

DESS en analytique de données massives sur infonuagique

Projet de création d'un programme de grade par
l'École de technologie supérieure

2019/12/01



Table des matières

1.	Identification du programme et ses objectifs	4
1.1	Résumé.....	4
1.2	Description	Error! Bookmark not defined.
1.3	Grade.....	4
1.4	Discipline	4
1.5	Intitulé du programme et du diplôme.....	4
1.6	Nombre de crédits.....	5
1.7	Département responsable.....	5
1.8	Établissement et lieu(x) prévus d’offre du programme	5
1.9	Particularités du programme	5
1.10	Objectifs généraux.....	5
1.11	Objectifs spécifiques.....	6
2.	Motifs de création du programme.....	7
2.1	Opportunité socioéconomique/socioculturelle	7
2.2	Opportunité systémique	11
2.3	Opportunité institutionnelle	14
3.	Structure du programme et cheminement des étudiants	14
3.1	Structure générale du programme.....	14
3.2	Description des activités.....	15
3.3	Description des formules pédagogiques	24
3.4	Grille de cheminement.....	25
4.	Ressources humaines et matérielles.....	25
4.1	Ressources professorales	25
4.2	Professionnels, techniciens, etc.	26
4.3	Ressources matérielles.....	27
5.	Cadre réglementaire	28
5.1	Conditions / Processus d’admission	28
5.2	Durée / régime d’études	29
5.3	Équivalences.....	29
5.4	Supervision / Évaluation des études	30

Annexes 30

1. Identification du programme et ses objectifs

1.1 Résumé

Proposition d'un nouveau programme de formation de spécialistes en mise en œuvre d'infrastructures et services infonuagiques pour l'analyses de données massives.

1.2 Description

L'objectif principal de ce programme est de répondre aux besoins de perfectionnement des intervenants responsables du développement d'applications d'analyses de données massives à l'aide de services infonuagiques. Le programme permet de conjuguer l'acquisition de connaissances et de compétences de haut niveau, d'une part en analyse de données massives, et d'autre part, en mise en place d'une infrastructure TI sur le nuage pour effectuer ces analyses.

Programme qui pourrait être offert à temps complet et à temps partiel. Automne 2022.

1.3 Grade

Diplômé en études supérieures spécialisées

1.4 Discipline

Le programme se situe dans le domaine de l'informatique, du développement logiciel et de la mise en place d'infrastructure TI. Il sera rattaché au département de génie logiciel et TI.

1.5 Intitulé du programme et du diplôme

DESS en analytique de données massives sur infonuagique

1.6 Nombre de crédits

Le nombre de crédits sera de 30 crédits

1.7 Département responsable

Le département de génie logiciel et TI sera responsable de la gestion de ce programme.

1.8 Établissement et lieu(x) prévus d'offre du programme

Programme offert sur le campus de l'ÉTS

1.9 Particularités du programme

Aide

Particularités du programme. La cas échéant, pour un baccalauréat en ingénierie, tableau des unités d'agrément (en annexes).

Exemple : Dans le cas d'un programme conduisant à l'exercice d'une profession réglementée, indiquer par quel ordre professionnel ou organisme de réglementation le programme sera reconnu.

À compléter par le promoteur

Cliquez ou appuyez ici pour entrer du texte.

1.10 Objectifs généraux

Les objectifs généraux de ce programme sont de former des spécialistes possédant une connaissance approfondie des techniques et technologies propres à l'analytique des données massives sur infonuagique et qui sont capables de mettre en œuvre et gérer une infrastructure infonuagique capable de gérer efficacement l'analyse de données massives autant en lot qu'en

temps réel et d'être en mesure de sélectionner et exploiter les services infonuagiques efficacement pour une solution efficace aux problèmes des organisations.

1.11 Objectifs spécifiques

Les objectifs spécifiques du programme sont d'acquérir les compétences suivantes :

- Reconnaître et communiquer efficacement les avantages de l'utilisation de l'infonuagique au niveau des dépenses d'investissement, des économies d'échelles, de l'agilité de l'adaptation de la capacité et de la portée mondiale des services pour l'analyse de quantités massives de données;
- Connaître l'impact de la présence d'une des cinq caractéristiques du domaine des technologies qui exploitent les données massives (c.-à-d. Volume, Vitesse, Variété, Véracité et Valeur) sur des infrastructures et services infonuagiques afin de mettre en œuvre pour une solution efficace d'analytique de données;
- S'initier aux défis de la mise en œuvre d'infrastructures et de services infonuagiques capables de traiter efficacement des quantités massives de données dans le but de les analyser. Éléments des compétences :
 - se familiariser avec les différents services d'infrastructure TI infonuagiques de compression, stockage et de diffusion d'objets;
 - choisir judicieusement, mettre en œuvre et sécuriser des instances de calcul et des services sur demande, à l'aide de patrons architecturaux infonuagiques, qui adressent les problématiques d'adaptation de la capacité et de montée/descente en charge selon le type de vitesse des données massives impliquées dans chaque cas de figure d'analyse de données (lots, temps réel);
 - identifier convenablement quelle technologie/services de base de données est adaptée au cas de figure impliquant des volumes, vitesses et variétés de données massives à analyser;
 - expérimenter et maîtriser les algorithmes et moteurs d'analyses infonuagiques permettant le traitement en parallèle sur grappes d'ordinateurs pour les cas de figure d'analyse de données en lots et en temps réel;
 - s'initier au domaine de l'analytique à grande échelle permettant de se familiariser avec les technologies d'analyses lors de l'ingestion de Lac de Données, de production de tableaux de bords interactifs et de requêtes efficaces sur des données massives.

