

Module 106

Interroger, traiter et assurer la maintenance des bases de données

Karl Legrand

Qui suis-je ?

Activités



Formations

- ✓ CIE 105 - Langage SQL – VD & GE
- ✓ CIE 130 – Réseau – VD & GE
- ✓ CIE 340 – Virtualisation – VD
- ✓ CIE 106 - Langage SQL et maintenance DB – VD
- ✓ 12Harmos – Culture générale informatique



Développement Full-stack

- ✓ CRM
- ✓ Base de données SQL
- ✓ Outils DotNet spécifiques



Support

- ✓ Pour nos applications, Sage, etc.



CIE 106 : Déroulement



4.5 jours : cours présentiel (sauf ordre DGEP) / participation obligatoire



- ✓ Théorie : 0.5 jour
- ✓ Exercice et correction : 4 jours
- ✓ Pause matin, midi et après-midi

0.5 jour = examen individuel en ligne sur Moodle



- ✓ Avec document
 - QCM
 - Requête SQL à écrire à partir de MPD
 - Questions ouvertes



Pourquoi une base de données

(<https://www.oracle.com/fr/database/pourquoi-utiliser-base-de-donnees.html>)

Au commencement les feuilles de calcul

Elles sont d'excellents outils ... pour faire des calculs. Mais si vous avez beaucoup de données, par exemple sur vos clients, vos employés ou vos stocks, cela se complique (duplication de données, partage des données, sécurité des données, etc.)

Avantages d'une base de données

- Optimisation taille disque / taille des données
- Sécurité
- Partage des informations
- Cloud
- Performance



A quoi sert une base de données ?

