

Formation Big Data, Sécurité des données

Présentation

La sécurisation des clusters Hadoop est une problématique de haute importance pour chaque organisation qui l'utilise. Permettant à une organisation de conserver des données en toute sécurité, la sécurisation des clusters Hadoop permet également de s'assurer que l'organisation est conforme à diverses normes comme PCI ou HIPAA.

Cette formation vous apprendra à mettre en œuvre des clusters Hadoop sécurisés en utilisant l'authentification, l'autorisation et le chiffrement.

Durée : 14,00 heures (2 jours)

Tarif INTRA : [Nous consulter](#)

Objectifs de la formation

- Comprendre la qualification complexe des données
- Identifier les principaux risques touchant les solutions de traitement des données massives
- Maîtriser le cadre juridique (CNIL et PLA (Privacy Level Agreement))
- Connaître les principales solutions techniques de base pour se protéger des risques
- Savoir mettre en œuvre une politique de sécurité pour traiter les risques, les menaces, les attaques

Prérequis

- Avoir de bonnes connaissances dans la sécurité réseau et système,
- Connaître les plateformes Hadoop.

Public

- Ingénieurs de données,
- Développeurs logiciel,
- Consultants sécurité et SI,
- Administrateurs système.

Programme de la formation



Aperçu de la sécurité Hadoop

- Qu'est-ce que Hadoop Security ?
- Pourquoi est-ce important ?
- Aspects de sécurité : Éléments clés à prendre en compte
- Sécuriser l'écosystème Hadoop
- Mettre en place un cluster multi-noeuds de Cloudera

Authentication

- Approches, avantages et inconvénients de l'authentification
- Introduction à Kerberos
- Comment travailler avec Kerberos ?
- Comment activer Kerberos
- Authentification dans Cloudera Manager (CM)
- Exécution des commandes HDFS
- Travailler avec les applications YARN
- Effectuer une analyse ad hoc en utilisant Hive
- Intégration de Kerberos avec Windows Active Directory (AD)
- Avantages de travailler avec l'AD
- Installation de Windows Active Directory Server
- Configuration du serveur AD
- Intégration de Hue avec Active Directory
- Préparer un cluster avec Kerberos
- Authentification
- Intégration de Kerberos avec Active Directory
- Activation de l'identification unique (Single Sign On)

Autorisation

- Principaux frameworks d'autorisation
- Qu'est-ce qu'Apache Sentry ?
- Travailler avec l'autorisation Sentry
- Intégration de Sentry avec HUE
- Interroger Hive
- Listes de contrôle (ACLs) étendues HDFS
- Limitations de Sentry
- Qu'est-ce que Cloudera Record Service ? Son utilité.
- Mise en œuvre du service d'enregistrement (Record Service)

Chiffrement

- Types de chiffrement
- Chiffrement au niveau du système d'exploitation
- Chiffrement HDFS
- Mise en place d'une zone de chiffrement HDFS
- Travailler avec les zones de chiffrement HDFS
- Chiffrement des données en mouvement
- Introduction aux outils SSL
- Utilisation d'un certificat auto-signé avec une autorité de certification racine (Self-Signed Root CA)
- Activation et validation du SSL pour Hadoop Core
- SASL Hive et HiveServer2
- SSL avec Hue

Autres sujets liés à la sécurité (si le temps le permet)

- Audit avec Cloudera Navigator
- Les ACLs HBase
- Usurpation d'identité
- Masquage des données sensibles
- Connexion à l'aide d'une clé
- Fichier à onglets
- Fondamentaux UserGroupInformation

Organisation**Formateur**

Les formateurs de Docaposte Institute sont des experts de leur domaine, disposant d'une expérience terrain qu'ils enrichissent continuellement. Leurs connaissances techniques et pédagogiques sont rigoureusement validées en amont par nos référents internes.

Moyens pédagogiques et techniques

- Apports des connaissances communes.
- Mises en situation sur le thème de la formation et des cas concrets.
- Méthodologie d'apprentissage attractive, interactive et participative.
- Equilibre théorie / pratique : 60 % / 40 %.
- Supports de cours fournis au format papier et/ou numérique.

- Ressources documentaires en ligne et références mises à disposition par le formateur.
- Pour les formations en présentiel dans les locaux mis à disposition, les apprenants sont accueillis dans une salle de cours équipée d'un réseau Wi-Fi, d'un tableau blanc ou paperboard. Un ordinateur avec les logiciels appropriés est mis à disposition (le cas échéant).

Dispositif de suivi de l'exécution et de l'évaluation des résultats de la formation

En amont de la formation

- Recueil des besoins des apprenants afin de disposer des informations essentielles au bon déroulé de la formation (profil, niveau, attentes particulières...).
- Auto-positionnement des apprenants afin de mesurer le niveau de départ.

Tout au long de la formation

- Évaluation continue des acquis avec des questions orales, des exercices, des QCM, des cas pratiques ou mises en situation...

A la fin de la formation

- Auto-positionnement des apprenants afin de mesurer l'acquisition des compétences.
- Evaluation par le formateur des compétences acquises par les apprenants.
- Questionnaire de satisfaction à chaud afin de recueillir la satisfaction des apprenants à l'issue de la formation.
- Questionnaire de satisfaction à froid afin d'évaluer les apports ancrés de la formation et leurs mises en application au quotidien.

NB : dans le cadre d'une Action collective, chaque stagiaire bénéficiaire sera contacté par un prestataire choisi par l'Opco Atlas afin d'évaluer « à chaud » la qualité de la formation suivie.

Accessibilité

Nos formations peuvent être adaptées à certaines conditions de handicap. Nous contacter pour toute information et demande spécifique.

