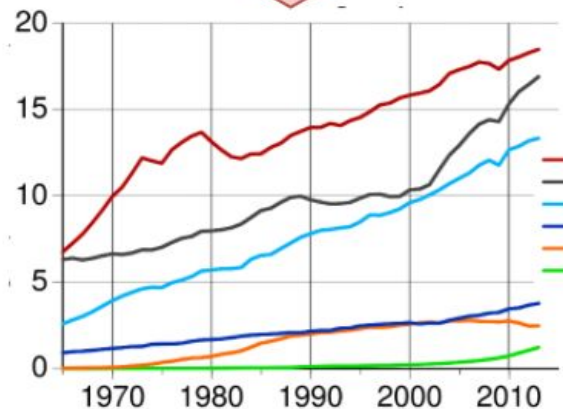
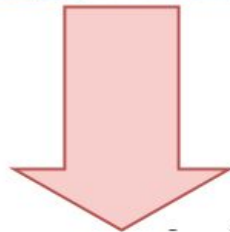


Collecte et Analyse de données Scientifiques

TOURE	Adama Alpha	SAR
HA VINH	Vincent	STL

Encadrant : THIERRY-MIEG Yann

Problématique



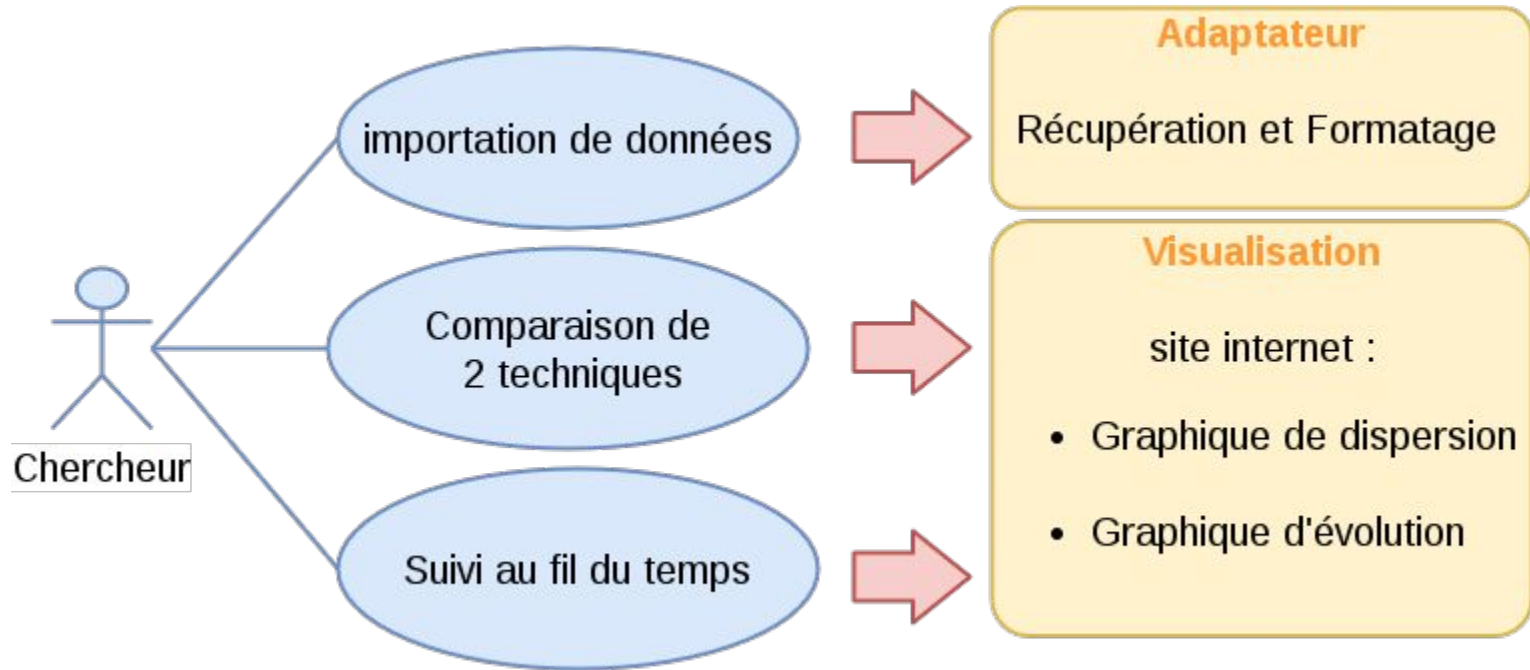
Input :

