

---

# INF8200

## Syst. et infrastructures pour les données massives

### Plan de cours

#### Responsable(s) du cours

---

COTI, Camille  
DS-4420  
[coti.camille@uqam.ca](mailto:coti.camille@uqam.ca)  
Groupes : 020

#### Description du cours

---

##### Objectifs

Ce cours prépare les étudiants aux systèmes et infrastructures pour les données massives et les familiarise avec les activités fondamentales liées aux données massives. Ses objectifs sont principalement :

- Comprendre les fondements des systèmes répartis et parallèles, systèmes indispensables à la mise en œuvre de solutions pour le stockage et le traitement des données massives.
- Comprendre les principes, méthodes et mécanismes des plateformes de stockage et de traitement de données massives (par ex., MapReduce, Hadoop, Spark) et les outils associés.
- Identifier et expérimenter certaines des solutions technologiques disponibles.
- Se familiariser avec certains systèmes à la fine pointe de la recherche dans le domaine.

##### Sommaire du contenu

- Fondements des systèmes répartis et parallèles : modèles, architectures, communications, nommage, coordination, cohérence, fiabilité, réplication ;
- Infrastructures de stockage des données massives : systèmes et services de stockage de données, systèmes de fichiers répartis, infonuagique, gestion de transactions, entrepôts de données, intégration, fragmentation et duplication de données ;
- Infrastructures de traitement des données massives : collecte et pré-traitement des données, nettoyage et intégration, analyse et visualisation, modèles de traitement, évaluation des performances, plateformes et outils de traitement distribué, systèmes de traitement de flux de données.

##### Modalité d'enseignement

Séances magistrales ; Exercices pratiques ; Études de cas ; Projets ; Questionnaires à choix multiples.

---

## Modalités d'évaluation

---

Outil d'évaluation	Pondération	Échéance
Projet 1	20%	S7
Projet 2	20%	S12
Projet 3	20%	S15
Quiz 1	7%	S5
Quiz 2	7%	S9
Quiz 3	7%	S11
Quiz 4	7%	S14
Présentation d'article	12%	à déterminer

## INFORMATIONS IMPORTANTES

La qualité du français constitue un critère d'évaluation (pour un maximum de 10%). En cas de retard dans la remise des travaux, une pénalité de 5% par jour ouvrable sera appliquée. Un retard de plus de cinq jours ouvrables ne sera pas accepté. Les règlements concernant le plagiat seront strictement appliqués. Pour plus de renseignements, consultez le site suivant : <http://www.sciences.uqam.ca/etudiants/integrite-academique.html>

---

## Calendrier détaillé du cours

---

### Semaine 1

- Introductions aux systèmes de stockage de données massives
- Lien avec les bases de données NoSQL
- Infrastructures matérielles
- Introduction à Redis
- Modèle publish-subscribe

### Semaine 2

- Introduction au stockage distribué
- Mise en œuvre dans Hadoop HDFS
- Architecture de stockage de HDFS
- Approches pour la tolérance aux pannes

### Semaine 3

- Présentation des fonctions map et reduce
- Introduction du modèle MapReduce

#### **Semaine 4**

- Approfondissement du modèle MapReduce
- MapReduce multiétapes
- Interfaçage avec les systèmes de stockage vus précédemment
- Ordonnancement de travaux
- YARN

#### **Semaine 5**

- Quizz
- Algorithme des k-moyennes
- Mise en place avec MapReduce

#### **Semaine 6**

- Notion de graphe orienté acyclique
- Introduction des bases de données distribuées résilientes (RDD)
- Introduction du modèle de Spark
- Architecture de Spark

#### **Semaine 7**

- Présentation de projets

#### **Semaine 8**

- Approfondissement de Spark
- Introduction de SparkSQL
- Composition d'une requête SparkSQL

#### **Semaine 9**

- Quizz
- Introduction de Cassandra
- Modèle de stockage des données dans Cassandra
- Présentation du langage de requête de Cassandra CQL
- Introduction d'ElasticSearch
- Modèle de requêtes dans ElasticSearch

#### **Semaine 10**

- Présentation de Hbase
- Opérations CRUD sur Hbase

- Utilisation avec MapReduce
- Présentation de Hive
- Présentation de la syntaxe de requêtes HQL

### **Semaine 11**

- Quizz
- Présentation de Pig Latin
- Modèle de données manipulées par Pig
- Forme des requêtes Pig

### **Semaine 12**

- Présentation de projets

### **Semaine 13**

- Présentation d'Apache Kafka
- Représentation des données dans Kafka
- Modèle publish-subscribe dans Kafka, modèle producteur-consommateurs
- Lire des données dans Kafka avec un consommateur

### **Semaine 14**

- Quizz
- Présentations d'articles

### **Semaine 15**

- Présentations de projets

## **Médiagraphie**

---

### **Obligatoire**

- Diapositives de cours disponibles sur le dépôt Git du cours
- Tutoriels disponibles sur le dépôt Git du cours

### **Ressources complémentaires**

- Hadoop : The Definitive Guide, 4th Edition - Tom White, O'Reilly Media, Inc. 2015.

