

Sigle : INF6323 Gr. 01
Titre : Programmation infonuagique avancée
Session : Automne 2021 Horaire et local
Professeur : Abdul Wahab, Omar

1. Description du cours paraissant à l'annuaire :

Objectifs

Apprendre et maîtriser les concepts et les techniques de l'infonuagique et des mégadonnées. Concevoir et implémenter des applications pratiques de science des données sur des plateformes infonuagiques.

Contenu

Modélisation des données avec XML et JSON. Services Web de type SOAP. Services Web de type REST. Introduction à l'infonuagique. Modèles de services en infonuagique (logiciel-service, plateforme-service, infrastructure-service, fonction-service, etc.). Modèles de déploiement de l'infonuagique (privé interne, privé externe, public, communautaire, multi-cloud et hybride). Techniques de virtualisation en infonuagique (virtualisation par machines virtuelles et virtualisation par conteneurs). Programmation infonuagique (Amazon Web Services et Google Cloud Platform). L'écosystème Hadoop : le système de fichiers distribué HDFS, le gestionnaire de ressources YARN, le modèle de programmation MapReduce. L'écosystème Apache Spark pour l'analyse des données en temps réel. Bases de données non relationnelles NoSQL. Le système MongoDB de gestion de bases de données orientées documents.

[Descriptif – Annuaire](#)

2. Objectifs spécifiques du cours :

- Maîtriser et appliquer les principaux concepts et outils du *cloud computing*.
- Se familiariser avec les principales plateformes de traitement, de stockage et de déploiement du *cloud computing*.
- Comprendre les principaux concepts, principes et outils liés au *Big Data*.
- Appliquer les services du *cloud computing* et d'API REST pour concevoir et mettre en œuvre des applications de science des données.
- Se familiariser avec les services d'Amazon Web Services et de Google Cloud.
- Se familiariser avec l'écosystème Hadoop, en particulier le système de fichiers distribués Hadoop (HDFS), le gestionnaire de ressources YARN, et le modèle de programmation MapReduce.
- Se familiariser avec l'écosystème Apache Spark et ses différents composants.
- Se familiariser avec les bases de données non relationnelles et développer des applications avec MongoDB.
- Concevoir et mettre en œuvre des applications de science des données sur des plateformes infonuagiques en utilisant les dernières plateformes du *Big Data*.

3. Stratégies pédagogiques :

Au cours de cette activité, diverses formules pédagogiques seront utilisées, notamment : diapositives, démonstrations et devoirs. Les attentes sont que les étudiant(e)s investissent au moins 90 heures de travail personnel en plus des 45 heures de cours.

Les cours auront lieu en présentiel.

4. Heures de disponibilité ou modalités pour rendez-vous :

Disponible avant ou après les cours, et sur rendez-vous.

Courriel : omar.abdulwahab@uqo.ca

5. Plan détaillé du cours sur 15 semaines :

Semaine	Thèmes	Dates
1	Aperçu du cours, description du projet et modèles de données Modèle de données XML. Les règles de syntaxe des documents XML. Documents XML bien formés. Éléments et attributs dans les documents XML. Parseurs des documents XML : parseurs orientés événement et parseurs orientés hiérarchie. Le langage de requêtes XQuery. Propriétés des documents XML. Modèle de données JSON. Types de données en JSON. Les règles de syntaxe des documents JSON.	07 sept. 2021
2	Services Web de type SOAP et services Web de type REST Raison d'être des services Web. Services Web SOAP : invocation, interaction et découverte. Approches de programmation des services Web SOAP : approche ascendante (bottom-up) et approche descendante (top-down). L'API REST et le protocole HTTP. Les caractéristiques et contraintes des services Web REST. Cas d'utilisation des services Web REST. Développement d'APIs REST avec Python Flask.	14 sept. 2021

