

# Documentation sur le sujet

Sonny Klotz - Jean-Didier Pailleux - Malek Zemni

UVSQ

22/02/2017

- 1 Introduction
- 2 Analyse descriptive de données
- 3 Big Data et Machine Learning
  - Big Data
  - Machine Learning
- 4 Graphe de flux et Graph Mining
  - Graphe de flux
  - Graph Mining
- 5 Conclusion

DCbrain est un logiciel qui permet de visualiser ce qui se passe sur les **réseaux physiques (de fluide)** pour pouvoir trouver/prédire les problèmes et optimiser ces réseaux.

DCbrain est un logiciel qui permet de visualiser ce qui se passe sur les **réseaux physiques (de fluide)** pour pouvoir trouver/prédire les problèmes et optimiser ces réseaux.

Cette visualisation provient de données collectées à partir des réseaux physiques (à l'aide de mesures, de capteurs IOT, etc.) puis analysées grâce aux technologies du **Big Data**.

DCbrain est un logiciel qui permet de visualiser ce qui se passe sur les **réseaux physiques (de fluide)** pour pouvoir trouver/prédire les problèmes et optimiser ces réseaux.

Cette visualisation provient de données collectées à partir des réseaux physiques (à l'aide de mesures, de capteurs IOT, etc.) puis analysées grâce aux technologies du **Big Data**.

## Réseaux physiques

Les réseaux qu'on va traiter dans le cadre de ce logiciel sont des réseaux industriels physiques (des réseaux de fluide, de distribution). Il s'agit des réseaux industriels tels les réseaux de distribution pétrolière, gazière, électrique, etc.

## Big Data

Décrit des ensembles de très gros volumes de données, à la fois structurées, semi-structurées ou non structurées, qui peuvent être traitées et exploitées dans le but d'en tirer des informations intelligibles et pertinentes.

## Big Data

Décrit des ensembles de très gros volumes de données, à la fois structurées, semi-structurées ou non structurées, qui peuvent être traitées et exploitées dans le but d'en tirer des informations intelligibles et pertinentes.

Exemple de sources :

## Big Data

Décrit des ensembles de très gros volumes de données, à la fois structurées, semi-structurées ou non structurées, qui peuvent être traitées et exploitées dans le but d'en tirer des informations intelligibles et pertinentes.

Exemple de sources :

- Capteurs utilisés pour collecter les informations climatiques, de trafic, consommation (Smart cities, Internet des Objets).



## Big Data

Décrit des ensembles de très gros volumes de données, à la fois structurées, semi-structurées ou non structurées, qui peuvent être traitées et exploitées dans le but d'en tirer des informations intelligibles et pertinentes.

Exemple de sources :

- Capteurs utilisés pour collecter les informations climatiques, de trafic, consommation (Smart cities, Internet des Objets).
- Messages sur les réseaux sociaux

## Big Data

Décrit des ensembles de très gros volumes de données, à la fois structurées, semi-structurées ou non structurées, qui peuvent être traitées et exploitées dans le but d'en tirer des informations intelligibles et pertinentes.

Exemple de sources :

- Capteurs utilisés pour collecter les informations climatiques, de trafic, consommation (Smart cities, Internet des Objets).
- Messages sur les réseaux sociaux
- Enregistrements transactionnels d'achat en ligne

## Big Data

Décrit des ensembles de très gros volumes de données, à la fois structurées, semi-structurées ou non structurées, qui peuvent être traitées et exploitées dans le but d'en tirer des informations intelligibles et pertinentes.

Exemple de sources :

- Capteurs utilisés pour collecter les informations climatiques, de trafic, consommation (Smart cities, Internet des Objets).
- Messages sur les réseaux sociaux
- Enregistrements transactionnels d'achat en ligne
- Signaux GPS de téléphones mobile

Comment utiliser et donner du sens à ces masses de données enregistrées sur ces réseaux ?

- 1 Introduction
- 2 Analyse descriptive de données
- 3 Big Data et Machine Learning
  - Big Data
  - Machine Learning
- 4 Graphe de flux et Graph Mining
  - Graphe de flux
  - Graph Mining
- 5 Conclusion

## Définition

Ensemble de techniques de statistique descriptive.

## Définition

Ensemble de techniques de statistique descriptive.

- **Objectifs** : une description succincte, regrouper les données.

## Définition

Ensemble de techniques de statistique descriptive.

