

Infrastructure de données 2

noSQL

Infrastructure de données 2

- Laurent Berthelot : laurent.berthelot@heig-vd.ch
- Mail & Teams uniquement

Préambule

Infrastructure de données 1

- Connaître la problématique générale des infrastructures de données et du stockage de l'information en fonction des différents **types de données**.
- Maîtriser la **transformation** de la modélisation **conceptuelle** de données dans un modèle **physique**.
- Maîtriser les bases de l'**administration d'une base de données relationnelles**.
- S'initier aux principes de bases liées à la **sécurité des infrastructures de données**: contrôle d'accès, sauvegarde
- Connaître les problématiques liées à l'**optimisation d'une base de données relationnelles** dans le contexte d'une application Web.
- Pratiquer la **mise en place d'une base de données relationnelle**.

Infrastructure de données 2

- Connaître les avantages des bases de données documentaires et distribuées par rapport aux systèmes relationnels dans le contexte des **applications web**.
- Comprendre les enjeux de **haute-disponibilité/réplication** dans les infrastructures de données
- Maîtriser le choix du type d'**infrastructure de stockage** d'information en fonction du type de données à stocker (alphanumérique, images, son, vidéo)
- Savoir exploiter une **source de données NoSQL** d'objets structurés sans schéma prédéterminé (ex. avec MongoDB)
- Pratiquer les techniques d'**indexation et de recherche d'information** (ex. avec Elasticsearch)

Infrastructure de données 2

- **Limites du SQL**, SQL versus NoSQL, **BASE versus ACID**, types d'infrastructure de données distribuées - 8 périodes
- **Bases de données NoSQL** (ex. CouchDB ou MongoDB) - 10 périodes
- **Indexation et recherche** d'information (ex. avec Elasticsearch) - 12 périodes
- Base de données et **haute disponibilité** - 10 périodes
- **Données multimédia et binaires** - 8 périodes

Organisation

- 12 cours : total de 48 périodes soit 36 heures
- Organisation d'une session :
 - 1 heure de théorie
 - 2 heures de pratique
- Evaluations
 - Mode projet - Evaluation continue
 - AB Test / QCM

Organisation

- Appréhender la théorie par la pratique
- Autonomie VS Copy/Paste



Base de données relationnelle

Forces et Faiblesses

Base de données relationnelle - Base de données ACID

Système de Gestion de Base de données relationnelle dit base de données **ACID**

- **Atomicity** - atomicité
 - Les transactions sont atomiques, soit elles sont complètement exécutées soit elle ne le sont pas du tout
 - Si je dois mettre à jour deux éléments, la transaction est valide si et seulement si les deux éléments sont à jour.
- **Consistency** - cohérence
 - Contrainte d'intégrité
- **Isolation** - Isolation
 - Exécution en série
 - Une transaction après l'autre
- **Durability** - Durabilité
 - Une transaction valide est ferme et définitive

Base de données relationnelle - Forces

- **Typage fort**
- **Contrainte d'intégrité** - Schéma relationnel
- **Langage SQL**
 - Langage commun
 - Précision des requêtes de transactions
- Forte **communauté (?)** et **solutions éprouvées**

Base de données relationnelle - Forces

- **Typage fort**
- **Contrainte d'intégrité** - Schéma relationnel
- **Langage SQL**
 - Langage commun
 - Précision des requêtes de transactions
- Forte **communauté (?)** et **solutions éprouvées**
- **Les bases de données relationnelles se focalisent principalement sur la véracité des données en assurant cohérence et donc fiabilité de celle-ci**

Base de données relationnelle - Faiblesse

- Difficulté d'évolution des schémas
 - Ajout de nouvelles contraintes d'intégrité
 - Changement de type de données
- Problématique de coût machine
 - Limite du nombre de requêtes
- Problématique d'accessibilité
 - Isolation et exécution en série
- Problématique d'évolution des infrastructures techniques - structure dite verticale

Evolution du marché

Evolution du marché

- Augmentation de la volumétrie des données - GAFA
- Besoin de réactivité forte
- Tolérance quant à la véracité des données
- Amélioration des capacités des devices

noSQL

no SQL / not only SQL

noSQL - Base de données BASE

Système de Gestion de Base noSQL dit base de données **BASE**

- **Basically Available** - Accessible avant tout
 - Le système doit toujours être accessible (non isolation)
 - Mise à jour des données simultanée
- **Soft state** - état souple
 - Un jeu de données, et donc la base de données peut être incohérent (not consistency)
- **Eventually Consistent** - Cohérent après coup
 - Malgré le Soft State, la cohérence peut être retrouvée

