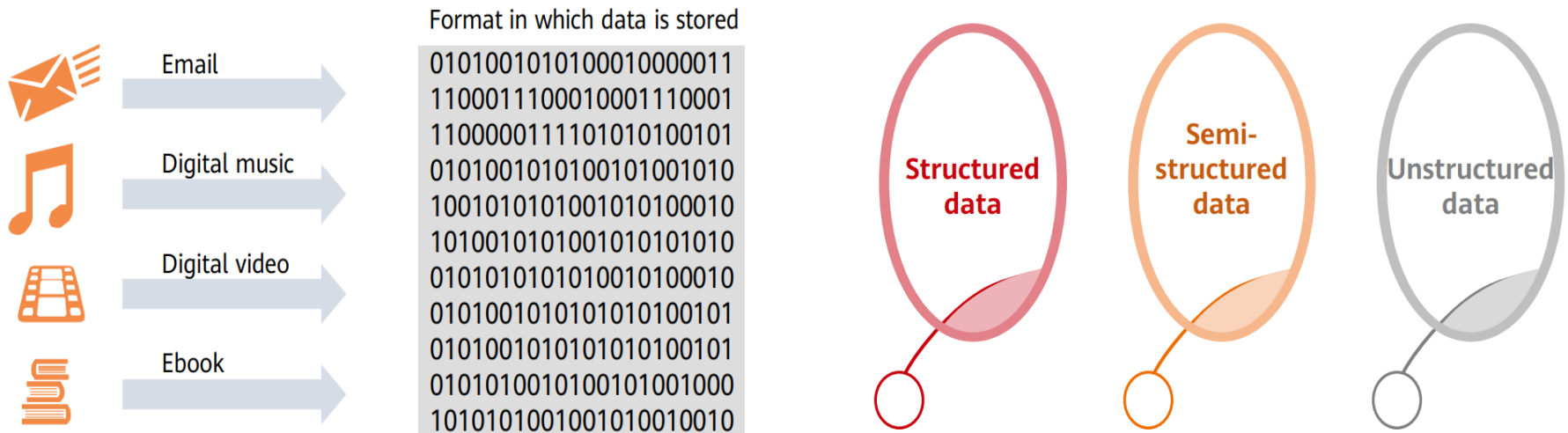
L'importance du stockage

Quantité de données



L'importance du stockage

Les données



SNIA (Storage Networking Industry Association) définit les données comme la représentation numérique de quoi que ce soit à n'importe quelle forme

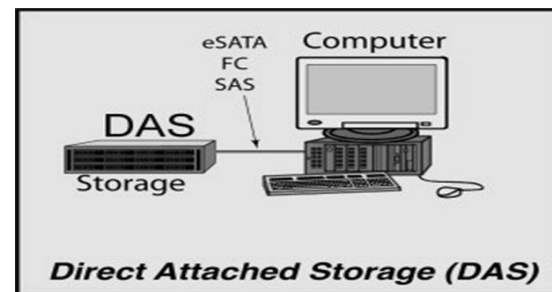
Section 2 : Les techniques de stockage

Les techniques de stockage

Direct Attached Storage

- Le stockage à connexion directe (DAS) est un stockage numérique directement connecté à l'ordinateur qui y accède.
- Un système DAS typique est constitué d'un périphérique de stockage de données connecté directement à un ordinateur via un adaptateur de bus hôte (HBA).
- Entre ces deux points, il n'y a pas de périphérique réseau (comme un concentrateur, un commutateur ou un routeur), et c'est la principale caractéristique du DAS.
- Les principaux protocoles utilisés pour les connexions DAS sont ATA, SATA, eSATA, NVMe, SCSI, SAS, USB, USB 3.0 et IEEE 1394.
- DAS est une relation 1: 1 entre le stockage et l'hôte.
- Le DAS est également généralement plus rapide que les stockage réseau, la plupart des réseaux étant limités à 1 Gbit / s, tandis que l'USB 3.0, commun sur les disques externes, prend en charge jusqu'à 5 Gbit / s.

