

# Big Data



Stephane Rose  
Carnet de projet

Arena-si  
9, rue de clichy 75009 Paris,  
France  
01 55 32 00 00  
14/10/2013

# Carnet de projet :

---

Ce carnet de projet a pour but de définir l'état d'avancement global de la maquette pour le projet Big Data. Elle est disponible dans le dossier dropbox correspondant. La documentation sera mise à jour régulièrement afin de faciliter la transparence et la communication durant toute la période de stage.

## Table des matières

Sujet.....	2
Technologies utilisées : .....	2
Présentation graphique : .....	2
Schéma : .....	2
Planning.....	3
Octobre :.....	3
Novembre :.....	4
Décembre : .....	4
Licence.....	5

# Sujet

Mise en place d'une architecture de données distribuée dans le cadre d'un projet Big Data.

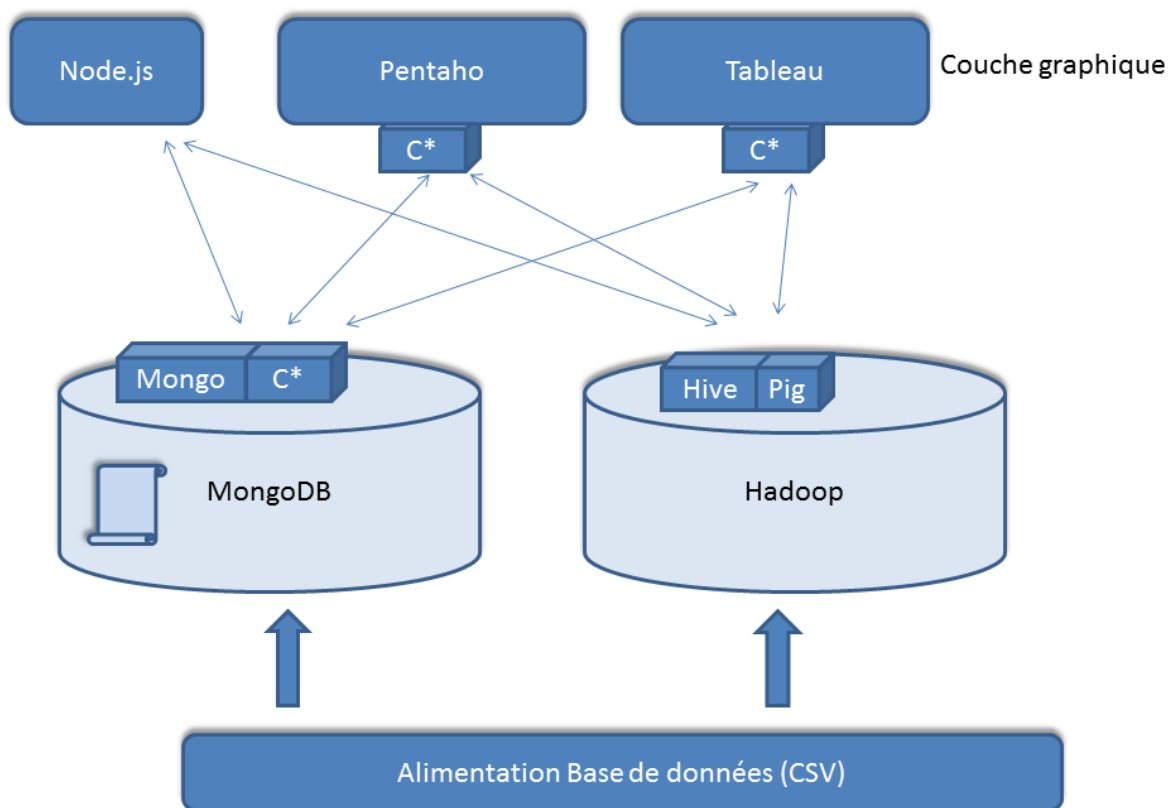
## Technologies utilisées :

- Hadoop
- MongoDB (noSQL)
- Node.js
- Java Script

## Présentation graphique :

- Pentaho
- Tableau
- Node.js

## Schéma :



Projet Big-Data

\*Connecteurs

## Planning

Représentation du travail effectué / restant. La colonne de gauche représente les mois, celle du milieu représente les semaines et la colonne de droite les trois jours de la semaine durant lesquels le stagiaire sera présent (respectivement lundi, mardi, mercredi). Les cases grisées dans la colonne de droite représentent les jours passés et les cases blanches représentent les jours à venir. Les cases vides seront remplies au fur et à mesure de l'état d'avancement du stage.

