



Collège
de Bois-de-Boulogne

Projet session

Minimiser temps d'attente à la douane



prof. Hafed Benteftifa

Yvon Cadieux

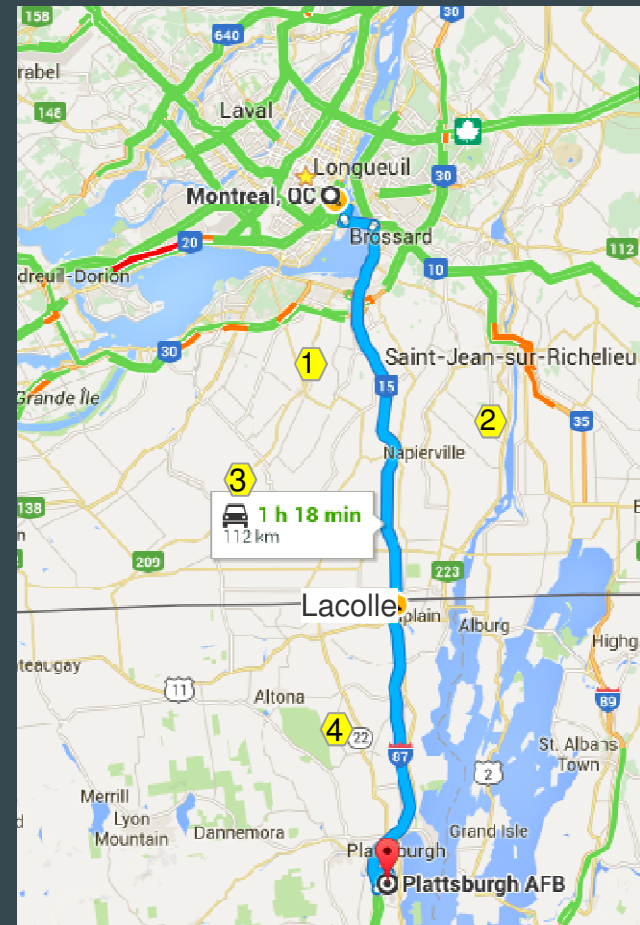
Angelo Fernandes

Février - 2016

Besoin d'affaire

Objectif: Minimiser temps d'attente à la douane

1. Livraison de produits
2. Horaire: 7:00 - 18:00
3. Parcours: Montreal - Plattsburgh
4. Arrêts: multiples
5. Goulot : Douane Lacolle



Cluster utilisé: caractéristiques



WorkStation
18 Gig
2x Xeon 2860
2.1 Ghz
CentOS
6.4



CdhMaster

12 Gig
5 coeur
192.168.1.200

- Name Node
- Secondary NameNode
- jobTracker



CdhNode 1

192.168.1.201

6 Gig
3 coeur

1

- Data Node
- task tracker



CdhNode 2

192.168.1.202

6 Gig
3 coeur

2

cloudera

Ask Bigger Questions

- HBase
- HDFS
- Hive
- Hue
- Oozie
- YARN (MR2 Included)
- ZooKeeper

Cloudera Manager

Accueil

30 minutes avant 8 mars 2016, 18:01 EST

Etat Tous les problèmes d'état d'intégrité Configuration Toutes les commande récentes

Ajouter un cluster

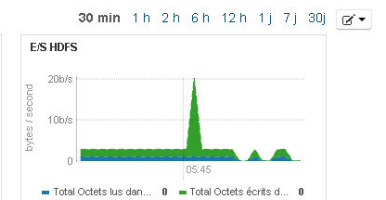
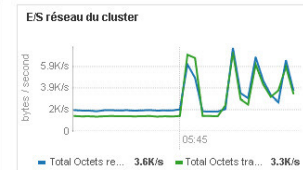
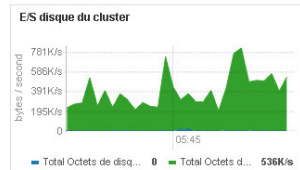
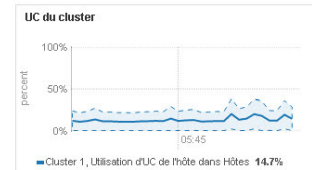
Cluster 1 (CDH 5.5.2, Parcels)

	Hôtes	
	HBase	
	HDFS	1
	Hive	
	Hue	1
	Oozie	
	YARN (MR2 Incl...)	
	ZooKeeper	1

Cloudera Management Service

	Clouder...	
--	------------	--

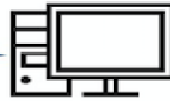
Tableaux



30 min 1 h 2 h 6 h 12 h 1 j 7 j 30j



Btw-taf-2010-2014.csv
origine: Statistique Canada
taille: 50 MB

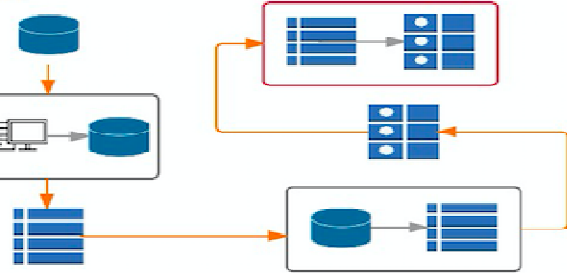


format: String
Bureau: "St-Benard-de-Lacolle"
Emplacement: "Lacolle, QC"
Mis-a-jour: YY-MM-DD HH:MM EST
Délai: HH:MM



