



## Traitement des Données Big Data

» Modalité: en ligne

» Durée: 6 mois

» Qualification: TECH Université Technologique

» Intensité: 16h/semaine

» Horaire: à votre rythme

» Examens: en ligne

Accès au site web: www.techtitute.com/fr/informatique/diplome-universite/diplome-universite-traitement-donnees-big-data

## Sommaire

O1 O2

Présentation Objectifs

page 4 page 8

03 04 05
Direction de la formation Structure et contenu Méthodologie

page 12 page 16

page 22

06 Diplôme

# **Présentation**

Le *Big Data* est en pleine expansion. Le volume de données circulant sur le réseau est immense et son intérêt est dans de de nombreux cas, encore plus grand. Le traitement de ces informations nécessite un plan d'action pour faire face à toute défaillance, notamment en matière de sécurité. Le respect des exigences légales et réglementaires est également essentiel pour les entreprises. Ce programme spécialise les professionnels de l'informatique dans l'Azure *Storage Cloud* et la gouvernance des données, ainsi que dans l'analyse, le traitement et la gestion des données en *Streaming*. Tout cela dans un programme 100% en ligne, avec un contenu disponible dès le premier jour et téléchargeable depuis n'importe quel appareil disposant d'une connexion Internet



1010101 1010101010

0110101100101100101

Exceller dans le domaine du Traitement des Données Big Data grâce à ce Certificat Avancé. Inscrivez-vous et apprenez-en davantage sur le Data Governance et le Cloud Streaming"

## tech 06 | Présentation

Ce Certificat Avancé cible les professionnels de l'informatique qui souhaitent progresser dans un domaine technologique en pleine expansion. Bien que le terme *Big Data* soit largement utilisé par la population générale, peu de gens sont réellement conscients de l'importance d'un bon travail effectué par un spécialiste du traitement des données sur le web.

Dans ce programme, les professionnels acquerront les compétences nécessaires, grâce à un contenu théorique et pratique, pour mettre en œuvre des sauvegardes, établir une approche de la gouvernance des données, appliquer des politiques pour garantir que les organisations et les entreprises se conforment aux réglementations légales, et analyser le processus de collecte, de structuration, de traitement et d'interprétation des données en *Streaming*.

Ce programme abordera la technologie *Big Data* la plus avant-gardiste. Une équipe d'enseignants spécialisés ayant une formation académique dans ce domaine innovant encadrera les étudiants pendant les six mois de ce diplôme.

C'est une opportunité de progresser professionnellement tout en conciliant le travail et la vie personnelle, grâce à la modalité 100% en ligne proposée par TECH. De plus, le système de *Relearning*, basé sur la répétition du contenu, et une grande variété de ressources multimédia faciliteront l'apprentissage et l'acquisition de connaissances solides.

Ce **Certificat Avancé en Traitement de Données Big Data** contient le programme académique le plus complet et le plus actuel du marché. Les principales caractéristiques sont les suivantes:

- Le développement d'études de cas présentées par des experts en Transformation Numérique
- Les contenus graphiques, schématiques et éminemment pratiques avec lesquels il est conçu, fournissent des informations pratiques sur les disciplines essentielles à la pratique professionnelle
- Les exercices pratiques d'auto-évaluation pour améliorer l'apprentissage
- Les méthodologies innovantes
- Des cours théoriques, des questions à l'expert, des forums de discussion sur des sujets controversés et un travail de réflexion individuel
- La disponibilité d'accès aux contenus à partir de tout dispositif fixe ou portable doté d'une connexion internet



Renforcez vos connaissances en matière de programmation de Cloud en temps réel grâce à ce Certificat Avancé"



Analisez les différentes options de Cloud disponibles et confrontez avec des garanties les risques subis par une entreprise avec ce Certificat Avancé"

Le corps enseignant comprend des professionnels du secteur qui apportent à cette formation l'expérience de leur travail, ainsi que des spécialistes reconnus de sociétés de référence et d'universités prestigieuses.

Son contenu multimédia qui est développé avec les dernières technologies éducatives, permettra aux professionnels d'apprendre de manière située et contextuelle, c'est-à-dire dans un environnement simulé qui offrira une formation en immersion programmée pour s'entraîner aux situations de la vie réelle.