eRacks/DAS24



Les techniques de stockage

Network Attached Storage

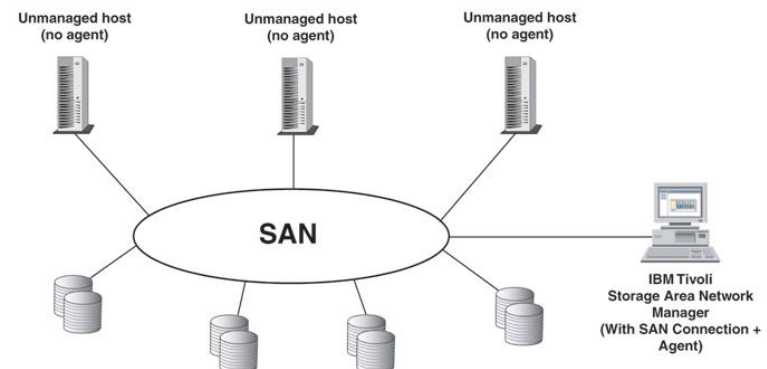
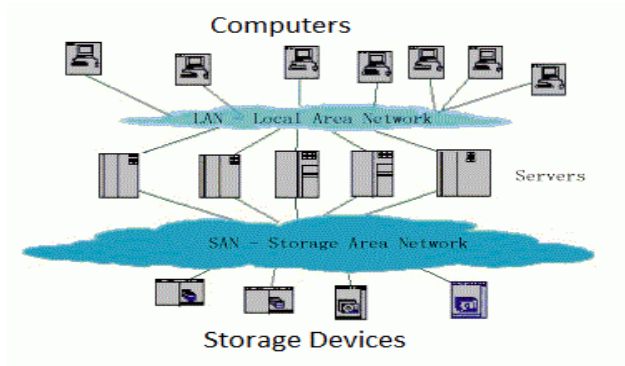
- Le stockage en réseau (NAS) est un serveur de stockage de données informatiques au niveau fichier (par opposition au stockage au niveau bloc) connecté à un réseau informatique fournissant un accès aux données à un groupe hétérogène de clients.
- Une unité NAS est un ordinateur connecté à un réseau qui fournit uniquement des services de stockage de données basés sur des fichiers à d'autres périphériques sur le réseau.
- Un système d'exploitation complet n'est pas nécessaire sur un périphérique NAS, donc souvent un système d'exploitation allégé est utilisé. Exemple: FreeNAS ou NAS4Free.
- Le NAS utilise des protocoles basés sur des fichiers tels que NFS (populaire sur les systèmes UNIX), SMB (Server Message Block) (utilisé avec les systèmes MS Windows), AFP (utilisé avec les ordinateurs Apple Macintosh). Les unités NAS limitent rarement les clients à un seul protocole.



Les techniques de stockage

Storage Area Network

- Un réseau de stockage (SAN) ou réseau de stockage est un réseau informatique qui donne accès à un stockage de données consolidé au niveau des blocs.
- Un réseau de stockage se différencie des autres systèmes de stockage par un accès bas niveau aux disques.
- Les baies de stockage n'apparaissent pas comme des volumes partagés sur le réseau. Elles sont directement accessibles en mode bloc par le système de fichiers des serveurs.
- chaque serveur voit l'espace disque d'une baie SAN auquel il a accès comme son propre disque dur.
- La plupart des réseaux de stockage utilisent le protocole SCSI pour la communication entre les serveurs et les unités de disque. Une couche de mappage vers d'autres protocoles est utilisée pour former un réseau.



Les techniques de stockage

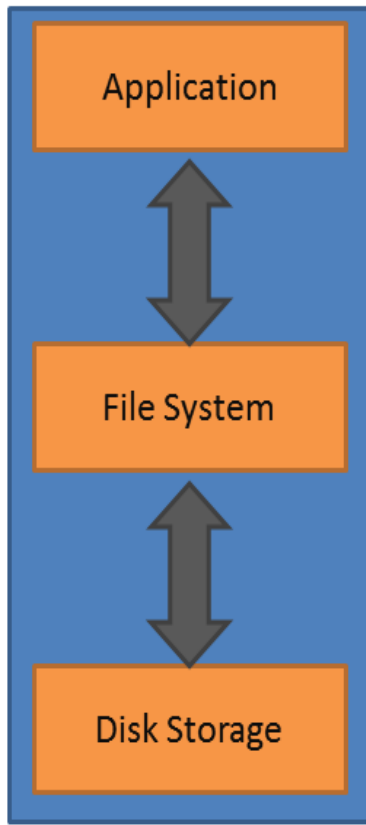
Comparaison

- La principale différence entre DAS et NAS est que le stockage DAS n'est directement accessible qu'à partir de l'hôte auquel le DAS est connecté. Un DAS n'incorpore aucun matériel réseau ni environnement d'exploitation associé pour fournir une installation permettant de partager les ressources de stockage de manière indépendante.
- Dans le cas du NAS, la ressource de stockage est directement connectée au réseau IP de l'entreprise. Le serveur NAS intègre le support de multiples systèmes de fichiers réseau (ex: NFS).
- Dans le cas du SAN, les baies de stockage n'apparaissent pas comme des volumes partagés sur le réseau. Elles sont directement accessibles en mode bloc par le système de fichiers des serveurs.
- Un SAN a plus en commun avec un DAS qu'un NAS, la principale différence étant que le DAS est une relation 1: 1 entre le stockage et l'hôte alors que le SAN est multiple.