Procès

Fichier



Bureau: "St-Benard-de-Lacolle" <string>
Emplacement: "Lacolle, QC" <string>
Mis-a-jour: YY-MM-DD HH:MM EST <string>
Anne: YY <int>
Mois: MM <int>
Jour: DD <int>
Heure: HH <int>
Délai: HH:MM

Manipulation et résultats



```
Select
anne,mois,jour,heure,avg(tempsattente)
tmp_attente
From douane
Where heure between 7 and 19
Group by anne,mois,jour,heure;
```

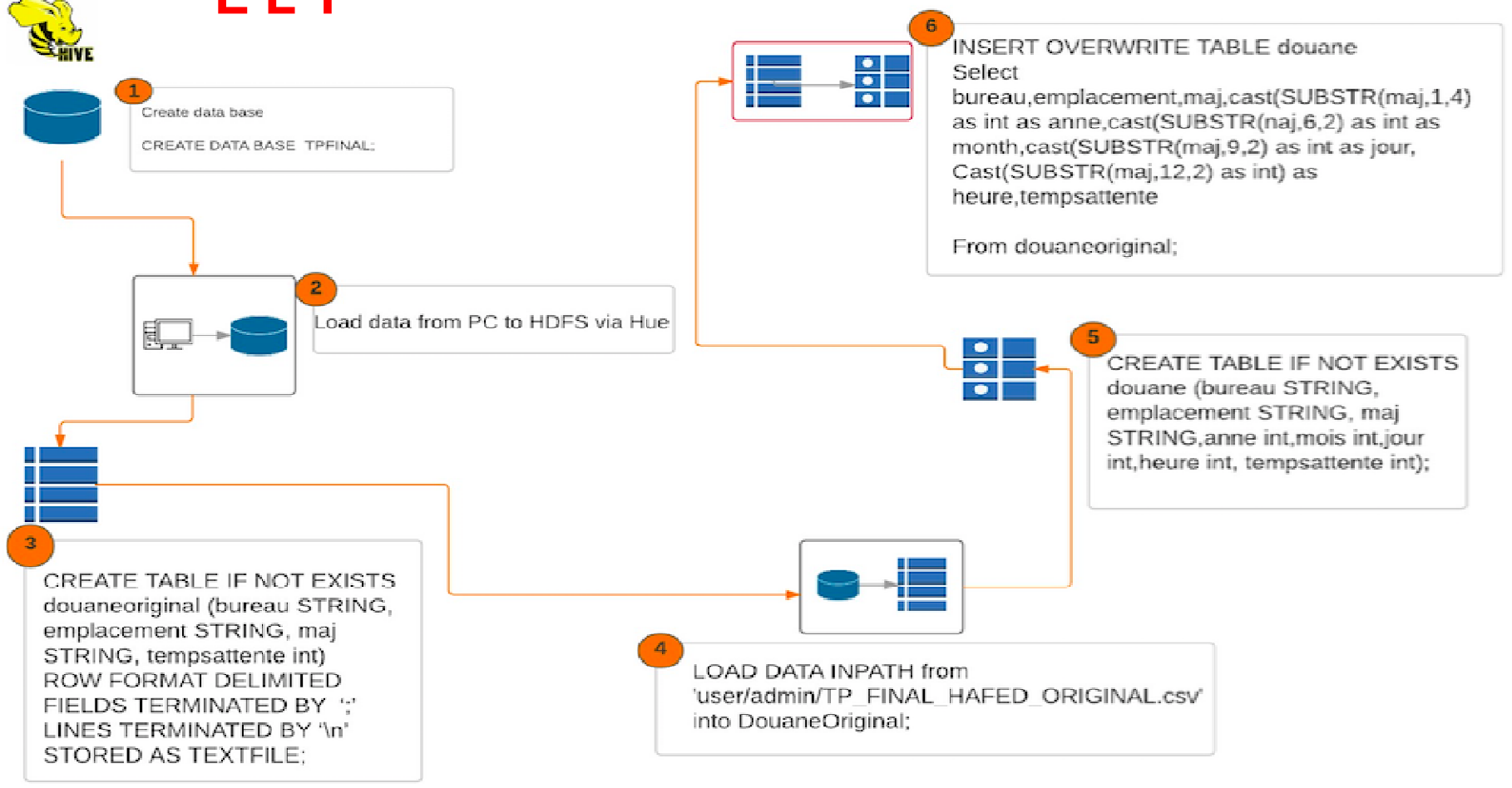
```
1. SELECT `anne`,`mois`,`jour`,`heure`,`avg(tempsattente)` `tmp_attente`
2. FROM `douane`
3. WHERE `heure` BETWEEN 7 AND 19
4. GROUP BY `anne`,`mois`,`jour`,`heure`;
```

