

S.A.003 BIG DATA

1. Informations sur l'unité de cours/module

| | | | | | |
|----------------------------------|--|--------------------------|----------------------------------|---------------------------------------|---------------------|
| Faculté | Informatique, Microélectronique et Calculatrices | | | | |
| Département | Informatique et Ingénierie des Systèmes | | | | |
| Cycle d'études | Études universitaires de licence, Cycle I | | | | |
| Programul de studiu | 0612.3 Science des données | | | | |
| Année d'études | Semestre | Type d'évaluation | Catégorie formative | Catégorie d'option Optionnelle | Crédits ECTS |
| III (enseignement en présentiel) | V | E | S – unité de cours de spécialité | A - unité de cours au choix | 4 |

2. Temps total estimé

| | | | | | |
|--|--------------------|-------------------|--------------------|-----------------------------|------------------------------|
| Total d'heures dans le plan d'enseignement | Parmi lesquelles | | | | |
| | Heures d'auditoire | | Travail individuel | | |
| | Cours | Travaux pratiques | Projet de l'année | Étude du matériel théorique | Préparation des applications |
| 120 | 30 | 30 | | 30 | 30 |

3. Préconditions d'accès à l'unité de cours/module

| | |
|-------------------------------|--|
| According to the curriculum | L'étude repose sur les disciplines suivantes : Mathématiques discrètes, Algèbre, Théorie des probabilités et statistiques mathématiques, Méthodes de programmation, Méthodes d'analyse des données. De plus, les étudiants doivent maîtriser la structure de l'ordinateur, le système d'exploitation Windows 10, les langages de programmation C#, C++, HTML, PHP, Python, et l'utilisation d'un éditeur de texte. Des connaissances avancées en systèmes modernes de gestion de bases de données, tels que : MySQL, MS SQL Server 2019, CMS Wordpress, frameworks Laravel, CodeIgniter, Symphony, etc., sont également requises. Ils doivent avoir des compétences dans la création d'applications Desktop ou Web, Front-End et Back-End, en utilisant des bases de données opérationnelles, ainsi que dans la modélisation multidimensionnelle des données, les entrepôts de données, la technologie OLAP, les outils BI (VS, Power BI), la technologie Data Mining, le travail avec la technologie Cloud, les plateformes Azure, AWS (Amazon Web Services) et Google Cloud. |
| According to the competencies | Les étudiants doivent avoir les connaissances et compétences suivantes pour maîtriser la discipline académique "Big Data" : la capacité d'utiliser correctement les TIC pour résoudre des problèmes professionnels appliqués, en utilisant les compétences de modélisation mathématique des phénomènes socio-économiques, de l'environnement ambiant, du domaine technique, ainsi que dans les domaines : financier-bancaire, enseignement, médecine, assurances, marketing, etc., compétences acquises dans les disciplines mentionnées ci-dessus. Les résultats principaux de la discipline "Big Data" pourront être utilisés à l'avenir pour étudier les disciplines : "Analyse des données", "Méthodes de protection de l'information", "Science des données", "Analyse Big Data", "Informatique en entreprise", "Traitement Big Data", "Analyse Big Data", "Technologies Big Data". |

4. Conditions for conducting the educational process for

| | |
|--------------------------------|---|
| Cours | Pour la présentation du matériel théorique, la salle de cours doit être équipée d'un projecteur et d'un ordinateur. Les retards des étudiants ainsi que les conversations téléphoniques pendant le cours ne seront pas tolérés. |
| Laboratoire/séminaire/pratique | Les étudiants devront perfectionner les rapports conformément aux conditions prévues par les indications méthodologiques. La date limite de soumission du travail de laboratoire/pratique, individuel ou en groupe, sera lors de la prochaine séance de travail, sous format électronique. Pour toute remise tardive du travail effectué, une pénalité de 1 point par semaine de retard sera appliquée. |

5. Compétences spécifiques acquises

| | |
|-------------------------------------|--|
| Compétences professionnelles | CP1. Gestion du niveau de service (A.2)** <ul style="list-style-type: none"> ➤ K2 Comment comparer et interpréter les données de gestion. CP2. Conception et développement d'applications (B.1) <ul style="list-style-type: none"> ➤ K4 Technologies de pointe. ➤ K5 Langages de programmation. ➤ K6 Systèmes de gestion de bases de données (SGBD). ➤ K7 Systèmes d'exploitation et plateformes logicielles. |
|-------------------------------------|--|