La conception de ce programme est axée sur l'Apprentissage Par les Problèmes, grâce auquel le professionnelle devra essayer de résoudre les différentes situations de pratique professionnelle qui se présentent tout au long du programme. Pour ce faire il sera assisté d'un système vidéo interactif innovant créé par des experts dans ce domaine.

Exploitez au mieux vos connaissances et apprenez grâce à TECH comment développer une machine virtuelle avec Azure.

Maîtrisez Apache Spark Streaming, Kafka Stream ou Flink Stream et offrez un bon service à vos clients et entreprises.





À l'issue du Certificat Avancé en Traitement des Données Big Data, les professionnels de l'informatique sauront établir quel est le *Cloud* le plus adapté à l'entreprise à laquelle le service est proposé, détecter les principaux risques qui peuvent être encourus et les solutions à appliquer dans chaque situation. De plus, les étudiants seront en mesure d'identifier les principaux langages de programmation en matière de *Big Data*. De cette manière, les professionnels acquièrent une formation dont le contenu est actualisé et conforme aux exigences du secteur.



## tech 10 | Objectifs



### Objectifs généraux

- Analyser les différentes approches de l'adoption du cloud et leurs contextes
- Acquérir des connaissances spécialisées pour déterminer le Cloud approprié
- Développer une machine virtuelle dans Azure
- Déterminer les sources de menaces dans le développement d'applications et les meilleures pratiques à appliquer
- Évaluer les différences dans les implémentations concrètes des différents fournisseurs de *Cloud* publique
- Déterminer les différentes technologies appliquées aux conteneurs
- Identifier les aspects clés dans l'adoption d'une stratégie d'adoption Cloud-Native
- Connaître les fondamentaux et évaluer les langages de programmation les plus utilisés dans le domaine du *Big Data*, nécessaires à l'analyse et au traitement des données





#### Objectifs spécifiques

#### Module 1. Storage dans le Cloud Azure

- Parcourir une machine virtuelle dans Azure
- Établir les différents types de stockage
- Évaluer les fonctions dans Backup
- Gérer les ressources Azure
- Analyser les différents types de services
- Examiner les différents types de Sécurité
- Examiner les différents types de sécurité
- Générer des réseaux virtuels

#### Module 2. Programmation Cloud. Data Governance

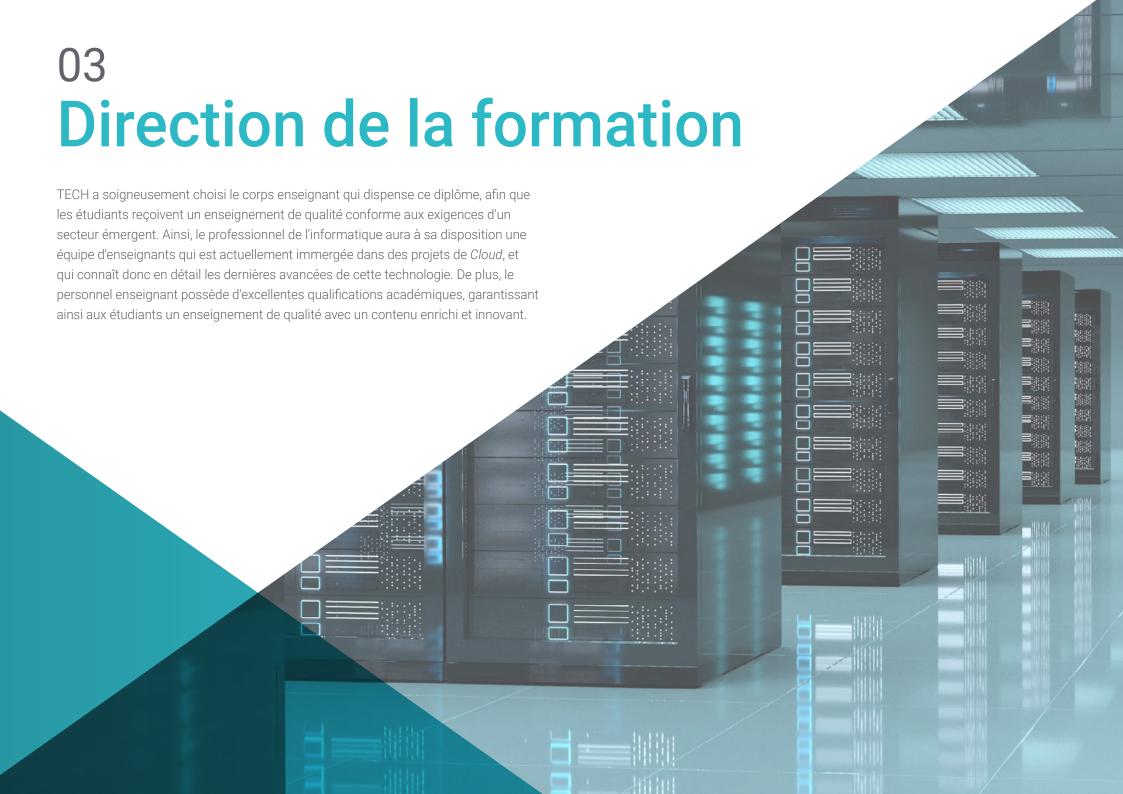
- Générer des connaissances spécialisées sur la gestion des données, les stratégies et les techniques de traitement
- Développer des stratégies de gouvernance des données visant les personnes, les processus et les outils
- Réaliser la gouvernance des données depuis l'ingestion jusqu'à la préparation et l'utilisation
- Déterminer les techniques permettant de régir la transmission des données
- Établir la protection des données pour l'authentification, la sécurité, la sauvegarde et les moniteurs