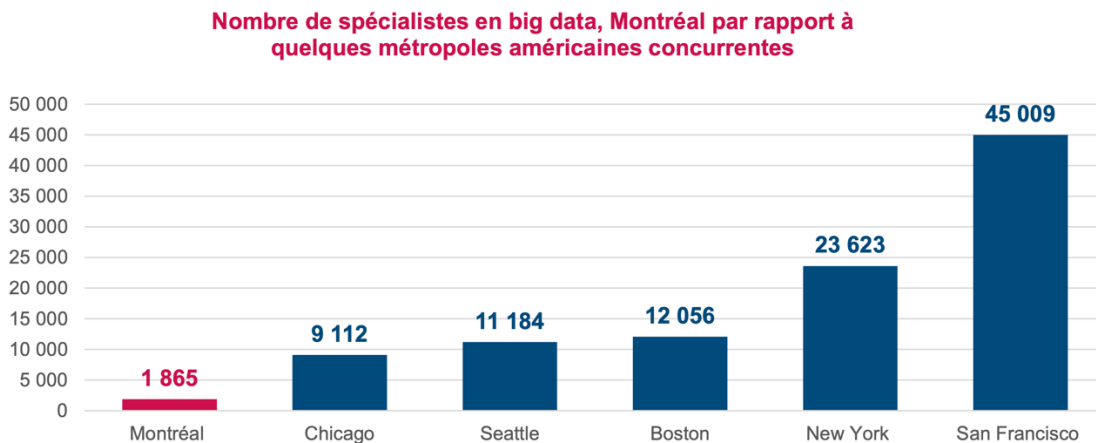
2. Motifs de création du programme

2.1 Opportunité socioéconomique/socioculturelle

L'étude « Profil du Big Data au Québec (53 p.12) identifie l'opportunité de poursuivre le développement de formations spécialisés en big data car ils identifient la demande croissante de spécialistes en forte hausse depuis 2015 (53 p94) et la proportion de diplômés au Québec est relativement faible (53 p95).

La figure suivante image l'opportunité dans le domaine car Montréal accuse un retard dans le nombre de spécialistes du domaine disponibles. Cette pénurie se situe entre 10,500 et 19,000 spécialistes (53 p86).

LE GRAND MONTRÉAL ACCUSE UN RETARD SUR LES PRINCIPAUX PÔLES NORD-AMÉRICAINS



Ainsi il y a une opportunité de « poursuivre le développement de formations spécialisées et adaptées aux besoins des entreprises » (53 p79).

En ajoutant une composante infonuagique forte à ce DESS, d'autres opportunités viennent s'ajouter au dossier, l'affranchissement des entreprises et des gouvernements de leur dette technique d'infrastructure désuète.

Une tendance forte actuelle est que « les services de TI consacrent la majeure partie de leur temps à régler des problèmes plutôt que de trouver des moyens nouveaux et plus efficaces de fournir les services » (54 p12) et selon l'IDC « les administrateurs des centres de données canadiens ont passé 74 % de leur temps à « garder les lumières allumées » en 2018,

comparativement à 70 % trois ans auparavant. Cette situation a été soulevée par les dirigeants d'entreprises qui progressivement prennent un virage majeur vers l'infonuagique au lieu d'investir dans leur centre de données.

C'est ainsi que le gouvernement du Québec, en 2015, a émis l'orientation suivante concernant ses infrastructures TI's « Les organismes publics privilégient les solutions infonuagiques lorsque celles-ci offrent un meilleur rapport qualité-prix et permettent une gestion efficace des risques, notamment à l'égard des renseignements personnels. » (51 p1)

Cette intention de rénover l'état Québécois par les technologies de l'information, le gouvernement s'est fixé l'objectif de « mettre à profit l'infonuagique » (49 p24), « Parce qu'elle a le potentiel d'accroître l'agilité gouvernementale et de diminuer les coûts d'acquisition et d'exploitation des infrastructures, des plateformes et des applications technologiques, l'infonuagique est perçue comme un important levier pour accroître la performance de l'État et améliorer les services aux citoyens et aux entreprises du Québec.»

Donc la mesure #17 de ce rapport du conseil du trésor vise à encourager les recours à l'infonuagique par les organismes publics.

Le gouvernement du Canada a rapidement suivi le pas « À l'automne 2017, SPC a commencé à attribuer des marchés à des fournisseurs de services infonuagiques qualifiés pour l'hébergement de données non classifiées. » (50 section 4.3)

« De plus, vos fournisseurs de technologie font le saut vers le nuage. Les déploiements de services infonuagiques sont de plus en plus omniprésents en raison de l'offre de produits commerciaux standards, et les déploiements sur place deviennent des exceptions.» (52 p3)

Il y a aussi l'impact sur l'environnement afin de réduire l'acquisition et la disposition d'équipement informatique et de se créer son propre centre de données privé, l'utilisation de l'infonuagique publique est souhaitable. Une première initiative est la mise en place d'une politique « Infonuagique publique en premier ». L'infonuagique publique est offerte par les grands fournisseurs tels que AWS, Azure et Google Cloud. Ces fournisseurs ont une préoccupation verte, par exemple, en 2017, Google a annoncé avoir atteint son objectif de 100% d'énergies renouvelables. C'est un exploit de taille, surtout si l'on considère que Google a lancé ces projets alors que la parité réseau n'était qu'un rêve pour l'énergie éolienne, et que l'énergie solaire était une technologie que seuls les riches Californiens et les Allemands ont installée sur leurs toits.

De son côté, AWS publie que son empreinte énergétique est 28% moins importante que le centre de données typique grâce à son plan d'approvisionnement en énergie verte et renouvelable. Amazon participe aussi à de nombreux programmes, notamment les principes de l'acheteur visant à accroître le pouvoir d'achat de sources d'énergie à faibles émissions de carbone et à l'American Council on Renewable Energy. Donc AWS tente de suivre l'exemple de Google. AWS a récemment annoncé que le centre de données de la région EST

(Ohio) a été construit en même temps que le parc éolien Amazon Central Park, un projet éolien de 100 mégawatts d'une capacité de 29 000 foyers américains par an. Ce projet rejoint le parc éolien Amazon de 208 mégawatts de la région EST de la Caroline du Nord, le parc éolien Fowler Ridge de 150 mégawatts et le parc éolien de 80 mégawatts Solar en Virginie. L'objectif d'AWS pour la fin de l'année 2020 est de faire en sorte que son infrastructure mondiale repose à 60% sur de l'énergie provenant de sources renouvelables.

Microsoft, de son côté publie l'objectif de 60% d'énergie renouvelable pour 2020. En 2019 Microsoft a doublé son engagement de réduire de 75% ses émissions de dioxyde de carbone au cours des onze prochaines années.

