

Formation BIG DATA Les Fondamentaux

Durée de la formation : 3 jour(s)

OBJECTIFS

A l'issue de la formation, le participant sera capable de :

- Comprendre le rôle stratégique de la gestion des données pour l'entreprise ou l'organisation
- Identifier ce qu'est la donnée, et en quoi consiste le fait d'assurer la qualité de données
- Synthétiser le cycle de vie de la donnée
- Assurer l'alignement des usages métiers avec le cycle de vie de la donnée
- Découvrir les bonnes pratiques en matière de contrôle de qualité des données
- Assurer la mise en oeuvre de la gouvernance de la donnée
- Disposer d'un premier aperçu des possibilités de traitement proposé par MapR et Hadoop

PREREQUIS

Une connaissance de SQL est un plus pour suivre cette formation

Cette formation ne peut être financée que dans le cadre d'un projet d'entreprise (prise en charge entreprise ou OPCO). Les dossiers à financement personnel et CPF ne sont pas pris en compte.

PUBLIC

MOA, chef de projet, urbaniste fonctionnel, responsable de domaine, analystes, développeurs, data miners ...

Futurs data scientists, data analysts et data stewards

Formation BIG DATA Les Fondamentaux

Durée de la formation : 3 jour(s)

PROGRAMME

Les nouvelles frontières du Big Data (Introduction)

Immersion
L'approche des 4 Vs
Cas d'usages du Big Data
Technologies
Architecture
Master-less vs Master-Slaves
Stockage
Machine Learning
Data Scientist et Big Data
Compétences
La vision du Gartner
Valeur ajoutée du Big Data en entreprise

La collecte des données Big Data

Typologie des sources
Les données non structurées
Typologie 3V des sources
Les données ouvertes (Open Data)
Caractéristiques intrinsèques des sources
Nouveau paradigme de l'ETL à l'ELT
Du "schema On Write" au "Schema on Read"
Le concept du Data Lake
La vision d'Hortonworks
Les collecteurs Apache on Hadoop
SQOOP versus NIFI
Apache SQOOP - Présentation
Apache NIFI - Présentation
Les API de réseaux sociaux
Lab : Ingestion de données dans un cluster avec Apache NIFI

Le calcul massivement parallèle

Genèse et étapes clés
Hadoop : Fonctions coeurs
HDFS - Différenciation
HDFS - Un système distribué
HDFS - Gestion des blocs et réplication
Exemples de commandes de base HDFS
MapReduce : aspects fonctionnels et techniques
Apache PIG et Apache HIVE
Comparatif des 3 approches
Les limitations de MapReduce
L'émergence de systèmes spécialisés
Le moteur d'exécution Apache TEZ
La rupture Apache SPARK
SPARK point clés principaux
SPARK vs Hadoop Performance
L'écosystème SPARK

Formation BIG DATA Les Fondamentaux

Durée de la formation : 3 jour(s)

IMPALA - Moteur d'exécution scalable natif SQL

Le moteur d'exécution Apache TEZ

Hive in Memory : LLAP

Big Deep Learning

La rupture Hardware à venir

Labs : Exemples de manipulations HDFS + HIVE et Benchmark moteurs d'exécutions HIVE

Les nouvelles formes de stockage

Enjeux

Le "théorème" CAP

Nouveaux standards : ACID => BASE

Les bases de données NoSQL

Panorama des solutions

Positionnement CAP des éditeurs NoSQL

Les bases de données Clé-Valeur

Focus Redis

Les Bases de données Document

Focus mongoDB

Les bases de données colonnes

Focus Cassandra et HBase

Les bases de données Graphes

Tendances 1 : Le NewSQL

Tendances 2 : OLAP distribué

Lab : Exemple d'utilisation d'une base NoSQL (HBASE)

Le Big Data Analytics (Partie I - Fondamentaux)

Analyse de cas concrets

Définition de l'apprentissage machine

Exemples de tâches (T) du machine learning

Que peuvent apprendre les machines ?