beaucoup de données
hétérogènes

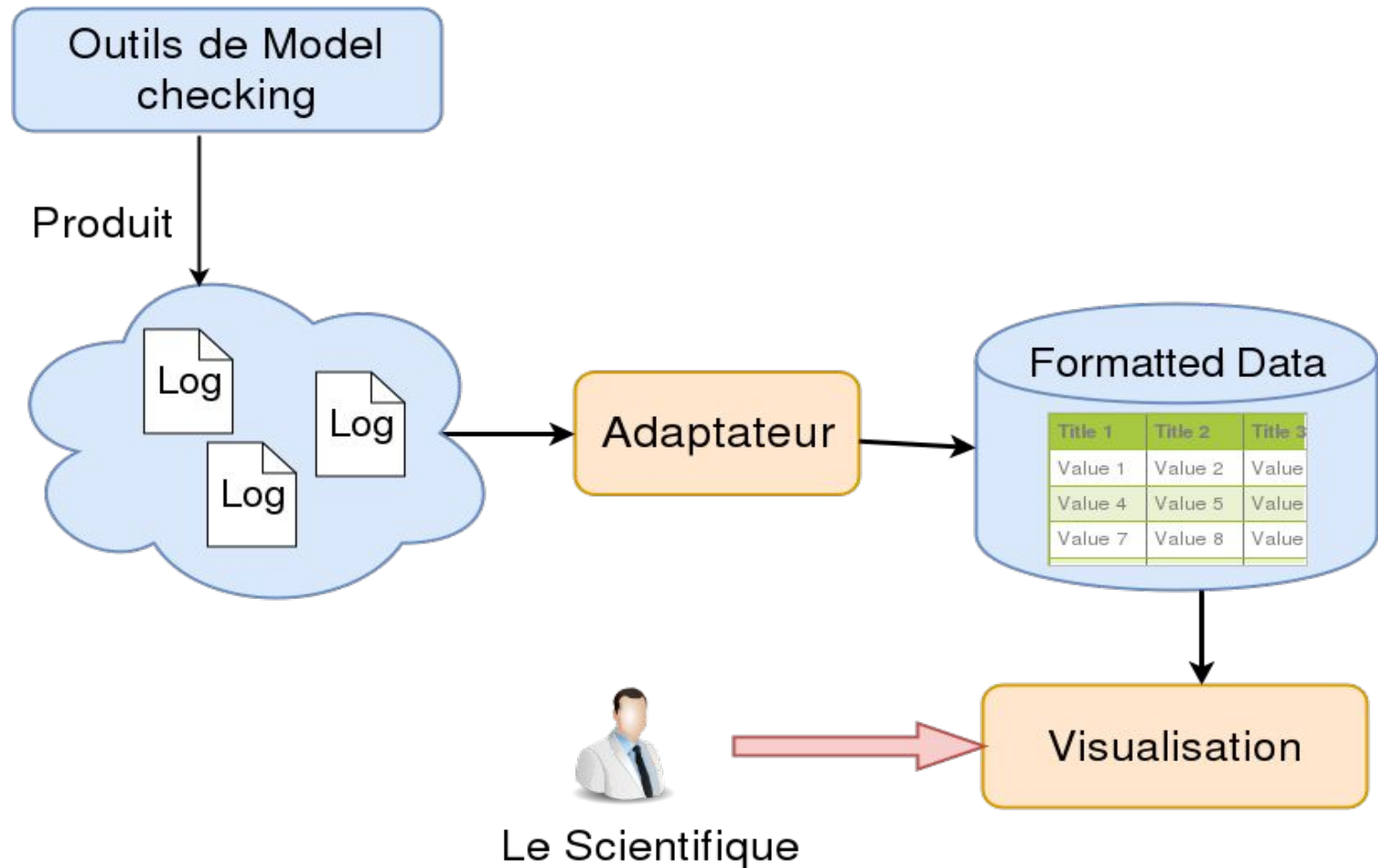
Contraintes :