| | |
|---------------------------|--|
| | <ul style="list-style-type: none"> ➤ K8 Environnement de développement intégré (IDE). <p>CP3. Intégration des composants (B.2)</p> <ul style="list-style-type: none"> ➤ K2 L'impact de l'intégration d'un système sur l'organisation ou sur le système existant. ➤ K4 Techniques de test d'intégration. ➤ K5 Outils de développement (par exemple, environnement de développement, gestion, contrôle des modifications et accès au code source). ➤ K6 Bonnes pratiques de conception. <p>CP5. Fourniture de services (C.3)</p> <ul style="list-style-type: none"> ➤ K1 Mode d'interprétation des exigences concernant la prestation de services informatiques. ➤ K2 Meilleures pratiques et normes pour la prestation de services informatiques. ➤ K6 Spécificités des technologies liées au web, au cloud et aux outils mobiles. <p>CP6. Gestion de l'information et des connaissances (D.10)</p> <ul style="list-style-type: none"> ➤ K1 Méthodes d'analyse de l'information et des processus métiers. ➤ K3 Défis liés à la taille des grands volumes de données (Big Data). ➤ K4 Défis liés aux données non structurées (par exemple, l'analyse des données). <p>CP7. Identification des besoins (D.11)</p> <ul style="list-style-type: none"> ➤ K2 Besoin d'affaires. ➤ K3 Processus et structure de l'organisation. ➤ K4 Techniques d'analyse des besoins des clients. <p>CP8. Marketing numérique (D.12)</p> <ul style="list-style-type: none"> ➤ K1 Stratégies de marketing. ➤ K2 Technologies web. ➤ K9 Problèmes/exigences juridiques. <p>CP9. Gestion de la relation client (E.4)</p> <ul style="list-style-type: none"> K1 Processus organisationnels, y compris la prise de décisions, les budgets et la structure de gestion. |
| Compétences transversales | <p>CTL1 Autonomie et responsabilité Démontre l'exécution responsable des tâches professionnelles dans des conditions d'autonomie.</p> <p>CTL2 Interaction sociale Exécute les rôles et activités spécifiques au travail en équipe et distribue les tâches entre les membres à des niveaux subordonnés.</p> <p>CTL3 Développement personnel et professionnel Prend conscience de la nécessité de la formation continue en utilisant efficacement les ressources et les techniques d'apprentissage pour le développement personnel et professionnel.</p> |

6. Objectifs de l'unité de cours/module

| | |
|---------------------------|---|
| Objectif général | Familiarisation, initiation et formation des étudiants afin qu'ils comprennent et utilisent dans la pratique l'essence et les spécificités du cours "Big Data", comment l'utilisation pratique des connaissances et des compétences acquises dans les cours précédents, "Bases de données" et "Gestion des bases de données", s'intègre dans ce processus, comment elles s'appliquent à la gestion de Big Data, aux bases de données opérationnelles (BDO), aux entrepôts de données (DWH) et aux systèmes modernes de gestion de ces derniers, ainsi que le développement d'habiletés, de compétences pratiques et l'acquisition de compétences tant sur le plan analytique (modélisation statistique, méthodes prédictives, techniques d'apprentissage automatique) que technologique (langages de programmation, architectures et plateformes). Cela permettrait aux étudiants, futurs spécialistes, d'appliquer les connaissances acquises tout au long des cours mentionnés pour résoudre les problèmes les plus divers dans l'économie nationale, en utilisant la simulation de processus et de phénomènes dans des domaines d'activité proches des futurs spécialistes dans l'application des TIC, formés au sein de la faculté d'Informatique et Microélectronique, dans le programme d'études - 0612.2 Gestion de l'information. |
| Les objectifs spécifiques | <p>Au niveau des connaissances :</p> <ul style="list-style-type: none"> • La capacité de travailler avec des concepts fondamentaux dans le domaine de la modélisation mathématique et de l'analyse statistique, ainsi que la capacité de les utiliser dans des contextes pratiques. • La capacité d'identifier, de mettre en œuvre et d'utiliser des algorithmes d'extraction de modèles à partir des données en utilisant des méthodes statistiques et des techniques d'apprentissage automatique. • La capacité de comprendre et d'appliquer les principes du traitement distribué des données et de l'utilisation des architectures de calcul haute performance. • La capacité d'utiliser des plateformes et des technologies spécifiques au traitement de grands volumes de données et de mettre en œuvre des applications évolutives. • La capacité d'adapter des solutions basées sur des approches data-driven pour résoudre des |

