

Catalogue des formations



Nos formations

- ❖ Veille sur internet
- ❖ Cybersécurité
- ❖ Big Data
- ❖ Data science et Intelligence Artificielle

Choisir H-Intelligence pour vous assurer de l'efficacité de votre formation !

Qui sommes-nous ?

- ❖ Hamis Intelligence est une entreprise de conseil et de formation créée en 2017, à Roissy en France.
- ❖ Nous proposons des formations sur la Veille et l'Intelligence Economique, la Cybersécurité, le Big Data, la Data science et l'Intelligence Artificielle et, enfin, nous proposons des formations « sur-mesure » en langues étrangères destinées aux entreprises.
- ❖ Toutes nos formations sont dispensées en présentiel car le contact humain et le partage direct entre formateurs et stagiaires nous semblent essentiel. Les formations peuvent se dérouler, selon vos souhaits, en inter ou intra-entreprises. Nous disposons de plusieurs salles de formation à Villepinte, dans le Parc d'activité de Paris Nord 2, à deux pas du RER.

Nos choix pédagogiques

- ❖ Les enseignements dispensés ont été conçus et développés avec des formateurs spécialistes ayant tous au minimum 7 ans d'expérience dans leurs domaines de compétences.
- ❖ Nos formations font l'objet d'une pré-évaluation, par questionnaire, des connaissances et des besoins précis de chaque stagiaire, et ce, plusieurs semaines avant la formation et par une vérification de niveau, en début de stage, par le formateur.
- ❖ Ces étapes peuvent être précédées d'un échange avec le Responsable Formation ou le RRH de l'entreprise, ainsi qu'avec le Responsable hiérarchique « Métier » du ou des futurs stagiaires le cas échéant.
- ❖ La transmission de la connaissance ne tient pas uniquement dans l'information donnée. L'empathie du formateur, la reformulation dont il constate la nécessité en observant son ou ses stagiaires, sont des vecteurs uniques de la bonne compréhension, de la mémorisation et de la réussite du stagiaire.

Nos pratiques

- ❖ Durant nos formations, chaque stagiaire dispose d'un poste informatique propre. Le matériel utilisé est de dernière génération, équipé des environnements Windows et Linux. Pour l'organisation des formations intra-entreprises dans vos locaux, le matériel devra être fourni par vos soins. Nous vous offrons, néanmoins, la possibilité de vous mettre à disposition le matériel sur devis.
- ❖ À l'issue de chaque formation, les participants reçoivent un support formation numérique. Une assistance mail est proposée à l'issue de nos stages pratiques.

Notre équipe est à votre disposition pour répondre à toutes vos problématiques spécifiques car

Pour vous comme pour nous, chaque client est unique !

FORMATIONS

BIG DATA

Les fondamentaux

Big data / état de l'art – Enjeux pour l'entreprise	59
---	----

Mise en œuvre

Mise en œuvre de solutions d'analyse de données de masse	62
--	----

Big data Analytics

Analyse statistique avec R - Les fondamentaux	65
Analyse Big data avec Python	67
Analyse Big data avec R	69

BID GATA

Les fondamentaux

INFOS PRATIQUES

- Réf : BEA
- Durée : 2 jours
- Prix : 1590€ HT
- Horaires : 09h00 – 17h30

DATES 2019

Formations :

- 13 et 14 juin 2019
- 03 et 04 octobre 2019

PUBLIC VISE

- Toute personne s'intéressant au Big data et ayant des connaissances de base en architecture technique

PRÉ-REQUIS

- Aucun

RESSOURCES

- Supports pédagogiques

OBJECTIFS :

- Découvrir les concepts du Big Data
- Savoir évaluer les avantages et les inconvénients du Big Data
- Identifier les problèmes et les solutions apportés par le Big Data
- Connaître les méthodes de base et les champs d'application du Big Data

PROGRAMME DE FORMATION:

Historique et contexte de l'explosion des usages autour des données

- Les origines du Big Data : un monde de données numériques, chronologie
- Une définition par les trois V, et les V complémentaires
- La valeur de la donnée : une prise de conscience
- La donnée en tant que matière première : pourquoi et comment comparer pétrole et données
- La prise de décision : origine et évolution des outils d'aide à la décision

L'Economie des données

- Principes de l'économie des données et des algorithmes
- Etudes de cas liés à la transformation numérique
- Analyse illustrée de différents modèles économiques liés à l'analyse de données
- Les métiers liés à la donnée
- Les entreprises du domaine de la donnée
- Exercice pratique de réflexion sur l'analyse de données non structurées