#### Module 3. Programmation Cloud en Temps Réel. Streaming

- Analyser le processus de collecte, de structuration, de traitement, d'analyse et d'interprétation des données en Streaming
- Développer les principes du traitement *Streaming*, le contexte actuel et les cas d'utilisation actuels
- Développer les fondamentaux clés des statistiques, Maching Learning, du data mining et de la modélisation prédictive pour comprendre l'analyse et le traitement des données
- Analyser les principaux langages de programmation Big Data
- Examiner les principes fondamentaux de *Apache Spark Streaming, Kafka Stream* y *Flink Stream*



Maîtrisez à la perfection les principaux langages de programmation Big Data. Devenez le professionnel que chaque entreprise recherche dans son équipe"





## tech 14 | Direction de la formation

#### Direction



#### M. Bressel Gutiérrez-Ambrossi, Guillermo

- Spécialiste en Administration des Systèmes et Réseaux Informatiques
- Administrateur de Réseaux de Stockage et SAN à Experis IT (BBVA)
- Administrateur de Réseaux à l'IE Business School
- Diplôme Supérieur en Administration des Systèmes et Réseaux Informatiques à ASIF
- \* Cours en Ethical Hacking en OpenWebinar
- Cours en Powershel en OpenWebinar



## Direction de la formation | 15 tech

#### **Professeurs**

#### M. Bernal de la Varga, Yeray

- Architecte de Solutions Big Data à Orange Bank
- Architecte Big Data à Bankia
- Ingénieur Big Data chez Hewlett-Packard
- Professeur Adjoint dans le Master en Big Data de l'Université de Deusto
- Licence en Informatique de l'Université Polytechnique de Madrid
- Expert en Big Data par U-TAD

#### Mme Rodríguez Camacho, Cristina

- Consultante Apis et Développeuse Microservices à Inetum
- Diplôme en Génie de la Santé, avec une spécialisation en Génie *Biomédica* de l'Université de Malaga
- Master en Blockchain et Big Data à l'Université Complutense de Madrid
- Experto en Devops & Cloud à UNIR





## tech 18 | Structure et contenu

#### Module 1. Storage dans le Cloud Azure

- 1.1. Installation de MV dans Azure
  - 1.1.1. Commandes de création
  - 1.1.2. Commandes de visualisation
  - 1.1.3. Commandes de modification
- 1.2. Blobs dans Azure
  - 1.2.1. Types de *Blob*
  - 1.2.2. Conteneur
  - 1.2.3. Azcopy
  - 1.2.4. Suppression réversible des Blobs
- 1.3. Disque géré et stockage dans Azure
  - 1.3.1. Disque géré
  - 1.3.2. Sécurité
  - 1.3.3. Stockage à froid
  - 1.3.4. Réplication
    - 1.3.4.1. Redondance locale
    - 1.3.4.2. Redondance dans une zone
    - 1.3.4.3. "Georredundante"
- 1.4. Tables, files d'attente, fichiers dans Azure
  - 1.4.1. Tables
  - 1.4.2 Files d'attente
  - 1.4.3. Archives
- 1.5. Cryptage et sécurité dans Azure
  - 1.5.1. Storage Service Encryption (SSE)
  - 1.5.2. Codes d'accès
    - 1.5.2.1. Signature d'accès partagé
    - 1.5.2.2. Politiques d'accès au niveau du conteneur
    - 1.5.2.3. Signature d'accès au niveau du Blob
  - 1.5.3. Authentification Azure AD