1

Mémorisation de données

- Humain
- Ordinateur
- Internet Of Things

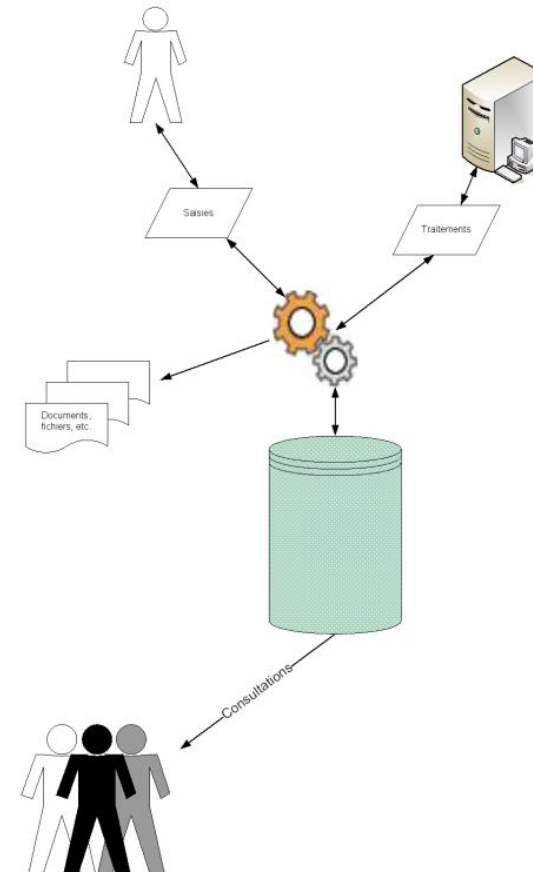
2

Traitement

Restitution

3

- Humain
- Internet Of Things
- IA
- Futurs calculs rétroactifs



Historique des bases de données



Quand	Organisation	Exemple	Avantages	Limites
Début de l'informatique et jusqu'aux années 60	Organisation classique en fichiers	Tableur	<ul style="list-style-type: none"> - Mise en œuvre simple - Pas ou peu d'accès multiutilisateurs 	<ul style="list-style-type: none"> - Particularisation de la saisie et des traitements en fonction des fichiers; un ou plusieurs programmes par fichier - Contrôle en différé des données ; Augmentation des délais et du risque d'erreur; - Grande redondance des données
Fin des années 60	Apparition des premiers SGBD (Systèmes de Gestion de Bases de Données), les systèmes réseaux et hiérarchiques	Adabas	<ul style="list-style-type: none"> - Centralisation des données - Rapide ; plus de 300 000 transactions par seconde 	<ul style="list-style-type: none"> - Organisation en arbre - Très limités en cardinalité 0-n ou 1-n car structure arborescente (hiérarchique) - Problème de rééquilibrage de l'arbre
À partir de 1970	Deuxième génération de SGBD, les systèmes relationnels – SGBDR	SQL serveur MySQL	<ul style="list-style-type: none"> - Adaptés aux cardinalités 0-n ou 1-n - Intégrité référentielle au niveau de la base de données ; facilité de codage - Orienté Gestion (données alphanumérique) - Les plus utilisés ; Standard SQL de fait 	<ul style="list-style-type: none"> - Plus complexes à mettre en œuvre - Nécessite AGL pour les bases de données complexes
Début des années 80	Troisième génération de SGBD, les systèmes orientés objet	Annuaire LDAP	<ul style="list-style-type: none"> - Représentation sous formes d'objet - Orienté multimédia (ex : radio, texte annotés, mp3, etc.) 	- Intérêt des grands constructeurs (Microsoft, Oracle, etc.) ?
Futur (proche ?)	Cloud Computing	Google « Fusion tables »	-Intégration simple de données multi-source et multi-format	



Les types de bases de données utilisées



Base de données orientée objets

- Les objets créés à l'aide de langage de programmation orientés objets sont généralement stockés sur des bases de données relationnelles. Toutefois, en réalité, les bases de données orientées objets sont plus adaptées pour stocker ce type de contenu.



Base de données relationnelle

- Inventées en 1970 par E.F. Codd de IBM.
- Constituées d'un ensemble de tables avec des colonnes classées par catégorie (type de données).
- Le langage pour parler à ces bases de données est le SQL (Structured Query Language).
- Facilement extensibles, et de nouvelles catégories de données peuvent être ajoutées après la création de la database originale sans avoir besoin de modifier toutes les applications existantes.



