

Paul MOCHKOVITCH
Martin PRILLARD
Guillaume PERRIN-HOUDON



NoSQL

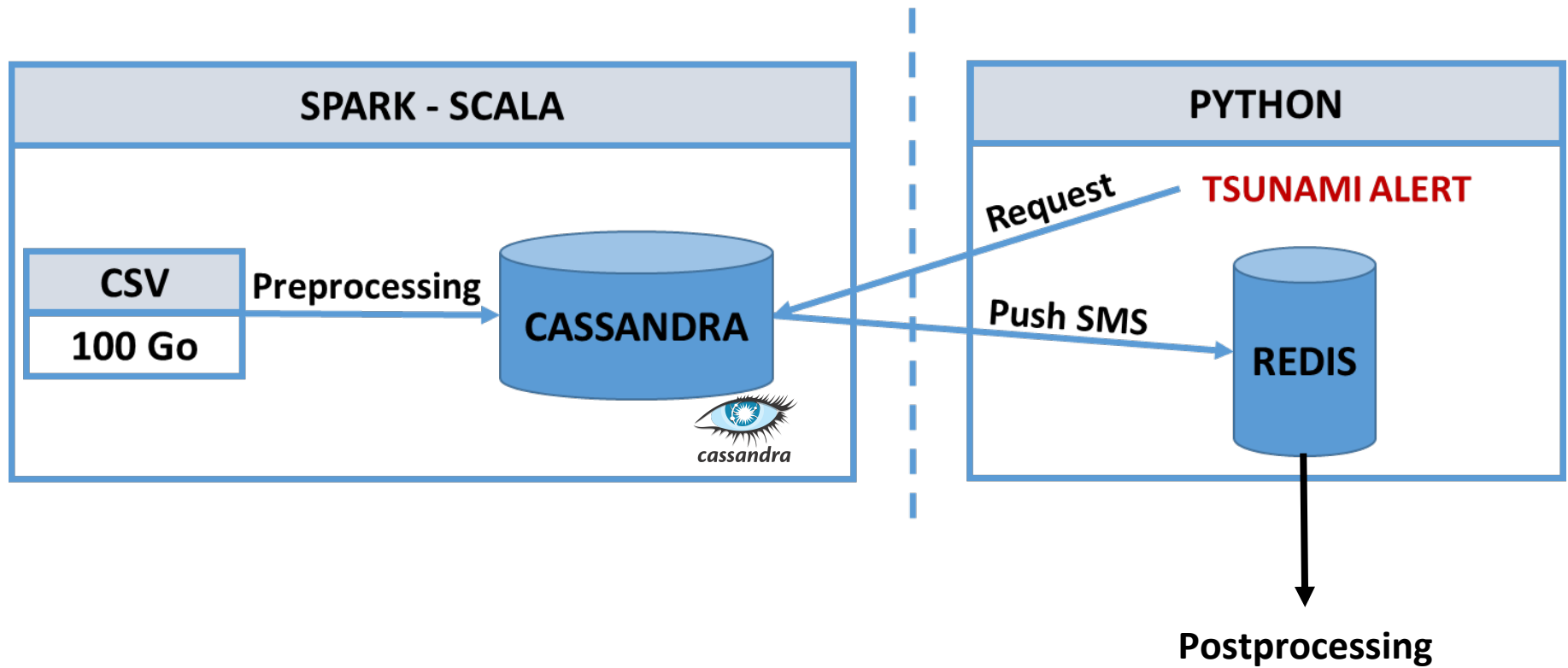
JAPAN TSUNAMI PROJECT



Notre approche

1. Disposer à **intervalles réguliers** de la **dernière localisation connue** (zone GSM) de chaque utilisateur
2. A partir des coordonnées de l'épicentre: **déterminer les zones GSM les plus proches à avertir** (dans les 500 Km)
3. Requêter la base de données pour **recupérer les numéros présents dans les zones à cibler**
4. **Envoyer** les messages par ordre de distance par rapport à l'épicentre
5. **Visualiser** les résultats
6. **Simuler** des répliques

Workflow



Preprocessing via Spark : l'escalier inversé

- Découpage de la fenêtre de temps en intervalles de 10 mn
- Avoir l'ensemble des numéros de téléphones répartis par code GSM pour chaque fenêtre de temps

