

Informatique

Bases de données avancées (36 heures)

26_27_4IRC_08_COMSC_UE8_081_MAJ

ACQUIS

- Comprendre les besoins actuels en données, les limitations des bases de données relationnelles, et les raisons du passage au NoSQL
- Connaître les différentes propriétés du Big Data : modèles ?Vs? (3V, 5V, ...)
- Connaître les différentes formes de stockage (data warehouse, data lake, data lakehouse) et leurs avantages respectifs
- Connaître le théorème CAP et savoir choisir quelles propriétés sont importantes dans un contexte donné
- Connaître différents types de bases de données non-relationnelles (clé-valeur, objet, document, graphe, séries temporelles, ...) et savoir choisir le plus approprié pour un problème
- Configurer et utiliser différents stockages (?objet? : Amazon S3, Azure Blob, ... ; ?graphe? : Neo4J)
- Utiliser des systèmes distribués (Hadoop, HDFS) pour augmenter la tolérance aux pannes et permettre le traitement de données massives
- Configurer un cluster MongoDB (sharding)
- Connaître et utiliser l'algorithme MapReduce pour traiter des données distribuées

CONTENU

- Introduction au BigData
- Datafication, ombre digitale et jumeaux numériques
- Architectures de stockage : Data warehouse, data lake, data lakehouse
- Caractéristiques du BigData (modèles ?Vs?) : 3V, 5V, ...
- Types de stockage
- Hot vs cold data
- Service Level Agreements (SLA) : durabilité, disponibilité, ...
- Types de stockage, exemples de logiciels, et leurs cas d'usage
- Clé-valeur : Redis
- Objet : Amazon S3, Azure Blob
- Document : MongoDB
- Graphe : Neo4J
- Séries temporelles : InfluxDB
- Polystores
- Systèmes distribués
- Architecture d'Hadoop
- Principe de fonctionnement d'HDFS
- Tolérance aux pannes

- Flux de contrôle vs flux de données
- MapReduce
- Lazy evaluation
- Graphe d'exécution

PRÉREQUIS

- Connaissance basique de Linux (ligne de commande)
- Bases de données relationnelles (SGBDr, SQL)

PÉDAGOGIE

ÉVALUATION

BIBLIOGRAPHIE