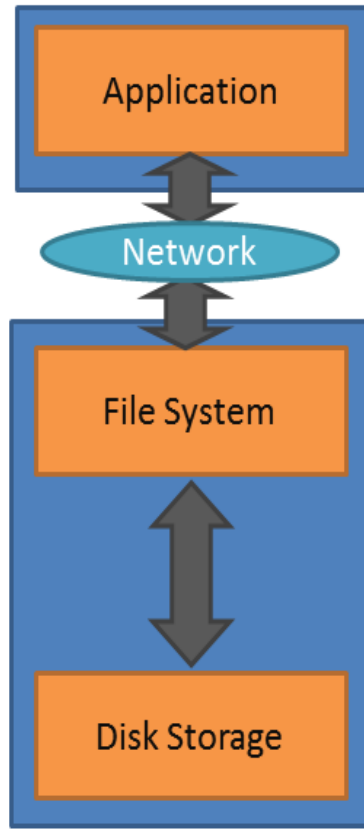
Les techniques de stockage

Comparaison

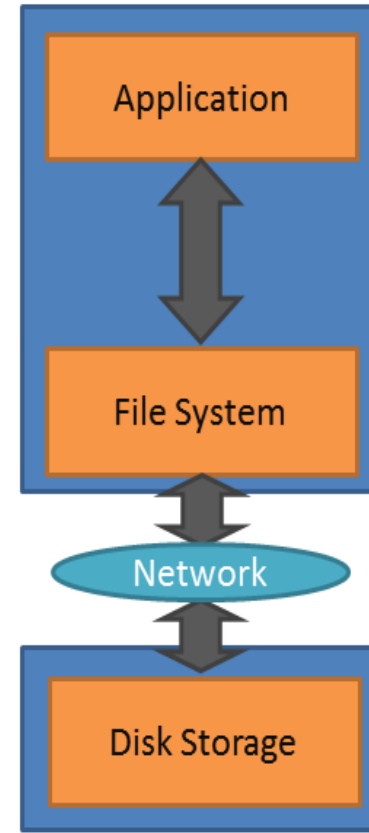
Direct Attached Storage



Network Attached Storage



Storage Area Network



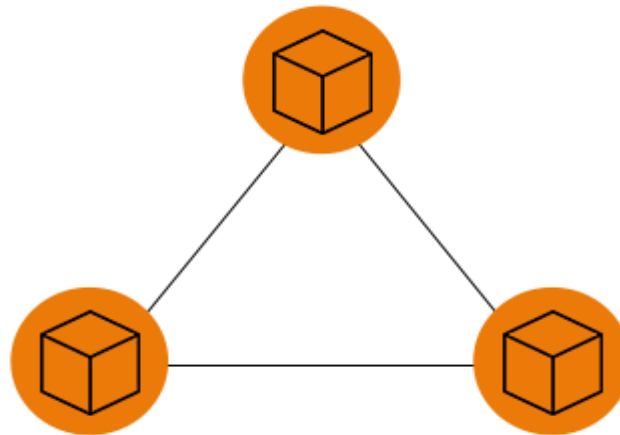
Section 3 :

Différents modes de stockage

Différents modes de stockage

Stockage en mode blocs

- **Qu'est ce que le stockage en mode blocs**
 - Le stockage en mode bloc rassemble les données en blocs qui sont stockés en tant qu'éléments séparés.
 - Chaque bloc de données se voit attribuer un identifiant unique, qui permet au système de stockage de conserver les petits éléments à l'emplacement le plus pratique.
 - Le stockage en mode bloc est souvent configuré pour dissocier les données de l'environnement de l'utilisateur.
 - Lorsqu'un utilisateur demande à accéder à des données, le logiciel de stockage sous-jacent réassemble les blocs de données et les lui représente.
 - Le stockage en mode bloc est généralement déployé dans des environnements de réseau de stockage (SAN) et doit être lié à un serveur en fonctionnement.



Différents modes de stockage

Stockage en mode blocs

- **Les avantages du stockage en mode blocs**
 - Performances: le stockage en blocs est extrêmement rapide et est le plus souvent utilisé par les applications nécessitant des opérations d'entrée-sortie rapides par seconde (IOPS).
 - Flexibilité: les volumes de données sont extrêmement flexibles, les volumes peuvent être étendus facilement sans temps d'arrêt à mesure que votre entreprise se développe, et de nouveaux volumes de blocs peuvent être ajoutés en fonction des besoins.
 - Agilité: Block Storage peut être déplacé d'un serveur à un autre avec une relative facilité. Avec une simple modification du chemin cible iSCSI, les données peuvent être dirigées vers un autre serveur.
 - Réduction de la charge du FS: Lorsque le logiciel peut écrire directement sur les blocs de stockage, il supprime le temps supplémentaire imposé aux activités de lecture / écriture du système de fichiers.
 - Démarrage: Les systèmes d'exploitation peuvent démarrer directement à partir d'un bloc de stockage présenté via des réseaux de stockage. Tout ce dont vous avez besoin est d'un serveur physique ou virtuel avec un BIOS capable de démarrer en mode SAN.

Différents modes de stockage

Stockage en mode blocs

- **Les inconvénients du stockage en mode blocs**
 - Verrouillé sur le serveur - le volume du bloc de stockage est verrouillé sur un serveur individuel à la fois et il est impossible d'y accéder ailleurs dans l'environnement à partir d'un autre serveur cible. Il existe certaines solutions logicielles pour lutter contre cela, mais tout cela ajoute à la surcharge IO.
 - Métadonnées limitées: la quantité limitée d'informations stockées dans les métadonnées de stockage en mode bloc est nettement inférieure à celle du stockage au niveau des fichiers ou des objets. Certaines applications qui utilisent des métadonnées détaillées peuvent en souffrir.
 - Coût : dans le cas du cloud, le stockage en bloc peut être une option abordable pour beaucoup; Cependant, pour les fournisseurs de services gérés, les coûts sont élevés, car les réseaux SAN (Storage Attached Networks) sont coûteux à l'achat et nécessitent souvent des experts en stockage spécialisés pour gérer l'optimisation des matrices de données.

