

Durée : 3 jours soit 21 heures

Référence : IF-BGDARCH



Public visé :

Durant cette formation, les participants découvriront le métier d'architecte Big Data, ses responsabilités, ses missions, les compétences et les qualités requises pour devenir Architecte Big Data.

Les profils visés :

Architectes SI,
Développeurs Big Data,
Chefs de projets,
BA.



Pré-requis :

- Aucune connaissance sur les technologies Big Data n'est requise.
- Des connaissances en architecture SI et en conception de bases de données serait un plus.



Objectifs pédagogiques :

- Maîtrise des aspects fondamentaux du Big Data
- Maîtrise des éléments clés liés aux architectures des SI
- Pouvoir proposer des architectures modernes, sur mesures, répondant aux besoins du client
- Comprendre les différents cas d'usages du Big Data et les outils correspondants
- Avoir une idée globale sur les différents outils et Frameworks Big Data et avoir un aspect critique qui permet de choisir le meilleur outil pour chaque cas d'usage
- Mettre en pratique les connaissances acquises dans le cours sur des problématiques réelles que nos experts ont rencontrées et rencontrent au sein de l'entreprise.



Compétences acquises à l'issue de la formation :



Modalités pédagogiques :

Session dispensée en présentiel ou téléprésentiel, selon la modalité inter-entreprises ou intra-entreprises sur mesure. La formation est animée par un(e) formateur(trice) durant toute la durée de la session et présentant une suite de modules théoriques clôturés par des ateliers pratiques validant l'acquisition des connaissances. Les ateliers peuvent être accompagnés de Quizz. L'animateur(trice) présente la partie théorique à l'aide de support de présentation, d'animation réalisée sur un environnement de démonstration. En présentiel comme en téléprésentiel, l'animateur(trice) accompagne les participants durant la réalisation des ateliers.

Note relative au programme : Les ateliers, démonstrations, travaux pratiques et scénarios présentés dans ce programme sont fournis à titre d'exemples illustratifs et ne sont pas exhaustifs. Afin de garantir une adéquation permanente avec les évolutions technologiques, le contenu pédagogique ainsi que les environnements de travaux pratiques (TP) sont susceptibles d'être adaptés, ajustés ou mis à jour au cours du temps.



Moyens et supports pédagogiques :

Cadre présentiel

Salles de formation équipées et accessibles aux personnes à mobilité réduite.

- Un poste de travail par participant
- Un support de cours numérique ou papier (au choix)
- Un bloc-notes + stylo
- Vidéoprojection
- Connexion Internet



- Accès extranet pour partage de documents et émargement électronique

Cadre téléprésentiel

Session dispensée via notre solution iClassRoom s'appuyant sur Microsoft Teams.

- Un compte Office 365 par participant
- Un poste virtuel par participant
- Un support numérique (PDF ou Web)
- Accès extranet pour partage de documents et émargement électronique



Modalités d'évaluation et de suivi :

Avant

Afin de valider le choix d'un programme de formation, une évaluation des prérequis est réalisée à l'aide d'un questionnaire en ligne ou lors d'un échange avec le formateur(trice), qui valide la base de connaissances nécessaires.

Les stagiaires disposent également, depuis leur extranet, d'un **outil d'auto-positionnement** leur permettant d'évaluer leur niveau de maîtrise des compétences visées par la formation. Cette auto-évaluation est réalisée avant le démarrage de la formation à l'aide d'un curseur de positionnement allant du niveau le plus faible au niveau le plus élevé.

Pendant

Après chaque module théorique, un ou plusieurs ateliers pratiques permettent de valider progressivement l'acquisition des connaissances. Un quiz peut accompagner les ateliers pratiques afin de mesurer la compréhension des notions abordées.

Après

À l'issue de la formation, les stagiaires réalisent une nouvelle **auto-évaluation** sur les mêmes compétences que celles évaluées avant la formation. Cette comparaison permet de mesurer leur perception de la progression réalisée et de visualiser les compétences développées au cours de la formation.

Lorsque le programme de formation le prévoit, un examen de certification est organisé conformément aux modalités définies par l'éditeur ou le centre de certification (TOSA, Pearson VUE, ENI, PeopleCert...).

Le formateur réalise également une **évaluation individuelle des compétences** de chaque stagiaire sur l'ensemble des compétences visées par la formation. Pour chacune d'elles, il indique si la compétence est :

- Acquis ;
- En cours d'acquisition ;
- Non acquise.

Lorsque la compétence est évaluée comme **en cours d'acquisition** ou **non acquise**, le formateur formule des observations précisant les difficultés rencontrées ainsi que les actions ou recommandations permettant au stagiaire de poursuivre sa montée en compétences.

Enfin

Un questionnaire de satisfaction permet au participant d'évaluer la qualité de la prestation de formation.



Informations sur l'accessibilité :

Nos formations sont, dans la mesure du possible, conçues pour être accessibles à toutes et à tous. Afin de garantir les meilleures conditions d'accueil et d'apprentissage pour les personnes en situation de handicap, nous vous invitons à contacter notre référente handicap certifiée :

Céline SOLATGES – 05 61 34 39 80 – csolatges@iform.fr

Nous vous remercions de bien vouloir nous communiquer toute information utile à ce sujet en amont de la formation, afin de mettre en place les adaptations nécessaires et d'assurer un accompagnement optimal.