Prévision d'effectifs étudiants

Le tableau ci-après montre la prédiction de l'évolution de l'effectif étudiant pour le programme court de 2e cycle en en analytique de données massives sur infonuagique.

Année	2020-21	2022-23	2023-24	2024-25	2025-26
# étudiant prévus	15	30	45	50	55

Perspectives d'emploi

L'industrie Québécoise ne doit pas rater le virage numérique et se mobilise en conséquence.

Plusieurs associations québécoises, ordres professionnels et organismes publics ayant partie prenante dans le secteur québécois des technologies de l'information se sont concertés en vue d'une action commune pour prendre le virage numérique. De fait « [...] sentant l'urgence d'agir, le gouvernement québécois et canadien a émis une politique infonuagique qui lance le virage solidement.

Les enjeux sur le plan juridique, technique, numérique, humain et organisationnel découlant de l'utilisation des données massives et de l'infonuagique ont été : Modéliser le changement », organisé en février 2019 par Contech Bâtiment.³⁷ En juin 2019, une matinée de conférences sur les réalités et les enjeux du BIM au Québec a été organisée par AluQuébec Grappe de l'aluminium³⁸.

En parallèle, des études et des rapports ont été publiés, visant à analyser le déploiement du BIM au Québec et à démontrer son rôle essentiel dans la croissance de la productivité. En 2015, le Groupe de recherche en intégration et développement durable en environnement bâti (GRIDD), en collaboration avec le Centre d'études et de recherches pour l'avancement de la construction au Québec (CERACQ) a effectué un sondage auprès de l'industrie de la construction québécoise dont l'un des objectifs était d'esquisser le portrait de l'adoption des nouvelles technologies et des pratiques innovantes³⁹. Suite aux réponses reçus de la part de 541 participants, on a constaté que le taux d'adoption du BIM était de 31%. La catégorie des ingénieurs se distinguait des autres répondants, affichant le pourcentage le plus élevé d'utilisation du BIM (62%), suivie par la catégorie des architectes (41%).

Toutes ces activités et initiatives ont porté fruit, car l'industrie québécoise de la construction est maintenant préparée au virage numérique. « À la Ville de Québec, toutes les nouvelles constructions font maintenant l'objet d'une maquette numérique. Deux centres communautaires sont actuellement en construction. Un centre de glace et un édifice de bureaux sont en conception, puis deux autres pavillons de parc suivront. En parallèle, le grand projet de l'usine de biométhanisation se fait lui aussi en mode BIM. » Quant à la SQI, d'ici 2021, elle va recourir au BIM-PCI dans une trentaine de projets majeurs (de plus de 50 M\$) et « en 2023, selon la nouvelle feuille de route qui sera élaborée, il est probable que tous les projets soient réalisés de cette façon. »⁴⁸.

Bien qu'il soit impossible de chiffrer aujourd'hui les embauches possibles, mais la conjoncture sera favorable pour les diplômés sera excellent. L'ÉTS ne prendrait donc pas un risque élevé en lançant cette formation qui vise à élargir les compétences du Programme court de 2e cycle en BIM, le besoin étant confirmé par la table ronde et le besoin en termes de travail collaboratif et de pratiques dans le domaine de la construction étant plus que favorable.

Actuellement, les employeurs n'exigent pas de connaissances poussées en BIM et en transformation numérique à l'embauche, faute de candidats pouvant répondre à cette exigence. Les individus possédant une telle expertise trouveront très rapidement un emploi. Le Québec devra combler son retard en ce domaine pour révolutionner le monde de l'architecture et l'ingénierie par la maîtrise de nouveaux outils.

(49)

https://www.tresor.gouv.qc.ca/fileadmin/PDF/ressources_informationnelles/strategie_ti/s_trategie_ti.pdf

(50) <https://www.canada.ca/fr/gouvernement/systeme/gouvernement-numerique/plan-strategique-operations-numerique-2018-2022.html>

(51)

https://www.tresor.gouv.qc.ca/fileadmin/PDF/ressources_informationnelles/architecture_entreprise_gouvernementale/AEG_3_3/Sous_arbre_infonuagique.pdf

(52)

https://www2.deloitte.com/content/dam/Deloitte/ca/Documents/consulting/ca_cloud_po_v_FR_doc.PDF

(53) https://cdn2.hubspot.net/hubfs/1913735/MyBeeye-Oct-2016/Pdf/2016-02_etude-profil_du_big_data_au_quebec.pdf

(54) <https://itac.ca/wp-content/uploads/2019/09/ITAC-govt-Cloud-paper-Fr-sept2019-final.pdf>

2.2 Opportunité systémique

Dans le guide de l'infonuagique, publié en initialement en 2015 par le dirigeant principal de l'information du gouvernement du Québec, on souligne que « L'infonuagique (de l'anglais - cloud computing) constitue une tendance mondiale en matière d'acquisition de services technologiques dont l'un des objectifs est de diminuer les coûts d'exploitation des infrastructures technologiques et des applications. Il s'agit d'un nouveau mode d'acquisition qui permet aux individus et aux organisations d'accéder, par les technologies d'Internet, à un bassin de ressources informatiques configurables, externalisées et qui sont proposées sous forme de services. Ce nouveau mode de livraison de services permet aux consommateurs de s'approvisionner en services de technologies de l'information (TI) auprès d'un fournisseur infonuagique de façon automatisée et sur demande. La consommation des services est mesurée et facturée selon l'utilisation. L'infonuagique procure plusieurs avantages et bénéfices aux utilisateurs. En effet, les ressources infonuagiques offrent une agilité et une flexibilité à l'utilisation, puisqu'elles s'acquièrent rapidement, s'adaptent facilement à la demande et permettent un délestage tout aussi rapide. De plus, l'infonuagique présente des possibilités de faire des économies substantielles, puisqu'elle favorise une meilleure utilisation des infrastructures technologiques, réduisant les coûts en capitalisation, en exploitation et en entretien à l'échelle gouvernementale. » 51 (p.1)

La problématique principale reliée à l'enseignement universitaire de l'analytique des données massives sur infonuagique tient au fait de l'expertise requise dans plusieurs champs de spécialité du domaine de L'infrastructure matérielle et du génie logiciel. C'est la tendance forte actuellement de voir que les deux spécialités sont en train d'en devenir une seule. Cette réalité ou les développeurs déploient eux-mêmes leurs infrastructures induit un nouveau paradigme qui se heurte au mode de spécialisation traditionnel, et ce, même au niveau de l'enseignement.

