Au cœur du monde connecté, les défis de l'IOT :

Réseaux, Cloud, **Plateformes**, Big Data, Sécurité, Blockchain, 5G….

IOT-IDO - Stockage des données

Le 8 octobre 2019

BASSET Jean-Claude

ASPROM

D’apres
HULFT

ASPROM

IOT et BIG DATA

L'internet des objets (ou IoT) et le Big Data sont deux technologies interconnectées et indissociables.

L'Internet des Objets (IoT) génère de la valeur ; il produit aussi des données dont le gigantesque volume entraîne une remise en question des modèles de stockage.
L'IoT accroît la pression sur l'infrastructure qui doit répondre aux deux conditions que posent le **volume des données** et **son traitement en temps réel**. Dans un contexte où il s'agit d'aller plus vite, de manière plus agile et plus sûre,

Quant au volume total de **données** stockées au niveau mondial, il pourrait atteindre **175 zettaoctets (10*21)** en **2025**, équivalent à 175 milliards de téraoctets (1O*12) , selon IDC. Le marché de l'IOT serait de **90 Zettaoctets**

En 2025,, le volume du stockage dans le **Cloud computing** serait de **49 %** du volume total des données
30% des données seront traitées en temps réel
20% seront des données critiques
le Edge computing utilisé par **82%** des entreprises

L'IOT est inclus dans le Big Data car il respecte la regle des **3V** : Volume , Vélocité , Variété

L'IoT impose la mise en place de solutions de Big Data pour extraire de l'intelligence de toute cette masse de données.

OBJETS CONNECTÉS

Les objets connectés sont des objets possédant des capteurs, connectés à un réseau pouvant être programmés et pilotés à distance via un ordinateur, une tablette ou un smartphone, et qui collectent des données.

Les Objets connectés permettent de:

– S'identifier entre eux
– Recueillir des données sur leur environnement
– Communiquer entre eux et avec les utilisateurs
– Anticiper les pratiques
– Interagir avec un écosystème en fonction des données

Les Objets Connectés : Consommateur

- Les particuliers
  - Automobile
  - Habitation,
  - Maison et environnement
- Santé,
  - bien-être
  - sport
- les entreprises
  - Logistique
  - Agriculture
  - Energie /télécommunications
  - Banque /assurance
- Les collectivités et services publics
  - Ville
  - Energie
  - Citoyen

ASPROM

OBJETS CONNECTÉS

Typologie des objets

On peut distinguer plusieurs types d'objets

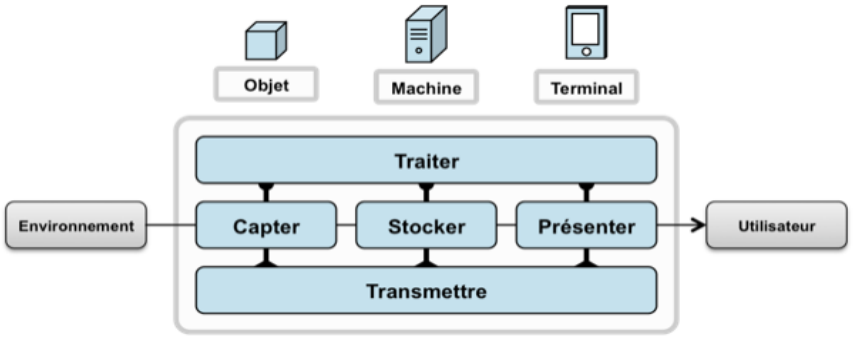
- Des **marqueurs passifs** (code barre, RFID, …) qui nécessitent le recours à un système de lecture
  - Puce RFID / code Barre
  - Faible capacité de stockage
  - Peuvent embarquer un capteur simple et être réinscriptifs
- Des **capteurs actifs spécialisés** capables de transmettre leurs données directement vers Internet ou indirectement via une passerelle
  - Plusieurs capteurs
  - Grande capacité de stockage
  - Peuvent communiquer sur le réseau et faire des traitements
- Des **objets "couteau suisse"**, généralement des Smartphones, Ces objets généralistes sont généralement moins précis dans leurs mesures que les capteurs spécialisés.

L'interaction avec le monde réel peut prendre plusieurs formes :

- Mesure de **données environnementales** : localisation, température, vent,…
- Mesure de **données comportementales** : mouvement de personnes, …
- Mesure industrielle M2M : commande d'actionneur
- Mesure de retro-action
- **piloter** les objets à distance, par exemple dans des applications domotiques
- **d'envoyer des notifications** au fournisseur
- **d'envoyer des notifications** aux usagers (ex : ralentir car l'autoroute est saturée)

Les DONNÉES dans l'architecture IOT

Structure d'un objet actifs



ce sont de dispositifs permettant de

- **collecter,**
- **stocker,**
- **transmettre**
- **traiter des données**

- issues du monde physique.

Cloud computing est souvent associé au concept d'IOT dont il offre des capacités de stockage et de gestion

Le cloud computing ou informatique en **nuage** est une infrastructure dans laquelle

- **la puissance** de calcul
- le stockage est géré par des ***serveurs***
- les usagers se connectent via une liaison **Internet** sécurisée.
- L' **ordinateur** de bureau ou portable, le **téléphone mobile** et autres objets connectés deviennent des points d'accès
- exécution d' **applications** ou consulter des données qui sont hébergées sur les serveurs.

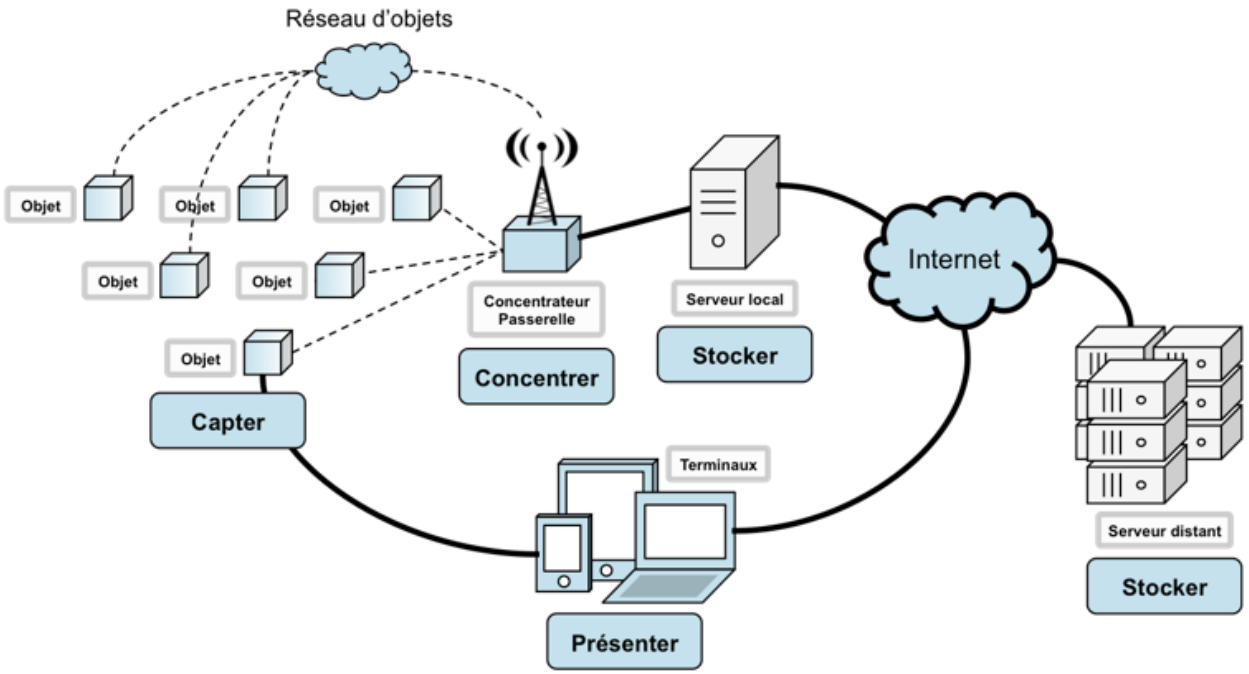
Le cloud se caractérise également par sa souplesse qui permet aux fournisseurs d'adapter automatiquement la capacité de stockage et la puissance de calcul aux besoins des utilisateurs

Les DONNÉES dans l'architecture IOT /2

Principe de l'architecture

le rôle des différents processus

- **Capter** désigne l'action de transformer une grandeur physique analogique en un signal numérique.
- **Concentrer** permet d'interfacer un réseau spécialisé d'objet à un réseau IP standard (e.g. WiFi) ou des dispositifs grand public.
- **Stocker** qualifie le fait d'agréger des données brutes, produites en temps réel, méta taguées, arrivant de façon non prédictible.
- Enfin, présenter indique la capacité de restituer les informations de façon compréhensible par l'Homme, tout en lui offrant un moyen d'agir et/ou d'interagir.



Fonctions globales

traitement des données
transmission des données

Fonctions de communication

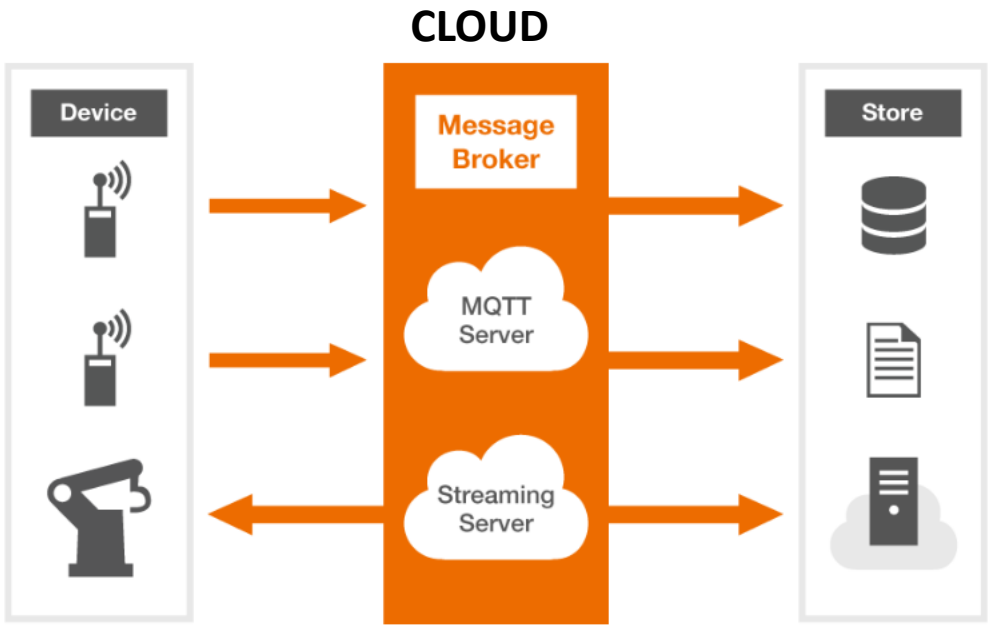
Réseau local de concentration
Réseau de transmission au serveur

ASPROM

Les DONNÉES dans l'architecture IOT /3

type d'architecture de reseau IOT

1 – DEVICE TO CLOUD



Avantages:

- Peut être mis en place relativement rapidement.
- L'architecture nécessite peu de modifications à mesure que le nombre de périphériques augmente.
- Permet une gestion flexible des données dans le cloud.

Désavantages:

- Les coûts de maintenance augmentent avec le nombre d'appareils.
- Lorsque les données de chaque appareil diffèrent, toute action ultérieure doit avoir lieu dans le cloud.
- Le renvoi des données doit être traité du côté de l'appareil.
- La consommation de la batterie est élevée, ce qui la rend peu pratique.

MQTT =protocole de transport
Messaging Queuing Telemetry Transport

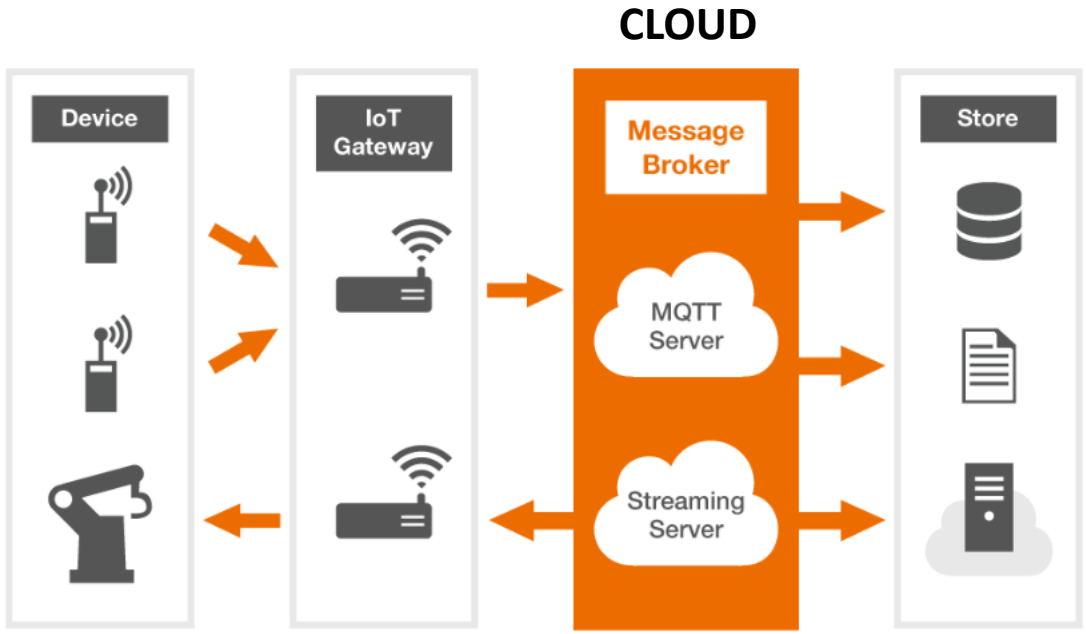
Etant donné que l'architecture Devise-to-Cloud requiert une certaine performance des périphériques, les cas dans lesquels elle peut être utilisée sont limités.

Dans ce contexte, le device utilise un réseau IP de télécommunication (terrestre ou satellitaire)

ASPROM

Les DONNÉES dans l'architecture IOT /3

2 – utilisation d'une GATEWAY (passerelle)



Avantages:

• La communication à courte portée peut être utilisée pour réduire la charge sur le périphérique.