3	L'infonuagique L'infrastructure Cloud. Centres de données. Définition et caractéristiques de l'infonuagique. Modèles de services infonuagiques (logiciel-service, plateforme-service, infrastructure-service, fonction-service, etc.). Modèles de déploiement en infonuagique (privé interne, privé externe, public, communautaire, multi-cloud et hybride). Techniques de virtualisation en infonuagique (virtualisation par machines virtuelles et virtualisation par conteneurs). Tendances d'adoption de l'infonuagique. Le rôle des APIs REST dans l'infonuagique.	21 sept. 2021
4	Programmation en infonuagique L'écosystème Amazon Web Services. Création des applications fonction-service avec AWS Lambda. L'écosystème Google Cloud. Étude de cas : analytique des données massives relationnelles sous Google Cloud avec BigQuery.	28 sept. 2021
5	L'écosystème Hadoop La notion de Big Data. Enjeux et défis du Big Data. Vue d'ensemble de l'écosystème Hadoop. Différences entre Hadoop et SQL. Le concept de localité de données. Le système de fichiers distribué de Hadoop (HDFS). Le gestionnaire de ressources de Hadoop (YARN).	05 oct. 2021
6	Semaine d'études	12 oct. 2021
7	Le modèle de programmation MapReduce Vue d'ensemble de MapReduce. Traitement par lots et traitement de flux de données. La logique de MapReduce. Les fonctions <i>map</i> et <i>reduce</i> . Analytique des mégadonnées avec MapReduce. Fonctionnalités avancées en MapReduce. Limites de MapReduce.	19 oct. 2021
8	Examen intra (en présentiel)	26 oct. 2021
9	L'écosystème Apache Spark Raison d'être de Spark. Différences entre Spark et MapReduce. Structures de données distribuées en Spark. Transformations et actions en Spark. Le langage SQL en Spark. Spark en streaming. Apprentissage machine en Spark.	02 nov. 2021
10	Présentations des devoirs	09 nov. 2021
11	Présentations des devoirs	16 nov. 2021
12	NoSQL et NewSQL Les limites des bases de données relationnelles. Mises à l'échelle horizontales et verticales des bases de données. <i>Clustering</i> des bases de données. Le concept de <i>Sharding</i> : partitionnement des bases de données. Réplication de données. La raison d'être des bases de données NoSQL. Les types des bases de données NoSQL (key-value, documents, <i>wide-column</i> , graphes). Théorème CAP. Le modèle BASE et le modèle ACID. Les limites des bases de données NoSQL. Raison d'être et état de l'art des bases de données NewSQL.	23 nov. 2021
13	MongoDB Vue d'ensemble de MongoDB. Conception des bases de données dans MongoDB : incorporer ou référencer? Partitionnement et réplication des données en MongoDB. Atomicité et consistance en MongoDB. Commandes de mises à jour : INSERT, UPDATE, FIND et REMOVE. Étude de cas et exemples pratiques.	30 nov. 2021
14	Examen final (en présentiel)	07 déc. 2021
15	Présentations des projets	14 déc. 2021

6. Évaluation du cours :

- Examen intra : 25 %
 - Devoir : 15 %
 - Projet : 25 %
 - Examen final : 35 %
- **Projet de session : conception et implémentation** d'une application de la science de données à l'aide d'une plateforme infonuagique. Pour plus de détails, prière de consulter l'échéancier et les consignes additionnelles qui vous seront donnés. Une description sera fournie durant le cours.
 - **Devoir** : Les étudiant(e)s devront **enquêter, comprendre et communiquer** des concepts et plateformes clés liés à l'infonuagique et à la science des données (p. ex : HBase, Cassandra, Pig, Hive, Docker, Kubernetes, apprentissage fédéré etc.). Des instructions supplémentaires seront discutées en classe et affichées sur Moodle.

- **N.B.** Pour chaque travail, une échéance est imposée. Il sera de la responsabilité des étudiant(e)s de rencontrer ces échéanciers. Tout retard dans la remise des travaux entraînera **nécessairement** une pénalité de **5 %** par jour sur la note attribuée.

7. Politiques départementales et institutionnelles :

- [Politique du département d'informatique et d'ingénierie relative à la tenue des examens](#)
- [Note sur le plagiat et sur la fraude](#)
- Absence aux examens : [cadre de gestion](#), [demande de reprise d'examen \(formulaire\)](#)

À l'UQO, **les violences à caractère sexuel, c'est tolérance zéro!**

La communauté universitaire s'engage à lutter contre les inconduites, le harcèlement et les violences à caractère sexuel : parce que **le respect, c'est l'affaire de tout le monde!**

N'oubliez pas de faire la formation obligatoire :

uqo.ca/bimi/formation-obligatoire

Pour de plus amples renseignements :

bimi@uqo.ca



8. Principales références :

1. Hadoop: *The Definitive Guide*, Tom White, 4^e édition, O'Reilly, avril 2015.
2. Spark: *The Definitive Guide*, Matei Zaharia et Bill Chambers, O'Reilly Media, février 2018.

9. Page Web du cours :

<https://moodle.uqo.ca>