Base de données distribuée

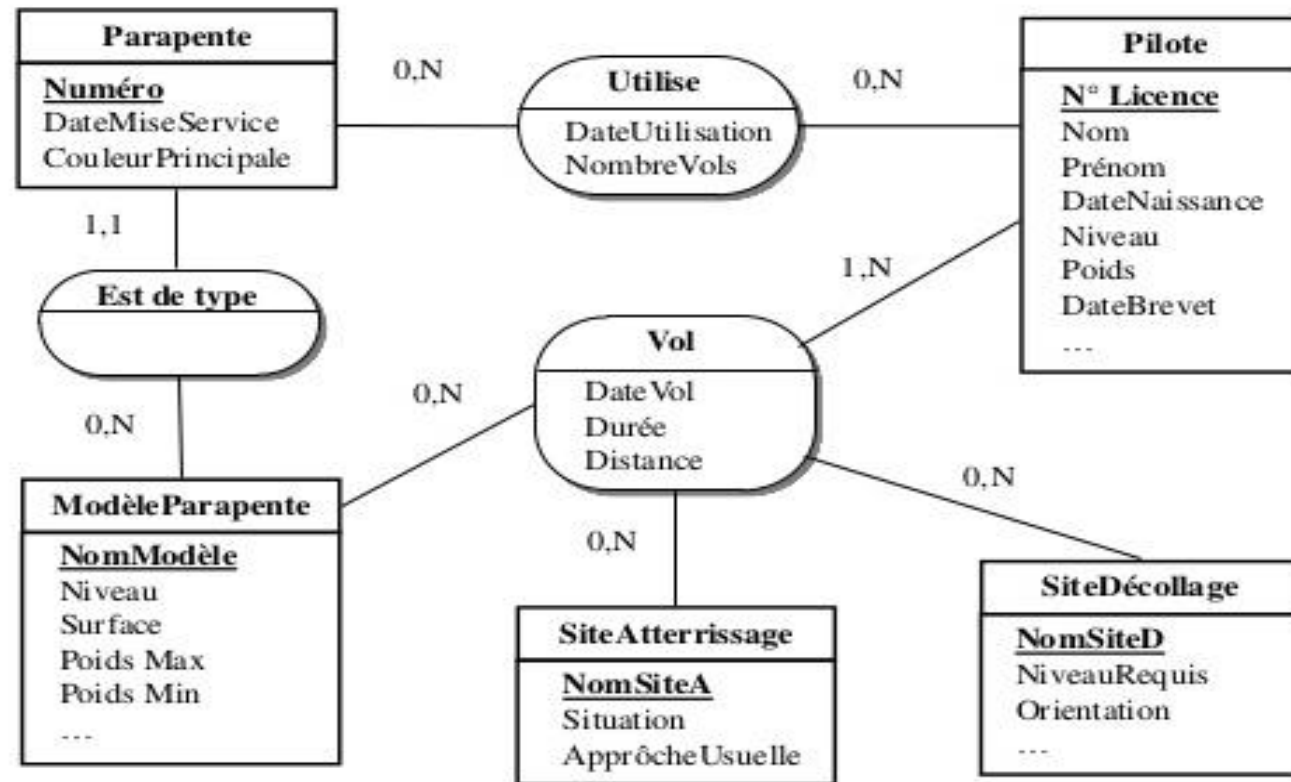
- Certaines portions sont stockées à plusieurs endroits physiques. Le traitement est réparti ou répliqué entre différents points d'un réseau.
- Peuvent être homogènes ou hétérogènes. Dans le cas d'un système de base de données distribuée homogène, tous les emplacements physiques fonctionnent avec le même hardware et tournent sous le même système d'exploitation et les mêmes applications de bases de données. Au contraire, dans le cas d'une database distribuée hétérogène, le hardware, les systèmes d'exploitation et les applications de bases de données peuvent varier entre les différents endroits physiques.

Base de données NoSQL

- Les bases de données NoSQL sont utiles pour les larges ensembles de données distribuées. En effet, les bases de données relationnelles ne sont pas conçues pour le Big Data, et les ensembles de données trop larges peuvent poser des problèmes de performances.
- Si une entreprise doit analyser d'importantes quantités de données non structurées, ou des données stockées sur plusieurs serveurs cloud virtuels, la database NoSQL est idéale. Avec l'essor du Big Data, les bases de données NoSQL sont de plus en plus utilisées.

Organisation des données (DB relationnelle)

MCD : Modélisation des données :



Organisation des données (NoSQL)

Organisation sous forme de collections

Redondance

Les bases de type NoSQL délaissent les points forts des bases relationnelles, que sont la notion d'enregistrement et les relations entre éléments, pour se focaliser sur la notion de "document".

La structure organisationnelle n'est plus liée à un schéma relationnel difficile à modifier, et la base peut donc croître sans contrainte.

D'autre part, l'orientation "document" facilite le déploiement de la base sur de multiples machines.

Lorsque la base devient trop importante, il suffit de définir de nouvelles machines connectées sur le réseau, un nouveau cluster, et la base NoSQL se débrouille.