Gestion du patrimoine informationnel et pilotage de la valeur

- Préparation d'un projet Big Data : définition des cas d'usage
- Construction d'un Business Model Canvas et d'une matrice d'évaluation
- Les architectures techniques : data warehouse, data lake, etc.
- Les acteurs du Big Data
- Analyse détaillée des évolutions technologiques dans les domaines de l'acquisition des données, du stockage et de la restitution
- Exercice pratique de rétro-analyse d'une architecture Big Data

Big Data : Acquisition des données

- L'enchaînement des opérations. L'acquisition.
- Le recueil des données : crawling, scraping.
- La gestion de flux événementiel (Complex Event Processing, CEP).
- L'indexation du flux entrant.
- L'intégration avec les anciennes données.
- La qualité des données

Big Data : Stockage des données

- Les différentes formes de stockage des données structurées : SQL, OLAP, Bases Graphes
- NoSQL : différences et apports
- Panorama des solutions NoSQL
- HADOOP : de quoi s'agit-il ?
- Détails du fonctionnement et de l'architecture d'un système HADOOP
- L'écosystème Hadoop : Hive, Pig. Les difficultés d'Hadoop
- HADOOP et Big Data sont-ils indissociables ?
- Les autres solutions de stockage de données (Data Lake)
- Le modèle d'architecture des Clouds publics et privés
- Les objectifs et avantages des architectures Cloud
- Les infrastructures
- Les égalités et les différences entre Cloud et Big Data
- Les Clouds de stockage
- Les outils de restitution : de la Data Visualization au Data Storytelling
- Panorama des outils tels que Qlik, Tableau, TIBCO Spotfire, Microsoft PowerBI.

Machine Learning, Deep Learning, Intelligence Augmentée

- Comprendre les concepts et les définitions
- Différences entre intelligence artificielle et intelligence augmentée
- Qu'est-ce que le machine learning
- Détails du processus d'apprentissage machine : modélisation, apprentissage, exécution, rétroaction
- Cas pratiques de mises en place
- Exercice pratique de conception d'une matrice d'apprentissage

Cas d'usage à travers des exemples et conclusion

- L'anticipation : besoins des utilisateurs dans les entreprises, maintenance des équipements.
- La sécurité : des personnes, détection de fraude (postale, taxes), le réseau.
- La recommandation. Analyses marketing et analyses d'impact.
- Analyses de parcours. Distribution de contenu vidéo.
- Big Data pour l'industrie automobile ? Pour l'industrie pétrolière
- Faut-il se lancer dans un projet Big Data ?
- Gouvernance du stockage des données : rôle et recommandations, le Data Scientist, les compétences d'un projet Big Data

BID GATA

Mise en œuvre

Mise en œuvre de solutions d'analyse de données de masse

INFOS PRATIQUES

- Réf : MSA
- Durée : 4 jours
- Prix : 2380€ HT
- Horaires : 09h00 – 17h30

DATES 2019

Formations :

- 08 au 11 octobre 2019
- 16 au 19 décembre 2019

PUBLIC VISE

- Data scientists
- Consultants
- Chefs de projet
- Développeurs
- Analystes et statisticiens

PRÉ-REQUIS

- Connaissances de base des modèles statistiques et des langages de programmation
- Langage SQL

RESSOURCES

- Supports pédagogiques
- 40% de théorie
- 60% de pratique
- 1 PC par personne / Internet

OBJECTIFS :

- Acquérir les compétences techniques nécessaires à la mise en œuvre d'analyses Big Data
- Comprendre le cadre juridique du stockage et de l'analyse de données
- Savoir utiliser des outils de collecte Open Source
- Être en mesure de choisir la bonne solution de stockage de données selon le type de projet
- Maîtriser l'analyse des résultats et comprendre la signification des données extraites

PROGRAMME DE FORMATION:

La collecte de données

- Où et comment collecter des données ?
- Les sources de données, les API, les fournisseurs, les agrégateurs...
- Les principaux outils de collecte et de traitement de l'information (ETL)
- Prise en main de Talend ETL et de Talend Data Preparation (outils libres)
- Les particularités de la collecte des données semi-structurées et non-structurées

Le stockage des données

- Les différentes formes de stockage des données : rappel de l'architecture relationnelle de stockage des données transactionnelles (SGBD/R) et multidimensionnelles (OLAP)
- Les nouvelles formes de stockage des données – compréhension, positionnement et comparaison : Bases orientées clé-valeur, documents, colonnes, graphes
- Panorama des bases de données NoSQL
- Prise en main d'une base de données orientée colonne (Hbase)
- Particularités liées au stockage des données non-structurées
- Comment transformer des données non structurées en données structurées

