

MASTER MIAGE 2ÈME ANNÉE
UNIVERSITÉ PARIS NANTERRE

MÉMOIRE

**Les architectures de traitement de
données dans le domaine du Big Data**



LUDWIG SIMON

Tuteur :
DR. EMMANUEL HYON

Septembre 2018 — Juillet 2018

Remerciements

Sommaire

1	Présentation	7
	Bibliographie	9

Présentation

Aujourd'hui le Big Data est devenu indispensable pour les entreprises. Malgré les barrières présentes pour son implémentation, le Big Data permet aux entreprises de s'améliorer grandement dans leurs activités, notamment avec l'utilisation d'intelligence ou bien le machine learning. [1] Le point commun de tout les outils concernant le Big Data est le traitement des données, c'est un point très important car il faut bien le choisir en fonction des besoins pour éviter un surcoût et donc une perte d'efficacité. Il existe trois types de traitement de données dans le Big Data :

- Batch
- Micro Batch
- Streaming

A travers ce mémoire, je vais parcourir ces trois types de traitements de données ainsi que les solutions déjà existante avant de pouvoir réaliser un tableau comparatif pour définir si une solution est meilleure que les autres ou bien pour choisir la bonne solution selon les besoins. [2] [3] [4] [5]

Dans un second temps, je vais essayer de proposer une nouvelle solution pour le type de traitement le plus adéquat par rapport aux comparaisons réalisés en amont.

Bibliographie

- [1] Christophe Parageaud. *Big Data : Panorama des solutions 2016*. url : <http://blog.ippon.fr/2016/03/31/big-data-panorama-des-solutions-2016/>.
- [2] Ph.D Iman Samizadeh. *A brief introduction to two data processing architectures— Lambda and Kappa for Big Data*. url : <https://towardsdatascience.com/a-brief-introduction-to-two-data-processing-architectures-lambda-and-kappa-for-big-data-4f35c28005bb>.
- [3] Pedro F. Pérez-Arteaga. *Cost Comparison of Lambda Architecture Implementations for Transportation Analytics using Public Cloud Software as a Service*. url : https://www.google.fr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=2ahUKEwidzsup_OfdAhVHXRoKHdZPBOYQFjAAegQIBxAC&url=http%3A%2F%2Fwww.insticc.org%2FPrimoris%2FResources%2FPaperPdf.ashx%3FidPaper%3D68693&usg=AOvVaw1ZMGNY87EFZq0-LXzcwfJr.
- [4] Uthayanath Suthakar. *Optimised Lambda Architecture for monitoring scientific infrastructure*. url : https://www.researchgate.net/publication/324493058_Optimised_Lambda_Architecture_for_monitoring_scientific_infrastructure.
- [5] Johannes KroB. https://research.spec.org/fileadmin/user_upload/documents/wg_devops/20150731_LambdaArchitectures.pdf. url : https://research.spec.org/fileadmin/user_upload/documents/wg_devops/20150731_LambdaArchitectures.pdf.

Table des matières

1	Présentation	7
	Bibliographie	9

Table des figures