| | |
|--|--|
| | <p>problèmes spécifiques à un domaine d'application particulier.</p> <ul style="list-style-type: none"> • La capacité de planifier et d'organiser l'activité de recherche dans un domaine par l'organisation du travail en groupe ; la compréhension des responsabilités et le respect des règles d'éthique professionnelle et de la confidentialité des données. • L'utilisation de méthodes et techniques efficaces d'apprentissage, d'information et de spécialisation ; la capacité à sélectionner et analyser de manière critique les ressources bibliographiques ; l'engagement dans un plan de développement professionnel continu en mettant à jour les connaissances et compétences en corrélation avec les progrès scientifiques et technologiques réalisés dans le domaine. <p>Au niveau des compétences :</p> <ul style="list-style-type: none"> • Élaborer des stratégies de Big Data, en fonction du contexte de l'entité qui les utilise, des objectifs définis pour son développement et de la création de valeur ajoutée. • Développer des flux et des pipelines de données nécessaires à la gestion du "Big Data", facilitant les activités de gestion de l'information nécessaires au processus décisionnel de l'entité. • Analyser l'utilisation du "Big Data" en termes d'efficacité et d'utilisation des données, lorsqu'il s'agit d'un projet informationnel concret lié au processus décisionnel d'une entité, soutenu par le "Big Data". • Utiliser Python, Power BI, Visual Studio, des technologies cloud et des plateformes cloud telles que AWS, Azure, Google Cloud pour l'analyse et l'analyse du Big Data. |
|--|--|

7. Le contenu de l'unité de cours/module :

| La thématique des activités pédagogiques | Le nombre d'heures | |
|--|----------------------------|----------------------------------|
| | enseignement en présentiel | enseignement à fréquence réduite |
| Thématique des conférences | | |
| TP1 Introduction au cours. Règles. Notions. Définitions. Caractéristiques. | 2 | |
| TP2 Rôle du Big Data dans le processus décisionnel. Stratégie de données Big Data. Big Data dans le SSD d'une entité publique/privée. | 2 | |
| TP3 Big Data. Types de données dans le Big Data. Données structurées, semi-structurées, quasi-structurées et non structurées. | 2 | |
| TP4 Big Data, Data Engineering. Big Data Engineer. | 2 | |
| TP5 Architecture Big Data – Vue d'ensemble et détaillée. | 2 | |
| TP6 Big Data. Lacs de données. | 2 | |
| TP7 Technologies Big Data. Outils. | 2 | |
| TP8 Plateforme Azure Databricks. Delta Lake. | 2 | |
| TP9 Plateforme Azure Databricks Lakehouse. | 2 | |
| TP10 Azure Databricks Lakehouse ETL. | 2 | |
| TP11 Applications Big Data. Conversion des données non structurées en données exploitables. Entrepôts de données en Big Data, Outils de gestion, NOSQL. | 2 | |
| TP12 Entrepôts de données en Big Data, Outils de gestion, NOSQL. Théorème CAP. MongoDB. Atlas MongoDB. | 2 | |
| TP13 Hadoop et MongoDB en Big Data. Analytique Big Data. | 2 | |
| TP14 Analytique Big Data et DSS. Outils. | 2 | |
| TP15 BI, Analytique Big Data et Science des Données. | 2 | |
| Total cours: | 30 | |

| Thématique des activités didactiques | Nombre d'heures | |
|---|----------------------------|------------------------------|
| | Enseignement en présentiel | Enseignement à temps partiel |
| Thématique des travaux pratiques | | |
| TP1. Big Data dans le SSD d'une entité commerciale/publique. Élaboration d'une stratégie de données Big Data. Étude de cas. | 2 | |
| TP2. Statistiques. Python. Applications en Big Data. Importance des statistiques pour le Big Data. Étude de cas/Exemples. | 2 | |
| TP3. Modélisation du Big Data. Projet. Éléments de base. Structure. Étude de cas. | 2 | |
| TP4. Flux de données. Pipeline de données. Étude de cas. | 2 | |
| TP5. Big Data. Projet. EDA (Exploratory Data Analysis). Étude de cas/Exemples. | 2 | |
| TP6. Cartographie des données. Conversion des données non structurées en données exploitables. Outils. Étude de cas/Exemple. | 2 | |
| TP7. Pipeline de données SSIS/ETL pour MS SQL SERVER et VS. Étude de cas/Exemple. | 2 | |