- Big Data Analytics Beyond Hadoop : Real-Time Applications with Storm, Spark, and More Hadoop Alternatives - Vijay Srinivas Agneeswaran, Ph.D, Pearson FT Press. 2014.
- Hadoop with Python - Zachary Radtka & Donald Miner, O'Reilly, 2015
- A gentle introduction to Apache Spark - Databricks, ebook
- Learning Spark : Lightning-Fast Data Analytics, 2nd edition - Jules S. Damji, Brooke Wenig, Tathagata Das, and Denny Lee, O'Reilly, 2020.
- Cassandra : The Definitive Guide, 2nd edition - Jeff Carpenter & Eben Hewitt, O'Reilly, 2016.
- Elasticsearch : The Definitive Guide - Clinton Gormley and Zachary Tong, O'Reilly, 2015.
- Programmign Hive - Edward Capriolo, Dean Wampler, and Jason Rutherglen, O'Reilly, 2017.
- Hbase : The Definitive Guide - Lars George, O'Reilly, 2011.
- Kafka : The Definitive Guide : Real-Time Data and Stream Processing at Scale - Neha Narkhede, Gwen Shapira, and Todd Palino, O'Reilly, 2017.

### Politique d'absence aux examens

*Reprise d'examen.* L'autorisation de reprendre un examen en cas d'absence est de **caractère exceptionnel**. Pour obtenir un tel privilège, l'étudiant.e doit avoir des motifs sérieux et bien justifiés.

*Conflits d'horaire.* Il est de la responsabilité de l'étudiant.e de ne pas s'inscrire à des cours qui sont en conflit d'horaire, tant en ce qui concerne les séances de cours ou d'exercices que les examens. **De tels conflits d'horaire ne constituent pas un motif justifiant une demande d'examen de reprise.**

*Procédure.* L'étudiant.e absent.e lors d'un examen doit, dans les cinq (5) jours ouvrables suivant la date de l'examen, présenter une demande de reprise en utilisant le formulaire prévu, disponible sur <http://info.uqam.ca/repriseexamen/>.

*Pièces justificatives.* Dans le cas d'une absence pour raison médicale, l'étudiant.e doit joindre un certificat médical original et signé par le médecin décrivant la raison de l'absence à l'examen. Les dates d'invalidité doivent être clairement indiquées sur le certificat. Une vérification de la validité du certificat pourrait être faite. Dans le cas d'une absence pour une raison non médicale, l'étudiant.e doit fournir les documents originaux expliquant et justifiant l'absence à l'examen ; par exemple, lettre de la Cour en cas de participation à un jury, copie du certificat de décès en cas de décès d'un proche, etc. Toute demande incomplète sera refusée. Si la direction du programme d'études de l'étudiant.e constate qu'un.e étudiant.e a un comportement récurrent d'absence aux examens, l'étudiant.e peut se voir refuser une reprise d'examen.

*Pour plus d'informations.* Consulter la page <http://info.uqam.ca/politiques>.

### Règlement numéro 18 sur les infractions de nature académique (extraits)

Tout acte de plagiat, fraude, copiage, tricherie ou falsification de document commis par une étudiante, un étudiant, de même que toute participation à ces actes ou tentative de les commettre, à l'occasion d'un examen ou d'un travail faisant l'objet d'une évaluation ou dans toute autre circonstance, constituent une infraction au sens de ce règlement.

La liste non limitative des infractions est définie comme suit :

- la substitution de personnes ;
- l'utilisation totale ou partielle du texte d'autrui en la faisant passer pour sien ou sans indication de référence ;
- la transmission d'un travail pour fins d'évaluation alors qu'il constitue essentiellement un travail qui a déjà été transmis pour fins d'évaluation académique à l'Université ou dans une autre institution d'enseignement, sauf avec l'accord préalable de l'enseignante, l'enseignant ;
- l'obtention par vol, manoeuvre ou corruption de questions ou de réponses d'examen ou de tout autre document ou matériel non autorisés, ou encore d'une évaluation non méritée ;
- la possession ou l'utilisation, avant ou pendant un examen, de tout document non autorisé ;
- l'utilisation pendant un examen de la copie d'examen d'une autre personne ;
- l'obtention de toute aide non autorisée, qu'elle soit collective ou individuelle ;
- la falsification d'un document, notamment d'un document transmis par l'Université ou d'un document de l'Université transmis ou non à une tierce personne, quelles que soient les circonstances ;
- la falsification de données de recherche dans un travail, notamment une thèse, un mémoire, un mémoire-crédation, un rapport de stage ou un rapport de recherche ;
- Les sanctions reliées à ces infractions sont précisées à l'article 3 du Règlement no 18.