•Nécessite une configuration minimale du périphérique lorsque le nombre de périphériques augmente.

• Le transfert de données est facile à contrôler (ré-envoi, filtrage, etc.)

• Les périphériques peuvent être gérés en groupe

Désavantages:

• Coûts plus élevés.

• La gestion des périphériques nécessite une structure distincte.

• Lorsque la passerelle échoue, tout le groupe de périphériques devient inutilisable.

Dans ce contexte

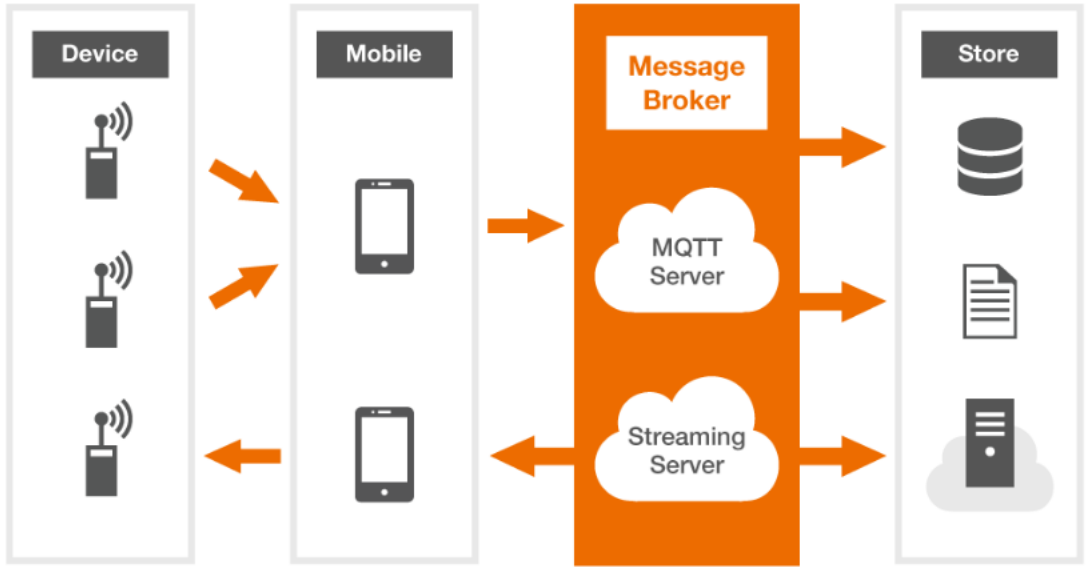
- La passerelle (Gateway, hub) utilise un reseau ip poue se connecter au Cloud
- Les devices (objets) utilisent des réseaux de telecommunications privés ou publics

Les DONNÉES dans l'architecture IOT /3

3 – utilisation de mobile

Avantages:

- Peut être redirigé vers le périphérique mobile d'un utilisateur.
- Peut être utilisé à d'autres fins que le relais.
- Contrôlez facilement les transferts.



Désavantages:

- Les appareils qui collectent des données n'ont pas toujours d'appareil mobile à proximité.
- Vous devez vérifier au préalable si le périphérique de collecte de données fonctionne.
- Un grand nombre de périphériques de collecte de données représente une charge importante pour le périphérique mobile.
- L'exemple le plus simple de ce type d'architecture serait probablement les services liés aux Smartphone
- les opérateurs de mobile développent des offres de connexion et de stockage dans le nuage

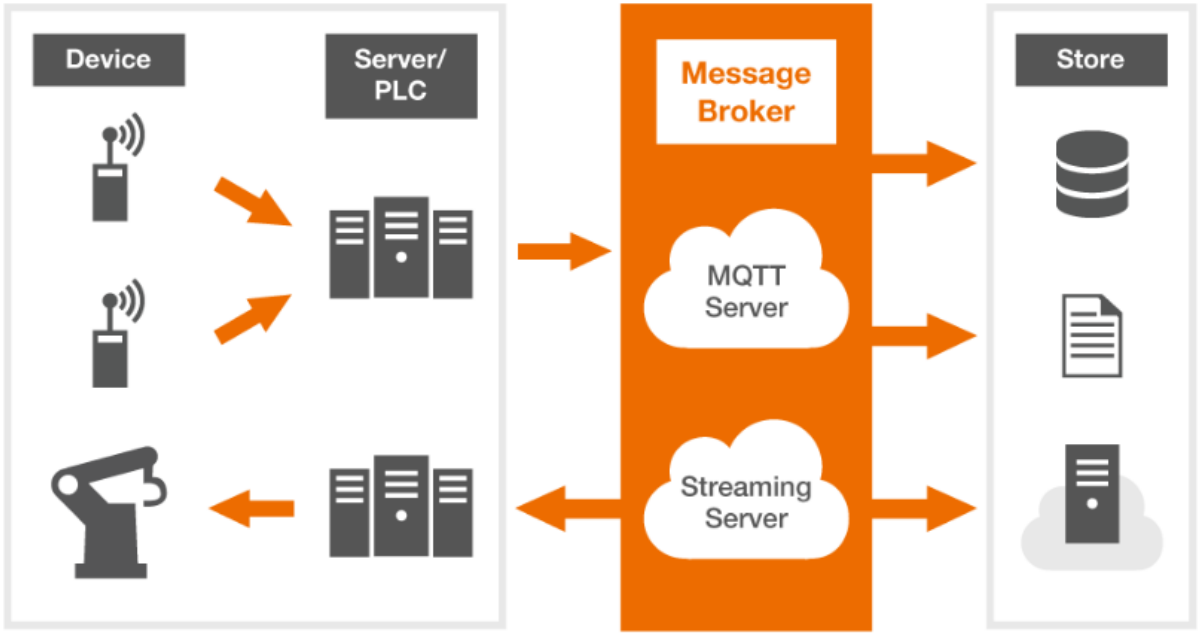
ASPROM

Les DONNÉES dans l'architecture IOT /3

utilisation d'un serveur

Le terme serveur sert à inclure des périphériques tels qu'unautomlate programmable **PLC**
(Programmable Logic Controller) , un serveur léger et compact destiné à être utilisé dans des situations spécifiques.

Avantages:
Permet d'effectuer des tâches telles que la conversion et le stockage de données pendant le relais de données.
Les actifs existants peuvent souvent être détournés.
Contrôlez facilement les transferts.



Désavantages

- Manque de mobilité (les appareils sont fixes)
- Ajouter des serveurs augmente les couts
- Peut causer des problèmes avec la construction de l'environnement réseau et les mesures de sécurité.

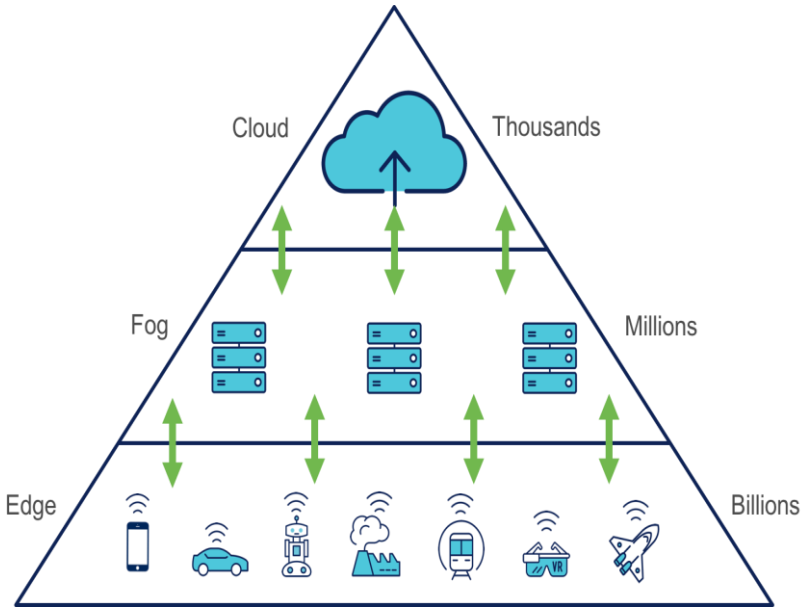
les données sont collectées à partir d'une ligne de production, il peut s'avérer plus efficace de traiter le transfert de données à l'aide de périphériques existants, tels que des automates programmables ou des serveurs, qui agrègent les données sans introduire de passerelles IoT.

ASPROM

EDGE COMPUTING

Le Edge Computing est une forme d'architecture informatique faisant office d'alternative au Cloud Computing. Plutôt que de transférer les données générées par des appareils connectés IoT vers le Cloud ou un Data Center, il s'agit de traiter les données en périphérie du réseau directement où elles sont générées

- ✓ le *Cloud* ne reçoit que les résultats des étapes de calculs réalisés par les extrémités du réseau et peut à son tour réaliser d'autres calculs ou simplement servir de relai vers les utilisateurs finaux
- ➢ Une autre p^topn consiste à utiliser des passerelles qui servent d'agrégateur de connectivité et traduisent des protocoles IoT vers des protocoles plus standards ou servent à chiffrer ou déchiffrer des données.



En bref

- L'Edge computing permet de **désengorge**r les réseaux et ainsi d'en limiter les pannes.
- La diminution de charge sur les serveurs centraux permet aussi de diminuer la latence.
- Dans le cas d'une défaillance réseau ou d'une indisponibilité du système centralisé, les appareils peuvent continuer de fonctionner de manière autonome jusqu'au retour à la normale. On parle ici alors **de résilience de l'architecture.**
- Les données qui arrivent au final vers le Cloud sont peu sensibles et chiffrées.

ASPROM

IIOT - EDGE COMPUTING

L'Internet des objets industriel (IIoT) est un moteur important de l'informatique de pointe. IIoT est la prochaine étape de l'expansion rapide de l'automatisation industrielle, qui se trouve être l'un de nos marchés les plus vastes et les plus robustes. Les entreprises prêtes à moderniser leurs infrastructures sont attirées par IIoT, car il offre des améliorations substantielles en termes d'efficacité, de productivité et d'analyse commerciale.

IIoT implique l'utilisation de la technologie Big Data et de l'apprentissage automatique en exploitant les données des capteurs et en adoptant des technologies de communication et d'automatisation machine à machine dans les usines de fabrication.

Edge computing permet aux appareils d'utiliser l'intelligence de la puissance de traitement et de la communication sur site. Ces appareils peuvent ensuite utiliser les capacités de calcul de périphérie pour déterminer quelles données doivent être stockées localement et quelles données doivent être envoyées au cloud pour une analyse plus approfondie. Au fur et à mesure que la capacité et la connectivité de l'Internet des objets industriels augmentent, on s'éloignera de l'informatique en nuage et de l'informatique de pointe.

Réseaux M2M

Les réseaux M2M sont des réseaux IOT , qui interconnecte directement des capteurs et actionneurs.la Technologie de communication et les capacités d'energie ont une influence directe sur le format des données

Typologie des réseaux

- **WLAN –short range** : Wireless Local Area Network, en français réseau local, ce terme désigne un réseau informatique local de faible distance de raccordement, qui relie des ordinateurs dans une zone limitée, [**WiFi , Blu**etooth , ZigBee]
- **LPWAN –long range/batterie longue durée -**réseau de télécommunication sans fil à grande échelle conçu pour permettre des communications à longue portée entre les objets connectés. Le taux de données LPWAN varie de 0,3 kbit/s à 50 kbit/s par canal. [**Lora, SigFox , Weightless**] **Ingenu** réseaux RPMA
- **Cellular –long range w/power**

  Une nouvelle technologie radio cellulaire standardisée par le 3GPP (3rd Génération Partnership Project), **NB-IOT** (**Narrowband Internet of Things**)- 127 Kb **[Free].** Et **LTE-M (**long term evolution – Machine) , plus rapide10 Mb **[Orange]**

Orbite satellite

- **Orbite géostationnaire** : 36000 KM - satellite de communications 2G/3G/4G/6G
- **Orbitre moyenne altitude** : 800/10 000 km MEO [Medium Earth Orbit]
- **Orbitre basse altitude** : 180/400 Km : LEO [low Earth Orbit]

Réseaux M2M

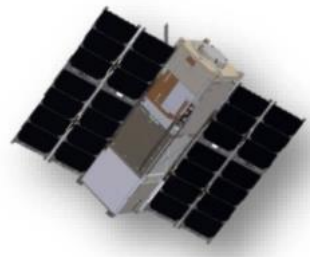
Satellite LEO – usage IOT:

- **IRIDIUM** / 780 KM – 66 satellites – iRIDIUM CloudConnect – adopté par **AMAZON AWS**
  - **Magnitude space (hiber)**
- **TELEDESIC**/ 700 KM /288 satellites /_projet abandonné
- **GLOBALSTER** / 1380 km/ 48 satellites/ gamme **SmartOne** / Spot
- **IMMARSAT** /1452/1492 GHZ – infrarouge & radio**/ BGAN M2M/Isat M2M/IsatDAtaPRO**
  - **IsatPro**

Classification des petits satellites

- Femtosatellite ; masse < 100 g.
- Picosatellite : masse < 1 kg
- Nanosatellite : masse < 10 kg (CubeSat)
- Microsatellite : masse < 100–150 kg (NASA < 100 kg)
- Minisatellite : masse < 500 kg (NASA small satellite < 180 kg)