Différents modes de stockage

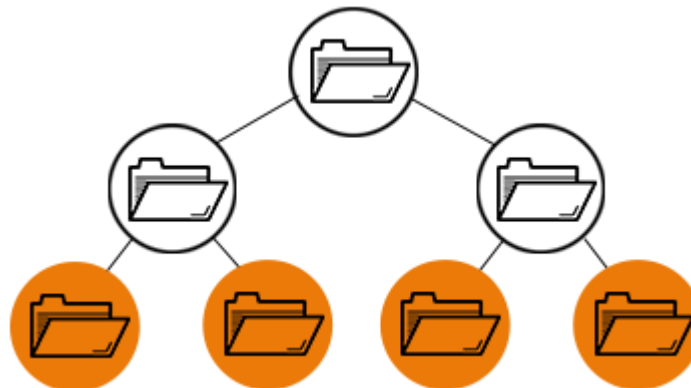
Stockage en mode blocs

- **Les cas d'utilisation du stockage en mode blocs**
 - Bases de données: Les bases de données nécessitent des opérations de lecture / écriture rapides et les volumes de stockage de blocs sont montés directement sur le système de fichiers, de sorte que le délai d'écriture et de récupération des données est minimal.
 - Les données temporaires: Diverses étapes du cycle de vie du développement logiciel peuvent créer de grandes quantités de données temporaires. La création d'un volume de stockage en mode blocs pour les données temporaires constitue un cas d'utilisation judicieux lorsque le stockage fourni avec votre instance n'est pas suffisant ou lorsque vous avez besoin d'espace supplémentaire pour une courte période.
 - Sauvegarder des données: Le stockage en blocs est un support de sauvegarde évolutif et rapidement accessible. Stockez tout ce que vous pourriez avoir besoin de transférer rapidement vers une autre instance ou tout ce dont vous pourriez avoir besoin à tout moment.
 - Vous pouvez démarrer à partir d'images de disque installées sur un volume de stockage en blocs.

Différents modes de stockage

Stockage en mode fichiers

- **Qu'est ce que le stockage en fichiers**
 - les données sont stockées en tant qu'élément unique d'information à l'intérieur d'un dossier.
 - Pour accéder à une donnée, l'ordinateur doit connaître le chemin à parcourir pour la trouver.
 - Les données stockées dans des fichiers sont organisées et récupérées à l'aide de quelques métadonnées qui indiquent à l'ordinateur où le fichier se trouve.
 - Ils permettent de stocker à peu près tous types de données.
 - Ils sont idéaux pour le stockage d'ensembles complexes de fichiers et rapides à parcourir.



Différents modes de stockage

Stockage en mode fichiers

- **Les avantages stockage en mode fichiers**
 - Simplicité: Le stockage de fichiers est l'approche la plus simple et la plus familière pour organiser les fichiers et les dossiers sur le disque dur ou le périphérique NAS d'un ordinateur. Il n'est pas nécessaire d'écrire des applications ou du code pour accéder à vos données.
 - Partage de fichiers: le stockage de fichiers est idéal pour centraliser et partager des fichiers sur un réseau local (LAN). Les fichiers stockés sur un périphérique NAS sont facilement accessibles par tout ordinateur du réseau disposant des autorisations appropriées.
 - Protocoles communs: le stockage de fichiers utilise des protocoles de niveau fichier courants tels que SMB (Server Message Block), CIFS (Common Internet File System) ou NFS (Network File System).
 - Protection des données: le stockage de fichiers sur un périphérique de stockage distinct, connecté au réseau local, offre une protection des données en cas de défaillance de serveur.
 - Abordable: le stockage de fichiers à l'aide d'un périphérique NAS vous permet de déplacer des fichiers d'un matériel informatique coûteux vers un périphérique de stockage connecté à un réseau local plus abordable.

Différents modes de stockage

Stockage en mode fichiers

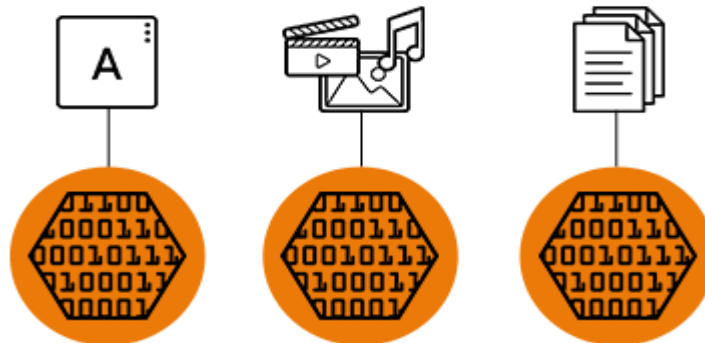
- **Les cas d'utilisation du stockage en mode fichiers**
 - Partage de fichiers local: si vos besoins en matière de stockage de données sont généralement cohérents et simples, tels que le stockage et le partage de fichiers avec les membres de l'équipe au bureau, considérez la simplicité du stockage au niveau des fichiers.
 - Collaboration de fichiers centralisée: si vous téléchargez, stockez et partagez des fichiers dans une bibliothèque centralisée (située sur le site, hors site ou dans le cloud), vous pouvez facilement collaborer sur des fichiers avec des utilisateurs internes et externes ou avec des invités extérieurs à votre ordinateur/réseau.
 - Archivage / stockage: vous pouvez archiver de manière rentable des fichiers sur des périphériques NAS dans un petit environnement de centre de données ou vous abonner à un service de stockage de fichiers basé sur le cloud pour stocker et archiver vos données.
 - Sauvegarde / reprise sur sinistre: vous pouvez stocker les sauvegardes de manière sécurisée sur des périphériques de stockage distincts, connectés au réseau local. Vous pouvez également vous abonner à un service de stockage de fichiers basé sur le cloud pour répliquer vos fichiers de données sur plusieurs centres de données dispersés géographiquement et bénéficier de la protection supplémentaire des données en termes de distance et de redondance.