- 1.6. Réseau Virtuel dans Azure
  - 1.6.1. Sous-réseau et jumelage
  - 1.6.2. Vnet to Vnet
  - 1.6.3. Lien privé
  - 1.6.4. Haute disponibilité
- 1.7. Types de connexions dans Azure
  - 1.7.1. Azure Application Gateway
  - 1.7.2. VPN site à site
  - 1.7.3. VPN point-à-site
  - 1.7.4. ExpressRoute
- 1.8. Ressources dans Azure
  - 1.8.1. Verrouillage des ressources
  - 1.8.2. Déplacement des ressources
  - 1.8.3. Retrait des ressources
- .9. Sauvegarde dans Azure
  - 1.9.1. Recovery Services
  - 1.9.2. Agent Azure Backup
  - 1.9.3. Azure Backup Server
- 1.10. Développement de solutions
  - 1.10.1. Compression, déduplication, réplication
  - 1.10.2. Recovery Services
  - 1.10.3. Disaster Recovery Plan

#### Module 2. Programmation Cloud. Data Governance

- 2.1. Gestion des données
  - 2.1.1. Gestion des données
  - 2.1.2. Éthique dans la gestion des données
- 2.2. Data Governance
  - 2.2.1. Classification Contrôle d'accès
  - 2.2.2. Règlement sur le traitement des données
  - 2.2.3. Data Governance. Valeurs
- 2.3. Gouvernance des données Outils
  - 2.3.1. Lignage
  - 2.3.2. Métadonnées
  - 2.3.3. Catalogue de données Business Glossary
- 2.4. Utilisateurs et processus de la gouvernance des données
  - 2.4.1. Utilisateurs
    - 2.4.1.1. Rôles et responsabilités
  - 2.4.2. Processus
    - 2.4.2.1. Enrichissement des données
- 2.5. Cycle de vie des données d'entreprise
  - 2.5.1 Création des données
  - 2.5.2. Traitement des données
  - 2.5.3. Entreposage de données
  - 2.5.4. Utilisation des données
  - 2 5 5 Destruction des données
- 2.6. Qualité des données
  - 2.6.1. La qualité dans la gouvernance des données
  - 2.6.2. Qualité des données analytiques
  - 2.6.3. Techniques de qualité des données
- 2.7. La gouvernance des données en transit
  - 2.7.1. La gouvernance des données en transit 2.7.1.1. Lignage
  - 2.7.2. La quatrième dimension

#### 2.8. Protection des données

- 2.8.1. Niveaux d'accès
- 2.8.2. Classification
- 2.8.3. Compliance. Règlementation
- 2.9. Suivi et mesure de la gouvernance des données
  - 2.9.1. Surveillance et mesure de la gouvernance des données
  - 2.9.2. Surveillance du lignage
  - 2.9.3. Surveillance de la qualité des données
- 2.10. Outils de gouvernance des données
  - 2.10.1. Talend
  - 2.10.2. Collibra
  - 2.10.3. Informatique