Cubesat – 6-10 kg5



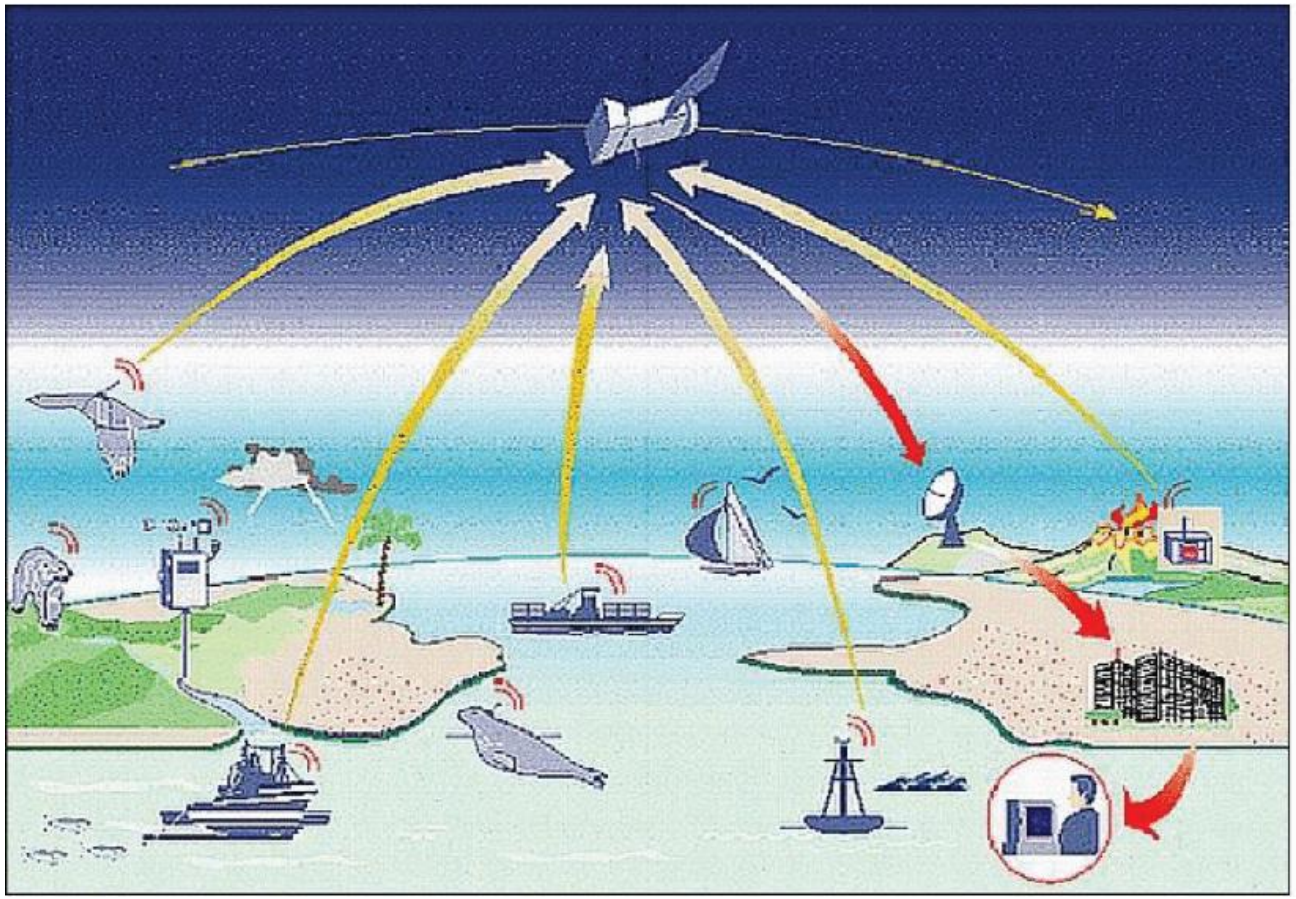
Nano satellite LEO

Les nanosatellites sont des satellites artificiels de petite taille, mesurant quelques dizaines de centimètres de côté.

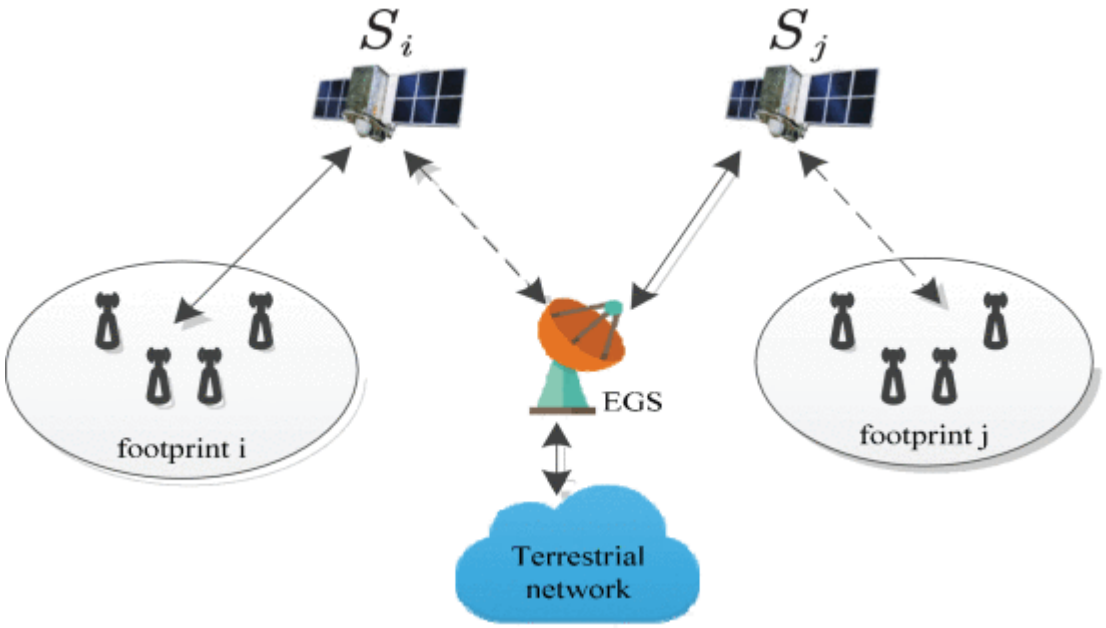
les satellites franchiront le même point après un certain intervalle de temps en jours

A. Applications à tolérance de retard (nanosatellite)

Le concept de DTA fait partie du DTN [Data Transmission Networks](DTL), une structure de communication innovante permettant de fournir des services de communication de données stockés et transférés automatisés dans des réseaux (RTT = temps transit A/R d 100 ms)



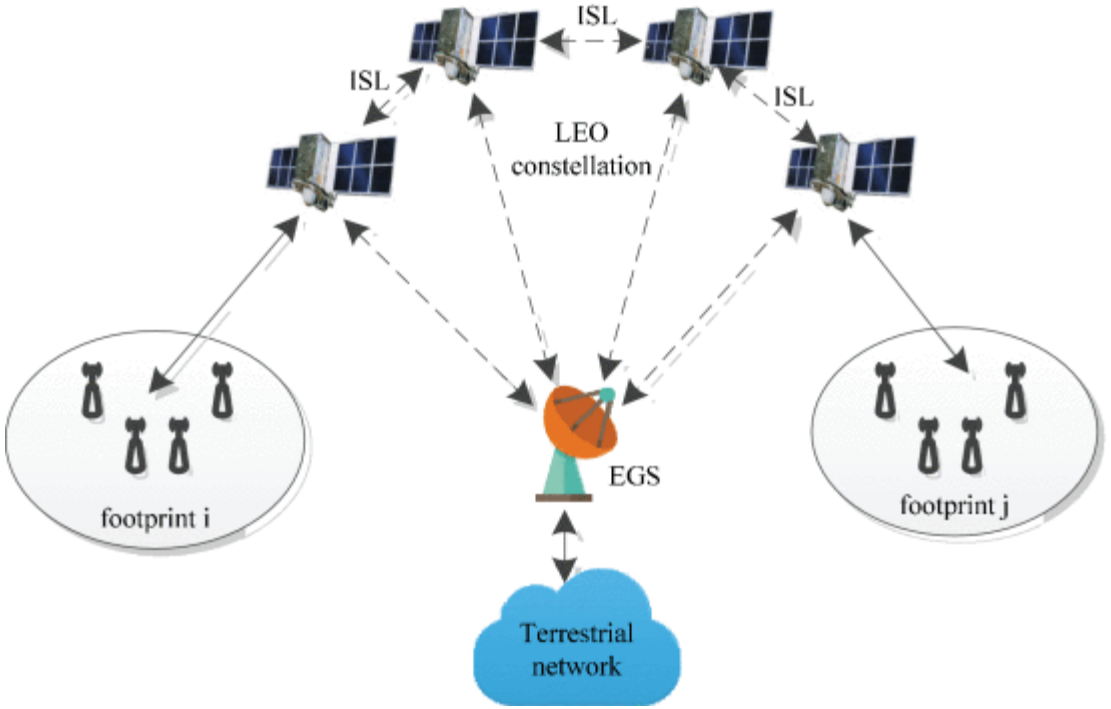
Réseau avec attente de passage



Une station de reception par zone (EGS)

Chaque satellite est un retransmetteur transparent permettant de relayer le trafic reçu des terminaux IoT et de la station de passerelle terrestre (EGS), et de restituer le trafic au sol.

Réseau avec continuité des données



Chaque satellite avec ISL est un commutateur de réseau permettant de communiquer avec les satellites voisins. Dans cette approche, les terminaux IoT situés dans une zone visible du satellite (appelée «empreinte de satellite») peuvent échanger du trafic avec les terminaux EGS et IoT situés dans les empreintes d'autres satellites sans la prise en charge des infrastructures terrestres.

ASPROM

Exemple de réseaux

Satellite EUTELSAT

L'opérateur français Eutelsat a annoncé le 24 septembre son intention de déployer une constellation de 25 nano-satellites. Une opération, pensée en partenariat avec Sigfox, qui doit permettre d'améliorer la transmission des données issues d'objets connectés toujours plus nombreux.

- 25 nano-satellite ELO en orbite basse (projet)
- Orbite basse LEO (Low Earth orbits)- 500-600 Km
- ELO pour Eutelsat LEO Objets
- ELO couverture mondiale pour les objets IOT
- Expérimentation avec satellite ELO ALPHA (Tyvak internationnal) - S1: 2020
- Lancement progressif jusqu'à 2021
- Partenariat initial avec Sigfox
- Latence inferieure à 10 minutes
- ELO1 et ELO2 sont fabriqués Par LORAL Orbital (USA)
- ELO32 et ELO4 sont fabriqués par CLYDE Space (GB)
- Si essais concluant 21 Satellite seront lancés

Chaque objet connecté sera survolé deux fois par jour par ELO. » Les données récoltées par ce dernier seront « *ensuite envoyées vers une station au sol située au Svalbard, un archipel de la Norvège situé dans l'océan Arctique, puis traitées et analysées par Sigfox* », qui possède un réseau terrestre mondial bas débit dédié à l'Internet des objets « *sur lequel s'appuiera Eutelsat* ».

BASES NoSQL

Le **NoSQL**, pour "**not only SQL**", désigne les bases de données qui ne sont pas fondées sur l'architecture classique des bases de données relationnelles. Développé à l'origine pour gérer du big data, l'utilisation de base de données NoSQL a explosée depuis quelques années.

Il existe pléthore de solutions NoSQL répondant plus ou moins bien à des besoins particuliers. Cependant, ces approches peuvent globalement être regroupées en quatre catégories

Les bases de données NoSQL

Représentent principalement les données du big data

Les propriétés **BASE** ont été proposées pour caractériser les bases NoSQL (au lieu du mode ACID :

- **Ba*sically Available*** : quelle que soit la charge de la base de données (données/requêtes), le système garantie un taux de disponibilité de la donnée
- **S*oft-state*** : La base peut changer lors des mises à jour ou lors d'ajout/suppression de serveurs. La base NoSQL n'a pas à être cohérente à tout instant
- **E*ventually consistent*** : À terme, la base atteindra un état cohérent

Avantages :

- L'évolutivité se fait de manière **horizontale** (pour augmenter les performances on ajoute des nouvelles machines)
- Les données sont distribuées sur plusieurs machines **(sharding**) de ce fait on évite les goulets d'étranglements lors de la récupération des données (fortes performances de lecture)
- La représentation des données est notable par l'absence de schéma **(schemaless**)
- La majorité des solutions est **Open Source**, néanmoins il existe des Support Pro pour répondre aux besoins des entreprises.

BASES NoSQL

Inconvénients :

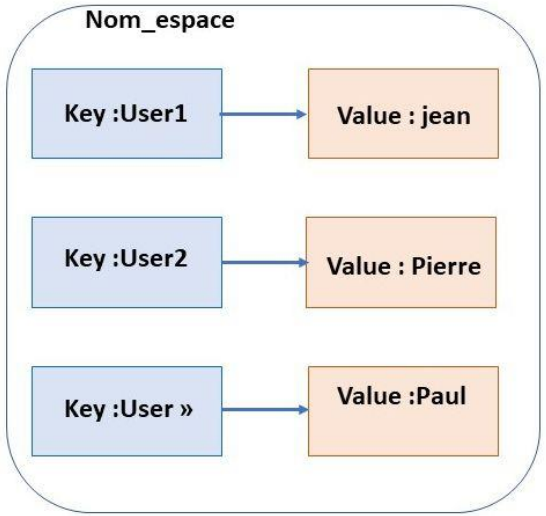
- Il n'existe pas de langage d'interrogation standardisé : chaque éditeur a mis en place le sien
- La mise en œuvre d'un environnement fortement transactionnel (fort besoin d'écriture) où le séquencement des écritures est primordial, reste complexe puisque l'architecture est distribuée compliquant l'atomicité et la cohérence des transactions
- L'écriture de requêtes complexes est difficile à mettre en oeuvre
- L'offre NoSQL est segmentée en plusieurs familles où chacune répond à un besoin précis.

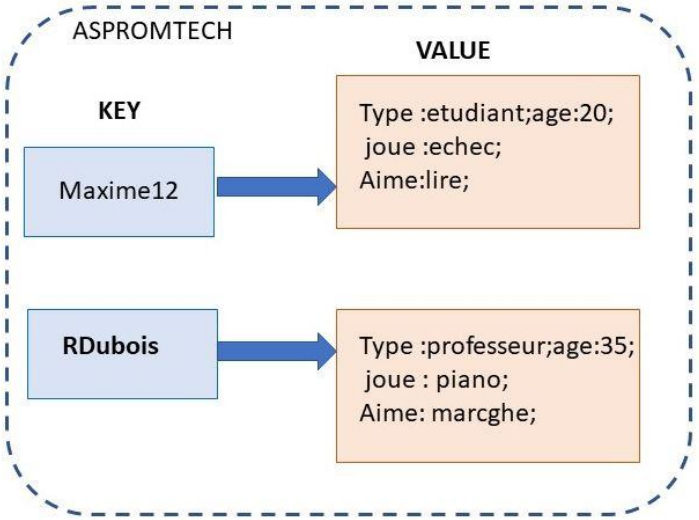
TYPE de BASES NoSQL

Un objet **Blob** représente un objet, semblable à un fichier, qui est immuable et qui contient des données brutes. Les **blobs** (pour B inary L arge Ob jects ...) sont des agrégats

1 – Type clé-valeur (Key value store)

Les données sont stockées en **clé-valeur** : une clé plus un BLOB (dans lequel on peut mettre : nombre, date, texte, JSON,XML, photo, vidéo, structure objet).