Différents modes de stockage

Stockage en mode objets

- **Qu'est ce que le stockage en mode objets**

- Le stockage en mode objets est une architecture de stockage de données permettant de traiter de grandes quantités de données non structurées.
- Les objets sont des unités discrètes de données stockées dans un environnement de données structurellement plat.
- Chaque objet est un référentiel simple et autonome comprenant les données, les métadonnées (informations descriptives associées à un objet) et un numéro d'identification unique (au lieu d'un nom de fichier et d'un chemin d'accès).
- Le stockage d'objets supprime les problèmes de complexité et d'évolutivité d'un système de fichiers hiérarchique avec des dossiers et des répertoires.
- Les objets (données) d'un système de stockage d'objets sont accessibles via des interfaces de programmation d'applications (API).



Différents modes de stockage

Stockage en mode objets

- **Les avantages du stockage en mode objets**
 - Mise à l'échelle illimitée: Les objets sont stockés dans un environnement de données structurellement plat, au sein d'un périphérique de stockage. Vous ajoutez simplement plus de périphériques / serveurs en parallèle à un cluster de stockage d'objets pour un traitement supplémentaire et pour prendre en charge les débits plus élevés requis par des fichiers volumineux tels que des vidéos ou des images.
 - Réduction de complexité: Le stockage d'objets supprime la complexité inhérente à un système de fichiers hiérarchique avec des dossiers et des répertoires. Cela améliore les performances, en particulier lors de la gestion de très grandes quantités de données.
 - Disponibilité: Vous pouvez configurer les systèmes de stockage d'objets afin qu'ils répliquent le contenu. Si un disque au sein d'un cluster échoue, un disque en double est disponible, garantissant que le système continue de fonctionner sans interruption ni dégradation des performances.
 - Gestion des données non structurées: Le stockage d'objets est de plus en plus adopté à l'ère du Cloud Computing et de la gestion des données non structurées qui, selon les analystes, représenteront dans un proche avenir la grande majorité de toutes les données dans le monde.

Différents modes de stockage

Stockage en mode objets

- **Les avantages du stockage en mode objets**
 - Le stockage objet est une solution idéale pour le stockage, l'archivage, la sauvegarde et la gestion de gros volumes de données statiques ou non structurées.
 - Applications natives dans le cloud: le stockage d'objets dans le cloud va de pair avec le développement d'applications dans le cloud.
 - Créez de nouvelles applications cloud natives ou transformez des applications héritées en applications cloud de nouvelle génération en utilisant le stockage d'objets basé sur le cloud en tant que stockage de données persistant.
 - Collectez et stockez de grandes quantités de données IoT et mobiles non structurées pour vos applications pour appareils intelligents.
 - Mettez à jour facilement et efficacement vos composants d'application.
 - Intelligence artificielle et analyse des données volumineuses:
 - Créez un référentiel de données centralisé, en exploitant un stockage d'objets économique et évolutif.
 - Collectez et stockez des quantités illimitées de données de tout type, de toute source.
 - Interrogez ces données pour effectuer des analyses de données volumineuses et obtenir des informations précieuses sur les clients, les opérations et le marché.

Section 4 :

Les modèles de cohérence

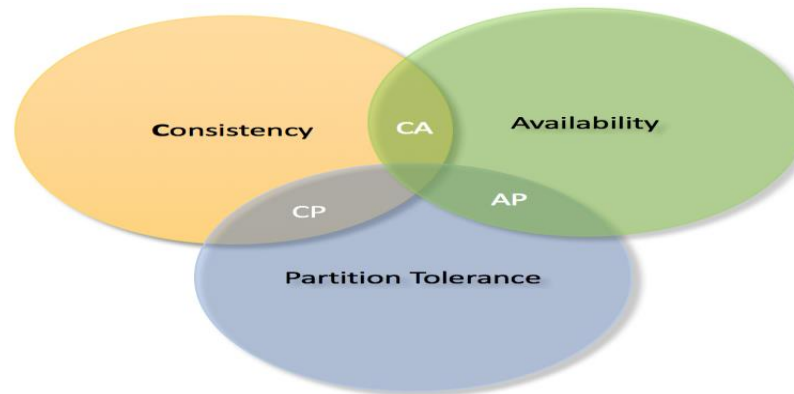
Les modèles de cohérence

Le théorème CAP

- Le théorème CAP a été présenté par Dr Brewer en 2000.
- Le théorème a été prouvé plus tard par Gilbert et Lynch en 2002.

Théorème CAP

Ce théorème stipule qu'au maximum seulement deux des trois propriétés suivantes peuvent être réalisées simultanément dans un système distribué : Cohérence (Consistency), disponibilité (Availability) et tolérance de partitionnement (Partition tolerance).