sample_supplies.sales

COLLECTION SIZE: 4.13MB TOTAL DOCUMENTS: 5000 INDEXES TOTAL SIZE: 68KB

Find

Indexes

Schema Anti-Patterns 0

Aggregation

FILTER {"filter": "example"}

QUERY RESULTS 1-20 OF MANY

```
_id: ObjectId("5bd761dcae323e45a93ccff3")
saleDate: 2015-07-22T02:45:20.727+00:00
items: Array
  0: Object
    name: "envelopes"
    tags: Array
      price: 21.46
      quantity: 5
    1: Object
      name: "notepad"
      tags: Array
        price: 21.82
        quantity: 1
    2: Object
    3: Object
storeLocation: "London"
customer: Object
couponUsed: false
purchaseMethod: "In store"
```

```
_id: ObjectId("5bd761dcae323e45a93ccfe8")
saleDate: 2015-03-23T21:06:49.506+00:00
items: Array
storeLocation: "Denver"
customer: Object
couponUsed: true
purchaseMethod: "Online"
```

Fonctionnalité d'une base de données

La persistance

Le système doit assurer la **persistance** des données. Les données sont persistantes si elles survivent à l'exécution des programmes.

La gestion du disque

Le système doit pouvoir efficacement gérer une hiérarchie de mémoire. Il doit le faire avec des **performances** acceptables. Pour cela, il doit utiliser des techniques éprouvées de regroupement, d'indexation, d'optimisation de requêtes et de gestion de zone tampon et de cache.

Le partage des données

Le système doit permettre à plusieurs utilisateurs de manipuler les données concurremment. Il doit au minimum assurer la sériabilité des **transactions**.

La fiabilité des données

Le système doit assurer la fiabilité des données. Une transaction doit être **atomique**, c'est-à-dire soit exécutée complètement soit pas du tout. Des mécanismes de reprise sur panne mémoire centrale et de reprise sur panne disque doivent être disponibles. On doit pouvoir spécifier des points de reprise et y retourner.

La confidentialité des données

Le système doit assurer la **confidentialité** des données. Pour cela, il doit reconnaître la notion d'utilisateur, lui associer des droits de lecture, d'écriture et d'exécution sur les entités de la base. Le système doit supporter des mécanismes de cession et de retrait de droits.

L'interrogation ad hoc / SQL / MQL proche du javascript.

Le système doit permettre à l'utilisateur d'interroger la base de façon simple et naturelle. Cela signifie que, pour extraire des informations de la base de données, on ne doit pas avoir à écrire un programme. Un **langage de requête** est une façon de satisfaire ce besoin, mais une interface graphique peut être une autre possibilité. Le langage de requête doit être déclaratif (on spécifie ce que l'on veut savoir et non la façon dont l'information doit être évaluée), générique (indépendant d'une application spécifique) et optimisable (un optimiseur doit pouvoir transformer les requêtes en un code efficace).



Le langage SQL

A l'origine, SQL était un langage développé pour System-R, un prototype de système de base de données relationnelles conçu au laboratoire de recherche IBM de San José entre 1974 et 1977.

Les travaux réalisés sur System-R ont été repris dans 2 produits d'IBM apparus vers le début des années 1980 : DB2 (MVS) et SQLJDS (VM).



SQL est un langage destiné à manipuler une base de données relationnelles, à l'interroger et à la mettre à jour. Sa syntaxe est simple et accessible à des non-informaticiens.



Une instruction SQL comprend un verbe suivi d'objets permettant de préciser la portée du verbe

Le langage SQL

Le langage SQL comporte trois sous-ensembles d'instructions :

DML - DATA MANIPULATION LANGUAGE

Ce module comprend les instructions de manipulation de données. Nous y trouvons des instructions comme :

SELECT, INSERT, DELETE, UPDATE, ...

DDL - DATA DESCRIPTION LANGUAGE

Ce module permet la définition des données avec des instructions comme :

CREATE TABLE, ...