Les systèmes NoSQL orientés **clé-valeur**

les plus connus sont Memcached, Amazon's Dynamo, Redis, Riak et Voldemort créé par LinkedIn , hazelcast,pickleDB, Apache Ignite, Ehcache.

2 – base orientée document

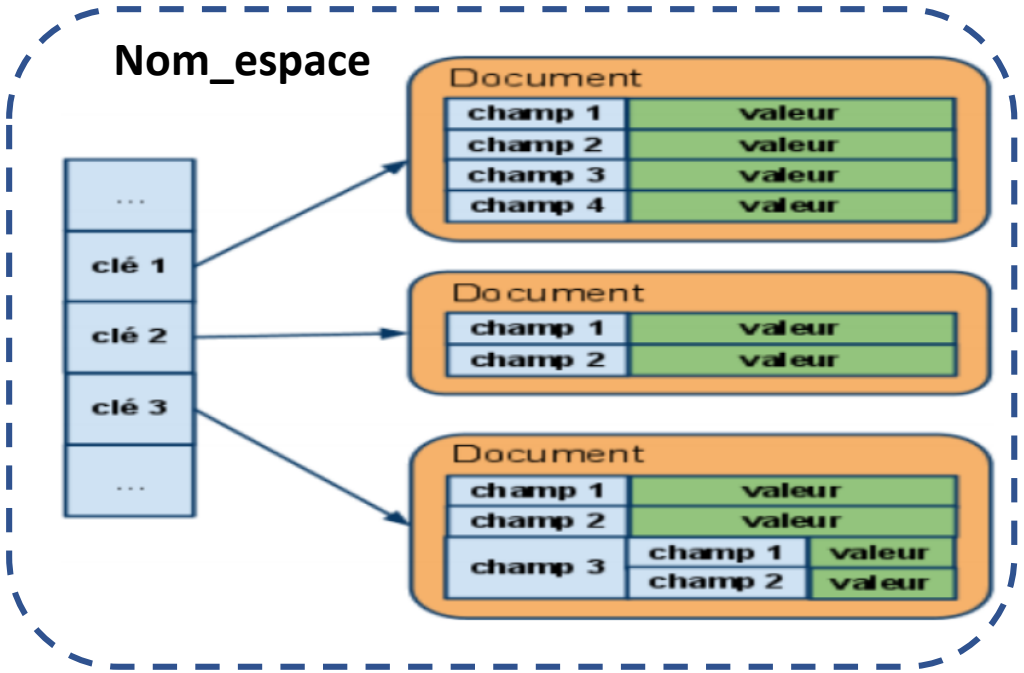
Ces bases de données stockent des données semi-structurées : le contenu est formaté JSON ou XML, mais la structure n'est pas contrainte.

Les plus

- Modèle de données simple mais puissant
- Requêtes plus complètes
- Flexibilité
- Evolutif au cours du temps

Les moins

- Duplication des données
- Cohérence difficile (pas de modèles interconnectées)
- Modelé limité à des clés



- MongoDB
- CouchDB
- OrientDB
- hRavenDB
- RethinkBD
- Amazon DynamoDB :

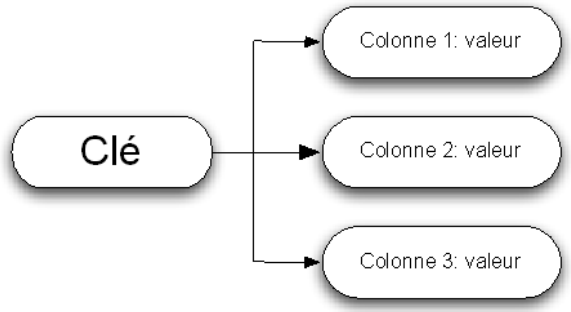
- IBM informix
- CrateDB
- DocumentDB :
- lusterpointDB :/
- Marklogic :

3 – base orientée colonne

Ces bases de données se rapprochent des bases de données relationnelles, à ceci près qu'elles permettent de remplir un nombre de colonnes variables.

Les principaux concepts associés sont les suivants

- **Colonne** :
  - o Entité de base représentant un champ de donnée
  - o Chaque colonne est définie par un couple clé / valeur
  - o Une colonne contenant d'autres colonnes est nommée **supercolonn**
- **Famille de colonnes :**
  - o Permettent de regrouper plusieurs colonnes (ou supercolonnes)
  - o Les colonnes sont regroupées par ligne
  - o Chaque ligne est identifiée par un identifiant unique (assimilées aux le modèle relationnel) et sont identifiées par un nom unique
- **Supercolonnes** :
  - o Situées dans les familles de colonnes sont souvent utilisées



	A	B	C	D	E
1	foo	bar	hello		
2		Tom			
3			java	scala	cobol

Organisation d'une table dans une BDD relationnelle

1	A foo	B bar	C hello
2	B Tom		
3	C java	D scala	E cobol

Organisation d'une table dans une BDD orientée colonnes

1	Choses	A foo	B bar	C hello
2	Choses	C texte12	D texte	Personnes: B Tom
3	Langages	C java	D scala	E cobol

Organisation d'une table dans une BDD orientée colonnes avec *super-colonnes*

BASES NoSQL

3 – base orientée colonne /2

Les plus

- Capacité de stockage accrue
- Accès rapide aux données

Les moins

- Requêtes limitées
- Limitation du nombre de types d’ objets

Utilisation

Les BD NoSQL type « Colonne » sont principalement utilisées pour :

- Netflix l'utilise notamment pour le login et l'analyse de sa clientèle
- Ebay l'utilise pour l'optimisation de la recherche
- Adobe l'utilise pour le traitement des données structurées et de Business Intelligence (BI)
- Des sociétés de TV l’utilisent pour cerner leur audience et gérer le vote des spectateurs (nb élevé d'écritures rapides et analyse de base en temps réel (Cassandra)
- peuvent être de bons magasins d'analyse des données semi-structurées
- utilisé pour la journalisation des événements et pour des compteurs

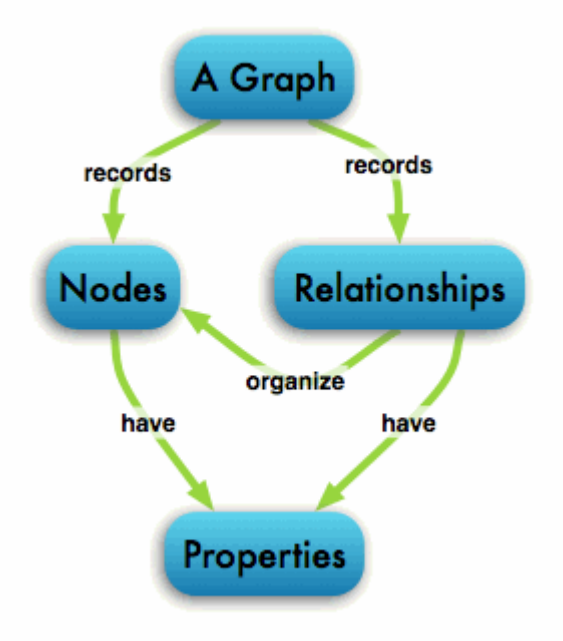
Implémentation

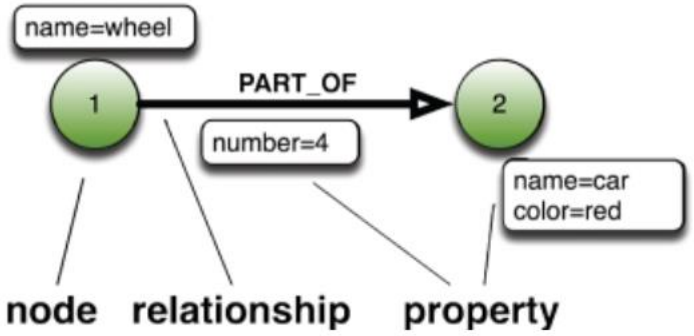
- Apache :HBASE
- BigTable (Google)
- Spark SQL (Apache)
- ElasticSearch (elastic)

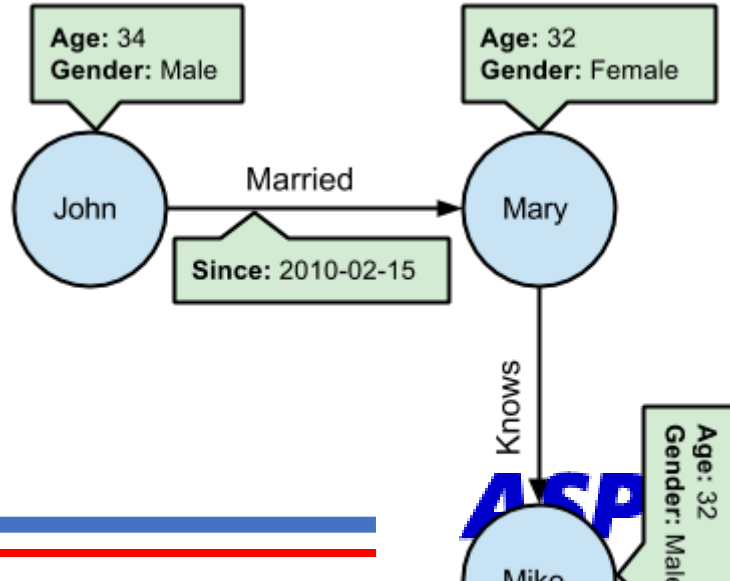
4 – base orientée Graph

Une base de données orientée graphe est une base de données orientée objet utilisant la théorie des graphes, donc avec des nœuds et des arcs, permettant de représenter et stocker les données en ce qui concerne le domaine des graphes. Il existe beaucoup de modèles de graphes différents. l'information dans un graphe attribué est modélisée grâce à trois blocs de base:

- le **nœud** ou sommet (*Node, vertex*)
- la **relation** ou arête (*relationship, edge*), avec une orientation et un type (orienté et marqué)
- la **propriété** ou attribut (*property, attribute*), portée par un nœud ou une relation







4 – base orientée Graph/2

Avantages des bases de données orientées graphe

Les bases de données orientées graphes apportent des avantages non négligeables :

Performances accrues :

- Traiter des données fortement connectées en évitant les multiples jointures très coûteuses qu'il faudrait mettre en œuvre dans les bases de données relationnelles traditionnelles et ainsi permettre des mises à jour très performantes, même pour un très grand ensemble de données.
- Offrir des performances exceptionnelles en termes de rapidité de temps de réponse pour les lectures locales, par parcours de graphe.

Développements simples :

- L'utilisation de langages de requêtes tels que Cypher ou Gremlin destinés au traitement des données connectées facilite les développements. Par exemple, la recherche d'amis prend une seule ligne de code en Cypher.

Modélisation facile :

- Gérer facilement un modèle complexe puisque la base de données ne s'appuie pas sur un schéma rigide.
- Permettre une modélisation parfois plus naturelle et plus lisible selon le cas d'utilisation.
- Découverte de nouveaux cas d'usages par une représentation naturelle des données.

- Neo4j - Open Source, Java, Modèle de Graphe Attribué
- AllegroGraph - Source fermée, RDF-QuadStore
- HypergraphDB - Open Source, Java, modèle Hypergraphe
- Sones - Source fermée, orienté .Net
- Virtuoso - Source fermée, orienté RDF

BASES NoSQL

Présence sur le marché

Vision **de Forrestear** ci-contre

Présence sur le marché

- MongoDB
- CouchDB
- Microsoft
- Amazon
- Google

Vision classement de **DB-engines**

		Rang	installation
NoSQL	• MongoDB	**5**	412
	• Redis	8	142
	• Cassandra	10	123
	• Neo4J	22	49
	• CouchDB	34	18
Multi-model	▪ Oracle	1	1355
	▪ Mysql	2	1283
	▪ Microsoft	3	1094
	▪ PostgreSQL	4	483
	▪ IBMdb2	6	170

ASPROM

BASES de DONNÉS AUTRES que SQL/NoSQL

L' évolution rapide du Big Data entraine l'utilisation de nouvelles données structurés ou non structurés,et un* accroissement des volumes traités, avec une emande de rapidité dans l'analyse des données.
De nombreux projets souvent associés à des start-up développent ces projets sur des bases nouvelles.
De nouvelles structures pour le stockage des données . Les grands fournisseurs proposent l'integration des dnnées brutes en complément des outils standards
sont utilisés , parfois sur la base de produits existant pour l'analyse et la gestion des données

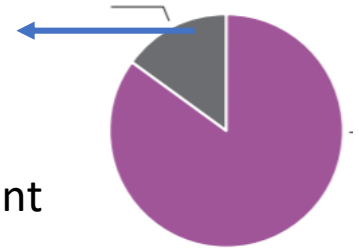
Nous présenterons succinctement quelques évolutions :

1 – DataLake (lac de données)

Un lac de données est une **collection de données**, pas une plate-forme pour les données. De la manière dont une base de données (définie comme une collection de données non structurées et d'éléments connexes) est gérée à l'égard d'un logiciel d'entreprise appelé système de gestion des bases de données relationnelles (RDBMS), un lac de données est une collection de données (ou de collections) qui est généralement gérée sur **Hadoop**, moins souvent avec un RDBMS.

Un lac de données est-il un problème ou une opportunité

15%-**Problème**,
parce que à datalake est difficile à sécuriser et à gouverner et nos compétences Hadoop sont immatures



85%- **Opportunité**, parce qu'elle modernise les données existantes écosystèmes et permet un plus large éventail de analytiquepour les utilisateur

ASPROM

1 – Data Lake (lac de données)/2

Avantages des lacs de données

- Analyse avancée
- Nouvelles pratiques des données
- Valeur commerciale du bigdata
- Modernisation des enterpots
- Diversrs structures des données

Obstacle des lacs de dionnées

- Gouvernance des données et confidentialité
- Intégration des données
- Experience du bigdata
- Analyse de la rentabilité
- Technologie immature

autre. Un certain nombre de répondants au sondage ont cité des obstacles fondés sur « la réticence [des utilisateurs professionnels] d'apprendre de nouveaux outils » et « la réticence des gens à changer et à permettre quelque chose de nouveau ».