Selon notre étude des cursus au niveau national et international, certains établissements d'enseignement ajoutent des cours portant sur l'infonuagique et le traitement de données massives au sein de leur cursus actuel au premier comme au deuxième cycle, alors que d'autres développent des cursus spécialisés ciblés seulement sur l'analytique de données. Il est clair que toutes ces connaissances ont un avantage d'être mise ensemble pour une exploitation efficace des quantités croissantes d'information.

Toutefois, l'offre de programmes en infonuagique, données massives et analytique de données des universités canadiennes reste assez limitée et son évolution accuse du retard par rapport aux besoins de l'industrie. Au Québec, HÉC Montréal offre une maîtrise en Gestion – Sciences des données et analytique d'affaires de 45 crédits avec projets supervisés. L'École Polytechnique offre un Programme court de 2e cycle en ingénierie et analytique de données. L'Université Laval et HÉC offrent aussi le certificat en analytique d'affaires/données massives.

Sinon il y a un cours en « analytique descriptive de données massives » à l'UQAM, deux cours « Intro to Big Data Technology » et « Big Data Analytics » à Concordia et le cours de « Mégadonnées » à l'Université de Shrebrooke. De plus, Téléq offre une formation en « analytique d'affaires ». À L'ÉTS nous offrons le MTI850 « Analytique de données massives » et le GTI778 « Infrastructures et services infonuagiques ».

Finalement, nous avons aussi développé, à l'ÉTS, une offre de formation (« les bases » et « l'architecture infonuagique ») pour ÉTS formation.

En somme :

- à la Polytechnique de Montréal ([Maitrise professionnelle \(OU DESS\) modulaire en génie informatique option ingénierie et analytique des données](#)).
- à l'UdeM, un microprogramme de 2^{ème} cycle ([en analyse des mégadonnées en santé](#))
- à l'ULaval, un microprogramme de 2^{ème} cycle ([en traitement des données massives](#))
- au HÉC, un microprogramme 1^{er} cycle ([en exploitation des données](#) en intelligence d'affaires) et un microprogramme de 2^{ème} cycle en ([analyses d'affaires](#))
- à l'ESG, un profil de spécialisation du DESS en TI ([analytique de données d'affaires](#))
- à l'USherbrooke, un microprogramme de 2^{ème} cycle en ([stratégie de l'intelligence d'affaires](#))
- à l'ULaval, un micorprogramme 2^{ème} cycle en ([analytique d'affaires](#))

Au niveau collégial

Un bon nombre de Cégep ont emboité le pas pour offrir des formations continue en infonuagique, mégadonnées et données massives.

C'est en 2015 que la direction du Service aux entreprises et de la formation continue du Cégep Limoilou a été la première institution d'enseignement dans le monde francophone à donner une formation préparatoire à l'obtention de la certification en sécurité de l'infonuagique (CCSK) reconnue par la Cloud Security Alliance (CSA). Ensuite plusieurs Cégeps ont suivi. Le Collège de Maisonneuve offre les cours Infonuagique et virtualisation et architecture infonuagique. Le CÉGEP Beauce-Appalaches offre une formation en ligne ' INITIATION AUX TECHNOLOGIES INFONUAGIQUES (CLOUD COMPUTING)'. Et le Cégep de Trois-Rivières avec sa formation en acquisition et traitement de données massives.

Ensuite des AEC sont apparus, comme ceux du Cégep de Sainte-Foy, du Collège Bois-De-BOULOGNE de spécialiste en mégadonnées et intelligence d'affaires.

Au Canada (à l'extérieur du Québec)

Au Canada, les programmes et les cours en BIM et l'industrie 4.0 se sont multipliés, les dernières années. En voici quelques-uns : Le Collège George Brown - Angelo DelZotto School of Construction Management de Toronto (Ontario) offre le Building Information Modeling

(BIM) Management Program (Postgraduate)- T412. Ce programme a une durée d'une année, à temps plein. Le Collège Algonquin (Ottawa, Ontario) propose un certificat en Building Information Modeling

- Nouvelle-Écosse - Dalhousie University (Institute for Big Data Analytics)
- Toronto - University of Toronto (Certificate in the Management of Enterprise Data Analytics) d'un maximum de 2 ans
- Toronto - Ryerson University (Certificate in Big Data Analytics, Big Data and Predictive Analytics)
- Toronto - Humber School of Media Studies & Information technology (Big Data Certificate)
- Vancouver - Simon Fraser University (Professional Master's Program in Big Data)

À l'étranger

À l'extérieur du Canada, les universités américaines offrent certains profils d'études au premier et/ou au deuxième cycle selon des modèles différents

- Arizona State University (Master degree in Business Analytics and Bachelor degree Business Data Analytics)
- Californie - Berkeley University (Master of Information and Data Science)
- Washington – U of Potomac (Master of Science in Data Analytics)
- Pennsylvanie – U of Pittsburgh (Big Data Analytics)

Caractère distinctif et apport spécifique du programme proposé :

Les programmes proposés par les écoles de gestion (HÉC, ESG et ULaVal) portent sur l'utilisation de techniques d'analyses de données (par exemple pour les analystes d'affaires). Le programme de l'université de Montréal vise spécifiquement l'analyse des données de la santé. Celui de polytechnique est un assemblage de cours existants. À L'ÉTS les cours proposés sont tous nouveaux, l'ensemble des travaux est exclusivement sur des services infonuagiques et nous désirons que nos finissant puissent non seulement utiliser les technologies analytiques mais le programme se concentre sur la capacité à mettre en œuvre les solutions, ce qui demande une maîtrise et une connaissance technique plus approfondie que n'importe quel autre programme offert au Québec actuellement. Il s'agit donc ici d'une compétence de concepteur de solution analytique de mégadonnées à l'aide de services infonuagiques par rapport à un utilisateur de ces technologies.