*Exemple sur le numéro de téléphone XX
pour 7 timestamps*

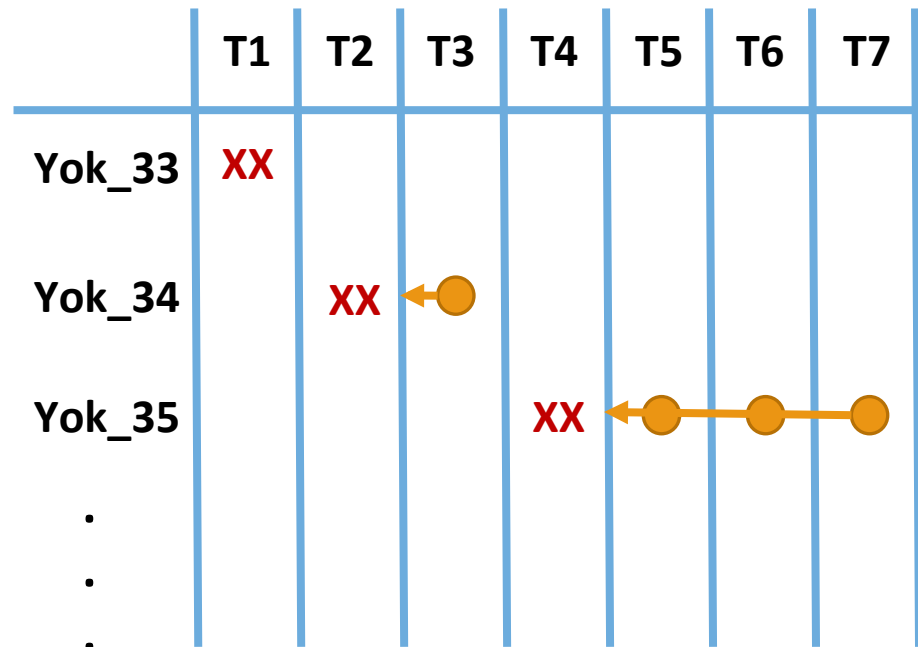


Schéma de la base de données



- Autant de lignes que de codes GSM
- Intervalle de temps de 10 minutes
- Pour chaque intervalle de temps, tous les numéros de téléphones

Résultats

Pre-processing

6 m3.2xlarge

10go : pas de 10 min
preprocessing : 23 min

100go – pas de 10 min
preprocessing : 29 min

100go – pas de 5 min
preprocessing : 31min

100 – pas de 1 min
preprocessing : 32min

10 m3.xlarge

100 – pas de 1 min
map : 32 min (2 fails)

Stages

Storage

Environment

Executors

Spark shell application UI

Spark Stages

Total Duration: 1.1 h
 Scheduling Mode: FIFO
 Active Stages: 0
 Completed Stages: 3
 Failed Stages: 0

Active Stages (0)

Stage Id	Description	Submitted	Duration	Tasks: Succeeded/Total	Input	Shuffle Read	Shuffle Write
----------	-------------	-----------	----------	------------------------	-------	--------------	---------------

Completed Stages (3)

Stage Id	Description		Submitted	Duration	Tasks: Succeeded/Total	Input	Shuffle Read	Shuffle Write
0	runJob at RDDFunctions.scala:24	+details	2015/02/09 18:25:36	36 min	20/20 (11 failed)		6.4 GB	
2	flatMap at <console>:100	+details	2015/02/09 18:12:52	13 min	20/20		22.4 GB	4.9 GB
1	map at <console>:100	+details	2015/02/09 17:56:20	17 min	20/20	100.0 GB		26.8 GB

Failed Stage

Stage Id	Reason
----------	--------

TSUNAMI : (40, 135) at 2015-02-01 00:00:00,000

OMG, the Tsunami destroyed Nagoya node (i-86a97d7c) !

Alert process started at 2015-02-08 23:46:24.270058

Number of GSM Zone to target : 934

Alert process completed at 2015-02-08 23:46:47.777509 in 0:00:23.507451 hours (90263 sent)

Post-processing in progress to calculate time to reach 80 percents of the population...

REDIS 80 percents of the population received the sms in 0:00:20.138750 hours

Visualisation

- CartoDB



- Lien:

<http://cdb.io/1KB1tmL>

Difficultés rencontrées

PROBLEMES MAJEURS



- Connecteur Cassandra – Spark (Save to Cassandra)

PROBLEMES MINEURS



Conf Spark

- Out of memory -> augmenter la taille RAM worker
- No left space on device -> Redirection des RDD de Spark sur l'autre partition

Des questions ?