Pour en savoir plus sur les dispositifs d'accompagnement existants, vous pouvez consulter les sites suivants :

- [AGEFIPH](#)
- [FIPHFP](#)
- MDPH de votre département

Description / Contenu

Module 1 : Comprendre le BIG Data

- Introduction (Jeu : Donnez un titre aux images)

IFORM : Agence de Balma : 16 avenue Charles de Gaulle, bâtiment 13, 31130 Balma

Siège social : Mbe 308 46 rue André Vasseur 31200 TOULOUSE

Tél : 05 61 34 39 80 • SIRET : 431 421 742 00096 • APE : 6202A

- De la donnée au Big Data (l'histoire de Google qui compare un grain de riz à un octet)
- Quelles avancées technologiques ont permis l'avènement du Big Data
- Quelles évolutions comportementales ont accéléré l'avènement du Big Data
- Les définitions du Big Data
- Les 5V du Big Data
- Quelles sont les sources de données ?

Module 2 : Connaissance du métier de l'architecte Big Data

- A la découverte du métier d'architecte Big Data
- Compétences et qualités requises
- Missions et responsabilités de l'architecte Big Data

Module 3 : Les questions clés que l'architecte Big Data doit se poser

- Limites des architectures classiques
- La nouvelle architecture big data et son impact sur le SI
- Quel est le cycle de vie de la donnée ? Par quelles transformations passe-t-elle ? Comment la gouverner ?
- Quelles sont les principaux modèles d'architecture d'un SI Big Data ? C'est quoi le "Data Lake" ?
- Le data Lake : une nouvelle philosophie pour le stockage et le traitement de la donnée
- Les avantages du data Lake et centralisation des données
- Comment dimensionner l'infrastructure et mesurer la scalabilité du système ?
- C'est quoi le cycle de vie d'un projet Big Data ?
- Cas d'usage, mise en situation

Module 4 : Le Cadre juridique des données et la CNIL

- Comprendre le champ d'application, les règles et les sanctions concernant la protection des données personnelles
- Les obligations légales des entreprises (anonymisation, droit à l'oubli)
- Comment concilier Big Data et réglementation CNIL ?
- Le nouveau règlement européen sur la protection des données personnelle applicable à partir du 25 mai 2018
- Cas d'usage / succès stories / retours d'expérience par secteurs et métiers

Module 5 : Les bases de données NOSQL

- L'origine du NOSQL
- Définition du NOSQL
- ACID (SQL) vs CAP (Big Data)
- Les différentes Bases de données NoSQL (Clé, Valeur, Clé colonne, Orientée document, Base graphe)
- Classification des bases de données NoSQL selon leurs types et leurs cas d'usages
- Zoom sur HBase, Cassandra et MongoDB

Module 6 : Formats de données

- CSV
- XML
- Json
- Avro
- Parquet
- RestFull et les API
- Couche Confluent Kafka

Module 7 : Panorama des outils Big Data

- Quand utiliser Hadoop ?
- Le paradigme MapReduce.
- Le système de gestion des fichiers distribués HDFS
- Echange de données via Sqoop, Flume, Kafka
- Analyse des données avec Pig et Hive
- Ordonnanceur Hadoop : Oozie
- Moteur de recherche : Solr, ElasticSearch
- Spark : framework de calcul distribué in memory
- La place des ETL & ELT dans le big data ainsi que leurs avantages.
- Zoom sur les ETL Informatica, Talend, ODI

Module 8 : Types d'architectures big data selon les exigences temporelles

- Architectures Batch
- Architectures Streaming
- Lambda Architectures
- Architecture réactive
- Quelques exemples d'architectures (google, LinkedIn, Netflix...)
- Les infrastructures dans l'air du temps
 - Nutanix
 - BDA
 - Gammas de serveurs chez HP, DELL,...
 - Le cloud (Amazon EMC, Microsoft azure, OVH)
- Les distributions (Cloudera, Hortonworks)

Module 9 : NIFI dataflow

- Pourquoi nous besoin d'accélérer la collecte de big data et de la gestion des flux de données
- Qu'est-ce que NIFI DataFlow ?
- Avantages de NIFI DataFlow
- Fonctionnalités de NIFI DataFlow
- Applications courantes de NIFI DataFlow
- Accélération de la collecte de big data et de la gestion des flux de données
- Démonstration

Module 10 : La sécurité et la gouvernance de données

- Dix signes d'alerte dont vous avez besoin pour mieux gérer la gouvernance et la sécurité des données
- Que faire si vous identifiez ces signes d'avertissement

- La sécurité et le contrôle d'accès Kerberos & Knox
- Atlas et Ranger pour la sécurité et la gouvernance

Module 11 : Les éléments indispensables dans un projet Big data

- Comment passer d'un POC à un cas d'usage industriel, les éléments à prendre en compte et les erreurs à ne pas faire.
- Comment passer les cas d'usage en production et MCO (Maintien en conditions opérationnelles)
- L'intérêt de la mise en place du DevOps et l'intégration continue dans les projets Big data
- Panorama sur les outils DevOps (Gitlab, Ansible, Jenkins..)

Module 12 : Conclusion

- Quelques livres
- Vos impressions
- Et après ?