### Octobre :

Octobre	Semaine 1	<ul style="list-style-type: none"> <li>- Installation du matériel, configuration de l'environnement de développement (Mises à jour système, création du compte, etc...)</li> <li>- Lecture de la doc MongoDB</li> <li>- Lecture sur le Big Data en général</li> <li>- Lecture sur node.js</li> <li>- Lecture sur Hadoop</li> </ul>
		<ul style="list-style-type: none"> <li>- Commande de la Dedibox chez online.net</li> <li>- Installation du système d'exploitation (Ubuntu-lts-12.0.4)</li> <li>- Configuration du serveur (Création des comptes développeurs, compte root, etc)</li> <li>- Création du dépôt GitHub pour la gestion du projet</li> <li>- Installation d'une première version de MongoDB pour faire les premiers tests de familiarisations avec l'environnement</li> </ul>
		<ul style="list-style-type: none"> <li>- Réunion avec Jean-Marie pour définir avec plus de précisions l'architecture de la maquette Big-Data à réaliser</li> </ul>
	Semaine 2	<ul style="list-style-type: none"> <li>- Installation de la nouvelle version de MongoDB-10gen</li> <li>- Création d'un script d'installation automatique de MongoDB (Mise à jour de la clé GPG, et de la liste des paquets, etc)</li> <li>- Installation de webmin sur le serveur pour une administration via le web</li> <li>- Création du script d'automatisation du processus d'installation de webmin + création du mot de passe root</li> <li>- Installation d'une première version node.js et npm (Egalement à remplacer par un script pour récupérer la version à jour)</li> </ul>
		<ul style="list-style-type: none"> <li>- Configuration de MongoDB (Fichier de conf, changement des path par défaut, etc)</li> <li>- Lecture de la doc MongoDB</li> </ul>
		<ul style="list-style-type: none"> <li>- Apprentissage de MongoDB via Mongo</li> <li>- Premiers tests de gestion de collection</li> <li>- Lecture de la doc MongoDB</li> </ul>
	Semaine 3	<ul style="list-style-type: none"> <li>- Ouverture du compte DropBox pour la documentation du projet</li> <li>- Apprentissage de MongoDB via Mongo</li> <li>- Création de plusieurs fichiers de conf pour pouvoir lancer plusieurs instances sur le même serveur</li> <li>- Modification du fichier de conf (Interface REST, chemin par défaut, auht, etc)</li> <li>- Lecture de la doc MongoDB</li> <li>- NOTE : Travail temporaire sur mon serveur personnel à cause d'un problème de disque sur la Dedibox online.net(Ticket en cours)</li> </ul>

		- Ecriture de la documentation (Carnet de projet)
		- Faire un script pour la création de l'arborescence des fichiers de conf MongoDB (lancement de plusieurs instances de la base)
		- Tester le serveur avec les différentes instances.
		- Lecture de la doc MongoDB
		- Réparation du serveur online (Mode Rescue)
	Semaine 4	- Lecture de la doc MongoDB
		- Tester la réplication entre les différentes bases de données sur le serveur
		- Mise à jour du fichier arena-log.pdf dans la Dropbox (Webmin)
		- Lecture de la doc MongoDB
		- Continuer les tests dans la base avec mongo Shell

## Novembre :

Novembre	Semaine 1	
	Semaine 2	
	Semaine 3	
	Semaine 4	

## Décembre :

Novembre	Semaine 1	
	Semaine 2	
	Semaine 3	
	Semaine 4	

## Licence

Les licences à obtenir pour la mise en place de la couche graphique sont :

- <http://www.tableausoftware.com>
- <http://www.pentaho.com/>

Le reste du projet est développé sur des technologies open source et n'est donc pas soumis à des droits d'auteurs dans la mesure où le projet respecte les critères établis par l'Open Source Initiative.