- Compréhensible
- Auditable
- Reproductible

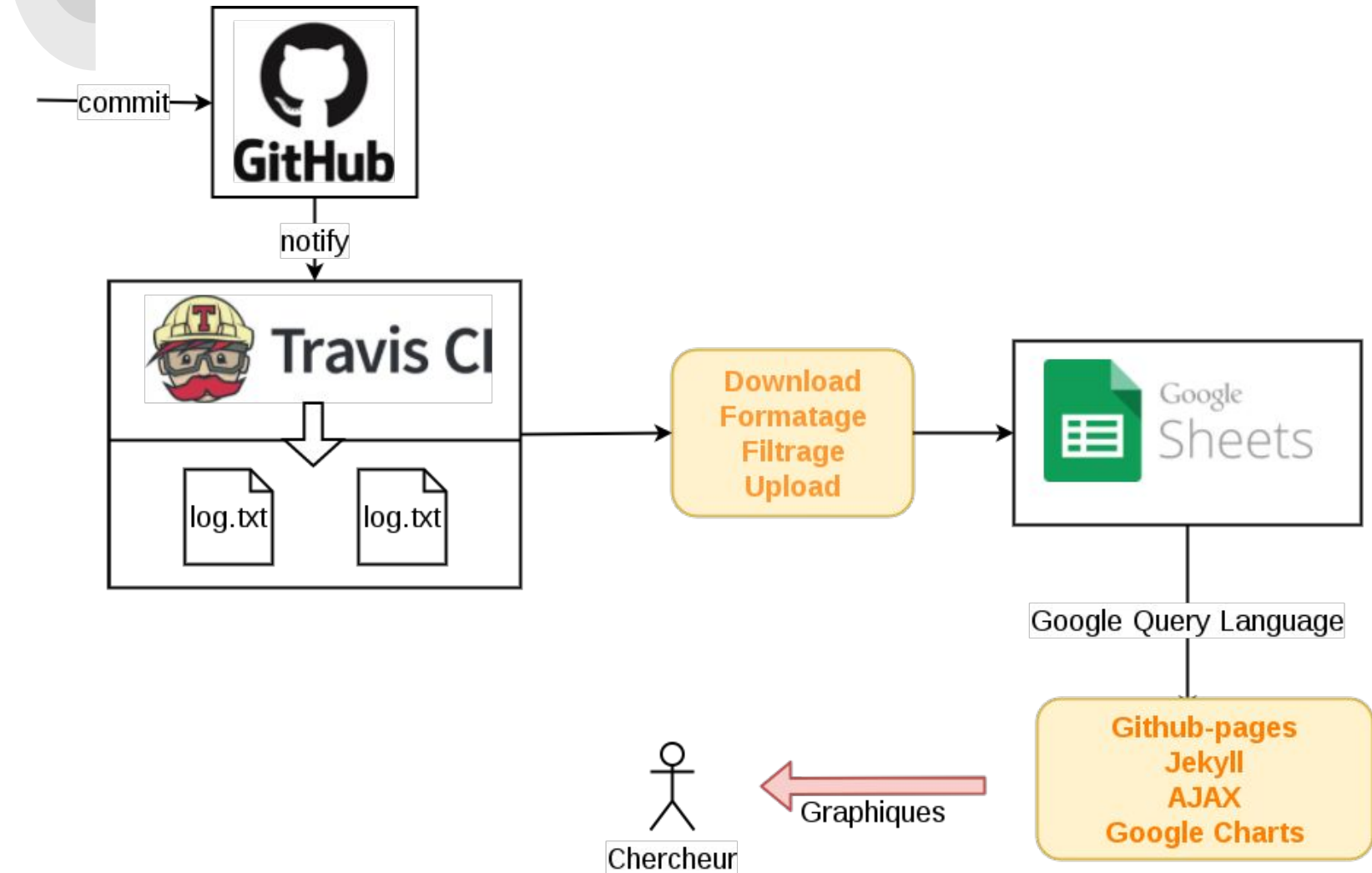
Cas d'utilisation



Architecture de notre solution



Instanciación





Adaptateur



Download

données brutes .TXT
API Travis V3

bash:

requêtes GET http (**curl**) sur l'**API RESTful** Travis-ci
parser **jq** pour lire les réponses au format **JSON**

Formatage

Parsing spécifique
Agrégation en .CSV

perl:

script de parsing des données métiers.
Peut être changé en fonction de la source des logs

Filtrage

Suppression des duplicats
Sélection sur critères

bash:

sort selon k colonnes, pour faire remonter le log à garder

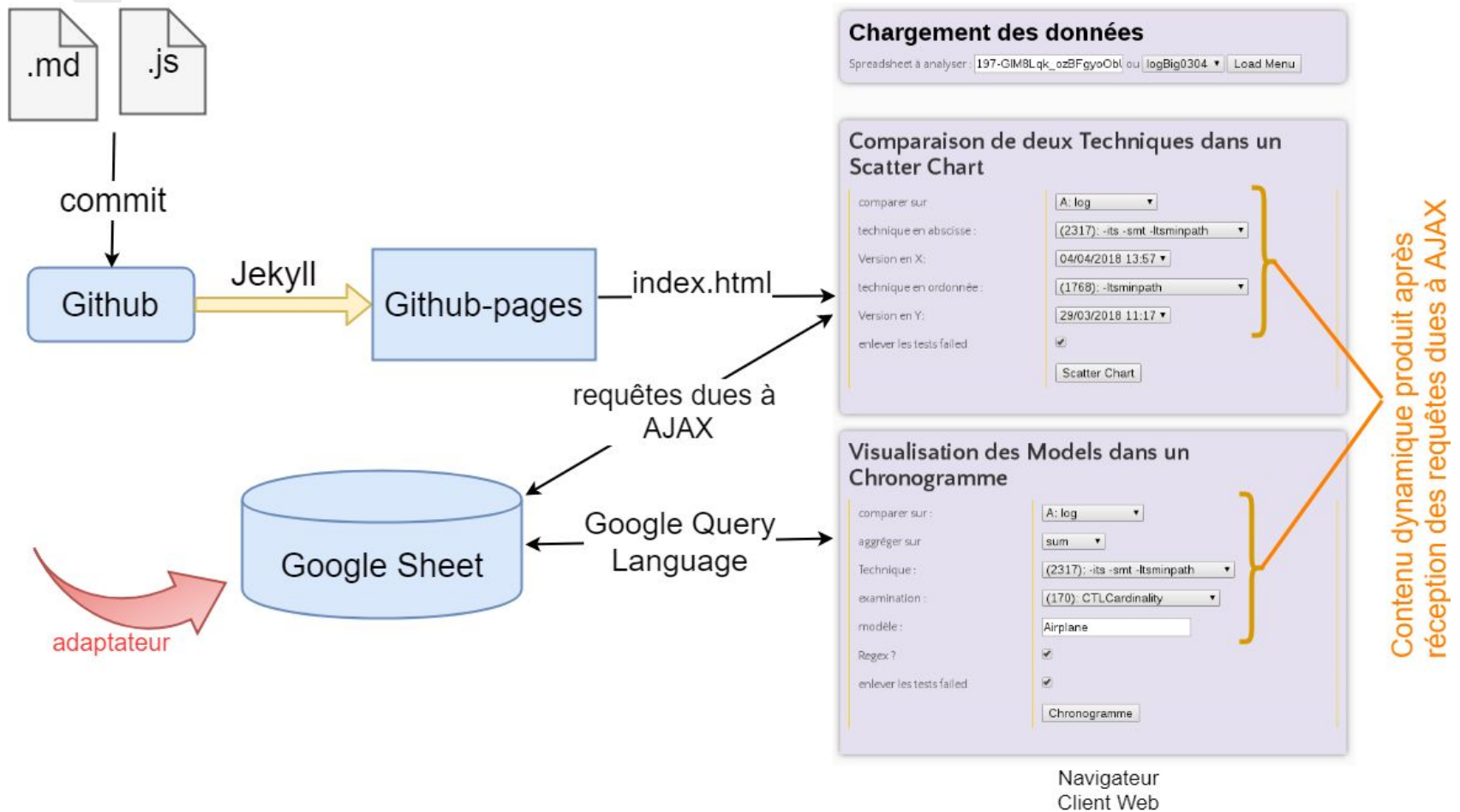
Upload

vers Google Sheets

manuel:

exportation du **.csv** formaté sur une **Google Sheet** en
lecture publique

Visualisation



Visualisation des Modèles dans un Chronogramme

comparer sur :

aggréger sur

Technique :

examination :

modèle :

Regex ?

enlever les tests failed

H: duration(ms) ▼

average ▼

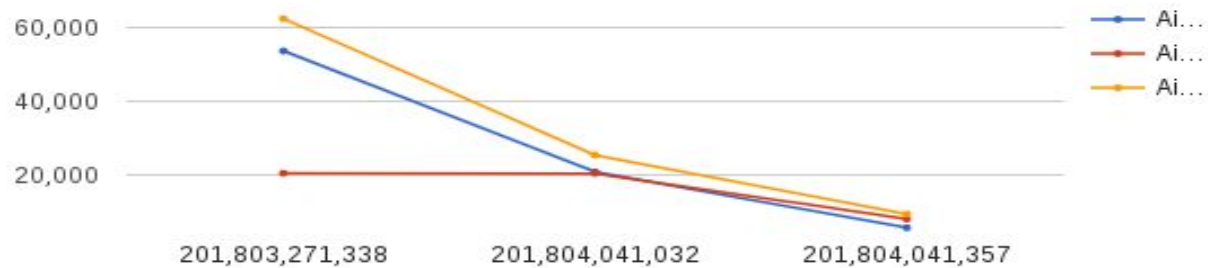
(2317): -its -smt -ltsminpath ▼

(1245): ReachabilityDeadlock ▼

AirplaneLD-COL-0



Chronogramme



		AirplaneLD-COL-0010	AirplaneLD-COL-0200	AirplaneLD-COL-0500
1	201,803,271,338	53,771	20,599	62,499
2	201,804,041,032	20,971	20,533	25,504
3	201,804,041,357	5,834	8,173	9,524