DCL - DATA CONTROL LANGUAGE

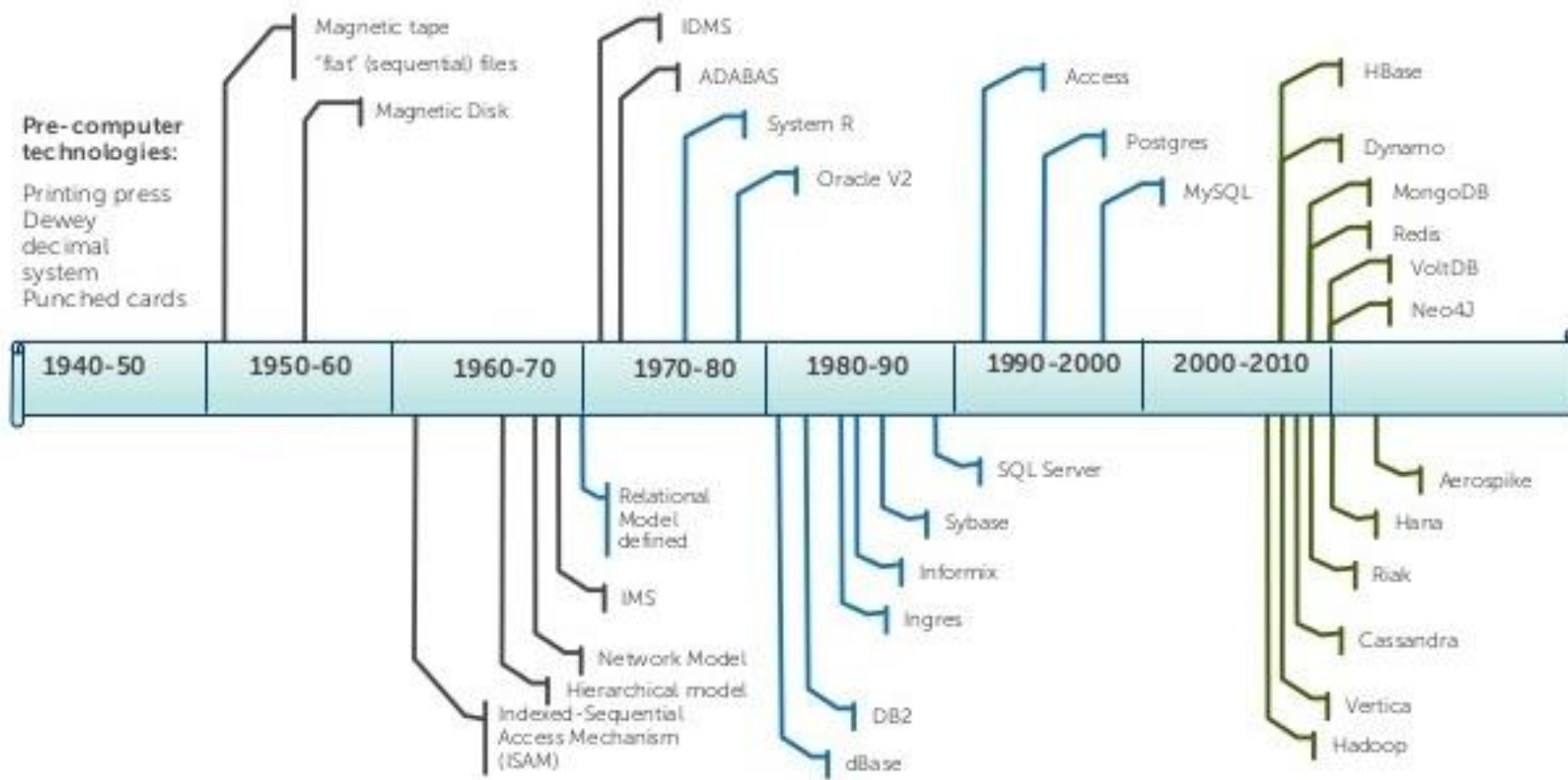
Ce module permet de garantir l'intégrité de la base en contrôlant les accès. Les principales instructions sont :

CREATE VIEW, GRANT, REVOKE, COMMIT, ROLLBACK,



SQL est un langage qui a été normalisé par l'ANSI (American National Standard Institut) et ISO. La dernière normalisation date de 2016.