2.3 Opportunité institutionnelle

L'École de technologie supérieure de Montréal est la seule au Canada à proposer un programme d'ingénierie des TI's et de génie logiciel qui conjugue la conception d'infrastructure logicielle et de logiciel applicatif. Le département de génie des TI's et de logiciel accueille plus de 1530 étudiants au baccalauréat en 2020. Par conséquent, le département se positionne très favorablement pour offrir un DESS en analytique de données massives sur infonuagique, une formation à caractère multidisciplinaire qui pourrait intéresser les étudiants de l'ÉTS et aussi d'autres étudiants de la région.

Le programme de DESS en analytique de données massives sur infonuagique serait unique au Québec. Il correspond à un profil de formation en demande de la part de l'industrie, mais qui pourtant ne correspond à aucune offre à l'heure actuelle. Il se distinguerait des programmes offerts et mentionnés précédemment :

- par la combinaison de professeurs-chercheurs à la fine pointe de la recherche dans ce domaine, et de chargés de cours actifs dans l'industrie et chefs de file dans l'utilisation de l'infonuagique au Québec ; et
- par une démarche globale et intégrée qui inclurait non seulement la conception d'infrastructures infonuagiques pouvant traiter des données massives, mais également leur gestion et analyse en entreprise.

Le nouveau DESS vise à élargir les compétences de l'ingénieur en logiciel et TI et former des agents du changement dans les entreprises Québécoises afin de les aider dans leur virage numérique et dans le développement de l'agilité organisationnelle axée sur l'analytique des données. Il s'adresse à toutes les industries.

3. Structure du programme et cheminement des étudiants

3.1 Structure générale du programme

Ce qui est proposé ici serait donc de se positionner rapidement dans ce marché pour lequel les autres universités ont déjà lancé des initiatives. Il s'agit d'un programme d'études supérieures spécialisées (DESS) comportant les options suivantes: Un programme de **DESS en analytique de données massives sur infonuagique** comportant 2 cheminements:

Profil avec projet de synthèse de 15 crédits:

- 2 cours obligatoires, 3 cours optionnels et un projet* de 15 crédits

Profil avec activité de synthèse de 6 crédits

- 2 cours obligatoires, 6 cours optionnels et un projet* de 6 crédits

**Soit un projet d'intervention en entreprise ou un projet d'application supervisé par un professeur du département de génie logiciel et TI.*

3.2 Description des activités

Les activités propédeutiques, hors programme, suivantes (imposées selon le dossier de l'étudiant):

- **INFXXX** Systèmes d'exploitation et services Internet (3cr.);
- **LOG660** Bases de données de haute performance (4cr.);
- **LOG121** Conception orientée objet (4cr.);

L'activité obligatoire, hors programme, suivante :

- **ATE800** Intégrité intellectuelle : un savoir-être et un savoir-faire (0 cr.)

Activités obligatoires :

- **DM810** Architecture d'infrastructure infonuagique (3cr.);
- **DM815** Introduction aux données massives (3cr.); (prérequis, corequis DM810)

Activités optionnelles :

- **DM820** Moteur d'analyse pour le traitement distribué de données massives (3cr.);
- **DM825** Traitement de données massives en temps réel (3cr.);
- **DM830** Moteurs de recherche sur données massives non structurées (3cr.);
- **DM835** Introduction aux services analytiques et d'apprentissage machine infonuagique (3cr.);
- **DM840** Bases de données infonuagiques (3 cr.);
- **DM900** Sujets spéciaux en infonuagique et données massives (3 cr.)
- **MTI840** Sujets avancés sur l'Internet et l'infonuagique (3 cr.)
- **MTI850** Analytiques de données massives (3 cr.)

Sur approbation préalable du directeur du programme, deux des activités optionnelles peuvent être remplacées par des activités pertinentes d'autres programmes de 2e cycles offerts par l'ÉTS ou par d'autres universités.

Description des cours proposés :

DM810-Architecture d'infrastructure infonuagique (3cr.)

À la suite de ce cours, l'étudiant(e) sera en mesure :

- d'expliquer les fondements et les avantages de l'infonuagique pour le traitement efficace des données massives ;
- d'expliquer les différents défis reliés au déploiement et à la gestion des infrastructures infonuagiques ;
- de concevoir une infrastructure infonuagique basée sur des requis non fonctionnels ;
- d'appliquer des patrons d'architecture infonuagique.

Ce cours présente les différentes caractéristiques de l'infonuagique et les différents modèles de service. Ensuite, il aborde la proposition de valeur et décrit les services clés infonuagiques qui sont les composants de base pour concevoir une infrastructure infonuagique. L'étudiant sera capable de caractériser les différents services selon leur fonction dans une architecture et choisir le meilleur service pour leurs besoins basé sur la performance, la sécurité, la vitesse et l'efficacité requise. Le cours présentera des notions de : sécurité et protection des données, options de stockage, options de calculs et de mise en réseau, options de déploiement et gestion d'une infrastructure, un survol des bases de données disponibles, les notions d'évolutivité et de dimensionnement automatique. Il pourra aussi et comprendre et mettre en oeuvre les différents patrons architecturaux infonuagiques ainsi que des cadres permettant d'améliorer la conception d'architecture infonuagique pour traiter efficacement des quantités massives de données.

DM815-Introduction aux données massives (3cr.) (prérequis ou Co requis DM810)

À la suite du cours, l'étudiant(e) sera en mesure :

- de comprendre les limites des bases de données relationnelles pour traiter, en lots, des grandes quantités de données non structurées et semi-structurées ;
- de maîtriser les techniques : 1) d'ingestion (collection); 2) de persistance (storage); 3) de traitement (processing); et 4) d'analyse de données massives (analyse and visualisation);
- de concevoir une stratégie de grappe de serveurs infonuagiques distribués permettant le traitement efficace de données massives ;
- de se familiariser avec les différentes composantes de l'écosystème NoSQL d'un système de gestion de base de données non relationnelle orientée colonne et des techniques de compression des données.

Ce cours se veut une introduction aux applications du domaine des données massives pour des traitements analytiques en lots qui couvre une synthèse de l'écosystème Hadoop/Spark. Le cours présente des aspects nouveaux et d'intérêt majeur en compression, stockage et traitement de données ainsi qu'une introduction à un système de gestion de base de données non relationnelle orientée colonne. Il permet à l'étudiant de concevoir une application et de justifier ses choix de technologies selon les requis non fonctionnels d'un projet typique. Le cours est une introduction aux techniques d'ingestion de données massives, compression des données, le traitement en lot et une initiation aux activités d'analytique sur des quantités massives de données en utilisant des services infonuagiques.

