

Procès-verbal n°8	TM: «Products recommender system using the social media networks»	Séance du 25.11.2015
Simone Cogno	079 / 810 20 14	
Lieu: Neuchâtel	Durée: 1 h	Prochaine séance: À définir

Membres du team
Ghorbel Hatem *
Punceva Magdalena e
David Jacot *
Simone Cogno*

* présent
+ partiellement présent
? absent
e excusé

Ordre du jour
<ul style="list-style-type: none"> Dataset : Le schéma du dataset peut être amélioré en supprimant les relation many-to-many et utiliser le collections Liste et Map (optimisation pour la lecture) Database : Pourquoi une database ? Facilité pour la construction d'un Dataset (recherche doublons, mis à jour des objets dans un deuxième moment, performances en lecture et écriture) Pourquoi Cassandra ? <ul style="list-style-type: none"> Flexible schéma ORM adapter en Python CQL similaire à SQL Possibilité de gestion de nombreuse donnée (j'jusqu'à de Petabyte) Simplicité à mettre en place et a gérer un cluster. Planning, état actuel : On est à la fin de la partie Design, première semaine d'implémentation. Une prototype sera effectué dans le prochaine 2 semaine

Tâche(s) effectué	Quand	Qui
<ul style="list-style-type: none"> Etude en détail de l'algorithme choisi Prise en main de Spark (lit le premier «Introduction a Spark») <ul style="list-style-type: none"> Installation 	25.11.2015	<ul style="list-style-type: none"> Simone

<ul style="list-style-type: none"> ○ Database, RDD ○ Mllib ○ GraphX ○ Opérations lambda • Installation, prise en main et configuration d'un cluster Cassandra • Réalisation d'un api pour le téléchargement des données en Python (optimisé) • Dataset courant: <ul style="list-style-type: none"> ○ ~200'000 utilisateur (avec la relation «amis») ○ ~80'000 livres ○ ~55'000 évaluation et commentaire 		
---	--	--

Tâche(s) à faire	Quand	Qui
<ul style="list-style-type: none"> • Continuer le téléchargement du dataset • Premier jet de l'algorithme Hybride avec Spark sur le Dataset courant • Etudier en détails les mesures d'affinités sociales que on veut utiliser et les tester • Maîtriser Spark (Mllib, GraphX, etc.) 	2.12.2015	<ul style="list-style-type: none"> • Simone

PV	Qui	Modifié	Accepté
<ul style="list-style-type: none"> • N°7: séance du 27.11.2015 	<ul style="list-style-type: none"> • Simone 	—	