Base de données - Time lapse

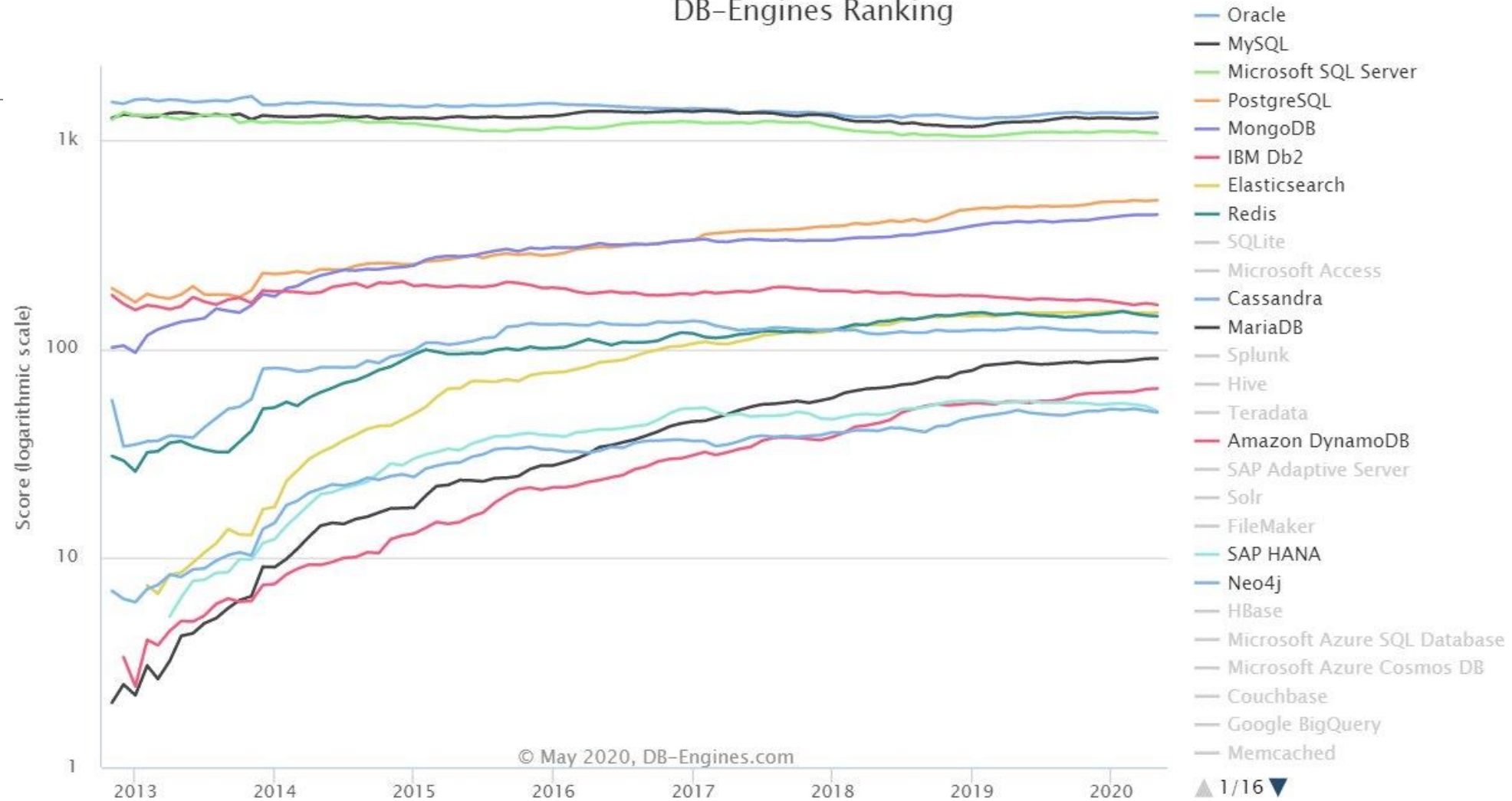


?