DM820-Moteur d'analyse pour le traitement distribué de données massives (3cr.)

À la suite de ce cours, l'étudiant(e) sera en mesure :

- de démontrer l'impact des principaux mécanismes de systèmes distribués en fonction de la variété de modèles de systèmes distribués et de pannes;
- d'analyser et concevoir des systèmes distribués selon les exigences fonctionnelles et non fonctionnelles requises;
- de concevoir et implémenter des méthodes de réplication, tolérance aux pannes, coordination dans les systèmes distribués;
- de comprendre les manipulations de trames de données en mémoire (RDD) et d'utiliser un moteur d'analyse qui permet la parallélisations des algorithmes (diviser pour régner) sur une grappe d'ordinateurs;

Ce cours présente les principaux fondamentaux des systèmes distribués tels que le réseau informatique, horloges, les modèles de tolérance aux pannes, coordination, la réplication et les transactions. Le cours vise aussi à introduire les principes avancés de systèmes distribués, tels que les patrons pour le traitement de données massives (p. ex. MapReduce, RDD), réseau pair-à-pair et décentralisés, et les notions de cohérence, de disponibilité et de tolérance au partitionnement. Les systèmes distribués à grande échelle pour données massives seront utilisés à titre d'exemple.

DM825-Traitement des données massives en temps réel (3cr.)

À la suite de ce cours, l'étudiant(e) sera en mesure :

- de comprendre les systèmes distribués de messageries;
- de concevoir des applications temps réels qui utilisent des données en streaming;
- de comprendre la notion de stockage par bloc nécessaire pour prendre en charge le débit des données en streaming;
- de concevoir des opérateurs et requêtes sur des flots de données en temps réel.

Ce cours permet de concevoir et d'expérimenter les traitements, en temps réel, sur des flux de données à l'aide de différentes plateformes distribuées qui centralisent les messages qui transitent entre différentes applications (par exemple : Spark Streaming, Flink, Zookeeper et Kafka, Storm et Samza). On y présentera les techniques de stockage de données ainsi que les modèles de pipelines de données comportant : des producteurs, des consommateurs, des flux, des connecteurs et des partitions de données. Il sera aussi question de comprendre les journaux de transactions afin d'optimiser le traitement et la rétention des données.

DM830-Moteurs de recherche sur données massives non structurées (3cr.)

À la suite du cours, l'étudiant(e) sera en mesure:

- de comprendre et expliquer le fonctionnement des différentes techniques et services infonuagiques de recherche dans des données massives non structurées;
- de comparer et mettre en œuvre les différentes technologies et services infonuagiques de recherche dans des données massives non structurées.

Une des décisions importantes que le spécialiste en données massives doit prendre est le choix d'une technologie de recherche qui permet d'effectuer des recherches dans les données de toutes sortes. Ce cours présente les techniques permettant des recherches efficaces, en temps réel, de dans de grandes quantités de données non structurées (documents, textes, images, voix et vidéos). Ces techniques permettent d'identifier des objets, des personnes, des messages vocaux, du texte, des scènes et des activités dans des documents, des images et des vidéos. Ce cours présentera les techniques de recherche dans des données massives, incluant: des outils de recherche efficaces, des analyseurs de textes, d'images de voix et de vidéos, une évaluation de la pertinence, la gestion des index et le partitionnement.

DM835- Sciences des données et analytique à grande échelle (3cr.)

À la suite de ce cours, l'étudiant(e) sera en mesure:

- de mettre en œuvre un flux de données complet sur infonuagique pour l'analytique sur une quantité massive de données ;
- de comprendre les aspects théoriques du domaine de l'apprentissage machine;
- de préparer les données, concevoir un modèle et entraîner un modèle de prédiction à l'aide d'un cas d'étude comportant des données massives.

Un premier demi cours présentes un survol des techniques du domaine analytique sur des données statiques à l'aide de la conception d'un flux de données ne nécessitant pas l'étape d'ETL classique en utilisant une architecture matérielle infonuagique « serverless » comportant : l'ingestion de données massives, le dépôt des données dans un « Data Lake », l'analyse des données massives (c.-à-d. des requêtes interactives) lors de l'ingestion, la production d'un tableau de bord interactif et ensuite comment effectuer des requêtes.

DM840- Bases de données infonuagiques (3cr.)

À la suite de ce cours, l'étudiant(e) sera en mesure :

- de distinguer les différents types de bases de données et de choisir une technologie adaptée à l'usage visé par votre cas d'utilisation ;
- de comprendre les différents services infonuagiques de bases de données ;
- d'identifier les facteurs impactant les performances d'une base de données infonuagique.

Ce cours approfondit les notions et connaissances de base de données infonuagiques introduites au cours DM815. Il présente, comment choisir parmi les différents moteurs de bases de données infonuagiques disponibles selon votre cas d'utilisation, c'est-à-dire la base de données relationnelle, clé-valeur, en mémoire, de documents, à larges colonnes, graphiques, de séries chronologiques et de registre. Une fois choisis, des exemples pratiques de mise en œuvre pour les applications de mise à l'échelle d'Internet, de temps réel, d'applications du logiciel libre et d'applications métiers sont présentés. Le cours aborde aussi comment modéliser une base de données comportant un schéma flexible, comment utiliser les outils d'explication de requêtes pour faire l'optimisation : des index, de la réplication et la distribution des données dans le contexte de données massives.

DM900-Sujets spéciaux (3cr.)

À la fin de ce cours, l'étudiant sera en mesure :

- de décrire, présenter et critiquer les travaux récents dans les domaines de l'infonuagique et du traitement de données massives ;
- d'identifier les défis et les opportunités de recherche liés à l'adoption de l'infonuagique et au traitement de données massives.

