



Flux de données avec Storm

CB038

Durée: 2 jours

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 regroupant des flux entre différents noeuds,
communication entre flux en JSON, lecture de flux d'origines diverses (JMS, Kafka, ...)



— Phirio —

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. Elasticité: traitements concurrents dans un cluster storm, ajouts de
noeuds, commande 'storm rebalance'