RDBMS pour les besoins relationnels

- Données de haute valeur reconnue
- Données calculées pour la plupart affinées
- Entités connues, suivies au fil du temps
- Les données sont conformes aux normes de l'entreprise
- Intégration des données d’avance
- Données transformées a priori
- Typiquement schéma sur écrire
- Amélioration des métadonnées A priori

- **Lac de données sur la base Hadoop**

- Hadoop pour des données diverses, évolutivité, faible coût
- Données des candidats de la valeur potentielle
- La plupart des données sources détaillées
- Matériel brut pour la découverte d'entités et de faits
- Fidélité au format et à l'état originaux
- Préparation des données à la demande
- Données réaménagées ultérieurement, au fur et à mesure que les besoins se font sentir
- Typiquement schéma sur lecture
- Les métadonnées développées à la lecture dans de nombreux cas

1 – DataLake (lac de données)/3

Les Data Lakes disposeront de dizaines de milliers de tables / fichiers et de milliards d'enregistrements Pire encore , ces données sont non structurées et très variables.

Dans cet environnement, la recherche est un outil indispensable:

Pour trouver les tables dont vous avez besoin - en fonction du schéma et du contenu de la table

Pour extraire des sous-ensembles d'enregistrements pour un traitement ultérieur

Travailler avec des ensembles de données non structurés (ou structurés de manière inconnue)

Et surtout, gérer l'analyse à grande échelle

Seuls les moteurs de recherche peuvent effectuer des analyses en temps réel à une échelle d'un milliard d'enregistrements à un coût raisonnable.

Les moteurs de recherche sont l'outil idéal pour gérer le lac de données d'entreprise car:

Les moteurs de recherche sont faciles à utiliser - Tout le monde sait comment utiliser un moteur de recherche.

Les moteurs de recherche sont sans schéma -

Les schémas n'ont pas besoin d'être prédéfinis.

Les moteurs de recherche peuvent gérer des enregistrements avec différents schémas dans le même index.

Les moteurs de recherche atteignent naturellement des milliards de disques.

La recherche peut parcourir un contenu totalement non structuré.

Les principaux acteurs

- Coudera (Accenture)
- Amazon AWS
- Microsoft Azure
- Google Cloud

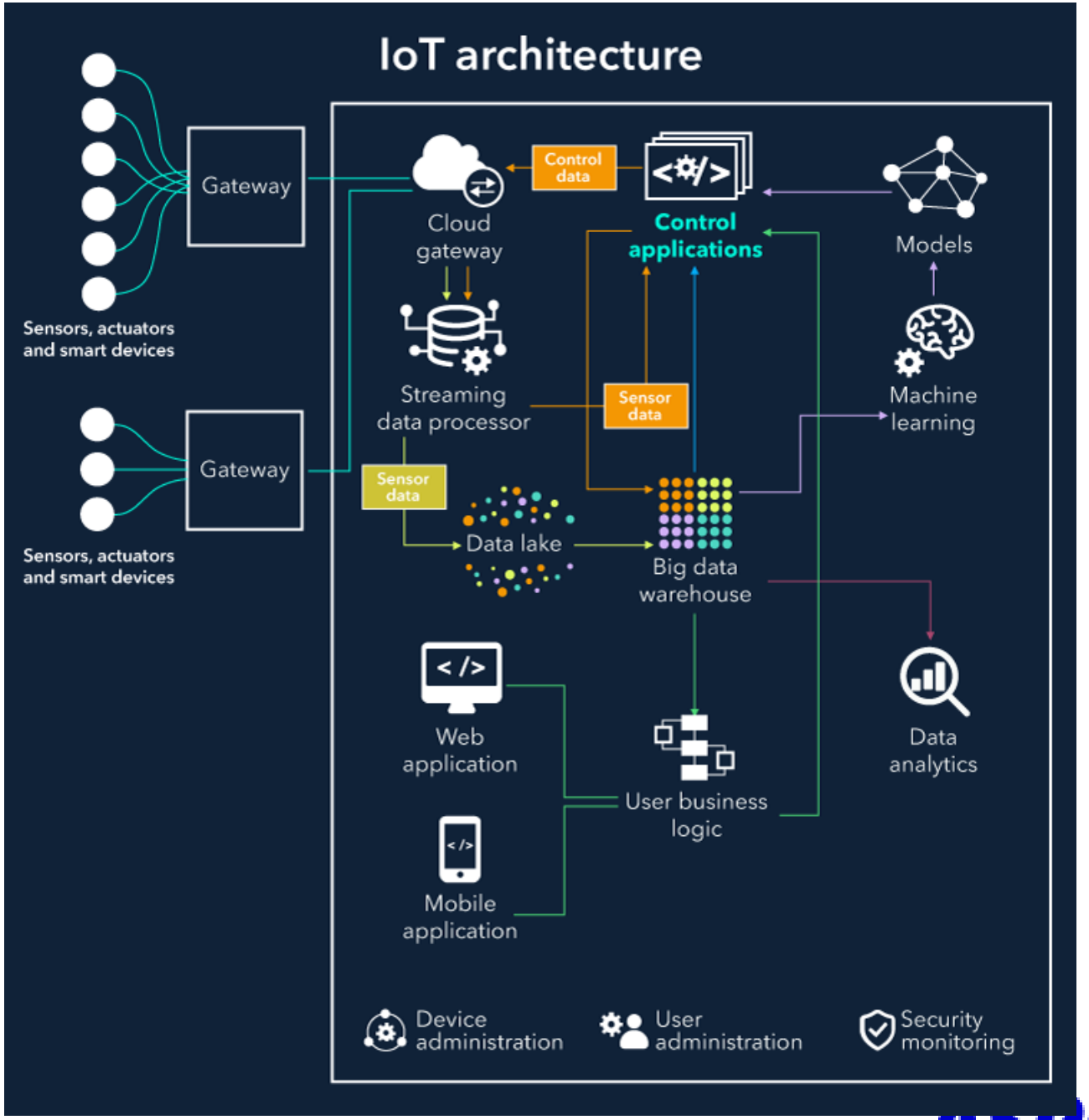
- Aspire d'Accenture
- Atacama
- Talent
- DataBricks (Apache)

- Indexima
- Search technologie
- Snowflake
- Zaloni

- Hortonworks/couldera
- IBM/Couldera
- Informatica

1 – Data Lake (lac de données)/4

Réalisation d’une architecture IOT
Basée sur un data lake



2 - NewSQL

NewSQL désigne une catégorie de systèmes de gestion de base de données (SGBD)relationnelles modernes qui cherchent à fournir la même puissance évolutive que les systèmes NoSQL pour letraitement en ligne transactionnel (lecture-écriture).

NewSQL est une catégorie de SGBD relationnelle moderne qui cherche à fournir :

- **La même puissance évolutive** (c'est à dire le faite de s'adapter à un changement d'ordre de grandeur, par exemple une forte demande) que le système NoSQL pour les applications concernant les traitement transactionnel en ligne (type d'application qui sert a modifier des informations en temps réel, par exemple des applications bancaire)
- **maintient les propriété ACID** d'un système de gestion de base de donnée traditionnel (atomicité, cohérence, isolation et durabilité).
- Elle tire aussi partie des évolutions du matériel et des nouvelle **architectures distribuées**.

caractéristiques

- Le SQL comme langage commun de requétage
- Transaction ACID
- Un mécanisme qui évite la pause de verrous lors d'opérations concurrentes de lecture avec les opérations d'écritures. La lecture en temps réel en est ainsi facilitée (moins de perte de temps).
- Une architecture qui a de meilleures performances par nœud que les solutions classiques de type SGBDR.
- Architecture distribuée
- la plupart utilise des **bases de données en mémoire**

fournisseurs

Clustrix; NuoDB; CoocroachDB; Pivarel ; Altibase;;MemSQL; VoltDB; C-tree ACE;Peercona; TokuDB; Apache Trafaddior; TiDCO; Active Spaces;ActorDB; Microsoft seerver

3 – In-memory data base– bases de données en mémoire

Une base de données dite « en mémoire » **(in-memory)**, ou **IMDB** (**In Memory DataBase**), ou encore MMDB (Main Memory DB), désigne une base de données dont les informations sont stockées en mémoire centrale **afin d'accélérer les temps de réponse**.

Les données source sont chargées dans la mémoire vive (RAM) du système, dans un format compressé non relationnel. Les bases de données en mémoire rationalisent les tâches qu'implique le **traitement des requêtes**.

Les bases IMDB volatiles basées sur la mémoire peuvent, et le font souvent, prendre en charge les trois autres propriétés ACID que sont l'atomicité, la cohérence et l'isolation

- Fichiers de capture instanée
- Journalisation ds transaction
- DIMM non volatile (partie DRAM avec mémoire NAND)
- DIMM non volatile en RAM sauvegardé par batterie
- Implantation à haute disponibilité

Solutions hybrides

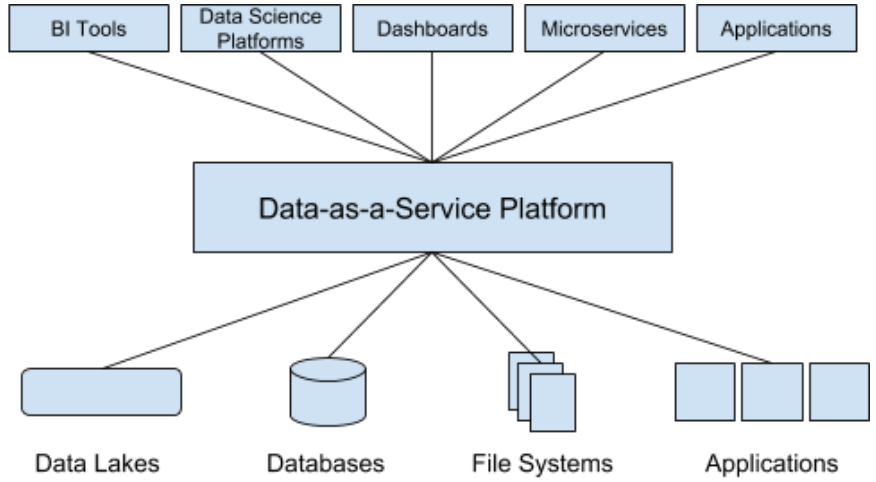
Certaines données peuvent être transférées de maniéré transparente sur disque ou sur le cloud ou adoptées une solution Edge Computing

La flexibilité des approches hybrides permet de trouver un équilibre entre: **performances** (améliorées par le tri, le stockage et la récupération des données spécifiées entièrement en mémoire plutôt que sur le disque)

- **coût**, car un disque dur moins coûteux peut être remplacé par plus de mémoire
- **facteur de forme**, car les puces de RAM ne peuvent pas atteindre la densité d'un petit disque dur
- **persistance**

Data As A Service -DaaS

Les données en tant que service (**DaaS)** est un modèle de fourniture et de distribution d'informations dans lequel des fichiers de données (y compris des textes, images, sons et vidéos) sont mis à la disposition des clients via un réseau, généralement Internet. Le modèle utilise une technologie sous-jacente basée sur le **cloud-Computing** Les informations DaaS sont stockées dans le cloud et sont accessibles via différents appareils. Le service transfère également les inconvénients de la gestion des données au fournisseur de cloud.



Les tarifications sont fonctions de :

- En fonction de la quantité des données
- En fonction de l'utilisation du service
- En fonction du type et des attributs des données
- Des sauvegardes (Backup As A Service)
- Parfois louer les données

Avantages

- Agilité, accès facile
- Données de haute qualité
- Rentabilité
- Interface utilisateur

inconvenient

- Confidentialité
- Sécurité
- gouvernance

intervenant

- Fournisseur de cloud
- Société de service spécialisés
- Fournisseur de données

Grands fournisseurs de solution Daas su Cloud

- Amazon
- Microsoft
- Google
- Oracle
- Cisco

ASPROM

TIME SERIES DATABASE – base de données chronologiques

Une **Time Series DataBase (TSDB)** est une base de données optimisée pour le stockage de données horodatées, telles que les données générées par l'internet des objets. **Time Series Data,** s'agit d'une séquence de points de données collectés à intervalles réguliers sur une période de temps de n'importe quelle donnée horodatée.

Les objets IOT du Big Data sont souvent équipés de capteurs horodateurs

- Véhicules
- Les équipements d'energie et de climatisation
- Appareils électroménagers
- Vêtements
- Machines de production
- Les humains

Dans ce contexte, le nombre de données **time series** a littéralement explosé et ces données sont produites en flux ininterrompu et avoir des temps de reponse brefs.
C'est la raison pour laquelle les bases de données **Time Series** sont devenues très importantes et très utilisées. Pour **prendre en charge l'immense volume de données horodatées** en provenance de multiples sources, les infrastructures de données doivent évoluer au même titre que le développement, la surveillance, le contrôle et la gestion des systèmes informatiques.