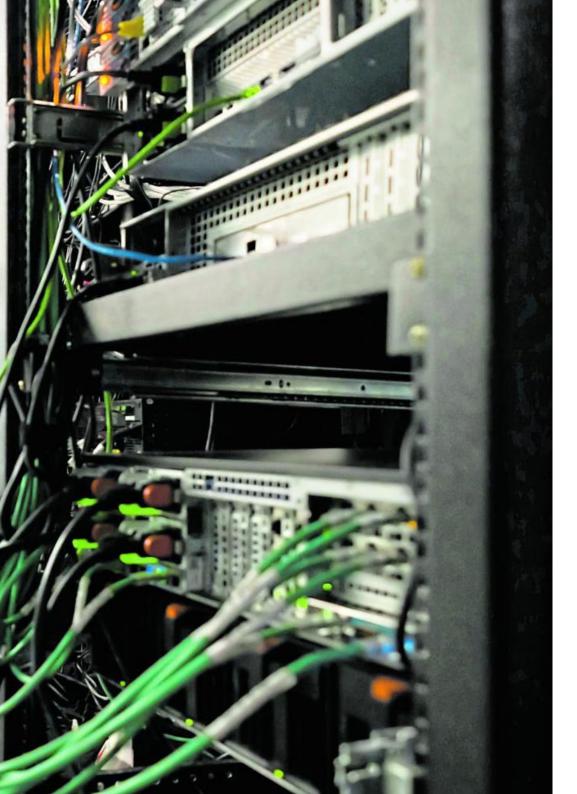
#### Module 3. Programmation Cloud en Temps Réel. Streaming

- 3.1. Traitement et structuration de l'information en Streaming
  - 3.1.1. Processus de collecte, de structuration, de traitement, d'analyse et d'interprétation des données
  - 3.1.2. Techniques de traitement des données en Streaming
  - 3.1.3. Traitement en Streaming
  - 3.1.4. Cas d'utilisation du traitement en Streaming
- 3.2. Statistiques pour comprendre le flux de donnée Streaming
  - 3.2.1. Statistiques descriptives
  - 3.2.2. Calcul des probabilités
  - 3.2.3. Inférence
- 3.3. Programmation avec Python
  - 3.3.1. Typologie, conditionnels, fonctions et boucles
  - 3.3.2. Numpy, Matplotlib, Dataframes, fichiers CSV y formats JSON
  - 3.3.3. Séguences: listes, boucles, fichiers et dictionnaires
  - 3.3.4. Mutabilité, exceptions et fonctions d'ordre supérieur

## tech 20 | Structure et contenu

- 3.4.1. Programmation avec R
- 3.4.2. Vecteurs et facteurs
- 3.4.3. Matrices et *Arrays*
- 3.4.4. Listes et Data Frame
- 3.4.5. Fonctions
- 3.5. Base de données SQL pour le traitement des données en Streaming
  - 3.5.1. Base de données SQL
  - 3.5.2. Modèle entité-relation
  - 3.5.3. Modèle relationnel
  - 3.5.4. SQL
- 3.6. Base de données NO SQL pour le traitement des données en Streaming
  - 3.6.1. Base de données NO SQL
  - 3.6.2. MongoDB
  - 3.6.3. Architecture de MongoDB
  - 3.6.4. Opérations CRUD
  - 3.6.5. Find, projections, Indexes Aggregation et curseurs
  - 3.6.6. Modèle de données
- 3.7. Exploration de données et modélisation prédictive
  - 3.7.1. Analyse multivariée
  - 3.7.2. Techniques de réduction de la dimensionnalité
  - 3.7.3. Analyse en grappes
  - 3.7.4. Séries
- 3.8. Maching Learning pour le traitement des Données en Streaming
  - 3.8.1. Maching Learning et la modélisation prédictive avancée
  - 3.8.2. Réseaux neuronaux
  - 3.8.3. Deep Learning
  - 3.8.4. Bagging et Random Forest
  - 3.8.5. Gradient Bosting
  - 3.8.6. SVM
  - 3.8.7. Méthodes d'assemblage





## Structure et contenu | 21 tech

- 3.9. Technologies de traitement des données en continu
  - 3.9.1. Spark Streaming
  - 3.9.2. Kafka Streams
  - 3.9.3. Flink Streaming
- 3.10. Apache Spark Streaming
  - 3.10.1. Apache Spark Streaming
  - 3.10.2. Composents de park
  - 3.10.3. Architecture de Spark
  - 3.10.4. RDD
  - 3.10.5. SPARK SQL
  - 3.10.6. Jobs, Stages et Task



Prenez de l'avance sur vos concurrents. Grâce à ce Certificat Avancé, vous disposez d'un enseignement au contenu actualisé et d'une méthodologie en ligne flexible"





## tech 24 | Méthodologie

#### Étude de Cas pour mettre en contexte tout le contenu

Notre programme offre une méthode révolutionnaire de développement des compétences et des connaissances. Notre objectif est de renforcer les compétences dans un contexte changeant, compétitif et hautement exigeant.