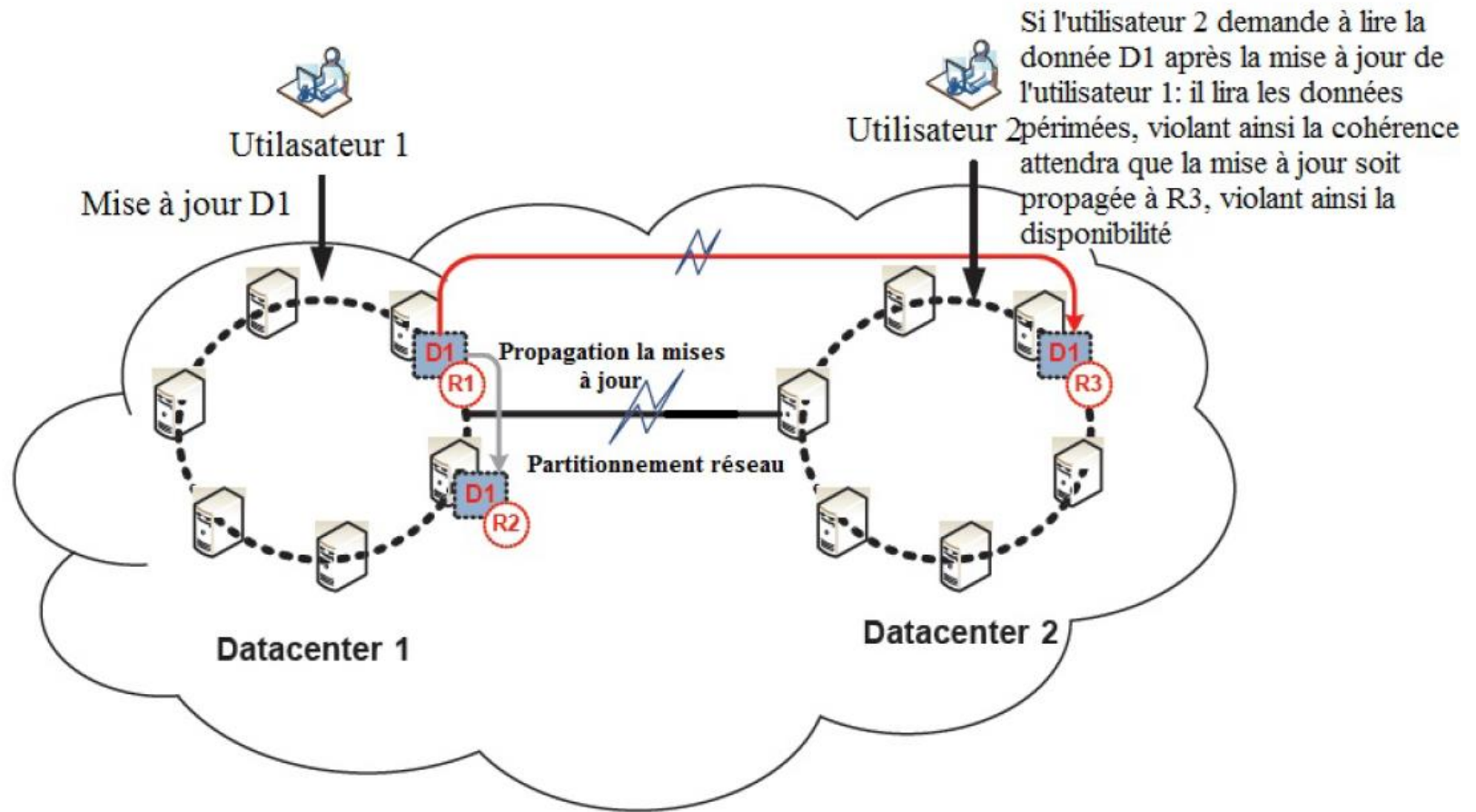
Les modèles de cohérence

Le théorème CAP

- **Cohérence** : La propriété de cohérence garantit qu'une opération ou une transaction est effectuée de manière atomique et laisse les systèmes dans un état cohérent ou échoue à la place. Ceci équivaut à garantir à la fois l'atomicité des opérations et la cohérence des propriétés de données (AC) de la sémantique ACID (Atomicity, Consistency, Isolation et Durability) dans les systèmes de gestion de bases de données relationnelles (SGBDR).
- **Disponibilité** : Un système de stockage distribué est continuellement disponible si chaque requête reçue par un noeud non défaillant doit aboutir à une réponse. C'est le cas des SLA (Service Level Agreement) de nos jours proposés par les fournisseurs de Cloud Computing assurant des taux de disponibilité supérieurs à 99,99% des opérations.
- **Tolérance aux partitionnements** : Dans un système tolérant aux partitions, le réseau peut perdre des messages entre des noeuds provenant de différents composants (par exemple, des centres de données). Lorsqu'une partition réseau apparaît, la communication réseau entre deux composants (racks, centres de données, etc.) est désactivée et tous les messages sont perdus. Les répliques pouvant être réparties sur différentes partitions dans un tel cas, cette propriété a un impact direct sur la cohérence et la disponibilité.