La spécificité de ces données n »est pas directement conciliable avec les bases **SGBD** ou **NoSQl**

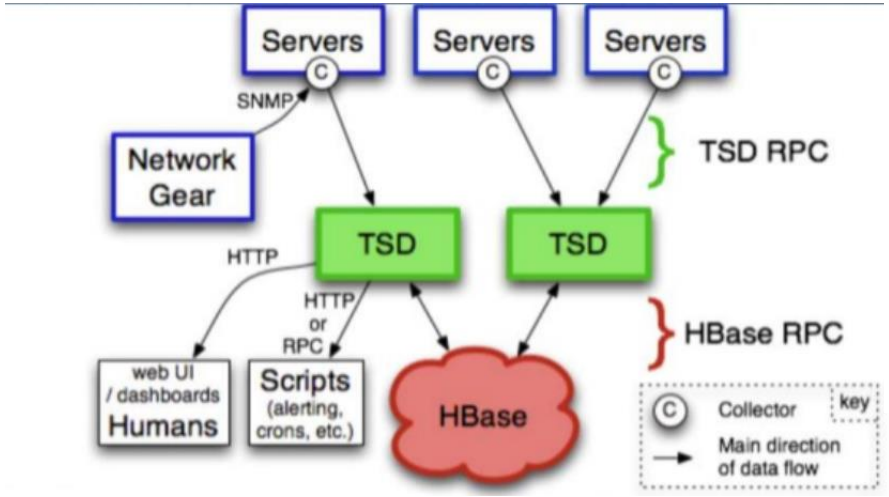
Les données sont souvent exprimées per une paire KVP: **clé-valeur** (Key-value pairs) , puis mises **en container**

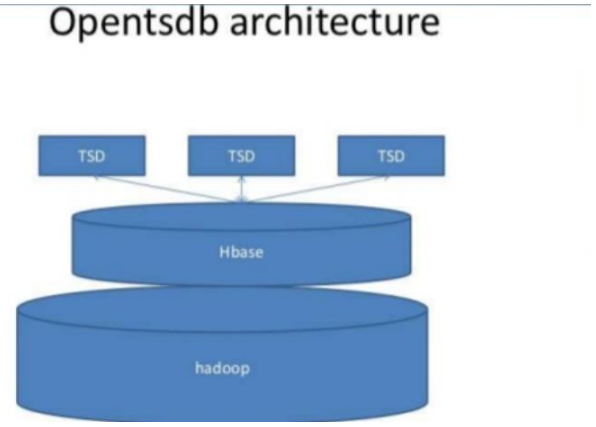
TIME SERIES DATABASE – base de données chronologiques

TSDB - Container

1 - HBASE

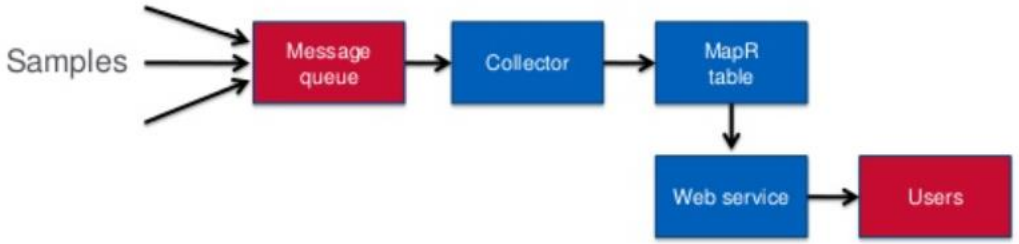
Apache HBase est un Magasin utilisant des paires clé/valeur. Il est conçu pour s'exécuter sur le système de fichiers **HDFS** **Hadoop** est une infrastructure qui permet de gérer des ensembles de données volumineux dans un environnement informatique distribué.





2 - MAPR

L'entreprise MapR propose une plateforme Big Data regroupant les composants Apache Hadoop et Spark, base de données en temps réel et un espace de stocka (Google, Amazon elasticSearch , MapReduce)



Key	13	43	73	103	blob
…					
series-uid.time-window	4.5	5.2	6.1	4.9	{t:[13,43,73,103], v=[4.5,5.2,6.1,4.9]}
…					

Will it Scale?

ASPROM

REPARTITION DES DONNÉES

1 - DATA CENTER –Centre de données

Un data center ou centre de données, est une infrastructure composée d'un réseau d'ordinateurs et d'espaces de stockage. Cette **infrastructure peut être utilisée par les entreprises pour organiser, traiter, stocker et entreposer de grandes quantités de données**.

Data centers de quelques grands acterurs du Big Data

- **Google data Center :** 7000 serveurs et 16 data center ' 1 M de serveurs
- **Facebook :** 5 Data Centers
- **Amazon :**7 Data Centers – 450 000 serveurs
- **Microsoft** : 1 million de serveurs

Classement des datacenter

- **Tier 1** : une alimentation, disponibilité :99,67%, 1 arrêt maintenance
- **Tier2** : alimentation redondée, disponibilité;99,74 %, arrêt/an :22 h
- **Tier3** :entièrement redondé, disponibilité: 99,98%, arrêt : 1,6 h/an;
- Tier4 : redondé composant sans influence, disponibilité :99,995%, arrêt ,8 h/an

Data Center de proximité

Des applications comme la voiture autonome, les villes intelligentes, la santé ou la virtualisation remettent en cause cette centralisation. Pour faire face à cette explosion de données à traiter dans un temps très court et sans embouteiller les autoroutes de l'information, il faut mettre en place une infrastructure décentralisée de traitement et de gestion de toutes les données « en bordure du réseau » : c'est l'« **Edge computing** ». Appelé parfois **« Fog computing »**

Datacenterrben conteneur ou inclus dans entreprise

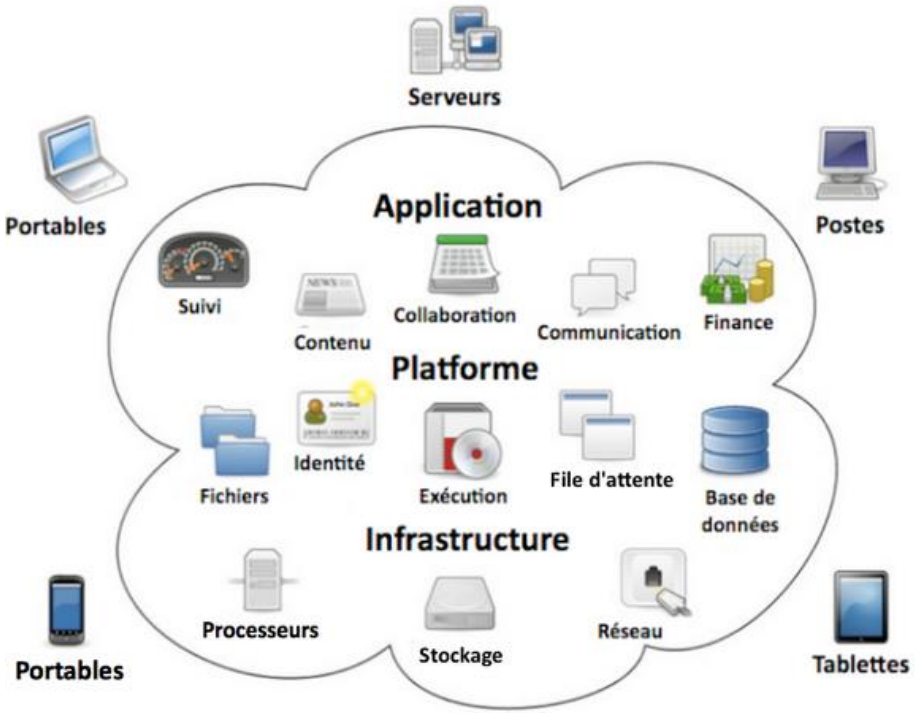
REPARTITION DES DONNÉES

2- CLOUD

Le ***cloud computing***, en français l'**informatique en nuage** (consiste à utiliser des serveurs informatiques distants par l'intermédiaire d'un réseau, généralement Internet, pour stocker des données ou les exploiter.
Un **serveur informatique** est un dispositif informatique (matériel ou logiciel) qui offre des services, à un ou plusieurs clients (parfois des milliers) en mode client-serveur.

Un nuage est caractérisé par sa disponibilité mondiale

- en libre-service, adaptation à la demande
- l'élasticité, adaptation capacité, vitesse
- l'ouverture, accès pc, tablette, téléphone
- la mutualisation , ressources hétérogènes pour plusieurs clients
- le paiement à l'usage



Nuage :

- **Public** :Un nuage public est mis à disposition du grand public. Les services sont généralement mis à disposition par une entreprise utilisant une infrastructure lui appartenant
- Privé :Un nuage privé est destiné exclusivement à une organisation qui peut le manipuler elle-même
- **Communautaire :**utilise une infrastructure provenant d'un ensemble de membres partageant un intérêt commun,.

ASPROM

REPLICATION des DONNÉES

La réplication des données est un élément très important de la maintenance de la base de données et du traitement des requêtes. Il existe généralement deux méthodes de protection des systèmes de stockage les plus récentes

- RAID _ gestion par des disques (reduncy Array of independent Disks)
- RAIN/RAIS – gestion par logiciel de nœuds (Reduncy Array of independant Nodes) /Servers

Répliquer, est un Outil indispensable, universel pour la robustesse des systèmes distribués notamment pour les dispositifs **NoSQL**

- **Tolérance** aux pannes Vous cherchez un document sur S1, qui est en panne ? On le trouvera sur S2
- **Distribution** des lectures Répartissons les lectures sur S1,S2,...,Sn pour satisfaire les millions de requêtes de nos clients.
- **Distribution des écritures** ? Oui, mais attention, il faut réconcilier les données ensuite
- **Resilience des données –**restoration des données auvegar
- et autres avantages, comme la construction d'un index sur un des serveurs sans affecter les autres, Le bon niveau de réplication?

R3 : Trois copies pour une sécurité totale (deux au minimum).

1 - Réseau redondant de disques (RAID)

•RAID1 - C'est ce qu'on appelle un «miroir de disque». Nécessite au moins deux lecteurs. (1X) pénalité d'écriture
• RAID5 - Nécessite au moins (3) disques pour créer une matrice RAID5. Peut supporter (1) lecteur perdu. (4x) écrire pénalité.
• RAID6 - Nécessite au moins (4) disques pour créer une matrice RAID6. Peut supporter (2) disques perdus. (5x) pénalité d'écriture.
• RAID10 - Nécessite au moins quatre disques. Combine le striping et la mise en miroir. N'a pas de parité et est une simple réplique de l'écriture.

Voici quelques raisons pour lesquelles le RAID **n'est pas apprécié**:

- Le RAID nécessite plus de disques par serveur.
- RAID nécessite une maintenance rapide
- Le RAID dégrade les performances des lectures et des écritures
- Le RAID à serveur unique ne fournit pas de protection contre une panne de serveur.
- La technologie RAId est bien adapté aux bases SQL (SGBD classique)

REPLICATION DES DONNÉES

DISQUE RAID

La technologie RAID (acronyme de Redundant Array of Independent Disks, (Ensemble redondant de disques indépendants) permet de constituer une unité de stockage à partir de plusieurs disques durs L'unité ainsi créée (appelée grappe) a donc une grande tolérance aux pannes (haute disponibilité), ou bien une plus grande capacité/vitesse d'écriture. La répartition des données sur plusieurs disques durs permet donc d'en augmenter la sécurité et de fiabiliser les services associés

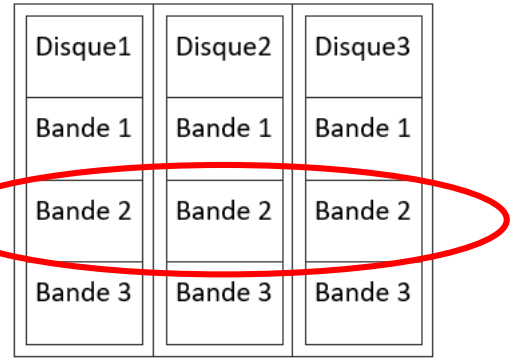
Niveau 0

Le niveau RAID-0, appelé striping (bande), consiste à stocker les données en les répartissant sur l'ensemble des disques de la grappe.

Niveau 1

Le niveau 1 a pour but de dupliquer l'information à stocker sur plusieurs disques, on parle donc de mirroring, ou shadowing pour désigner ce procédé. Ce type est obsolete (code erreur ECC dans disque=)

Niveau 0



Niveau 1

ASPROM

REPARTITION DES DONNÉES /2

DISQUE RAID /2

Niveau 3

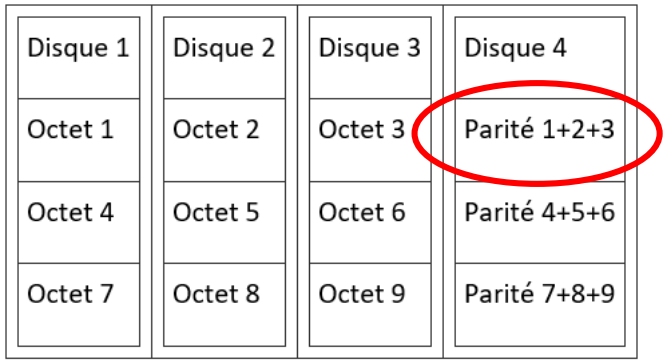
Le niveau 3 propose de stocker les données sous forme d'octets sur chaque disque et de dédier un des disques au stockage d'un bit de parité.

Niveau 4

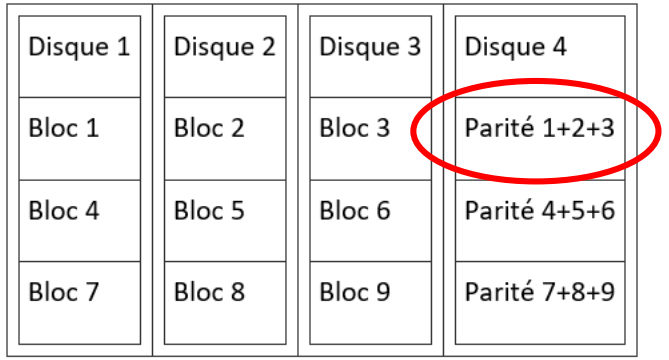
Le niveau 4 est très proche du niveau 3. La différence se trouve au niveau de la parité, qui est faite sur un secteur (appelé bloc) et non au niveau du bit, et qui est stockée sur un disque dédié

Niveau 5

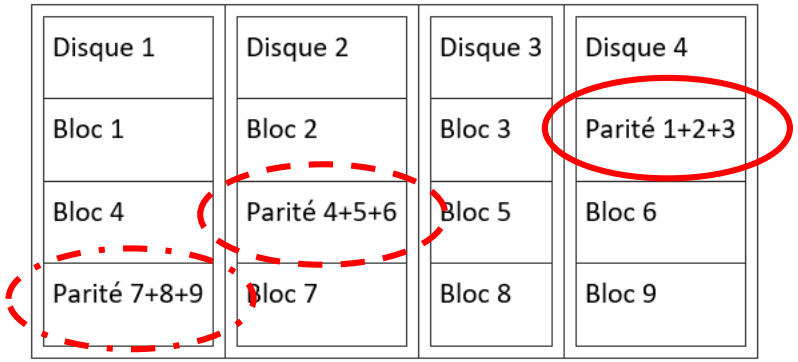
Le niveau 5 est similaire au niveau 4, c'est-à-dire que la parité est calculée au niveau d'un secteur, mais répartie sur l'ensemble des disques de la grappe.