Les différentes expériences (E)

L'apprentissage

Approche fonctionnelle de base

Les variables prédictives

Les variables à prédire

Les fonctions hypothèses

Pléthore d'algorithmes

Choisir un algorithme d'apprentissage machine

Sous et sur-apprentissage

La descente de gradient

Optimisation batch et stochastique

Anatomie d'un modèle d'apprentissage automatique

La chaîne de traitement standard

Composantes clés et Big Data

Trois familles d'outils machine Learning

Les bibliothèques de machine Learning standards et Deep Learning

Les bibliothèques Scalables Big Data

Les plates-formes de Data Science

Lab : Exemples de traitement Machine Learning avec Notebook

6 - Le Big Data Analytics (Partie II - L'écosystème SPARK)

Formation BIG DATA Les Fondamentaux

Durée de la formation : 3 jour(s)

- Les différents modes de travail avec Spark
- Les trois systèmes de gestion de cluster
- Modes d'écriture des commandes Spark
- Les quatre API Langage de Spark
- Le machine Learning avec Spark
- Spark SQL - Le moteur d'exécution SQL
- La création d'une session Spark
- Spark Dataframes
- Spark ML
- L'API pipeline
- Travail sur les variables prédictives
- La classification et la régression
- Clustering et filtrage coopératif
- Lab : Exemple d'un traitement machine learning avec Spark

Traitement en flux du Big Data (streaming)

- Architectures types de traitement de Streams Big Data
- Apache NIFI - Description, composants et interface
- Apache KAFKA - Description, terminologies, les APIs
- Articulation NIFI et KAFKA (NIFI ON KAFKA)
- Apache STORM - Description, terminologies, langage (agnostique)
- Articulation KAFKA et STORM (KAFKA ON STORM)
- Apache SPARK Streaming et Structured Streaming
- Articulation KAFKA et SPARK
- Comparatif STORM / SPARK
- Deux cas concrets
- Lab : Réalisation d'un traitement Big Data en Streaming (Big Data streaming analytics)

Déploiement d'un projet Big Data

- Qu'est ce que le Cloud Computing
- Cinq caractéristiques essentielles
- Trois modèles de services
- Services Cloud et utilisateurs
- Mode SaaS
- Mode PaaS
- Mode IaaS
- Modèles de déploiement
- Tendances déploiement
- Cloud Privé Virtuel (VPC)
- Focus offre de Cloud Public
- Caractéristiques communes des différentes offres de Cloud Public
- Focus Amazon AWS
- Focus Google Cloud Platform
- Focus Microsoft Azure
- Classement indicatif des acteurs
- Points de vigilance
- Lab : Visite d'une plate-forme de Cloud

Hadoop écosystème et distributions

Formation BIG DATA Les Fondamentaux

Durée de la formation : 3 jour(s)

L'écosystème Hadoop
 Apache Hadoop - Fonctions coeurs
 HDFS - Système de gestion de fichiers distribué (rappel)
 Map Reduce : système de traitement distribué (rappel)
 L'infrastructure YARN
 YARN - Gestion d'une application
 Docker on YARN
 Les projets Apache principaux et associés
 Les architectures types Hadoop
 Les distributions Hadoop
 Qu'est ce qu'une distribution Hadoop
 Les acteurs aujourd'hui
 Focus Cloudera
 Cloudera Distribution including Apache Hadoop (CDH)
 Focus Hortonworks
 Hortonworks Platforms HDP et HDF
 Nouvelle plate-forme Cloudera
 Vision Cloudera
 Cloudera Data Platform
 Cloudera Data Flow
 Lab : Visite d'une distribution Hortonworks dans le Cloud

Architectures de traitement Big Data

A - Traitement de données par lots (BATCH) : - le batch en Big Data - schéma de fonctionnement - usages types du batch processing - l'orchestrateur Apache OOZIE - les workflows OOZIE - les coordinateurs OOZIE (Coordinators) - limitations de OOZIE => FALCON - points de vigilance
 B - Traitement de données en flux (Streaming) : - principes - fonctionnement - rappel : modèles types de traitement de Flux Big Data - points de vigilance
 C - Modèles d'architecture de traitements de données Big Data : - objectifs - les composantes d'une architecture Big Data - deux modèles génériques : λ et K - architecture Lambda - les 3 couches de l'architecture Lambda - architecture Lambda : schéma de fonctionnement - solutions logicielles Lambda - exemple d'architecture logicielle Lambda - architecture Lambda : les + et les - - architecture Kappa - architecture Kappa : schéma de fonctionnement - solutions logicielles Kappa - architecture Kappa : les + et les -
 L'heure du choix
 Lab : Analyse architecturale de deux cas de figure