Les règlements concernant le plagiat seront strictement appliqués. Pour plus de renseignements :

- <http://www.infosphere.uqam.ca/rediger-un-travail/eviter-plagiat>
- <http://r18.uqam.ca/>

### **Politique no 16 visant à prévenir et combattre le sexisme et les violences à caractère sexuel**

Les violences à caractère sexuel se définissent comme étant des comportements, propos et attitudes à caractère sexuel non consentis ou non désirés, avec ou sans contact physique, incluant ceux exercés ou exprimés par un moyen technologique, tels les médias sociaux ou autres médias numériques. Les violences à caractère sexuel peuvent se manifester par un geste unique ou s'inscrire dans un continuum de manifestations et peuvent comprendre la manipulation, l'intimidation, le chantage, la menace implicite ou explicite, la contrainte ou l'usage de force.

Les violences à caractère sexuel incluent, notamment :

- la production ou la diffusion d'images ou de vidéos sexuelles explicites et dégradantes, sans motif pédagogique, de recherche, de création ou d'autres fins publiques légitimes ;
- les avances verbales ou propositions insistantes à caractère sexuel non désirées ;
- la manifestation abusive et non désirée d'intérêt amoureux ou sexuel ;
- les commentaires, les allusions, les plaisanteries, les interpellations ou les insultes à caractère sexuel, devant ou en l'absence de la personne visée ;
- les actes de voyeurisme ou d'exhibitionnisme ;
- le (cyber) harcèlement sexuel ;
- la production, la possession ou la diffusion d'images ou de vidéos sexuelles d'une personne sans son consentement ;
- les avances non verbales, telles que les avances physiques, les attouchements, les frôlements, les pincements, les baisers non désirés ;
- l'agression sexuelle ou la menace d'agression sexuelle ;
- l'imposition d'une intimité sexuelle non voulue ;
- les promesses de récompense ou les menaces de représailles, implicites ou explicites, liées à la satisfaction ou à la non-satisfaction d'une demande à caractère sexuel.

*Pour consulter la politique no 16*

[https://instances.uqam.ca/wp-content/uploads/sites/47/2018/05/Politique\\_no\\_16.pdf](https://instances.uqam.ca/wp-content/uploads/sites/47/2018/05/Politique_no_16.pdf)

*Pour obtenir de l'aide, faire une divulgation ou une plainte*

Bureau d'intervention et de prévention en matière de harcèlement  
514-987-3000, poste 0886

*Pour obtenir la liste des services offerts à l'UQAM et à l'extérieur de l'UQAM*

<https://harcelement.uqam.ca>

*Soutien psychologique (Services à la vie étudiante)*

514-987-3185  
Local DS-2110

*CALACS Trêve pour Elles – point de services UQAM*

514 987-0348  
[calacs@uqam.ca](mailto:calacs@uqam.ca)  
<http://trevepourelles.org>

*Service de la prévention et de la sécurité*

514-987-3131

### **Politique no 44 d'accueil et de soutien des étudiant.e.s en situation de handicap**

*Politique.* Par sa politique, l'Université reconnaît, en toute égalité des chances, sans discrimination ni privilège, aux étudiant.e.s en situation de handicap, le droit de bénéficier de l'ensemble des ressources du campus et de la communauté universitaire, afin d'assurer la réussite de leurs projets d'études, et ce, dans les meilleures conditions possibles. L'exercice de ce droit est, par ailleurs, tributaire du cadre réglementaire régissant l'ensemble des activités de l'Université.

*Responsabilité de l'étudiant.e.* Il incombe aux étudiant.e.s en situation de handicap de rencontrer les intervenant.e.s (conseiller.ère.s à l'accueil et à l'intégration du Service d'accueil et de soutien des étudiant.e.s en situation de handicap, professeur.e.s, chargé.e.s de cours, direction de programmes, associations étudiantes concernées, etc.) qui pourront faciliter leur intégration à la communauté universitaire ou les assister et les soutenir dans la résolution de problèmes particuliers en lien avec les limitations entraînées par leur déficience.

*Service d'accueil et de soutien aux étudiant.e.s en situation de handicap.* Le Service d'accueil et de soutien aux étudiant.e.s en situation de handicap (SASESH) offre des mesures d'aménagement dont peuvent bénéficier certains étudiant.e.s. Il est fortement recommandé aux de se prévaloir de ces services afin de réussir ses études, sans discrimination. Pour plus d'information, visiter le site de ce service : <https://vie-etudiante.uqam.ca/etudiant-situation-handicap/nouvelles-ressources.html> et celui de la politique institutionnelle d'accueil et de soutien aux étudiant.e.s en situation de handicap : [https://instances.uqam.ca/wp-content/uploads/sites/47/2018/05/Politique\\_no\\_44.pdf](https://instances.uqam.ca/wp-content/uploads/sites/47/2018/05/Politique_no_44.pdf)

Il est important d'informer le SASESH de votre situation le plus tôt possible :

- En personne : 1290, rue Saint-Denis, Pavillon Saint-Denis, local AB-2300
- Par téléphone : 514 987-3148
- Par courriel : [situation.handicap@uqam.ca](mailto:situation.handicap@uqam.ca)
- En ligne : <https://vie-etudiante.uqam.ca/>