Les modèles de cohérence

Le théorème CAP



Les modèles de cohérence

Le théorème CAP

- Les implications du théorème CAP ont introduit des compromis difficiles et fondamentaux pour les systèmes distribués et les concepteurs de services.
- Les systèmes conçus pour être déployés sur des entités uniques, tels qu'un SGBDR, visent à fournir à la fois des propriétés de disponibilité et de cohérence, car le partitionnement ne pose aucun problème.
- pour les systèmes distribués qui reposent sur la mise en réseau, tels que le Cloud, la tolérance de partitions est indispensable pour la grande majorité d'entre eux.
- Ceci introduit parmi d'autres compromis dérivés du théorème CAP, la Cohérence vs. Disponibilité comme compromis majeur.
- Ce théorème a introduit ainsi plusieurs compromis, qu'on appelle modèles de cohérence tels que : la cohérence faible, la cohérence forte, la cohérence séquentielle, la cohérence causale, etc.

Section 5 :

Les Bases de données distribuées

Les modèles de cohérence

Qu'est ce qu'un système de base de données distribués

Base de données distribuée :

Nous définissons une base de données distribuée comme un ensemble de plusieurs bases de données reliées logiquement situées aux nœuds d'un système distribué.

Système de base de données distribuée :

Un système de gestion de base de données distribuée (SGBD distribué) est alors défini comme le système logiciel qui permet la gestion de la base de données distribuée et rend la distribution transparente pour les utilisateurs.

Les Bases de données distribuées

Les promesses des SGBD distribués

- **Gestion transparente des données distribuées et répliquées** : La transparence fait référence à la séparation de la sémantique de niveau supérieur d'un système des problèmes de mise en œuvre de niveau inférieur. En d'autres termes, un système transparent «cache» les détails de mise en œuvre aux utilisateurs. L'avantage d'un SGBD totalement transparent est le haut niveau de support qu'il offre pour le développement d'applications complexes.
- **Fiabilité grâce aux transactions distribuées** : Les SGBD distribués sont destinés à améliorer la fiabilité car ils ont répliqué des composants et éliminent ainsi les points de défaillance uniques. La défaillance d'un seul site, ou la défaillance d'une liaison de communication qui rend un ou plusieurs sites inaccessibles, ne suffit pas à faire tomber l'ensemble du système. Dans le cas d'une base de données distribuée, cela signifie que certaines des données peuvent être inaccessibles, mais avec un soin approprié, les utilisateurs peuvent être autorisés à accéder à d'autres parties de la base de données distribuée.

Les Bases de données distribuées

Les promesses des SGBD distribués

- **Amélioration des performances** : Les arguments en faveur de l'amélioration des performances des SGBD distribués reposent généralement sur deux points.
 - Tout d'abord, un SGBD distribué fragmente la base de données, ce qui permet de stocker les données à proximité de ses points d'utilisation (également appelée localité des données). Cela présente deux avantages potentiels:
 - 1. Étant donné que chaque site ne gère qu'une partie de la base de données, les conflits pour les services CPU et E / S ne sont pas aussi graves que pour les bases de données centralisées.
 - 2. La localité réduit les délais d'accès à distance qui sont généralement impliqués dans les réseaux étendus.
 - Le deuxième point est que le parallélisme inhérent aux systèmes distribués peut être exploité pour le parallélisme inter-requêtes et intra-requêtes. Le parallélisme entre les requêtes permet l'exécution parallèle de plusieurs requêtes générées par des transactions simultanées, afin d'augmenter le débit transactionnel. La définition du parallélisme intra-requête est différente dans les SGBD distribués et parallèles.
 - Dans le premier cas, le parallélisme intra-requête est obtenu en divisant une seule requête en un certain nombre de sous-requêtes, chacune étant exécutée sur un site différent, accédant à une partie différente de la base de données distribuée.
 - Dans les SGBD parallèles, il est réalisé par parallélisme interopérateur et intra-opérateur. Le parallélisme interopérateur est obtenu en exécutant en parallèle différents opérateurs de requête sur différents processeurs, tandis qu'avec le parallélisme intra-opérateur, le même opérateur est exécuté par de nombreux processeurs, chacun travaillant sur un sous-ensemble des données.

Les Bases de données distribuées

Les promesses des SGBD distribués

- **Evolutivité (Scalability)** : Dans un environnement distribué, il est beaucoup plus facile de s'adapter à des tailles de base de données croissantes et à des charges de travail plus importantes. L'expansion du système peut généralement être gérée en ajoutant une puissance de traitement et de stockage au réseau. Évidemment, il peut ne pas être possible d'obtenir une augmentation linéaire de la «puissance», car cela dépend également de la surcharge de distribution. Cependant, des améliorations significatives sont encore possibles. C'est pourquoi les SGBD distribués ont suscité beaucoup d'intérêt pour les architectures évolutives dans le contexte du cluster et du cloud computing.