- **Objectifs** : une description succincte, regrouper les données.
- **Les données** : tableaux de données quantitatives et qualitatives.



## Définition

Ensemble de techniques de statistique descriptive.

- **Objectifs** : une description succincte, regrouper les données.
- **Les données** : tableaux de données quantitatives et qualitatives.
- **Avantages** : traitement en masse, représentations graphiques.

- 1 Introduction
- 2 Analyse descriptive de données
- 3 Big Data et Machine Learning
  - Big Data
  - Machine Learning
- 4 Graphe de flux et Graph Mining
  - Graphe de flux
  - Graph Mining
- 5 Conclusion

## Big Data

Le Big Data fait référence à la masse de données collectée. On considère du Big Data quand le traitement devient trop long pour une seule machine.

## Big Data

Le Big Data fait référence à la masse de données collectée. On considère du Big Data quand le traitement devient trop long pour une seule machine.

Le Big Data est caractérisé par les 3V :

## Big Data

Le Big Data fait référence à la masse de données collectée. On considère du Big Data quand le traitement devient trop long pour une seule machine.

Le Big Data est caractérisé par les 3V :

- le Volume de données considérable à traiter.

## Big Data

Le Big Data fait référence à la masse de données collectée. On considère du Big Data quand le traitement devient trop long pour une seule machine.

Le Big Data est caractérisé par les 3V :

- le Volume de données considérable à traiter.
- la Variété de ces données qui peuvent être brutes, non structurées ou semi-structurées

## Big Data

Le Big Data fait référence à la masse de données collectée. On considère du Big Data quand le traitement devient trop long pour une seule machine.

Le Big Data est caractérisé par les 3V :

- le Volume de données considérable à traiter.
- la Variété de ces données qui peuvent être brutes, non structurées ou semi-structurées
- la Vélocité qui désigne le fait que ces données sont produites, récoltées et analysées en temps réel.

## Big Data

Le Big Data fait référence à la masse de données collectée. On considère du Big Data quand le traitement devient trop long pour une seule machine.

Le Big Data est caractérisé par les 3V :

- le Volume de données considérable à traiter.
- la Variété de ces données qui peuvent être brutes, non structurées ou semi-structurées
- la Vitesse qui désigne le fait que ces données sont produites, récoltées et analysées en temps réel.

Les traitements de cette quantité importante de données est massivement "parallélisé" avec MapReduce/Hadoop.



Le Machine Learning est :

Le Machine Learning est :

- Une discipline scientifique centrée sur le développement, l'analyse et l'implémentation de méthodes automatisables, offrant la possibilité à une machine d'évoluer grâce à un processus d'apprentissage à partir des données et à effectuer des tâches de façon performante.

## Le Machine Learning est :

- Une discipline scientifique centrée sur le développement, l'analyse et l'implémentation de méthodes automatisables, offrant la possibilité à une machine d'évoluer grâce à un processus d'apprentissage à partir des données et à effectuer des tâches de façon performante.
- Un traitement statistique de masses de données réunissant à la fois mathématiques appliquées et informatique.

## Le Machine Learning est :

- Une discipline scientifique centrée sur le développement, l'analyse et l'implémentation de méthodes automatisables, offrant la possibilité à une machine d'évoluer grâce à un processus d'apprentissage à partir des données et à effectuer des tâches de façon performante.
- Un traitement statistique de masses de données réunissant à la fois mathématiques appliquées et informatique.
- Utilisé lorsque le Big Data rend inopérant les méthodes statistiques traditionnelles.

## Le Machine Learning est :

- Une discipline scientifique centrée sur le développement, l'analyse et l'implémentation de méthodes automatisables, offrant la possibilité à une machine d'évoluer grâce à un processus d'apprentissage à partir des données et à effectuer des tâches de façon performante.
- Un traitement statistique de masses de données réunissant à la fois mathématiques appliquées et informatique.
- Utilisé lorsque le Big Data rend inopérant les méthodes statistiques traditionnelles.

Le Machine Learning est composé de plusieurs types d'algorithmes d'apprentissage (Supervisé, non supervisé, semi-supervisé, par renforcement).

- 1 Introduction
- 2 Analyse descriptive de données
- 3 Big Data et Machine Learning
  - Big Data
  - Machine Learning
- 4 Graphe de flux et Graph Mining
  - Graphe de flux
  - Graph Mining
- 5 Conclusion

DCbrain emploie une approche basée sur une représentation des réseaux physiques en graphes de flux.

