

Flux de données avec Storm

Durée: 2 jours

1160 €

15 au 16 mars
28 au 29 juin

25 au 26 octobre
13 au 14 décembre

Public:

Chefs de projet, data scientists, développeurs.

Objectifs:

Savoir mettre en oeuvre Storm pour le traitement de flux de données.

Connaissances préalables nécessaires:

Connaissance d'un langage de programmation comme Java ou Python.

Programme:

- Introduction : Présentation de Storm:fonctionnalités, architecture, langages supportés
Définitions:spout, bolt, topology
- Architecture : Etude des composants d'un cluster Storm :master node 'nimbus' et worker nodes
Positionnement par rapport à un cluster Hadoop.Le modèle de données. Différents types de flux.
- Premiers pas : Configuration d'un environnement de développement.
Installation d'un cluster Storm. Travaux pratiques sur le projet storm-starter
- Flux de données : Définition du nombre de flux dans un noeud, création de topologies regroupants des flux entre différents noeuds, communication entre flux en JSON, lecture de flux d'origines diverses (JMS, Kafka, ...)
- Haute disponibilité : Tolérance aux pannes: principe de fiabilisation des master node, workers node, nimbus
Garantie de traitement des flux: principe,paramètres TOPOLOGY_MESSAGE_TIMEOUT_SECS, TOPOLOGY_ACKERS
Traitements temps réel avec Trident. Scalabilité: parallélisme dans un cluster storm, ajouts de noeuds, commande 'storm rebalance'