L'écosystème Hadoop

- Présentation des principaux modules de la distribution Apache Hadoop
- Présentation et comparaison des principales distributions commerciales (Cloudera, Hortonworks...)
- L'infrastructure matérielle et logicielle nécessaire au fonctionnement d'une distribution Hadoop en local ou dans le Cloud
- Les concepts de base de l'architecture Hadoop : Data Node, Name Node, Job Tracker, Task Tracker
- Présentation de HDFS (Système de gestion des fichiers de Hadoop)
- Prise en main et exercices pratiques dans HDFS

- Présentation de MapReduce (Outil de traitement de Hadoop)
- Les commandes exécutées au travers de PIG
- Utilisation de HIVE pour transformer du SQL en MapReduce

L'analyse de données

- Requête les données
- Analyser et comprendre la signification des données extraites
- Particularités liées à l'analyse des données non structurées
- Analyse statistique : notions de base
- Analyse prédictive : comment transformer des données du passé en prévisions pour le futur
- Calculer des tendances
- Développer des programmes simples d'automatisation des analyses (en Python) Machine Learning : les bases de l'apprentissage machine avec Spark
- Deep Learning : notions de base de l'analyse future automatisée de données non structurées

Mise en œuvre de projets Big data

- Automatisation de tâches avec Oozie
- Mise en production de programmes de Machine Learning
- L'utilisation des notebooks comme livrables
- Traitement du temps réel Gouvernance de données Big Data

BID GATA

Big Data Analytics

INFOS PRATIQUES

- Réf : ASR
- Durée : 3 jours
- Prix : 1890€ HT
- Horaires : 09h00 – 17h30

DATES 2019

Formations :

- 11 au 13 septembre 2019
- 13 au 15 novembre 2019

PUBLIC VISE

- Analystes
- Statisticiens
- Ingénieurs
- Toute personne intéressée par l'analyse statistique

PRÉ-REQUIS

- Environnement Windows
- Connaissances de base en statistiques et mathématiques

RESSOURCES

- Supports pédagogiques
- 40% de théorie
- 60% de pratique
- 1 PC par personne / Internet
- Environnement Windows

OBJECTIFS :

- Savoir installer l'environnement d'analyse R
- Etre en mesure d'importer et d'exporter des données
- Réaliser des analyses statistiques basiques avec R
- Etre capable de restituer des résultats à l'aide de graphiques

PROGRAMME DE FORMATION:

Savoir installer et utiliser l'environnement R et RStudio

- Motivation autour du langage R
- Installer R et RStudio
- Paramétrer efficacement : divers fichiers et options
- Travailler avec fluidité sur des projets Data : créer des fichiers .Rprofile, exécuter des scripts R, installer des packages tiers, utiliser le help de R, etc.
- Se servir des fichiers sources de la formation

Lire et écrire différents formats de données sur R

- Importer et exporter des données :
 - Fichiers plats : csv, tsv, txt, etc.
 - Fichiers Excel
 - A partir d'une base de données relationnelles (SQL)
- Diverses autres sources : SAS, Stata, xml, json...
- Comprendre les différences entre un dataframe, un tibble et un datatable
- Comprendre les différents types de colonnes d'un cadre de données
- Transformer le cadre des données vers le format adéquat : long Vs. wide
- Manipuler les autres objets de R : vecteurs, listes et matrices.

Réaliser des analyses statistiques de base avec R

- Exploration des données avec dplyr
- Traiter les valeurs manquantes et anormales
- Calculer les indicateurs de position, de dispersion et de forme
- Calculer les statistiques descriptives entre :
 - Une variable qualitative et une variable quantitative
 - Deux variables qualitatives
 - Deux variables quantitatives
- Comprendre les jointures et les réaliser : inner join, left, right, full, etc.

Elaborer des graphiques à partir des résultats obtenus

- Charger et contrôler ggplot2
- Visualiser les données avec ggplot2
- Rendre les graphiques interactifs : htmlwidgets, plotly
- Rendre les graphiques en norme avec la charte de l'entreprise pour diffusion
- Publier un graphique

Produire du code R plus poussé

- Maîtriser les boucles for et les conditions if, else
- Utiliser, de préférence, des fonctions vectorisées
- Faire de l'échantillonnage
- Maîtriser les formats dates.