La gouvernance des données Big Data

Challenges Big Data pour la gouvernance des données
 L'écosystème des outils de gouvernance Big Data
 Les 3 piliers de la gouvernance Big Data
 Mise en perspective dans une architecture Big Data
 Management de la qualité des données Big Data
 Tests de validation de données dans Hadoop
 Les acteurs face à la qualité des données Big Data
 Management des métadonnées Big Data
 Focus Apache HCatalog
 Focus Apache ATLAS
 Management de la sécurité, de la conformité et la confidentialité Big Data
 Focus Apache RANGER
 Tendances sécurisation des SI
 Points de vigilance
 Lab : Réflexion collective ou individuelle sur des opportunités de projets Big Data dans l'organisation et définition

Formation BIG DATA Les Fondamentaux

Durée de la formation : 3 jour(s)

des objectifs et des premiers jalons

Formation BIG DATA Les Fondamentaux

Durée de la formation : 3 jour(s)

Prise en compte du handicap

Pour les personnes en situation de handicap, afin de nous permettre d'organiser le déroulement de la formation dans les meilleures conditions possibles, contactez-nous via notre formulaire de contact ou par mail (formation@access-it.fr) ou par téléphone (0320619506). Un entretien avec notre référente handicap pourra être programmé afin d'identifier les besoins et aménagements nécessaires.

Modalités et moyens Pédagogiques, techniques et d'encadrement mis en œuvre

Répartition théorie/pratique : 45%/55%. Cette formation se compose d'une alternance d'apports théoriques, de travaux pratiques s'articulant autour d'une application fil rouge, de démonstrations, de phases d'échanges entre participants et de synthèses de la part du formateur.

Formation accessible à distance de n'importe où et n'importe quand, via un ordinateur type PC disposant d'une connexion à Internet à haut débit (ADSL ou plus).

Pour assurer un démarrage dans les meilleures conditions au premier jour de la formation, notre service logistique se met systématiquement en relation, en amont, avec vous afin de réaliser un test de validation technique et de vous présenter l'environnement de formation.

Pendant toute la durée de la formation, le stagiaire dispose d'une assistance technique et pédagogique illimitée, par e-mail, avec un délai de prise en compte et de traitement qui n'excède pas 24h. En complément, le stagiaire peut planifier un rendez-vous pédagogique avec un formateur expert afin d'échanger sur des éléments de la formation.

La durée de la formation affichée sur cette page est une durée estimée qui peut varier en fonction du profil du stagiaire et de ses objectifs (notamment s'il souhaite valider sa formation par le passage d'un examen de certification).

Durant la formation, le formateur prévoit :

Des démonstrations organisées en modules et en séquences découpées le plus finement possible, en suivant le programme pédagogique détaillé sur cette page ;

Des énoncés et corrigés de travaux pratiques à réaliser tout au long de la formation ;

Des travaux pratiques sont proposés ; la plateforme prévoit l'environnement technique nécessaire à la réalisation de l'ensemble des travaux pratiques ;

Le formateur valide les connaissances acquises après chaque TP ;

Il est proposé un ou plusieurs livres numériques faisant office d'ouvrage(s) de référence sur le thème de la formation.

Validation et sanction de la formation

Une attestation mentionnant les objectifs, la nature et la durée de l'action et les résultats de l'évaluation des acquis de la formation sera remise au stagiaire à l'issue de sa formation par courrier électronique.

A la demande, il sera délivré un certificat de réalisation.

Type de formation

Professionalisante ayant pour objectif le perfectionnement, l'élargissement des compétences

Moyens permettant de suivre l'exécution de l'action

Le contrôle de la présence des stagiaires sera assuré par la vérification de l'assiduité des participants. Le stagiaire signera une feuille de présence par demi-journée de formation. Celle-ci sera également signée par le formateur.

Modalité d'évaluation des acquis

Durant la formation, le participant est amené à mettre en pratique les éléments du cours par la réalisation de travaux pratiques réalisés sur PC.