Base de données - Etat des lieux



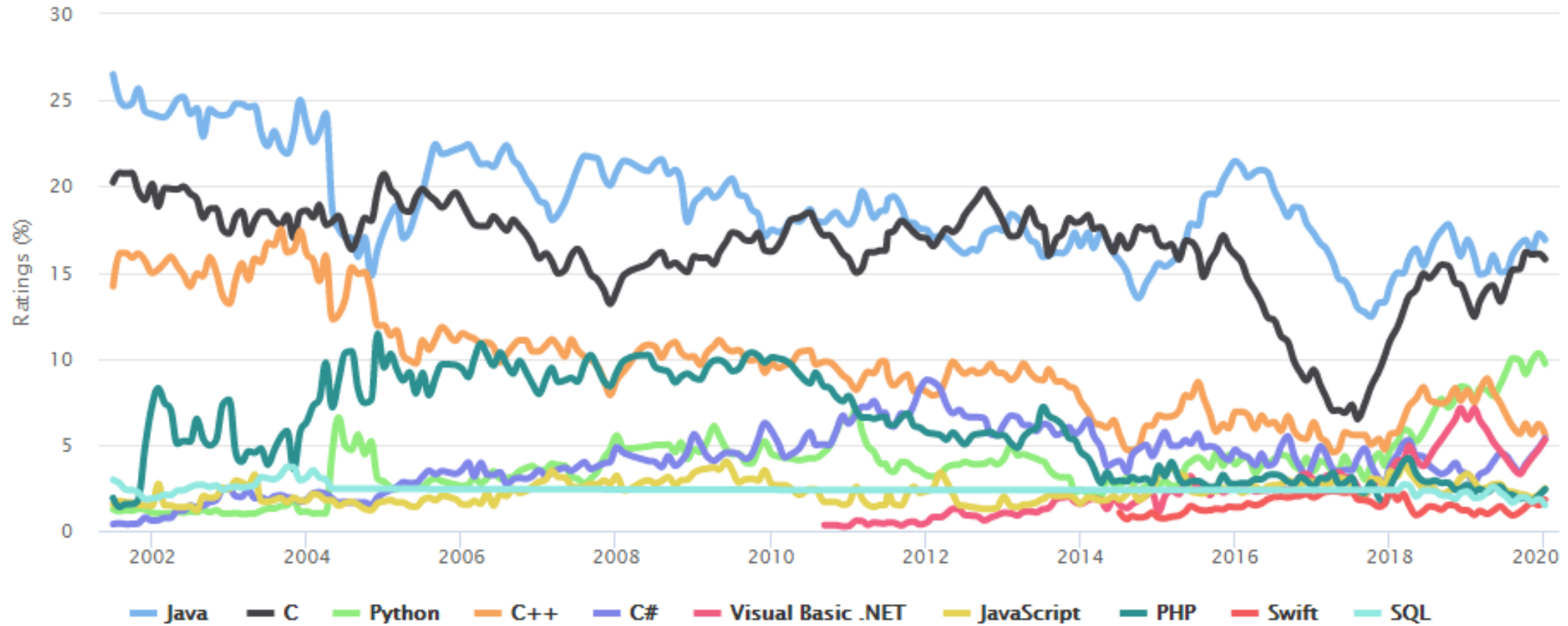
DB-Engines Ranking



SQL - Etat des lieux

TIOBE Programming Community Index

Source: www.tiobe.com



Base de données - Etat des lieux en Suisse

Le big data, littéralement « grosses données », ou mégadonnées désigne des ensembles de données devenus si volumineux qu'ils dépassent l'intuition et les capacités humaines d'analyse et même celles des outils informatiques classiques de gestion de base de données ou de l'information.