Comparaison de deux Techniques dans un Scatter Chart

comparer sur

technique en abscisse :

Version en X:

technique en ordonnée :

Version en Y:

enlever les tests failed

H: duration(ms) ▼

(1768): -ltsminpath ▼

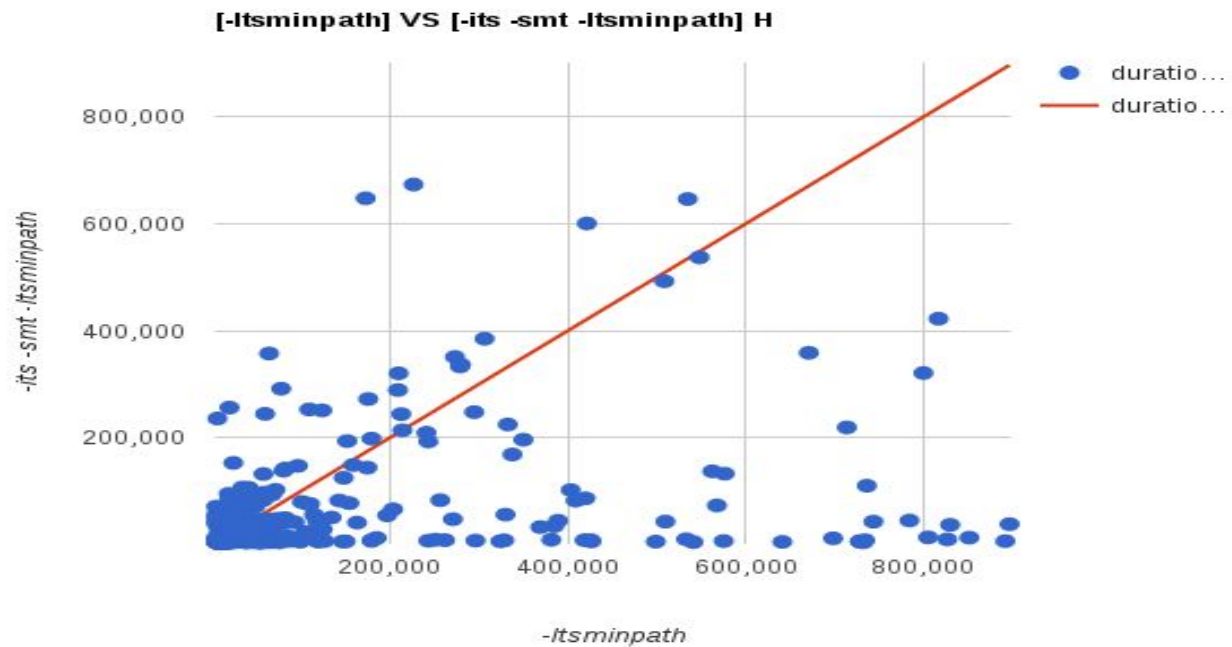
27/03/2018 13:38 ▼

(2317): -its -smt -ltsminpath ▼

27/03/2018 13:38 ▼



Scatter Chart





Mot Pour la fin

Merci

<https://adamaalpha.github.io/TestBeb/>



Volumétrie

Adaptateur :	logs.sh	27
	log2csv.pl	96
	filterRows.sh	91
Visualisation :	test.md (html+js)	135
	index.md (html)	185
	init.js	285
	functions.js	285
total lignes de code :		1104

Données : 156 084 lignes de logs sur Google Sheets



Algorithme de filtrage

#le premier sort trie avec en premier les colonnes clefs

[model,examination,techniques,version]

#puis si égalité dans ces colonnes, trie par ordre croissant les **Test Fail**

#puis si égalité pour Test Fail, trie par ordre décroissant les **Test Passed**

#puis si égalité pour Test passed, trie par ordre décroissant (?) **duration.**

#le 2e sort trie les colonnes **[model,examination,techniques,version]**

avec -u, à valeurs égales dans ces colonnes clefs, **ne garde que la**

première ligne rencontrée

(qui est la ligne filtrée selon le premier sort (fail min, passed max, duration max))

Attention: c'est la 1e ligne rencontrée dans les données brutes, avant tri (non triées)