Base de données noSQL - Forces

- Accessibilité des données
 - No isolation
 - Possibilité de réplication
 - Architecture dite horizontale
- Structure de données souples
 - Orientée Objet
 - Pas de typage fort

Base de données noSQL - Forces

- Accessibilité des données
 - Possibilité de réplication
 - Architecture dite horizontale
- Structure de données souples
 - Orientée Objet
 - Pas de typage fort
- **Les bases de données noSQL se focalisent principalement sur l'accessibilité des données, indépendamment de la volumétrie**
- **L'évolution de la structure des données est simplifiée**

Base de données noSQL - Faiblesse

- Risque d'incohérence des données
- Pas de SQL (!)
 - Absence de normalisation
- Logique métier à prendre en compte lors de l'implémentation
- Communauté éparse

Base de données noSQL - Type de base de données

- **Document**
- Key-Value
- Wide-column
- Graph

Base de données noSQL - Document DB

- **Les données sont stockées sous forme de collection de document**
- Clé unique d'index (id)
- Stockage sous forme JSON

Base de données noSQL - Document DB

+ Commencer une collection	+ Ajouter un document	+ Commencer une collection
categories	05EsGcpAoiX7faDMXnup >	+ Ajouter un champ
news	1PWKBeQpSwoKicKoNJwv	active: true
orders	3WGtReL0K2bWAAky2fSr	category: "N2ihyS8Oz1tGOimRpePm"
⋮ partners >	4s3DWYyUG24Tg5jb3Hhm	creation_date: "2024-05-29T14:07:07.161Z"
products	5pTWZqjSTqZs2pmr8lQx	id: "05EsGcpAoiX7faDMXnup"
users	6Q0FQfa1PykaA1uTxxsL	image: "https://firebasestorage.googleapis.com/v0/b/uncommon12-3c8c2.appspot.com/o/images%2Ffedfceef4-c20a-4188-a8e1-671e0583d8c3?alt=media&token=9011d730-3403-4fb1-bcfe-a9ff80d3b273"
	6b002txkLRaStePtkRho	logo_partner: "https://firebasestorage.googleapis.com/v0/b/uncommon12-3c8c2.appspot.com/o/images%2F997a8967-e680-49ed-896b-479a5f6752e4?alt=media&token=618a3339-857b-47a1-9c03-e88c3c2e9e3a"
	7Xzdovfu6jzW0xfpexyx	partner_address: "Rue François-Versonnex 5 1207 Genève"
	8j9qZ0P9CZT4nr0DgpJE	partner_description: "Qui ne connaît pas le Club Med ? Cette entreprise française, fondée en 1950, est renommée pour ses resorts de vacances haut de gamme à travers le monde. Avec des destinations dans plus de 40 pays, le Club Med offre une expérience tout compris alliant luxe, convivialité et une multitude d'activités pour tous
	92t0hWJqg12LWEkZqVe0	
	9Xj61bvEZFa3eMGx1lXa	
	AMIX10yZemNnoG6LGtHV	
	AT8H8b441vX0hHwsWSuJ	
	BVJRpo5RZL5XN4KPFVYF	
	BViBNyNDsf0FVJb5TSV	
	RiK7Fvd11M0FPC7nvRkH	

Base de données noSQL - Document DB

```
1 {
2   "_id": "4b1f10596bd54f8a070d6f4ef200439a",
3   "_rev": "5-b44965a8984ec72664ca3fa0d3247923",
4   "post_name": "Post title",
5   "post_content": "Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque lacus sem, placerat ac turpis non, aliq
6   "attributes": {
7     "creation_date": "today",
8     "author": "Main author"
9   },
10  "comments": [
11    {
12      "comment": "Comment 1",
13      "author": "Author 1"
14    },
15    {
16      "comment": "Comment 2",
17      "author": "Author 2"
18    },
19    {
20      "comment": "Comment 3",
21      "author": "Author 3"
22    }
23  ]
24 }
```


Software

- Gitlab / Github
- Installer CouchDB
- Bootstrap application VueJS
- Création d'une collection dans CouchDB
- Fetch des datas au sein de l'application VueJS