Réflexions

- Migros/Coop – carte de fidélité (Estimation à 30 To/an)
- GAFAM – aspire toutes nos vies (Taille = secret défense)
- CERN – 100 Pétaoctets de stockage
 - pétaoctet (Po) = 1 000 To



Base de données & SQL - Etat des lieux

La base de données de la NSA (Révélations de Snowden en juin 2013)

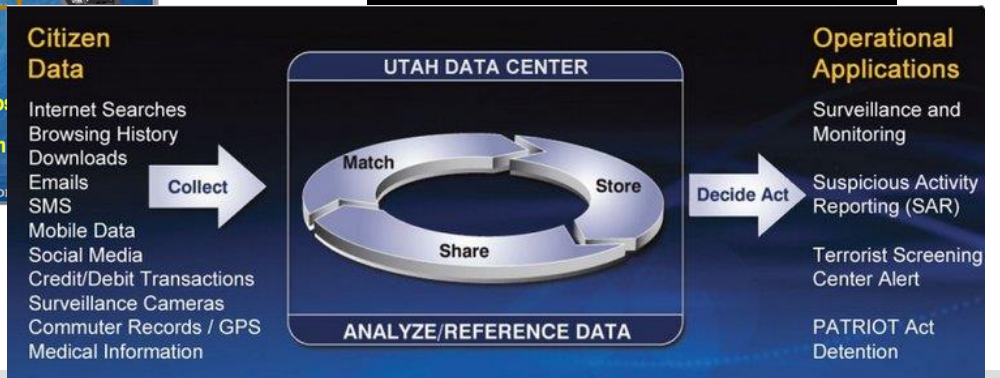
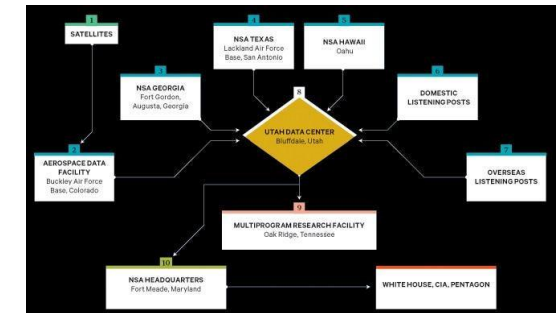
<https://www.businessinsider.com/pictures-of-the-nsas-utah-data-center-2013-6?IR=T>

https://www.lemonde.fr/big-browser/article/2013/06/10/prism-mais-ou-la-nsa-stocke-t-elle-ses-donnees_6001141_4832693.html



Réflexions

- NSA = big brother
 - 5 zéttaoctets
 - 1 pétaoctet (Po) = 1 000 To
 - 1 exaoctet (Eo) = 1000 Po
 - 1 zettaoctet (Zo) = 1 000 Eo
- = 1'073'741'824 To

















Base de données & SQL - Green

GREENPEACE

Rank 1-10

Clicking Clean Company Scorecard

		Final Grade	 Clean Energy Index	 Natural Gas	 Coal	 Nuclear	Energy Transparency	Renewable Energy Commitment & Siting Policy	Energy Efficiency & Mitigation	Renewable Procurement	Advocacy
1		A	100%	0%	0%	0%	A	A	A	A	A
2		A	83%	4%	5%	5%	A	A	A	A	B
3		A	67%	7%	15%	9%	A	A	A	A	B
4		A	56%	14%	15%	10%	B	A	A	A	A
5		B	43%	12%	16%	15%	B	A	C	B	B
6		B	32%	23%	31%	10%	B	B	C	B	B
7		B	20%	30%	29%	20%	B	B	B	C	B
8		B	16%	-	-	-	A	A	A	C	B
9		C	50%	17%	27%	5%	D	B	C	B	C
10		C	29%	29%	27%	15%	C	B	C	C	F

*Data compiled from the GreenPeace report available located in release below.

<https://www.prnewswire.com/news-releases/greenpeace-ranks-switch-highest-in-the-world-for-sustainability-among-all-internet-companies-300388310.html>