L'objectif de ce cours est d'acquérir une compréhension approfondie des enjeux de la recherche dans les domaines de l'infonuagique et du traitement de données massives. Ce cours permet de présenter des sujets à tendance récente de ces domaines (c.-à-d. algorithmes, architecture système, conception de centres de traitements, performance des applications, tests, et bien d'autres) sous forme d'invitation d'invités. Il y sera également sujet de recherche récente et avancements connexes tels que la planification des tâches de traitement de données, la configuration dynamique des équipements TI, la gestion des infrastructures virtuelles et de conteneurs ainsi que de développement/maintenance adapté à des domaines émergents.

MTI840-Sujets avancés sur l'Internet et l'infonuagique

Au terme de ce cours, les étudiants seront être en mesure :

- de maîtriser les technologies clés de l'Internet et des infrastructures infonuagiques du futur;
- de démontrer une compréhension approfondie des défis de recherche liés au déploiement de l'Internet et des infrastructures infonuagiques du futur;
- d'analyser, évaluer et présenter les travaux de recherche sur l'infonuagique et l'Internet du futur.

L'objectif est de présenter les technologies émergentes reliées à l'évolution de l'Internet et des infrastructures infonuagiques tels que la virtualisation des réseaux, la réseautique définie par logiciel, les réseaux 5G et l'Edge Computing. Il vise aussi à familiariser les étudiants avec les travaux de recherche récents dans ce domaine ainsi que les différents défis de recherche qui y sont reliés. Les objectifs seront atteints par la prestation hebdomadaire d'un cours, par des lectures obligatoires et suggérées, par des présentations préparées par les étudiants en classe, par l'évaluation d'articles scientifiques, par des séminaires, par la réalisation de projets reliés aux sujets abordés dans le cours.

MTI850-Analytiques des données massives

Au terme de ce cours, l'étudiant sera mesure :

- d'illustrer et expliquer la nature des systèmes d'analyse de données;
- d'appliquer les connaissances en apprentissage machine requise pour concevoir tels systèmes;
- d'utiliser les outils appropriés pour valider et évaluer la performance de tels systèmes, combiner narration, code et graphiques pour créer des documents analytiques convaincants;
- d'aborder les problèmes de la science des données à grande échelle avec créativité et esprit d'initiative.

Ce cours présente les concepts pour effectuer une analyse statistique de très grands ensembles de données qui ne tiennent pas sur un seul ordinateur. Ce cours vous permettra développer vos connaissances en analyse de données massives et améliorerez vos compétences en programmation et en mathématiques. Vous apprendrez à utiliser des outils analytiques essentiels pour l'analyse statistique des données massives. Plusieurs problèmes applicatifs seront étudiés et différentes méthodes et outils pour effectuer ce type d'analyse seront étudiés.

3.3 Description des formules pédagogiques

Aide

Description des formules pédagogiques retenues et justifier, au besoin (si trop volumineux, à joindre en annexe).

Note : Des règlements pédagogiques particuliers peuvent concerner le cheminement de l'étudiant dans le programme, l'agencement chronologique de certaines activités les unes par rapport aux autres et les modalités d'application de certains articles des Règlements généraux de l'UQ.

À compléter par le promoteur

Cliquez ou appuyez ici pour entrer du texte.

3.4 Grille de cheminement

Aide

Tableau (grille de cheminement selon les régimes d'études à temps complet et à temps partiel).

À compléter par le promoteur

Cliquez ou appuyez ici pour entrer du texte.

4. Ressources humaines et matérielles

4.1 Ressources professorales

Voici la liste des ressources qui ont démontré un intérêt pour enseigner au nouveau programme d'études graduées en données massives de l'ÉTS.

À l'ÉTS, le **professeur Alain April** auteur de cette initiative, prendrait la responsabilité de ce nouveau programme advenant que L'ÉTS lui donne un demi dégrèvement d'enseignement. Professeur April effectue des projets de recherche et de développement dans le domaine du BigData depuis 2007. Sa première publication du domaine, avec l'étudiant Jean-Daniel Cryans, participant au projet Hadoop et qui effectuait des travaux chez StumbleUpon, à Silicon Valley par la suite. Dr April travaille actuellement sur le projet Adam - BigData appliqué à la génomique. Il appuie aussi le groupe d'intérêt BigData Montréal, depuis 2012, et est représentant canadien de deux comités ISO dans ce domaine: 1) en normalisation du Cloud Computing (sc38); et, 2) en normalisation du BigData (wg9). Il serait prêt à enseigner les cours DM815, DM830 et DM840.

À l'ÉTS, **Abdelaoued Gherbi** professeur au département de génie logiciel et TI, a une expertise en infonuagique et intergiciels. Il serait prêt à enseigner les cours GDM825, DM830.

À l'ÉTS, **Christian Desrosiers** professeur au département de génie logiciel et TI, a une expertise en apprentissage machine et logiciels de recommandation. Il serait prêt à enseigner les cours DM840.

À l'ÉTS, **Allesandro Koerich** nouveau professeur au département de génie logiciel et TI, a une expertise en apprentissage machine, classification, biométrie et recherche de descripteurs pour l'audio et la musique. Il serait prêt à enseigner les cours DM835 et DM900.

À l'ÉTS, **Mohamed Faten Zhani**, nouveau professeur récemment arrivé au département de génie logiciel et TI. Ses travaux de recherches portent sur l'infonuagique et le traitement des données massives. Il serait prêt à enseigner les cours DM815 et DM900.

À l'ÉTS, **Kaiwen Zheng** est professeur au département de génie logiciel et TI depuis quelques années. Kaiwen est reconnu pour ses connaissances en programmation parallèle, et systèmes distribués. Il serait prêt à enseigner les cours DM820.

4.2 Autres ressources humaines

À l'ÉTS, **Mathieu Dupuis**, étudiant au Ph.D. a obtenu son diplôme d'ingénieur de l'ÉTS en génie logiciel. Il possède une accréditation AWS pour ce cours ce qui lui permet d'utiliser les contenus et laboratoires de AWS pour cette formation. Il enseigne présentement deux cours de conception d'infrastructures infonuagique AWS au sein d'ÉTS formation depuis quelques années. Il serait prêt à enseigner le cours DM810 et MTI850.

