

MASTER MIAGE 2ÈME ANNÉE
UNIVERSITÉ PARIS NANTERRE

MÉMOIRE DE FIN D'ÉTUDES

**Méthodes pour bien choisir son
architecture Big Data en fonction de
son processus métier**



Auteur :
LUDWIG SIMON

Tuteur :
MCF. EMMANUEL HYON

Février 2019 — Juin 2019

Remerciements

Remerciements

Résumé

Résumé

Motivations

Objectifs

Sommaire

1	Comment définir les critères de choix	9
2	Analyse des solutions existantes	11
2.1	Ingestion/Extraction de données	11
2.2	Traitement des données	11
2.3	Stockage des données	12
2.4	Requêtage	13
3	Méthodes d'analyse pour choisir la bonne architecture	15
3.1	Choisir le type d'outil	15
3.2	Choisir l'outil	15
4	Implémentation : Exemple avec un processus métier	17
4.1	Définition du processus métier	17
4.2	Application des méthodes sur le processus métier	17
4.3	Évaluation du résultat	17
	Bibliographie	19

Introduction

Le Big Data

Quand est-il pertinent d'utiliser le Big Data ?

Comment définir les critères de choix

Analyse des solutions existantes

Une fois nos critères définit, il faut décomposer le Big Data qui est un domaine très vaste, en plusieurs catégories [1]. Une fois cette décomposition effectuée, on va pouvoir s'intéresser aux outils existant permettant de réaliser chacune des tâches. On va noter leurs points faibles et leurs points forts ainsi que la manière dont ils sont censés être utilisés.

2.1 Ingestion/Extraction de données

La première catégorie, qui est aussi la première étape d'une architecture Big Data, c'est le récupération de données. Plus précisément comment nous allons récupérer des données, soit via des requêtes sur des sources externes, soit des sources externes nous envoient directement des données.

2.1.1 Avec système de messaging

2.1.1.1 Kafka

2.1.1.2 ActiveMQ

2.1.1.3 RabbitMQ

2.1.2 Sans système de messaging

2.2 Traitement des données

Une fois les données reçues, des traitements sont nécessaires afin de pouvoir stocker les données au format souhaité ou bien pour faire un tri des données utiles.

2.2.1 Visuel

2.2.1.1 Apache Nifi

2.2.1.2 Talend

2.2.2 Non visuel

2.2.2.1 Streaming

Spark Streaming

2.2.2.2 Micro Batch

Outil

2.2.2.3 Batch

Spark

MapReduce

2.3 Stockage des données

Une partie très importante du Big Data est le stockage des nombreuses données que l'ont reçoit. Il existe énormément de manières différentes de stocker des données selon la manière dont nous voulons les utiliser par la suite.

2.3.1 Time Series

2.3.1.1 OpenTSDB

2.3.1.2 InfluxDB

2.3.2 Graph

2.3.2.1 Neo4j

2.3.2.2 JanusGraph

2.3.3 Données Structurées

2.3.3.1 Hive

2.3.3.2 Second outil

2.3.4 Clé-Valeur

2.3.4.1 HBase

2.3.4.2 Redis

2.3.5 Index

2.3.5.1 ElasticSearch

2.3.5.2 Apache Solr

2.4 Requêtage

2.4.1 Visuel

2.4.1.1 Kibana

2.4.1.2 Banana

2.4.1.3 Grafana

2.4.1.4 Tableau

2.4.2 Non visuel

2.4.2.1 OLAP

2.4.2.2 OLTP

Méthodes d'analyse pour choisir la bonne architecture

3.1 Choisir le type d'outil

3.2 Choisir l'outil

Implémentation : Exemple avec un processus métier

4.1 Définition du processus métier

4.2 Application des méthodes sur le processus métier

4.3 Évaluation du résultat

Bibliographie

- [1] Christophe Parageaud. *Big Data, panorama des solutions*. url : <https://blog.ippon.fr/2016/03/31/big-data-panorama-des-solutions-2016/>.

Table des matières

1	Comment définir les critères de choix	9
2	Analyse des solutions existantes	11
2.1	Ingestion/Extraction de données	11
2.1.1	Avec système de messaging	11
2.1.1.1	Kafka	11
2.1.1.2	ActiveMQ	11
2.1.1.3	RabbitMQ	11
2.1.2	Sans système de messaging	11
2.2	Traitement des données	11
2.2.1	Visuel	12
2.2.1.1	Apache Nifi	12
2.2.1.2	Talend	12
2.2.2	Non visuel	12
2.2.2.1	Streaming	12
2.2.2.2	Micro Batch	12
2.2.2.3	Batch	12
2.3	Stockage des données	12
2.3.1	Time Series	13
2.3.1.1	OpenTSDB	13
2.3.1.2	InfluxDB	13
2.3.2	Graph	13
2.3.2.1	Neo4j	13
2.3.2.2	JanusGraph	13
2.3.3	Données Structurées	13
2.3.3.1	Hive	13
2.3.3.2	Second outil	13
2.3.4	Clé-Valeur	13
2.3.4.1	HBase	13
2.3.4.2	Redis	13

2.3.5	Index	13
2.3.5.1	ElasticSearch	13
2.3.5.2	Apache Solr	13
2.4	Requêtage	13
2.4.1	Visuel	13
2.4.1.1	Kibana	13
2.4.1.2	Banana	13
2.4.1.3	Grafana	13
2.4.1.4	Tableau	13
2.4.2	Non visuel	13
2.4.2.1	OLAP	13
2.4.2.2	OLTP	14
3	Méthodes d'analyse pour choisir la bonne architecture	15
3.1	Choisir le type d'outil	15
3.2	Choisir l'outil	15
4	Implémentation : Exemple avec un processus métier	17
4.1	Définition du processus métier	17
4.2	Application des méthodes sur le processus métier	17
4.3	Évaluation du résultat	17
	Bibliographie	19

Table des figures