niveau3



niveau4



niveau5

ASPROM

REPLICATION DES DONNÉES

DISQUE RAID /3

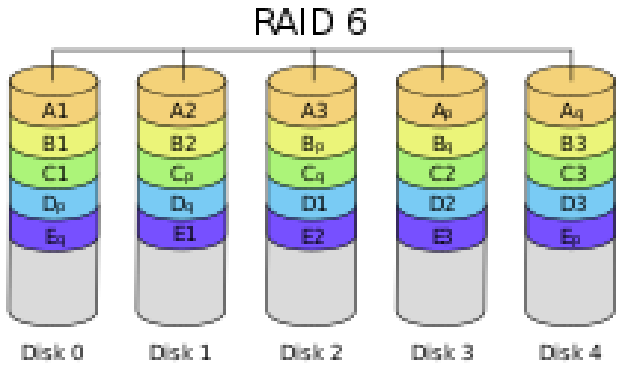
RAID 6

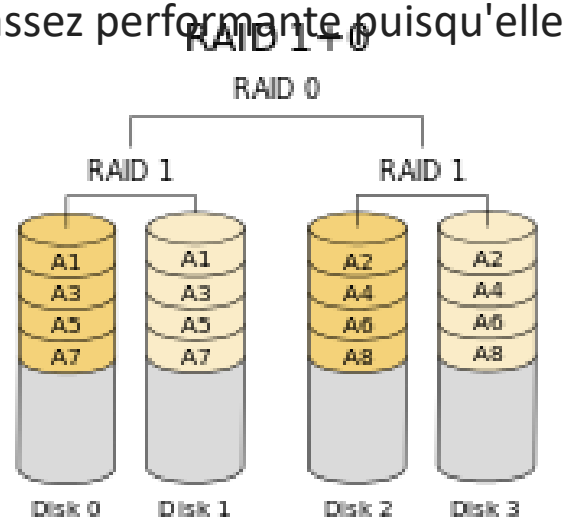
Le RAID 6 est une évolution du RAID 5 qui accroît la sécurité en utilisant ***n*** informations redondantes au lieu d'une. Il peut donc résister à la défaillance de ***n*** disques. Les fondements mathématiques utilisés pour les informations de redondance du RAID 6 sont beaucoup plus complexes que pour le RAID). Cette Structure améliore la résilience

RAID 10 (RAID 1+0)

Il permet d'obtenir un volume agrégé par bande avec un bon niveau de fiabilité (puisque basé sur des grappes répliquées). Chaque grappe contenant au minimum deux éléments et un minimum de deux grappes étant nécessaire, il faut au minimum quatre unités de stockage pour créer un volume RAID 1+0.

Sa fiabilité est assez grande puisqu'il faut que tous les éléments d'une grappe soient défectueux pour entraîner un défaut global. La reconstruction est assez performante puisqu'elle ne mobilise que les disques d'une seule grappe et non la totalité.





ASPROM

DISQUE RAID /3

Les solutions RAID généralement retenues sont le RAID de niveau 0, 1 et le RAID de niveau 5.

Le choix d'une solution RAID est lié à trois critères :

- Performances (niveaux 0 , le plus performant)
- Sécurité (niveaux 1 et 5)
- Le cout (lié aux volumes des données)

ABANDON des disques DUR- RAID pour le Big Data

Le Big Data concerne de grands volumes de **données non structurées** complétées par une **analyse rapide** et des informations sont notées en quelques secondes.

En outre, les technologies RAID qui voient de grands volumes de données et d'analyse sont essentiellement des données qui sont **structurées** et voient les opérations d'analyse en cours d'exécution sur une base de lots qui s'étend nécessairement sur des **périodes longues** incompatible avec les besoins du BigData

RAID sur CLOUD

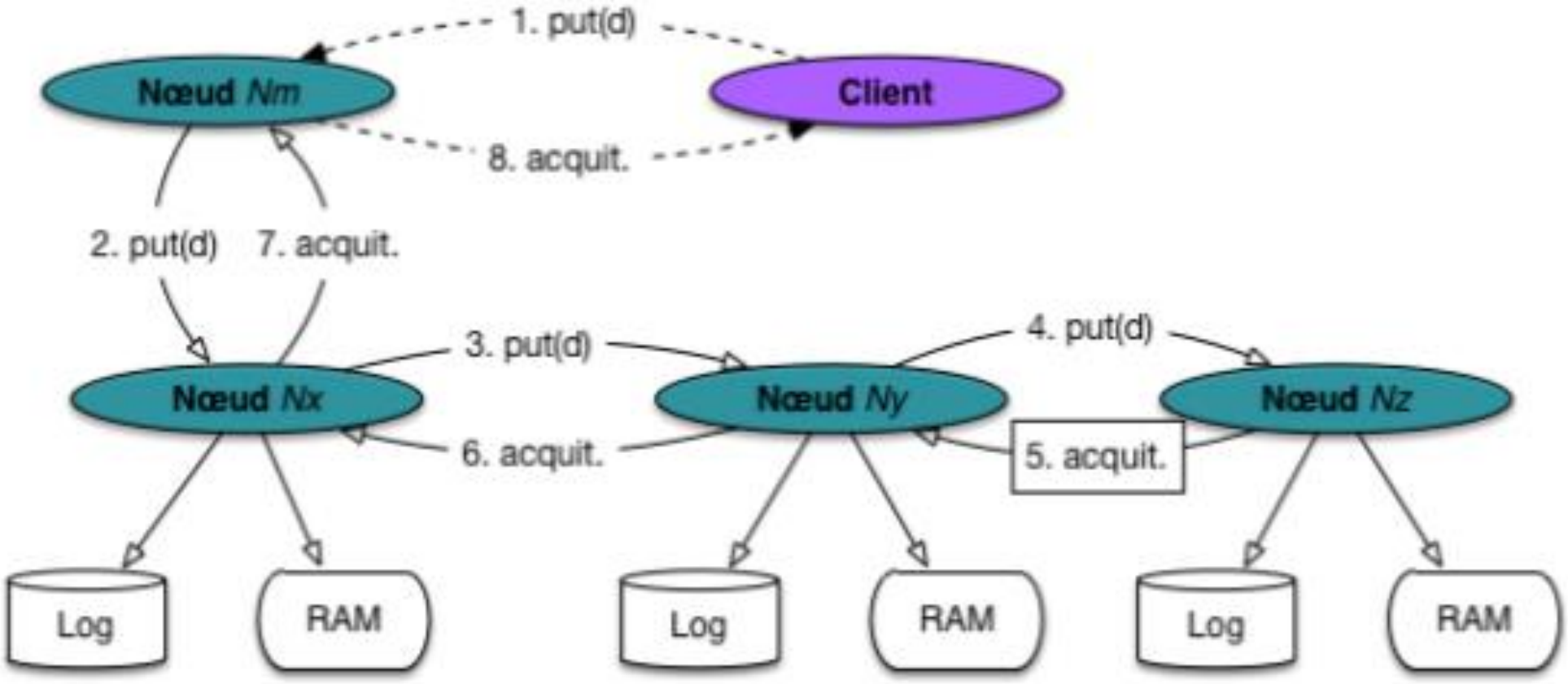
L'utilisation de la structure RAID sur le CLOUD supprmle les obstacle de l(utilisation des disques

Haute liaison :Plusieurs stockages en nuage sont utilisés comme un **système RAID5**, l'un des serveurs est utilisé comme un système de parité. Les données sont distribuées au système RAID5 des serveurs cloud. Même n'importe lequel des serveurs de nuages est en panne, vous pouvez toujours accéder à votre fichier en ligne.

Haute sécurité :Le fichier de données est compressé d'abord, puis crypté par AES 256bit; en outre, les données sont déshonorées au hasard aux blocs de données, puis les blocs de données sont téléchargés sur plusieurs serveurs cloud comme un système RAID5. Même les fournisseurs de services cloud ne peuvent pas divulguer vos données! Parce qu'ils ne détiennent qu'une partie de vos données.

2 – réseaux par gestion de nœuds :RAIN/RAINS

RAIN (réseau redondant de nœuds indépendants, ensemble de nœuds indépendants, est un groupe de nœuds connectés dans une topologie réseau avec de multiples interfaces et stockage redondant. RAIN est utilisé pour augmenter la tolérance aux défauts Il s'agit d'une implémentation de RAID à travers les nœuds au lieu de l'ensemble des disques.



R3 : Trois copies pour une sécurité totale (deux au minimum).

REPARTITION DES DONNÉES

NAS – Network Attache Storage

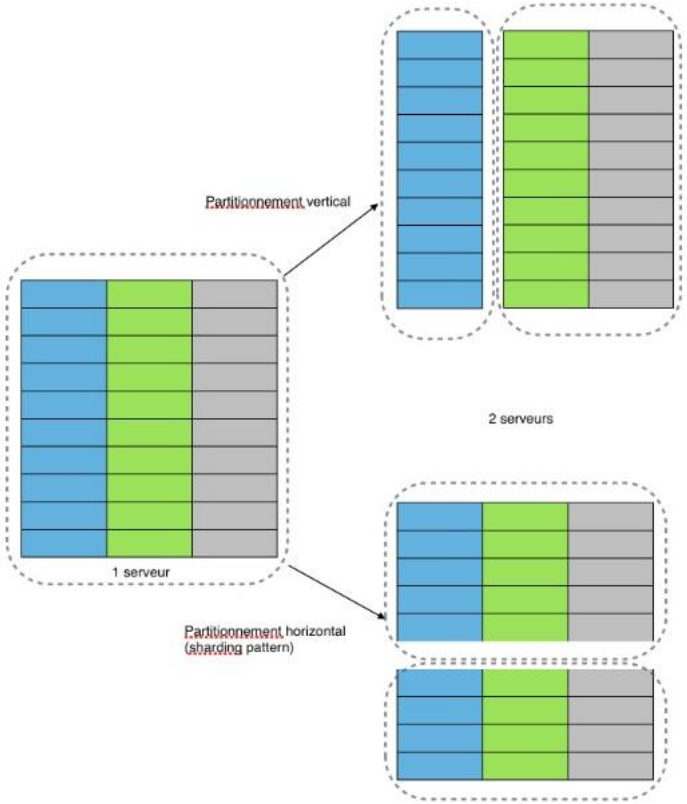
- Un serveur NAS est **un appareil de stockage de fichier connecté à un réseau local** (LAN). Ce type d'appareil permet le stockage et la récupération de données depuis une localisation centralisée pour les utilisateurs autorisés à accéder au réseau.
- Utiliser un Network Attached Storage revient à posséder un **Cloud privé sur site**. Le disque dur de ce type offre les mêmes avantages que le Cloud public **(cloud local).** Le NAS simplifie la sécurité
- Certains produits haut de gamme peuvent accumuler suffisamment de disques durs pour prendre en charge la technologie de **stockage RAIN à 3 Noeuds**
- Parmi les protocoles de partage de fichiers utilisés on compte sur le Network File System **(NFS),** le Common Internet File System (**CIFS**) ou l'Apple Filing Protocol **(AFP).**
- Le serveur NAS, et plus précisément le mode scale-out (ou en cluster) est idéal pour le Big Data pour plusieurs raisons. Tout d'abord, **ce type de configuration offre une grande flexibilité et scalabilité**. Si le volume de données à traiter augmente, il suffit d'ajouter des nœuds au cluster.
- Par ailleurs, **le stockage attaché en réseau en cluster permet de répondre à la problématique des grandes quantités de données non structurées** . En permettant de visualiser de larges volumes de données via un seul système fichier, le **NAS scale-out** s'avère idéal pour relever ce nouveau défi.
- NAS vs Cloud
  - NAS plus économique (service intensif)
  - NAS plus simple d'usage et olus fiable
  - Cloud plus de fonctionnalité

REPARTITION DES DONNÉES

SHARDING - ECLATEMENT

Les applications blockchain et les bases NoSQL sont à l'initiative Du développement de la technologie du SHARDING

L'éclatement consiste à fractionner ses données en deux ou plusieurs ***fragments*** plus petits, appelés *fragments logiques* . Les fragments logiques sont ensuite répartis sur des nœuds de base de données distincts, appelés *fragments physiques* , pouvant contenir plusieurs fragments logiques. Malgré cela, les données détenues dans tous les fragments représentent collectivement un ensemble de données logique complet.



le sharding de votre base de données implique de décomposer votre grande base de données en de nombreuses bases de données beaucoup plus petites qui ne partagent rien et peuvent être réparties sur plusieurs serveurs. Ces petites bases de données sont rapides, faciles à gérer, et sont souvent beaucoup moins chers à utiliser car ils sont souvent mis en œuvre en utilisant des bases de données sous licence open source.

La décomposition peut s »effectuer sur un ou plusieurs CLOUD

Le sharding n'est pas pris charge nativement dans les bases de données sauf MySQLCluster et MONGO Atlas

REPARTITION DES DONNÉES

SHARDING – ECLATEMENT

Architecture de sharding (ecaillage) : Architecture de partage- 3 méthodes

A) –éclatement à base de clés

Le partage basé sur les clés , également appelé *partage basé sur* ***le hachage*** , implique l'utilisation d'une valeur extraite de données nouvellement écrites - telles que le numéro d'identification du client et de le brancher à une *fonction de hachage* pour déterminer quel fragment les données doivent être gérer.

B – éclatement basé sur la portée

le sharding de données basé sur des plages **d'une valeur donnée**. Vous pouvez créer quelques fragments différents et diviser les informations de chaque produit en fonction de la gamme de prix dans laquelle elles se situent

C – éclatement basé sur un repertoire (annuaire)

il est nécessaire de créer et de gérer une *table de recherche* utilisant une clé de partage pour garder trace du fragment contenant les données. En un mot, une table de recherche est une table contenant un ensemble statique d'informations sur les endroits où trouver des données spécifiques.