Conclure cette formation d'analyse statistique avec R

INFOS PRATIQUES

- Réf : ABP
- Durée : 4 jours
- Prix : 2380€ HT
- Horaires : 09h00 – 17h30

DATES 2019

Formations :

- 24 au 27 septembre 2019
- 19 au 22 novembre 2019

PUBLIC VISE

- Analystes
- Statisticiens
- Gestionnaires de bases de données

PRÉ-REQUIS

- Connaissances de base en programmation
- Connaissances de base en techniques statistiques

RESSOURCES

- Supports pédagogiques
- 30% de théorie
- 70% de pratique
- 1 PC par personne / Internet

OBJECTIFS :

- Connaître l'environnement de développement en Python
- Acquérir les bases de la programmation en Python
- Savoir utiliser les bibliothèques spécialisées
- Maîtriser les méthodes et les outils pour visualiser et modéliser les données

PROGRAMME DE FORMATION:

Configurer efficacement son environnement de travail

- Motivation autour du langage Python et de la librairie PySpark pour le traitement des données massives
- Installer Python et la batterie des librairies nécessaires : numpy, pandas, Scikit-learn, PySpark
- Paramétrer efficacement : divers fichiers et options
- Travailler avec fluidité sur des projets où on distribue la Data et on parallélise les tâches
- Se servir des fichiers sources de la formation

Du data munging avec PySpark

- Quand distribuer les données et paralléliser les données ?
- Principaux types de données et des concepts de Resilient Distributed Datasets (RDD) et DataFrame
- Partition, transformations, lazy evaluations et les actions
- Gestion des RDDs :
 - Travailler sur les colonnes
 - Travailler sur les lignes
 - Travailler sur les dates
 - Travailler sur les jointures
 - Transformer un RDD : transformation map
 - Agrégation de RDDs : reduce, aggregate

- **PREMIER CHALLENGE**
 - Solution du 1er challenge

Analyse de données et régression logistique avec PySpark

- Bien faire la différence entre une action et une transformation
 - La transformation sample et l'action takeSample
- Présentation et fonctionnalités de MLlib
- Apprentissage Méthodes linéaires
- **SECOND CHALLENGE**
 - Solution du 2nd challenge

Requêtage, filtrage et exploration avec SparkSQL

- Construire un DataFrame depuis un dataframe de Pandas, depuis un csv, depuis un RDD
- Requêtes SQL
- Opérations sur un DataFrame

Définition des pipelines pour enchaîner les traitements

- Introduction et fonctionnalités de SparkML
- Estimator, Transformer et Param
- Pipeline : par ex Tokenzine, Hash et Régression logistique

Analyse Big data avec R

INFOS PRATIQUES

- **Réf :** ABR
- **Durée :** 4 jours
- **Prix :** 2380€ HT
- **Horaires :** 09h00 – 17h30

DATES 2019

Formations :

- 14 au 16 octobre 2019
- 04 au 06 décembre 2019

PUBLIC VISE

- Analystes
- Statisticiens
- Gestionnaires de bases de données

PRÉ-REQUIS

- Connaissances de base en programmation
- Connaissances de base en techniques statistiques

RESSOURCES

- Supports pédagogiques
- 30% de théorie
- 70% de pratique
- 1 PC par personne / Internet

OBJECTIFS :

- Connaître et apprendre à utiliser l'environnement R pour l'exploration des données issues de différentes sources
- Être capable de visualiser les données en utilisant des graphiques et des nuages de points
- Comprendre comment transformer et nettoyer des ensembles de données
- Être capable de construire et évaluer les modèles de régressions
- Être en mesure de créer, marquer et déployer des modèles de partition
- Savoir utiliser R dans les environnements SQL Server et Hadoop

PROGRAMME DE FORMATION:

Microsoft R server et R client

- Définition d'un serveur Microsoft R
- Utilisation du client Microsoft R
- Les fonctions ScaleR

Explorer de grandes données

- Comprendre les sources de données ScaleR
- Lecture de données dans un objet XDF
- Résumer des données dans un objet XDF

Visualisation d'un grand nombre de données

- Visualisation des données en mémoire
- Visualisation d'un grand nombre de données

Traitement d'un grand nombre de données

- Transformer un grand nombre de données
- Gestion d'un ensemble de données

Opérations d'analyse en parallèle

- Utilisation du contexte de calcul RxLocalParallel avec rxExec
- Utilisation du package revoPemaR

Création et évaluation de modèles de régression

- Clustering Big Data
- Générer des modèles de régression et faire des prédictions

Création et évaluation de modèles de partitionnement

- Création de modèles de partitionnement basés sur des arbres de décision
- Tester les modèles de partitionnement en effectuant et en comparant des prédictions

Traitement d'un grand nombre de données dans sql server et hadoop

- Utilisation de R dans SQL Server
- Utilisation de Hadoop Map / Reduce
- Utilisation de Hadoop Spark



Bâtiment Raphaël
Parc d'Activités Paris Nord 2
22, Avenue des Nations
BP 58425
93420 – Villepinte

Téléphone : 01.49.38.06.49
Email : contact@h-intelligence.com



H-INTELLIGENCE