| | | |
|--|-----------|--|
| TP8. Parsing et Scraping des données. Études de cas/Exemples. | 2 | |
| TP9. Construction d'un pipeline ETL avec Azure Databricks et Delta Lake. Étude de cas/Exemple. | 2 | |
| TP10. Traitement des données avec la plateforme Azure Databricks. Étude de cas/Exemple. | 2 | |
| TP11. Conversion des données non structurées en données exploitables. Entrepôts de données en Big Data, outils de gestion. NOSQL. MongoDB. Atlas MongoDB. | 2 | |
| TP12. Entrepôts de données en Big Data. Outils de gestion. NOSQL. MongoDB. Atlas MongoDB. Études de cas/Exemples. | 2 | |
| TP13. MongoDB dans le Big Data. Analyse Big Data. Étude de cas/Exemple. | 2 | |
| TP14. Analyse Big Data et DSS. Outils. Exemples. | 2 | |
| TP15. BI, Analyse Big Data et Science des Données. Étude de cas/Exemples. | 2 | |
| Total des travaux pratiques: | 30 | |

8. Références bibliographiques

| | |
|-------------|---|
| Principales | <ol style="list-style-type: none"> 1. Big Data Analytics with Spark, Copyright © 2015 by Mohammed Guller, ISBN-13 (pbk): 978-1-4842-0965-3, ISBN-13 (electronic): 978-1-4842-0964-6, www.apress.com/source-code/, OR www.it-ebooks.info 2. Thomas Erl, Wajid Khattak and Paul Buhler, Big Data Fundamentals, Concepts, Drivers & Techniques, Copyright © 2016 Arcitura Education Inc., ISBN-13: 978-0-13-429107-9 3. Raul Estrada Isaac Ruiz, Big Data SMACK: A Guide to Apache Spark, Mesos, Akka, Cassandra, and Kafka, Mexico City Mexico City, ISBN-13 (pbk): 978-1-4842-2174-7 ISBN-13 (electronic): 978-1-4842-2175-4, DOI 10.1007/978-1-4842-2175-4, Library of Congress Control Number: 2016954634, Copyright © 2016 by Raul Estrada and Isaac Ruiz 4. Wes McKinney, Python for Data Analysis, Copyright © 2013 Wes McKinney. All rights reserved. Printed in the United States of America. Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472. 5. John Wiley & Sons, Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, 10475 Crosspoint Boulevard, Indianapolis, IN 46256, www.wiley.com Copyright © 2015 by John Wiley & Sons, Inc., Indianapolis, Indiana, Published simultaneously in Canada ISBN: 978-1-118-87613-8, ISBN: 978-1-118-87622-0 (ebk) ISBN: 978-1-118-87605-3 (ebk) 6. Peters Morgan, Data analysis from scratch with Python. Step by step guide © Copyright 2016 by AI Sciences LLC, All rights reserved. First Printing, 2016, Edited by Davies Company Ebook Converted and Cover by Pixels Studio Published by AI Sciences LLC, ISBN-13: 978-1721942817, ISBN-10: 1721942815 7. Building Big Data and Analytics Solutions in the Cloud International Technical Support Organization December 2014 8. Big Data Principles and Paradigms, Edited by Rajkumar Buyya The University of Melbourne and Manjrasoft Pty Ltd, Australia Rodrigo N. Calheiros The University of Melbourne, Australia Amir Wahid Dastjerdi The University of Melbourne, Australia, Morgan Kaufmann is an imprint of Elsevier 50 Hampshire Street, 5th Floor, Cambridge, MA 02139, USA Copyright © 2016 Elsevier Inc. All rights reserved. 496 pag 9. 10 Best Big Data Books in 2023 [Learn Big Data ASAP] https://azure.microsoft.com/en-us/products https://realtoughcandy.com/best-big-data-books/ 10. 20 Best Big Data Books for Beginners https://bookauthority.org/books/beginner-big-data-books 11. Must Read Books for Beginners on Big Data, Hadoop and Apache Spark https://www.analyticsvidhya.com/blog/2015/10/books-big-data-hadoop-apache-spark/ https://dokumen.pub/big-data-concepts-technology-and-architecture-9781119701828.html 12. Microsoft Azure for students https://engineering.buffalo.edu/computer-science-engineering/information-for-students/information-technology/software-distribution/microsoft-azure-for-students.html https://learn.microsoft.com/en-us/azure/education-hub/azure-dev-tools-teaching/azure-students-program https://azure.microsoft.com/en-us/resources/students?activetab=pivot:githubtab https://learn.microsoft.com/en-us/azure/education-hub/azure-dev-tools-teaching/program-faq https://www.techopedia.com/definition/26434/azure-service-platform 13. Big Data Analytics - Statistical Methods. https://www.tutorialspoint.com/big_data_analytics/statistical_methods.htm 14. Майер-Шенбергер, В. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим / Виктор Майер-Шенбергер, Кеннет Кукьер ; пер. с англ. Инны Гайдюк. — М. : Манн, Иванов и Фербер, 2014. — 240 15. Гусева Е. Н. Теория вероятностей и математическая статистика : Учебники и учебные пособия для ВУЗов [Электронный ресурс] - Москва : ФЛИНТА , 2016 - 220 – Режим доступа: |
|-------------|---|

