Les architectures et infrastructures pour le Bigdata

INFORMATIONS GÉNÉRALES

Type de formation : Formation continue

Éligible au CPF : Non

Domaine : Décisionnel, collaboratif

Action collective : Non

Filière : Stratégie décisionnelle

Rubrique : Modèlisation & conception

PRÉSENTATION

Objectifs & compétences

A l'issue de la formation, le stagiaire sera capable d'initier la conception d'une architecture et d'une infrastructure
Big Data, en disposant d'une vue d'ensemble des différentes solutions dédiées au traitement des données de masse

Public visé

Chefs de projets, Architectes, Développeurs, Data Scientists ou toute personne souhaitant connaître les outils pour concevoir une architecture Big Data, …

Pré-requis

Avoir une bonne culture générale des systèmes d'information et plus particulièrement, avoir des connaissances de base des modèles relationnels, des statistiques et des langages de programmation.

PROGRAMME

Comprendre les principaux concepts du Big Data ainsi que l'écosystème technologique d'un projet Big Data

- L'essentiel du BigData : calcul distribué, données non structurées. Besoins fonctionnels et caractéristiques techniques des projets.
- La valorisation des données. Le positionnement respectif des technologies de cloud,
- BigData et noSQL, et les liens, implications.
- Concepts clés : ETL, Extract Transform Load, CAP, 3V, 4V, données non structurées, prédictif, Machine Learning.
- L'écosystème du BigData : les acteurs, les produits, état de l'art. Cycle de vie des projets BigData.

● Atelier : Amazon Rekognition, Polly, EMR.

Savoir analyser les difficultés propres à un projet Big Data

- Rôle de la DSI dans la démarche BigData. Gouvernance des données : importance de la qualité des données,fiabilité, durée de validité, sécurité des données
- Emergence de nouveaux métiers : Data-scientists, Data labs, Hadoop scientists, CDO, …
- Intégration avec les outils statistiques présents et les outils BigData futurs.

Déterminer la nature des données manipulées

- Les différents modes et formats de stockage.
- Les types de bases de données : clé/valeur, document, colonne, graphe. Besoin de distribution. Définition de
- la notion d'élasticité. Principe du stockage réparti.
- Données structurées et non structurées, documents, images, fichiers XML, JSON, CSV, …

● Atelier : démonstrations avec une base MongoDB et une base Cassandra sur des données de différents types

Code de formation : RMCN304

€ Tarifs

Prix public : 1610 €

Tarif & financement :
Nous vous accompagnons pour trouver la meilleure solution de financement parmi les suivantes :
Le plan de développement des compétences de votre entreprise : rapprochez-vous de votre service RH.
Le dispositif FNE-Formation.
L'OPCO (opérateurs de compétences) de votre entreprise.
France Travail: sous réserve de l'acceptation de votre dossier par votre conseiller Pôle Emploi.
CPF -MonCompteFormation
Contactez nous pour plus d'information : contact@aston-institut.com

Lieux & Horaires

Durée : 14 heures

Rythme : 2 jours. 10 au 11 Septembre 20269 au 10 Novembre 2026

Délai d'accès : Jusqu'a 8 jours avant le début de la formation, sous condition d'un dossier d'insciption complet

Prochaines sessions

Consultez-nous pour les prochaines sessions.

Informations non contractuelles - PDF généré le 13/11/2026

Appréhender les éléments de sécurité, d'éthique et les enjeux juridiques

○ Les risques et points à sécuriser dans un système distribué.
○ Aspects législatifs et éthiques : sur le stockage, la conservation de données, ..., sur les traitements, la commercialisation des données, des résultats
● Atelier : mise en évidence des problèmes liés à la réplication inter-régions et concernant les aspects juridiques des données : droit d'exploitation, propriété intellectuelle, ...
○ Etude des failles de sécurité sur une infrastructure Hadoop.

Exploiter les architectures Big Data

○ Les objectifs de la supervision, les techniques disponibles. La supervision d'une ferme BigData.
○ Objets supervisés. Les services et ressources. Protocoles d'accès. Exporteurs distribués de données.
○ Définition des ressources à surveiller. Journaux et métriques.
○ Application aux fermes BigData : Hadoop, Cassandra, HBase, MongoDB
○ Besoin de base de données avec agents distribués, de stockage temporel (timeseriesDB)
○ Produits : Prometheus, Graphite, ElasticSearch.
○ Présentation, architectures.
○ Les sur-couches : Kibana, Grafana.

● Atelier : mise en oeuvre de prometheus pour la supervision d'une ferme
● Cassandra sur une infrastructure distribuée multi-noeuds.

Mettre en place des socles techniques complets pour des projets Big Data.

○ Etude des différents composants d'une infrastructure BigData :
○ Stockage distribué : réplication, sharding, gossip, hachage,
○ Principe du schemaless, schéma de stockage, clé de distribution, clé de hachage
○ Systèmes de fichiers distribués : GFS, HDFS, Ceph. Les bases de données : Redis, Cassandra, DynamoDB,
○ Accumulo, HBase, MongoDB, BigTable, Neo4j, ...
○ Calcul et restitution :Apport des outils de calculs statistiques
○ Langages adaptés aux statistiques, liens avec les outils BigData.
○ Outils de calcul et visualisation :R, SAS, Spark, Tableau, QlikView, ...
○ Caractéristiques et points forts des différentes solutions.
● Atelier : mise en oeuvre du sharding avec une base de données MongoDB sur une infrastructure distribuée

MODALITÉS

Modalités

Modalités : en présentiel, distanciel ou mixte . Toutes les formations sont en présentiel par défaut mais les salles sont équipées pour faire de l'hybride. – Horaires de 9H à 12H30 et de 14H à 17H30 soit 7H – Intra et Inter entreprise.
Pédagogie : essentiellement participative et ludique, centrée sur l'expérience, l'immersion et la mise en pratique. Alternance d'apports théoriques et d'outils pratiques.
Ressources techniques et pédagogiques : Support de formation au format PDF ou PPT Ordinateur, vidéoprojecteur, Tableau blanc, Visioconférence : Cisco Webex / Teams / Zoom.
Pendant la formation : mises en situation, autodiagnostics, travail individuel ou en sous-groupe sur des cas réels.

Méthode

Fin de formation : entretien individuel.
Satisfaction des participants : questionnaire de satisfaction réalisé en fin de formation.
Assiduité : certificat de réalisation.
Validations des acquis : grille d'evalution des acquis établie par le formateur en fin de formation.

Informations non contractuelles - PDF généré le 13/11/2026