



NoSQL : les fondamentaux

CB002

Durée: 2 jours

1 670 €

3 au 4 avril

3 au 4 juillet
2 au 3 septembre
20 au 21 novembre

Public :

Responsables informatique, chefs de projet, architectes, développeurs, décideurs, ...

Objectifs :

A l'issue de la formation, le stagiaire sera capable d'appréhender de façon opérationnelle les principales caractéristiques des bases de données NoSQL.

Connaissances préalables nécessaires :

avoir des connaissances de base des architectures techniques, du management SI et des bases de données.

Objectifs pédagogiques :

Identifier les différences entre SGBD SQL et SGBD NoSQL
Évaluer les apports et les inconvénients inhérents aux technologies NoSQL
Identifier les principaux acteurs et solutions du marché pour chaque modèle de données
Connaître les champs d'application des SGBD NoSQL en opérationnel et en analytique
Comprendre les différentes architectures, modèles de données et implémentations techniques
Identifier les critères de choix

Programme :

Identifier les différences entre SGBD SQL et SGBD NoSQL

Introduction :
origine des bases de données, les notions de transaction,
les SGBD, la standardisation SQL,
l'arrivée de nouveaux besoins : volumes importants liés aux technologies et aux nouveaux usages,
traitements optimisés de flux de données au fil de l'eau.
BigData et NoSQL : les 5 V
Présentation du Théorème CAP.
Développement des techniques sur différents aspects : stockage, indexation/recherche, calcul.
Définition ETL : Extract Transform Load.



Évaluer les apports et les inconvénients inhérents aux technologies NoSQL

Caractéristiques NoSQL.

Définition de Datalake : présentation, sources des données, intérêt

Les familles de bases NoSQL (Clé/valeur, orientées documents, orientées colonnes, graphes)

Structure de données proches des utilisateurs, développeurs:

sérialisation, tables de hachage, JSON.

Priorité au traitement du côté client.

Protocoles d'accès aux données, interfaces depuis les langages classiques.

Données structurées et non structurées, documents, images,

Stockage réparti : réplication, sharding, protocole gossip, hachage,...

Parallélisation des traitements : implémentation de map/reduce.

Cohérence des données et gestion des accès concurrents :

"eventual consistency" et multi-versions concurrency control.

Atelier : démonstration d'un système de stockage réparti avec MongoDB

Identifier les principaux acteurs et solutions du marché pour chaque modèle de données

Principaux acteurs

Les solutions NoSQL et leurs choix techniques :

Redis, MongoDB, Cassandra, HBase (Hadoop), Neo4j, Elasticsearch, ...

Les modèles de données de chaque solution.

Atelier : démonstrations et tests avec Cassandra, MongoDB, Elasticsearch et HBase.

Connaître les champs d'application des SGBD NoSQL en opérationnel et en analytique

Présentation des modes 'opérationnel' et 'analytique'.

Notions de données brutes, issues de flux de production, logs, ...

et d'analytique : traitement Hadoop, MapReduce, ElasticSearch

Etude des étapes de traitement des données : ETL, intégration sqoop

Atelier : étude de cas d'applications concrètes des SGBD NoSQL, analyse des données de base, et des possibilités offertes par les solutions NoSQL

Comprendre les différentes architectures, modèles de données et implémentations techniques

Etude des points à vérifier :

méthode d'utilisation des données, format de stockage JSON, XML, choix de la clé, notion de clé composite,

aspects matériels, besoins en mémoire, disques, répartition, ...

Import des données : outils et méthodes selon les moteurs NoSQL

Atelier : manipulations sur une base Cassandra pré-chargée et mise en évidence des points forts et points faibles



Identifier les critères de choix

Importance de l'analyse du besoin :

types de données à stocker, utilisation de la données, types de traitements, environnement dans lequel s'intègre les données, performances attendues (temps réel, calcul à long terme, flux continu, ...)

Critères de choix :

- performance,
- extensibilité
- flexibilité
- complexité

Positionnement des différents modèles de données pour chacun de ces critères

Atelier : étude d'un cas pratique de choix d'une solution NoSQL pour le stockage de données mise à disposition sur l'infrastructure de travaux pratiques