| | |
|---------------|---|
| | <p>http://biblioclub.ru/index.php?page=book_red&id=83543</p> <p>16. Гутова С. Г., Алтемерова О. А. Теория вероятностей и математическая статистика: Учебники и учебные пособия для ВУЗов [Электронный ресурс] - Кемерово : Кемеровский государственный университет , 2016 - 216 - Режим доступа: http://biblioclub.ru/index.php?page=book_red&id=481538</p> <p>17. Кочетков Евгений Семенович. Теория вероятностей и математическая статистика: Учебник [Электронный ресурс] : Форум , 2018 - 240 - Режим доступа: http://znanium.com/go.php?id=944923</p> <p>18. Миркин Б. Г. Введение в анализ данных. Учебник и практикум [Электронный ресурс] : М.:Издательство Юрайт , 2019 - 174 - Режим доступа: https://biblioonline.ru/book/vvedenie-v-analiz-dannyh-432851</p> |
| Supplimentare | <p>1. Белько Иван Васильевич. Теория вероятностей, математическая статистика, математическое программирование : Учебное пособие [Электронный ресурс] , 2016 - 299 -Режим доступа: http://znanium.com/go.php?id=542521</p> <p>2. Волкова Полина Андреевна. Статистическая обработка данных в учебно-исследовательских работах : Учебное пособие [Электронный ресурс] : Форум , 2019 - 96 - Режим доступа: http://znanium.com/go.php?id=1030246</p> <p>3. Непомнящая Наталья Васильевна. Статистика: общая теория статистики, экономическая статистика. Практикум : Учебное пособие [Электронный ресурс] , 2015 - 376 -Режим доступа: http://znanium.com/go.php?id=549841</p> <p>4. Основы теории вероятностей и математической статистики [Электронный ресурс] , 2018 - 90 - Режим доступа: https://lib.rucont.ru/efd/684477</p> |

9. Utilisation de l'IA générative

| | |
|-----------------------------------|---|
| Permission d'utilisation | <p>L'utilisation de l'IA générative dans le cadre des travaux et projets est autorisée, à condition que les étudiants respectent les règles suivantes :</p> <ul style="list-style-type: none"> • L'IA générative peut être utilisée pour générer des idées, des structures de texte ou du code, mais tous les matériaux générés doivent être révisés et ajustés par l'étudiant afin de garantir qu'ils respectent les exigences académiques. • Toute utilisation de l'IA générative doit être déclarée dans la section des annexes de chaque travail, en utilisant la phrase suivante : "Lors de la préparation de ce travail, l'auteur a utilisé [NOM DE L'OUTIL / SERVICE] dans le but de [MOTIF]. Après l'utilisation de cet outil/service, l'auteur a révisé et édité le contenu selon les besoins et assume l'entière responsabilité du contenu du travail." |
| Restrictions d'utilisation | <p>Les étudiants ne doivent pas considérer l'IA générative comme une source fiable d'informations, car elle ne fournit pas de références claires ou de sources documentées.</p> <ul style="list-style-type: none"> • Il n'est pas permis de citer directement le contenu généré par l'IA dans les travaux académiques comme s'il s'agissait d'une source primaire. • Les activités dans lesquelles l'utilisation de l'IA générative est interdite sont spécifiées par le professeur et concernent généralement les évaluations intermédiaires et finales, ou celles qui ne nécessitent pas de développement des compétences professionnelles. |

10. Évaluation

| Périodique | | Continue | Étude individuelle | Projet/Mémoire | Examen |
|---|------|----------|--------------------|----------------|--------|
| EP 1 | EP 2 | | | | |
| 15% | 15% | 15% | 15% | - | 40% |
| Norme minimales de performance | | | | | |
| <p>La présence et la participation aux conférences et travaux pratiques ; Obtenir la note minimale de "5" pour chaque évaluation et travail pratique individuel ou en groupe ; Démontrer, lors de l'examen final, les connaissances théoriques et les compétences pratiques nécessaires pour l'utilisation du Big Data dans le développement d'un projet concret.</p> | | | | | |