À l'ÉTS, **David Lauzon**, travaille chez TickSmith Corp. à titre d'architecte logiciel. TickSmith a développé une plateforme de données massives fondée sur Hadoop dans le domaine de la gestion des données financières et transactions boursières. Il a enseigné le cours d'introduction au BigData, à la maîtrise en génie logiciel, depuis l'automne 2015. Il a développé ce cours et donné un demi cours, le GTI780, sur ce même sujet au baccalauréat du GTI de l'ÉTS par le passé. David serait prêt à enseigner les cours DM815, DM820 et DM825.

Olivier Mirandette travaille à titre de VP ingénierie chez Plusgrade. Il a développé et donné des cours d'introduction aux données massives au CÉGEP Bois-de-Boulogne dans le cadre de la formation des adultes. Olivier est un des fondateurs et responsables du groupe d'intérêt BigData Montréal. Olivier serait prêt à enseigner le cours DM825.

Khaled Tannir est consultant spécialisé et auteur en BigData. Il a publié deux livres dans ce domaine: *Optimizing Hadoop for MapReduce* et *RavenDB Beginner's guide*. Khaled a développé et enseigné des cours d'introduction aux données massives au Cégep Bois-de-Boulogne dans le cadre de la formation des adultes. Khaled est prêt à enseigner le cours DM830.

4.3 Ressources matérielles

La partie expérimentale/laboratoire/devoirs sera effectuée sur les services infonuagiques des fournisseurs : Google, AWS et MsAzure à l'aide de crédits des programmes d'éducation qui seront émis aux étudiants. Les essais successifs, depuis 2013, au département de génie logiciel et TI démontrent que l'ÉTS serait prête à initier un programme opéré complètement sur les services infonuagiques de ces fournisseurs en prenant bien soins de rendre responsables les étudiants de leur crédits pour les cours.


Services infonuagiques populaires dans le marché (en 2020):


Selon le rapport de Gartner de 2019, trois grands fournisseurs de plateformes infonuagiques publiques sont leaders de cette industrie. En première place : **Amazon Web Services**, en seconde place : **Microsoft Azure** et en troisième place **Google Cloud**. Il serait souhaitable que les services des deux leaders de cette industrie soient utilisés et

expérimentés par les étudiants dans le cadre de ce programme de formation. Afin d'avoir accès à ces plateformes gratuitement, des ententes sont requises:

AWS : depuis 2016, le professeur Alain April est l'administrateur autorisé de l'ÉTS (voir lettre annexée à ce document) concernant les deux programmes éducationnels de AWS.

Le premier, **AWS Educate**  donne accès à un portail académique qui fournit aux professeurs et étudiants les ressources dont ils ont besoin pour l'apprentissage

de l'infonuagique d'AWS. Le second programme est **AWS academy**,  pour

lequel l'étudiant Mathieu Dupuis est accrédité à titre d'instructeur certifié,  ce qui donne le droit exclusif d'utiliser le matériel de AWS pour enseigner des cours pour lequel il a été accrédité (*c.-à-d. en date d'aujourd'hui les contenus des cours AWS Cloud Foundation et AWS Cloud Architecting*). Ces accréditations sont données seulement à la suite : 1) d'une formation spécialisée offerte par AWS; 2) du succès à l'examen de certification; et 3) au succès d'une prestation de formation démontrant les connaissances acquises devant les experts de AWS academy.

Microsoft Azure : Des essais et de la formation ont eu lieu avec cette plateforme depuis 2018. Un accès monazure.etsmtl.ca est disponible aux étudiants.

Google Cloud : Le programme CloudEDUGrants@google.com de Google offre 4 programmes : 1) Crédits Google pour les cours (le prof. April a fait une demande pour un cours d'essai été 2020, qui a été acceptée - programme Google Cloud Education Platform Grants for Computer Science \$50 par étudiant; 2) Contenus de formation Coursera et plateforme d'essais Qwicklabs, une formation en ligne, à votre rythme; et 3) formation de certification Carrer Readiness : Associate Cloud Engineer. Finalement il est possible d'obtenir de crédits pour la recherche

Le professeur Alain April a été accepté par Google à titre d'administrateur du programme de formation : [coursera.org](https://www.coursera.org), Coursera program: **CA-ACE Track-École de technologie supérieure-AAp**.

5. Cadre réglementaire

5.1 Conditions / Processus d'admission

Ce programme s'adresse donc aux professionnels en informatique, en logiciel et en génie logiciel qui possèdent déjà une expérience en programmation de logiciels, base de données et systèmes d'exploitation (** Linux/Unix, SQL, programmation orientée objet), et qui veulent s'orienter vers ce nouveau domaine. Il ne vise pas à former des experts, mais plutôt à rendre des professionnels aptes à connaître, comprendre, comparer, choisir judicieusement, configurer, concevoir, développer et utiliser les services infonuagiques publics et les technologies appropriées, pour des cas de figure comportant des données massives, dans les projets qui les concernent. Ils seront les seuls à détenir cette vue d'ensemble et expertise au Québec alors l'ÉTS profitera de cette réputation.

** des cours de propédeutiques seront offerts pour les candidats qui n'ont pas ces qualifications

5.2 Durée / régime d'études

Durée des études au DESS

Temps complet : 4 sessions (1 an et 4 mois), soit :

- 1 à 3 sessions pour les cours;
- 1 à 2 sessions pour l'activité de synthèse.

Temps partiel : 8 sessions (2 ans et 8 mois), soit :

- à 7 sessions pour les cours;
- 1 à 3 sessions pour l'activité de synthèse.

5.3 Équivalences

Aide

Équivalences

À compléter par le Décanat des études

Cliquez ou appuyez ici pour entrer du texte.

5.4 Supervision / Évaluation des études

Comme les autres DESS du département LOGTI

Annexes

Aide

- Recommandations des instances

À compléter par le Décanat des études et l'UQ

- Les lettres d'appui reçues par les promoteurs;
- Liste des personnes consultées au cours du processus d'élaboration du projet;
- CV des professeurs;
- Tableau synthèse des contributions à l'avancement de la recherche et des contributions à la formation;
- Descriptifs des activités (cours, laboratoires, stages, activités de recherche);
- Liste d'experts (pourront être consultés au besoin et dont les compétences sont reconnues dans la discipline du programme proposé;
- Toute autre information utile.

À compléter par le promoteur/autres (voir précédemment)