

Sigle : INF6083 Gr. 01

Titre : Sujets spéciaux (Programmation avancée en infonuagique)

Session : Automne 2019 Horaire et local

Professeur : Abdul Wahab, Omar

1. Description du cours paraissant à l'annuaire :

Objectifs

Permettre à l'étudiant d'acquérir des connaissances sur un (ou des) sujet(s) spécifique(s) pertinent(s) à son programme.

Contenu

Présentation d'une activité portant sur un (ou des) sujet(s) non couvert(s) dans les autres cours du programme. Activité offerte par un professeur ou une équipe de professeurs. Cette activité traite d'un ou de sujets d'intérêt et apporte une contribution particulière à la formation de l'étudiant. Le contenu de ce cours doit faire l'objet d'une approbation préalable par le Comité de programme.

Descriptif – Annuaire

2. Objectifs spécifiques du cours :

Apprendre et maîtriser les concepts et les techniques de l'infonuagique et du Big Data. Concevoir et implémenter des applications pratiques de la science des données à l'aide 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 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). Programmation en 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 mégadonnées en temps réel. Bases de données non-relationnelles NoSQL. Le système MongoDB de gestion de bases de données non-relationnelles orientées documents.

3. Stratégies pédagogiques :

Cours magistraux

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

Mercredi de 15 h à 17 h

Courriel : omar.abdulwahab@uqo.ca

5. Plan détaillé du cours sur 15 semaines :

Semaine	Thèmes	Dates
1	<p>Aperçu du cours, description du projet et modèles de données</p> <p>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.</p>	04 sept. 2019

2	<p>Services Web de type SOAP et services Web de type REST</p> <p>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.</p>	11 sept. 2019
3	<p>L'infonuagique</p> <p>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.</p>	18 sept. 2019
4	<p>Programmation en infonuagique</p> <p>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.</p>	25 sept. 2019
5	<p>L'écosystème Hadoop</p> <p>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).</p>	02 oct. 2019
6	<p>Le modèle de programmation MapReduce</p> <p>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.</p>	09 oct. 2019
7	<p>Semaine d'études</p>	16 oct. 2019
8	<p>Examen intra</p>	23 oct. 2019
9	<p>L'écosystème Apache Spark</p> <p>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.</p>	30 oct. 2019
10	<p>Présentations des devoirs</p>	06 nov. 2019
11	<p>Présentations des devoirs</p>	13 nov. 2019
12	<p>NoSQL et NewSQL</p> <p>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, wide-column, 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.</p>	20 nov. 2019
13	<p>MongoDB</p> <p>Vue d'ensemble de MongoDB. Conception des bases de données dans MongoDB : incorporer ou référencer? Partitionnement et réplication des données en MongoDB. Atomicité et</p>	27 nov. 2019

	consistance en MongoDB. Commandes de mises à jour : INSERT, UPDATE, FIND et REMOVE. Étude de cas et exemples pratiques.	
14	Présentations des projets	04 déc. 2019
15	Examen final	11 déc. 2019

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, 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
- Politique relative à la qualité de l'expression française écrite chez les étudiants et les étudiantes de premier cycle à l'UQO
- Absence aux examens : cadre de gestion, demande de reprise d'examen (formulaire)

• **SANS OUI
C'EST NON!**

Travaillons ensemble pour développer une culture du respect ! La communauté universitaire de l'UQO se mobilise et lance un message haut et fort de **tolérance zéro en matière de violence à caractère sexuel** (pour de plus amples renseignements, veuillez consulter la page Web : uqo.ca/sansouicestnon).

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>