DCbrain emploie une approche basée sur une représentation des réseaux physiques en graphes de flux.

Cette représentation en graphe permet de prendre en compte les spécificités des flux du réseau, c'est-à-dire :



DCbrain emploie une approche basée sur une représentation des réseaux physiques en graphes de flux.

Cette représentation en graphe permet de prendre en compte les spécificités des flux du réseau, c'est-à-dire :

- retranscrire les données du réseau sous forme de flux

DCbrain emploie une approche basée sur une représentation des réseaux physiques en graphes de flux.

Cette représentation en graphe permet de prendre en compte les spécificités des flux du réseau, c'est-à-dire :

- retranscrire les données du réseau sous forme de flux
- analyser (calculer) et représenter les données liées au flux du réseau

DCbrain emploie une approche basée sur une représentation des réseaux physiques en graphes de flux.

Cette représentation en graphe permet de prendre en compte les spécificités des flux du réseau, c'est-à-dire :

- retranscrire les données du réseau sous forme de flux
- analyser (calculer) et représenter les données liées au flux du réseau

La représentation du réseau en graphe de flux a pour avantage de :

DCbrain emploie une approche basée sur une représentation des réseaux physiques en graphes de flux.

Cette représentation en graphe permet de prendre en compte les spécificités des flux du réseau, c'est-à-dire :

- retranscrire les données du réseau sous forme de flux
- analyser (calculer) et représenter les données liées au flux du réseau

La représentation du réseau en graphe de flux a pour avantage de :

- repérer beaucoup plus facilement des anomalies dans le réseau

DCbrain emploie une approche basée sur une représentation des réseaux physiques en graphes de flux.

Cette représentation en graphe permet de prendre en compte les spécificités des flux du réseau, c'est-à-dire :

- retranscrire les données du réseau sous forme de flux
- analyser (calculer) et représenter les données liées au flux du réseau

La représentation du réseau en graphe de flux a pour avantage de :

- repérer beaucoup plus facilement des anomalies dans le réseau
- simuler des évolutions du réseau

Ce genre de graphe peut être utilisé pour tout réseau physique de fluide, par exemple les réseaux électriques :

Ce genre de graphe peut être utilisé pour tout réseau physique de fluide, par exemple les réseaux électriques :

### Graphe de flux : d'un réseau électrique :

- Nœuds : des connections
- Arcs : canaux pour acheminer l'électricité (câbles)

Ce genre de graphe peut être utilisé pour tout réseau physique de fluide, par exemple les réseaux électriques :

### Graphe de flux : d'un réseau électrique :

- Nœuds : des connections
- Arcs : canaux pour acheminer l'électricité (câbles)

L'analyse des données à partir d'un graphe de flux est réalisée grâce à la méthode de **graph mining**.



Les graphes sont un outil très efficace pour la représentation de structures complexes comme les réseaux physiques dans notre cas.

Les graphes sont un outil très efficace pour la représentation de structures complexes comme les réseaux physiques dans notre cas.

Le graph mining est une méthode d'analyse de données représentées par un graphe.

Les graphes sont un outil très efficace pour la représentation de structures complexes comme les réseaux physiques dans notre cas.

Le graph mining est une méthode d'analyse de données représentées par un graphe.

Il s'agit d'extraire des sous-graphes qui décrivent l'information recherchée du graphe. Ces informations peuvent ensuite être utilisées pour caractériser et classifier le graphe et effectuer des regroupement, des analyses de fréquence et des recherches de similarités de graphes (dans une base de données de graphes par exemple).

Cette méthode est utilisée dans plusieurs domaines comme les données web (graphes de réseaux sociaux), la chimie (réseaux biologiques), etc.

Cette méthode est utilisée dans plusieurs domaines comme les données web (graphes de réseaux sociaux), la chimie (réseaux biologiques), etc.

Le graph mining est une forme d'analyse de donnée structurée (*structured data mining* dont le processus consiste à trouver et extraire des informations utiles à partir d'une masse de **données semi-structurées**).

- 1 Introduction
- 2 Analyse descriptive de données
- 3 Big Data et Machine Learning
  - Big Data
  - Machine Learning
- 4 Graphe de flux et Graph Mining
  - Graphe de flux
  - Graph Mining
- 5 Conclusion

Et notre application ...