REPARTITION DES DONNÉES

HADOOP – fichiers HDSF

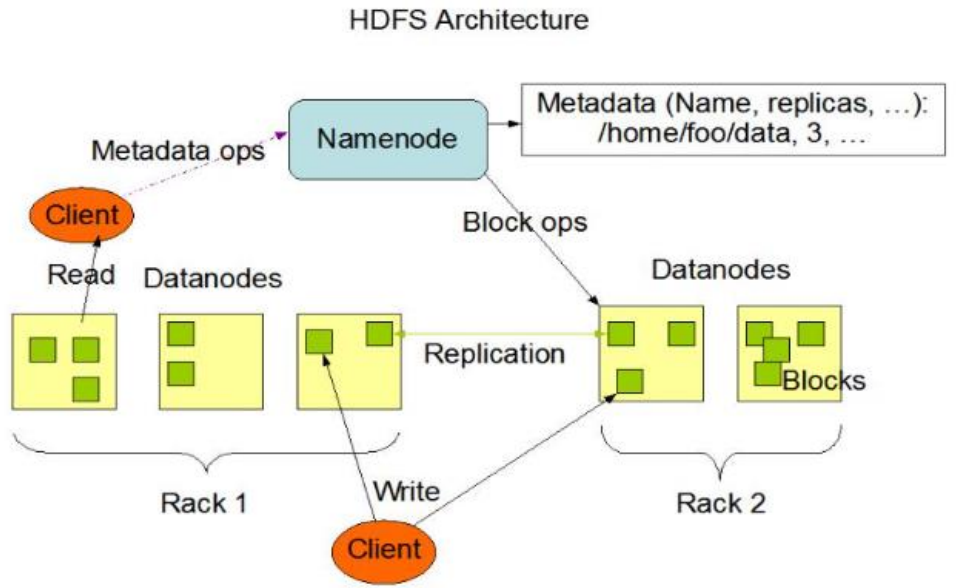
HDFS qui signifie "Hadoop Distributed File System" est un système de fichier, de stockage distribué émanant de **Hadoop Apache** et qui a été conçu pour prendre en charge de gros fichiers. Il est très adapté au **Big Data** et permet de procéder au traitement de ces grosses données.

Très efficace pour le traitement des données

- -Système très fonctionnel pour traiter les grosses données (Big Data)
- Chaque donnée est stockée à trois emplacements afin d'éviter la perte des données
- Déplacement des données afin d'éviter la congestion du réseau
- Sa portabilité
- Les données sont accessibles à partir de **MapReduce.**

architecture

- **NameNode** est le nœud maître de l'architecture Apache Hadoop HDFS qui maintient et gère les blocs présents sur les DataNodes (noeuds d'esclaves).
  - Enregistre les métadonnées
  - Gere la cartographie des blocks dans les nœuds
  - Gere la redondance (réplication)
- **DataNodes** sont les nœuds d'esclaves dans HDFS. Contrairement à NameNode, DataNode est un matériel de base



ASPROM

REPARTITION DES DONNÉES

BLACKCHAIN –objectif IOT

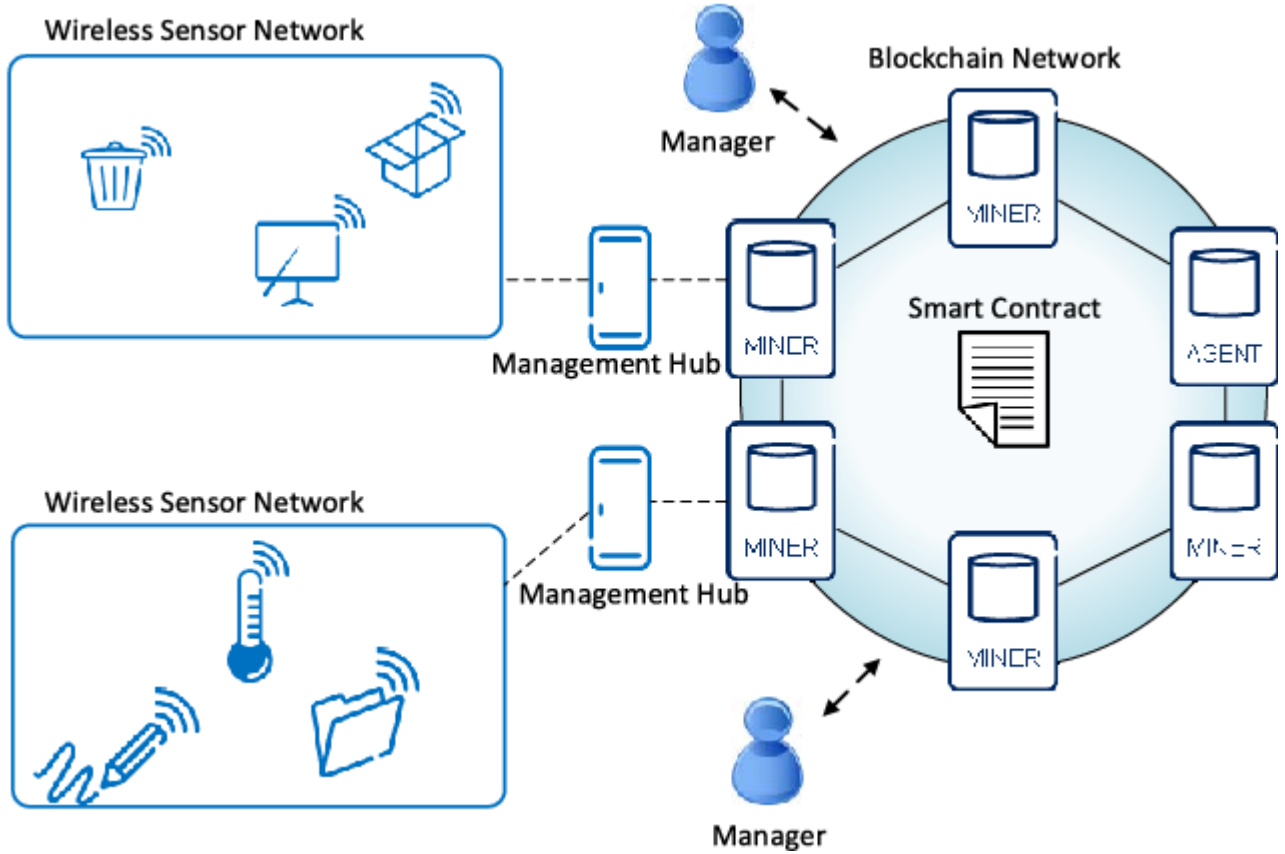
Logiquement, une blockchain est une chaîne de **blocs** contenant des informations spécifiques (base de données), mais de manière sécurisée et authentique, regroupée dans un réseau (d'égal à égal). En d'autres termes, la blockchain est une combinaison d'ordinateurs reliés les uns aux autres au lieu d'un serveur central, ce qui signifie que tout le réseau est décentralisé.

- Blockchain communautaire
- Blockgain privé
- Chain of things

- Désintermédiation (consensus)
- Sécurité
- Autonomie (puissance calcul)
- Evolution rapide vers IOT
- Grand livre (ledger)
- historisation

Nombreux projets et réalisation en IOT

- Filamant
- Iota
- Iotex
- IotChain
- IBM
- Ambrosus
- WaltonChain
- OriginTrail
- Stock.it
- BlockMesh
- Helium
- Moeco



ASPROM

OFFRE COMMERCIALES du cloud computing

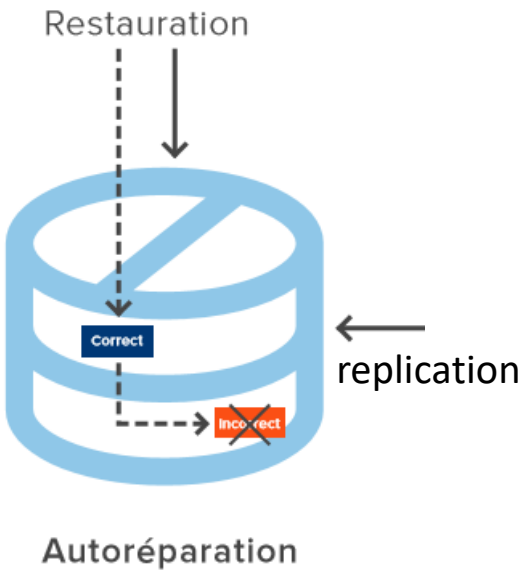
Les services *Cloud* représentent les ressources informatiques offertes. Les offres des éditeurs cloud dépendent du type de ressources offert. En fonction des types de ressources offert, on distingue 3 types d'offres commerciales Cloud (ou *taxonomie cloud*) : l'*IaaS*, le *PaaS* et le *SaaS*.

- **L'IaaS ou Infrastructure-as-a-Service** : est un niveau de service qui permet à des entreprises de louer des centres de données (*data Centers*), serveurs, réseaux, bref de s'équiper d'un système informatique complet sans s'inquiéter de créer et de maintenir la même infrastructure en interne ;
- **Le PaaS ou Platform-as-a-Service** est un niveau de service du *Cloud* dans lequel le fournisseur héberge et fournit un environnement de développement intégré (EDI) qu'un développeur peut utiliser pour créer et développer des logiciels, éliminant de ce fait la nécessité pour l'entreprise d'acheter constamment des EDI ;
- **Le SaaS ou Software-as-a-Service** : est un niveau de service du *Cloud* dans lequel le fournisseur héberge et met à disposition un logiciel ou une application, sans que l'entreprise n'ait à se soucier de l'installation et de la maintenance de cette application en interne ;

RESILIENCE des DONNÉES

Les entreprises stockent toujours plus de données, sur des périodes toujours plus importantes. De plus, ces données doivent être protégées. Plus spécifiquement, les entreprises doivent pouvoir restaurer leurs données de production depuis une sauvegarde en cas de perte ou de corruption, les sauvegardes doivent être aussi non corrompues

- **Evitement des défaillances** : Copie de sauvegarde
- **Détection de la corruption des données** :Les appliances série DR détectent les données corrompues selon les méthodes suivantes :



- La fonction **Verify on read** assure la vérification des données à chaque lecture
- **« Patrol Read** » est une technologie de détection en amont, active en permanence en arrière-plan. Elle vérifie les secteurs de tous les disques durs à la recherche d'anomalies.
- La fonction « **Vérification de cohérence » (CC)** est conçue pour détecter et corriger les incohérences entre les données des disques virtuels en RAID-6
- **La fonction ODIV (Online Data Integrity Verification**) vérifie de manière proactive et continue les données de l'appliance et les structures de données (datastores, cartes de blocs et serveur de métadonnées).
- La fonction **Replication** vérifie la cohérence des données avant leur réplication

- **Isolement des données corrompues** : suppression après création des données corrigées ecrassement du bloc « incorrect »tpar le bloc « correct »

RESILIENCE des DONNÉES /2

- **Autoréparation et correction :** corriger les données le plus rapidement :
  - **Self-healing** :(autoréparation) corrige les blocs de données identifiés comme corrompus et marqués comme incorrects
  - **Vérification de cohérence** :Corrige les incohérences dans les données des systèmes redondants
  - **Patrol Read** :agit au niveau du disque dur pour détecter et corriger les erreurs influant sur l'intégrité des données.
  - **Vérification de cohérence (CC)** est une vérification de l'intégrité au niveau des données, exécutée sur la base du RAID
  - **Vérification des opérations de lecture** est une vérification de l'intégrité des données de bout en bout de niveau du système, car la vérification est exécutée une fois que les données ont passé l'ensemble des couches physiques et
  - **Recovery manager** (RM: restaurer le logiciel système et les configurations définies par l'utilisateur.

- **Solution de stockage isolée au moyen d'un air GAP – attaque par rançongiciel** (ransonware) :

  Les solutions d'air gap permettent d'éviter la corruption des données en isolant une copie des données du réseau d'entreprise et en limitant l'accès à ces données. Cela permet de répliquer les données à la cible isolée pendant des périodes définies.

- **Respecter les regles sur kla confidentialité (GDPR et CCPA**

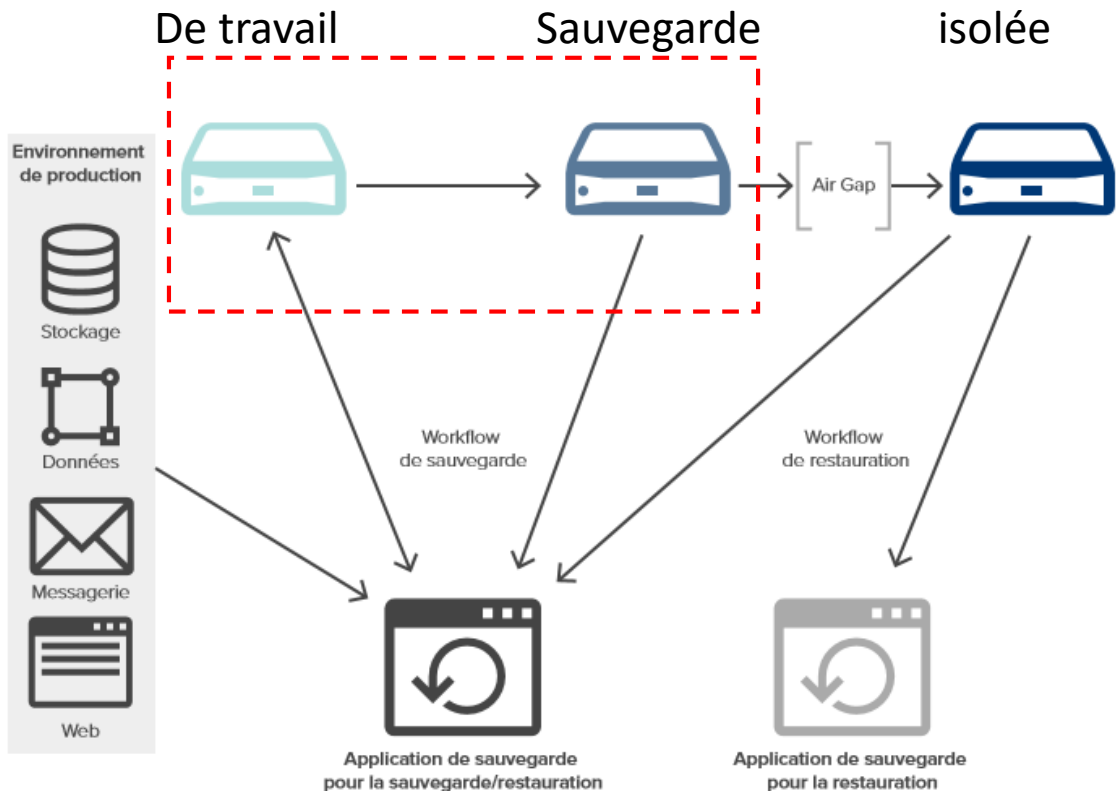
  Les réglementations sur la confidentialité, telles que GDPR et CCPA, stipulent comment les entreprises peuvent traiter, stocker et utiliser les données.

RESILIENCE des DONNÉES - 3

Fournisseur (Forester classement)

Architecture générale de principe

Stockage des données



ASPROM