La validation des acquis du stagiaire est faite par le formateur à la fin de chaque atelier. Cette validation individuelle est possible du fait du faible nombre de participants par session de formation (6 personnes maximum).

A la fin de la formation, le stagiaire a donc atteint les objectifs fixés par la formation.

En complément, pour les stagiaires qui le souhaitent, certaines formations peuvent être validées officiellement par l'éditeur en passant un examen de certification.

Access it étant centre d'examen ENI, les examens peuvent être réalisés sur demande à distance ou dans nos locaux de Villeneuve d'Ascq.

Les candidats à la certification doivent produire un travail personnel important en vue de se présenter au passage de l'examen, le seul suivi de la formation ne constitue pas un élément suffisant pour garantir un bon résultat et/ou l'obtention de la certification

Assistance Post-Formation

Toute personne ayant suivi une formation avec Access it bénéficie d'une assistance post-formation d'une durée de 1 mois. Ce nouveau service d'accompagnement permet aux stagiaires rencontrant des difficultés dans la mise en œuvre des connaissances acquises de solliciter l'aide de nos formateurs sur des aspects relatifs aux programmes de formation suivis. Pour en bénéficier, il suffit de se rendre sur la page contact de notre site web et de remplir le formulaire. Une réponse est apportée par mail ou par téléphone dans un délai de 48 heures.

Centre d'Examen

Notre centre de formation agréé Qualiopi bénéficie de l'agrément d'ENI et ICDL pour les certifications informatiques. C'est pour nos clients la garantie de pouvoir suivre des formations préparant à des certifications professionnelles.

Aide à l'orientation

Pour chacune des grandes thématiques couvertes par notre offre de formation, nous proposons via nos spécialistes un rendez-vous physique ou téléphonique qui via un diagnostic permettra aux personnes souhaitant être accompagnées dans le choix d'un programme ou dans la définition d'un parcours de formation une orientation vers les programmes les plus adaptés à leurs besoins et à leur niveau.

Aspects Pratiques

Dès leur inscription, les participants sont contactés par nos services qui s'assurent que les débits internet constatés sur le lieu depuis lequel ils souhaitent se former sont suffisants pour suivre la formation dans des conditions optimales.

À l'occasion de cet appel, nos experts s'assurent également qu'ils disposent du matériel nécessaire pour suivre la formation (PC Portable, webcam, Micro-casque..).

Avant le début de la formation, les participants reçoivent un lien leur permettant d'accéder à la classe virtuelle et leurs identifiants personnels de connexion. Un aide à l'utilisation de la solution de visioconférence utilisée leur est également proposée.

Le jour de la formation, ils se connectent à la classe virtuelle depuis leur navigateur internet. Ils voient et entendent le formateur ainsi que les autres participants et peuvent à tout moment communiquer avec eux.

Ils participent aux échanges et réalisent les ateliers dans les mêmes conditions que s'ils étaient en salle. Grâce à nos outils de prise en main à distance, les formateurs peuvent à tout moment prendre la main sur leurs postes pour les aider ou vérifier leurs TP.

Tout au long de la formation, les participants peuvent bénéficier de l'assistance immédiate de nos experts en composant le numéro qui leur a été communiqué avant la formation.

Des bilans intermédiaires ont lieu en présence des participants du formateur et du référent pédagogique d'Access it afin de vérifier l'état d'avancement de la session, les difficultés rencontrées et permettre d'éventuels actions correctives.

Bénéfices pour les participants

Se former depuis leur lieu de travail ou leur domicile,

Accéder sans se déplacer à la qualité d'une formation délivrée par un formateur consultant ayant une expérience probante sur le sujet animé.

Bénéficier à distance de la richesse d'une formation interentreprises : échanges avec le formateur et les autres participants, partages d'expériences, ateliers pratiques...

Pouvoir se former en toutes circonstances et notamment en cas d'imprévu.

Bénéfices pour l'entreprise

Optimiser ses budgets en limitant les frais de déplacement et d'hébergement.

Proposer à tous ses collaborateurs, quelle que soit leur situation géographique, des formations de qualité (en Inter comme en Intra).

Limiter les temps de déplacement.

Proposer davantage de choix dans les formations à des collaborateurs peu mobiles.

Assurer la montée en compétences de ses collaborateurs quelles que soient les circonstances