Avec TECH, vous pouvez expérimenter une manière d'apprendre qui ébranle les fondations des universités traditionnelles du monde entier"



Vous bénéficierez d'un système d'apprentissage basé sur la répétition, avec un enseignement naturel et progressif sur l'ensemble du cursus.



L'étudiant apprendra, par des activités collaboratives et des cas réels, à résoudre des situations complexes dans des environnements commerciaux réels.

### Une méthode d'apprentissage innovante et différente

Cette formation TECH est un programme d'enseignement intensif, créé de toutes pièces, qui propose les défis et les décisions les plus exigeants dans ce domaine, tant au niveau national qu'international. Grâce à cette méthodologie, l'épanouissement personnel et professionnel est stimulé, faisant ainsi un pas décisif vers la réussite. La méthode des cas, technique qui constitue la base de ce contenu, permet de suivre la réalité économique, sociale et professionnelle la plus actuelle.



Notre programme vous prépare à relever de nouveaux défis dans des environnements incertains et à réussir votre carrière"

La méthode des cas est le système d'apprentissage le plus largement utilisé dans les meilleures écoles d'informatique du monde depuis qu'elles existent. Développée en 1912 pour que les étudiants en Droit n'apprennent pas seulement le droit sur la base d'un contenu théorique, la méthode des cas consiste à leur présenter des situations réelles complexes afin qu'ils prennent des décisions éclairées et des jugements de valeur sur la manière de les résoudre. En 1924, elle a été établie comme méthode d'enseignement standard à Harvard.

Dans une situation donnée, que doit faire un professionnel? C'est la question à laquelle nous sommes confrontés dans la méthode des cas, une méthode d'apprentissage orientée vers l'action. Tout au long du programme, les étudiants seront confrontés à de multiples cas réels. Ils devront intégrer toutes leurs connaissances, faire des recherches, argumenter et défendre leurs idées et leurs décisions.

#### Relearning Methodology

TECH combine efficacement la méthodologie des Études de Cas avec un système d'apprentissage 100% en ligne basé sur la répétition, qui associe différents éléments didactiques dans chaque leçon.

Nous enrichissons l'Étude de Cas avec la meilleure méthode d'enseignement 100% en ligne: le Relearning.

En 2019, nous avons obtenu les meilleurs résultats d'apprentissage de toutes les universités en ligne du monde.

À TECH, vous apprendrez avec une méthodologie de pointe conçue pour former les managers du futur. Cette méthode, à la pointe de la pédagogie mondiale, est appelée Relearning.

Notre université est la seule université autorisée à utiliser cette méthode qui a fait ses preuves. En 2019, nous avons réussi à améliorer les niveaux de satisfaction globale de nos étudiants (qualité de l'enseignement, qualité des supports, structure des cours, objectifs...) par rapport aux indicateurs de la meilleure université en ligne.



### Méthodologie | 27 tech

Dans notre programme, l'apprentissage n'est pas un processus linéaire, mais se déroule en spirale (apprendre, désapprendre, oublier et réapprendre). Par conséquent, chacun de ces éléments est combiné de manière concentrique. Cette méthodologie a permis de former plus de 650.000 diplômés universitaires avec un succès sans précédent dans des domaines aussi divers que la biochimie, la génétique, la chirurgie, le droit international, les compétences en gestion, les sciences du sport, la philosophie, le droit, l'ingénierie, le journalisme, l'histoire, les marchés financiers et les instruments. Tout cela dans un environnement très exigeant, avec un corps étudiant universitaire au profil socio-économique élevé et dont l'âge moyen est de 43,5 ans.

Le Relearning vous permettra d'apprendre avec moins d'efforts et plus de performance, en vous impliquant davantage dans votre formation, en développant un esprit critique, en défendant des arguments et en contrastant les opinions: une équation directe vers le succès.

À partir des dernières preuves scientifiques dans le domaine des neurosciences, non seulement nous savons comment organiser les informations, les idées, les images et les souvenirs, mais nous savons aussi que le lieu et le contexte dans lesquels nous avons appris quelque chose sont fondamentaux pour notre capacité à nous en souvenir et à le stocker dans l'hippocampe, pour le conserver dans notre mémoire à long terme.