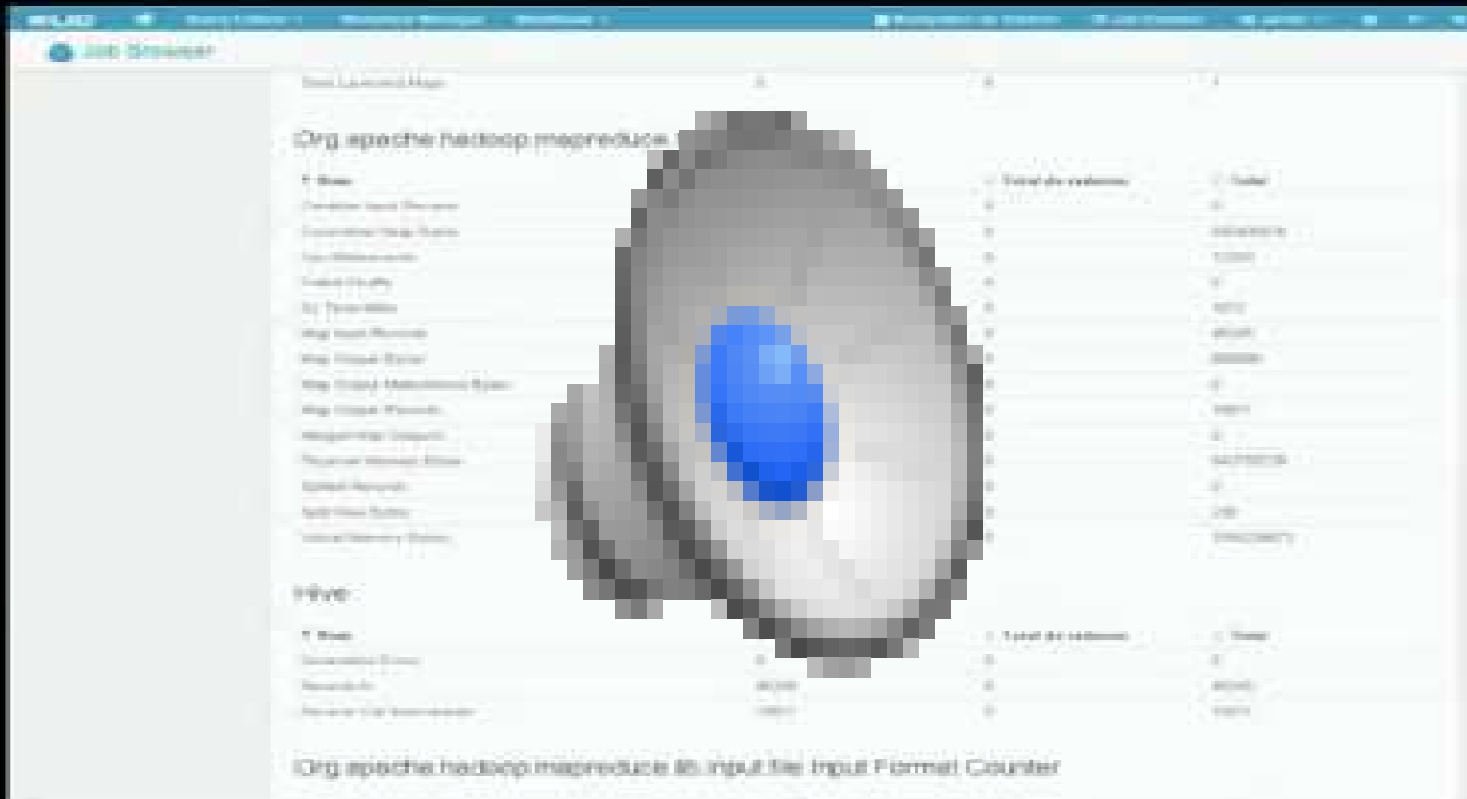
Exécuter Enregistrer Enregistrer sous... Expliquer ou créer une Nouvelle requête

Requêtes récentes	Requêtes	Journal	Colonnes	Résultats	Graphique
anne	mois	jour	heure	tmp_attente_moyen	
12	2010	4	2	7	0
13	2010	4	2	9	20
14	2010	4	2	10	20
15	2010	4	2	11	20
16	2010	4	2	12	20
17	2010	4	2	13	30
18	2010	4	2	14	30
19	2010	4	2	16	30
20	2010	4	2	18	30



ELT





Difficultés

- Installation Cloudera
- Installation carte réseau
- Administrer l'environnement
- Acquisition des données

Compétences Acquises

- Analyse des besoin
- Implantation architecture
- Programmation
- Analyse de résultat

Conclusion

Traitement

Force du traitement parallèle (Hadoop Mapreduce)

Visualisation

Les données peuvent être utiliser par un autre outils pour la visualisation