Section 5 : NoSQL DBs

- Pour gérer les données dans le cloud, on peut toujours s'appuyer sur un SGBD relationnel. Tous les SGBD relationnels ont une version distribuée, et la plupart d'entre eux fonctionnent dans le cloud.
- Bien qu'ils aient pu intégrer la prise en charge de tous les types de données (par exemple, objets multimédias, documents) et de nouvelles fonctions, cela a entraîné une perte de performances, de simplicité et de flexibilité pour les applications avec des exigences de performances spécifiques et strictes.
- Pour une application donnée, il existe un serveur de base de données, qui assure la tolérance aux pannes et la disponibilité des données via la réplication. À mesure que le nombre de clients Web augmente, il est relativement facile d'ajouter des serveurs Web. Cependant, le serveur de base de données devient le bottleneck et l'ajout de nouveaux serveurs de base de données nécessiterait de répliquer l'intégralité de la base de données, ce qui prendrait beaucoup de temps.

Qu'est ce que le NoSQL

- Un type important de nouvelle technologie de gestion de données est NoSQL, ce qui signifie «pas seulement SQL» pour contraster avec l'approche «taille unique» du SGBD relationnel.
- Les systèmes NoSQL sont des magasins de données spécialisés qui répondent aux exigences de la gestion des données Web et cloud.
- Ces nouvelles technologies de gestion des données ont conduit à une offre riche de services qui peuvent être utilisés pour créer des applications cloud gourmandes en données qui peuvent évoluer et présenter des performances élevées.
- Cependant, cela a également conduit à une large diversification des interfaces de stockage de données et à la perte d'un paradigme de programmation commun. Ainsi, cela rend très difficile pour un utilisateur de créer des applications qui utilisent plusieurs magasins de données, par exemple, un système de fichiers distribué, un SGBD relationnel et un SGBD NoSQL.
- le terme NoSQL (Not Only SQL) est apparu pour la première fois à la fin des années 1990 pour désigner les nouveaux magasins de données conçus pour répondre aux exigences de la gestion des données Web. Au lieu des bases de données relationnelles, ils prennent en charge différents modèles de données et langages autres que SQL standard.
- Ces systèmes mettent généralement l'accent sur l'évolutivité, la tolérance aux pannes et la disponibilité, parfois au détriment de la cohérence.

Les BD Clé - Valeurs

- Dans le modèle de données clé-valeur, toutes les données sont représentées sous forme de paires clé-valeur, où la clé identifie la valeur de manière unique.
- Les magasins de valeurs-clés sont sans schéma, ce qui offre une grande flexibilité et évolutivité.
- Ils fournissent généralement une interface simple telle que put (clé, valeur), valeur = get (clé), delete (clé).
- Une forme étendue de stockage clé-valeur est capable de stocker des enregistrements, sous forme de listes de paires attribut - valeur. Le premier attribut est appelé clé majeure ou clé primaire, et identifie de manière unique l'enregistrement parmi une collection d'enregistrements.
- Eg : DynamoDB, Cassandra, Riak, Oracle NoSQL Data Base.

Les BD basées documents

- Les magasins de documents sont des magasins de valeurs-clés avancés, dans lesquels les clés sont mappées à des valeurs de type de document, telles que JSON, YAML ou XML.
- Les documents sont généralement regroupés en collections, qui jouent un rôle similaire aux tables relationnelles.
- les documents sont différents des tuples relationnels. Les documents sont auto-descriptifs, stockent des données et des métadonnées (par exemple, des balises XML, des noms de champs dans des objets JSON) et peuvent être différents les uns des autres au sein d'une collection.
- De plus, les structures du document sont hiérarchiques, utilisant des constructions imbriquées, par exemple, des objets imbriqués et des tableaux dans JSON. Ainsi, la modélisation d'une base de données à l'aide de documents nécessite moins de collections qu'avec des tables relationnelles (plates), et évite également des opérations de jointure coûteuses.
- Eg : MongoDB, AsterixDB, CouchDB, RavenDB.

Les BD basées colonnes larges

- Les magasins de colonnes combinent certaines des propriétés intéressantes des bases de données relationnelles (par exemple, la représentation des données sous forme de tables) avec la flexibilité des magasins de valeurs clés (par exemple, des données sans schéma dans les colonnes).
- Chaque ligne d'une table à colonnes larges est identifiée de manière unique par une clé et comporte un certain nombre de colonnes nommées.
- Contrairement à une table relationnelle, où les colonnes ne peuvent contenir que des valeurs atomiques, une colonne peut être large et contenir plusieurs paires clé-valeur.
- Les magasins de colonnes larges étendent l'interface de magasin de clés-valeurs avec des constructions plus déclaratives qui permettent des analyses, des requêtes de correspondance exacte et de plage sur les familles de colonnes.
- À l'origine des magasins à colonnes larges se trouve Google Bigtable.

Les BD basées colonnes larges

- Les magasins de colonnes combinent certaines des propriétés intéressantes des bases de données relationnelles (par exemple, la représentation des données sous forme de tables) avec la flexibilité des magasins de valeurs clés (par exemple, des données sans schéma dans les colonnes).
- Chaque ligne d'une table à colonnes larges est identifiée de manière unique par une clé et comporte un certain nombre de colonnes nommées.
- Contrairement à une table relationnelle, où les colonnes ne peuvent contenir que des valeurs atomiques, une colonne peut être large et contenir plusieurs paires clé-valeur.
- Les magasins de colonnes larges étendent l'interface de magasin de clés-valeurs avec des constructions plus déclaratives qui permettent des analyses, des requêtes de correspondance exacte et de plage sur les familles de colonnes.
- À l'origine des magasins à colonnes larges se trouve Google Bigtable.