De cette manière, et dans ce que l'on appelle Neurocognitive context-dependent e-learning, les différents éléments de notre programme sont reliés au contexte dans lequel le participant développe sa pratique professionnelle.

Ce programme offre le support matériel pédagogique, soigneusement préparé pour les professionnels:



#### Support d'étude

Tous les contenus didactiques sont créés par les spécialistes qui enseigneront le cours, spécifiquement pour le cours, afin que le développement didactique soit vraiment spécifique et concret.

Ces contenus sont ensuite appliqués au format audiovisuel, pour créer la méthode de travail TECH en ligne. Tout cela, avec les dernières techniques qui offrent des pièces de haute qualité dans chacun des matériaux qui sont mis à la disposition de l'étudiant.



#### **Cours magistraux**

Il existe des preuves scientifiques de l'utilité de l'observation par un tiers expert.

La méthode "Learning from an Expert" renforce les connaissances et la mémoire, et donne confiance dans les futures décisions difficiles.



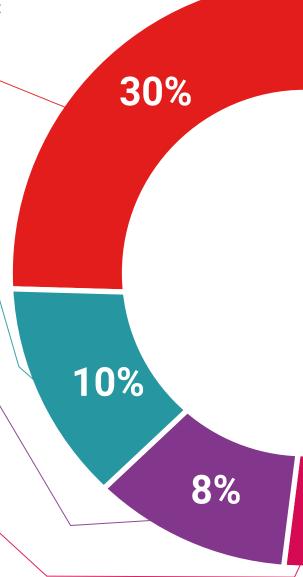
#### Pratiques en compétences et aptitudes

Les étudiants réaliseront des activités visant à développer des compétences et des aptitudes spécifiques dans chaque domaine. Des activités pratiques et dynamiques pour acquérir et développer les compétences et aptitudes qu'un spécialiste doit développer dans le cadre de la mondialisation dans laquelle nous vivons.



#### Lectures complémentaires

Articles récents, documents de consensus et directives internationales, entre autres. Dans la bibliothèque virtuelle de TECH, l'étudiant aura accès à tout ce dont il a besoin pour compléter sa formation.



Case studies
Ils réaliseront une sélection des meilleures études de cas choisies spécifiquement

pour ce diplôme. Des cas présentés, analysés et tutorés par les meilleurs spécialistes de la scène internationale.



Résumés interactifs

L'équipe TECH présente les contenus de manière attrayante et dynamique dans des pilules multimédia comprenant des audios, des vidéos, des images, des diagrammes et des cartes conceptuelles afin de renforcer les connaissances.

Ce système éducatif unique pour la présentation de contenu multimédia a été récompensé par Microsoft en tant que "European Success Story".



**Testing & Retesting** 

Les connaissances de l'étudiant sont périodiquement évaluées et réévaluées tout au long du programme, par le biais d'activités et d'exercices d'évaluation et d'autoévaluation, afin que l'étudiant puisse vérifier comment il atteint ses objectifs.



25%

20%





## tech 32 | Diplôme

Ce **Certificat Avancé en Traitement de Données Big Data** contient le programme le plus complet et le plus à jour du marché.

Après avoir réussi l'évaluation, l'étudiant recevra par courrier postal\* avec accusé de réception son correspondant diplôme de **Certificat Avancé** délivré par **TECH Université Technologique**.

Le diplôme délivré par **TECH Université Technologique** indiquera la note obtenue lors du Certificat Avancé, et répond aux exigences communément demandées par les bourses d'emploi, les concours et les commissions d'évaluation des carrières professionnelles.

Diplôme: Certificat Avancé en Traitement de Données Big Data

Nº d'heures officielles: 450 h.



<sup>\*</sup>Si l'étudiant souhaite que son diplôme version papier possède l'Apostille de La Haye, TECH EDUCATION fera les démarches nécessaires pour son obtention moyennant un coût supplémentaire.

technologique

## **Certificat Avancé** Traitement de Données Big Data

- » Modalité: en ligne
- » Durée: 6 mois
- » Qualification: TECH Université Technologique
- » Intensité: 16h/semaine
- » Horaire: à votre rythme
- » Examens: en ligne

