

Antonio
GOMES RODRIGUES

Raphael
DEBAN

Accélérer le Time To Market : Innovation et qualité grâce à
l'Analyze Performance Intelligence Pipeline

JOURNÉE
FRANÇAISE
DES TESTS
LOGICIELS



11 JUIN 2024
BEFFROI DE MONTROUGE



Focus sur l'équipe performance chez Carrefour

JOURNÉE
FRANÇAISE
DES TESTS
LOGICIELS

Accompagnement
sur la supervision

Campagne de tests de
charge en production
(shift right)

Mise en place de tests
automatiques de non
régression (shift left)



Accompagnement
des projets sur la
performance

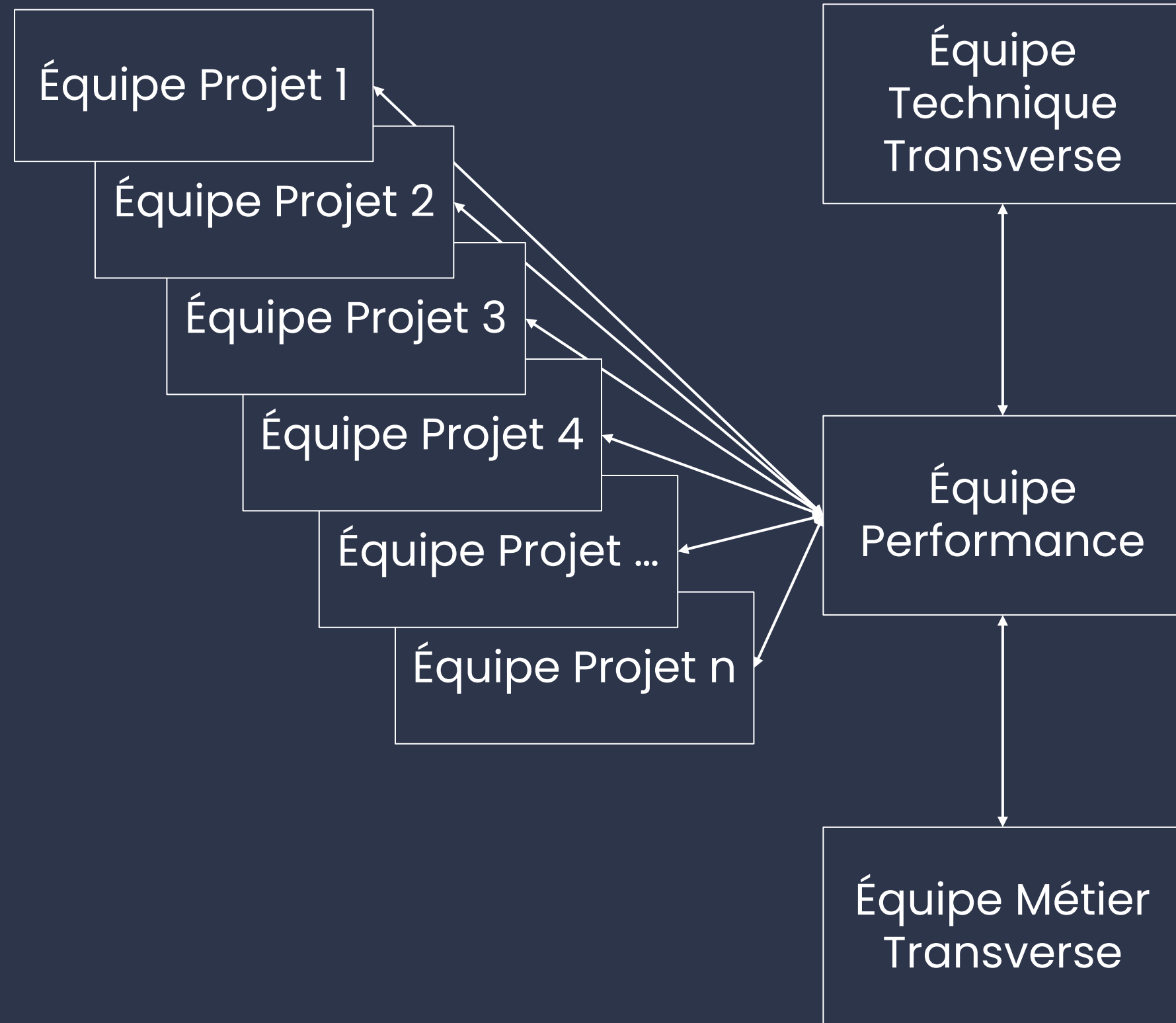
Aide au diagnostic
sur les problèmes
de performance

Accompagnement
lors des tests de
résilience (chaos
monkey)

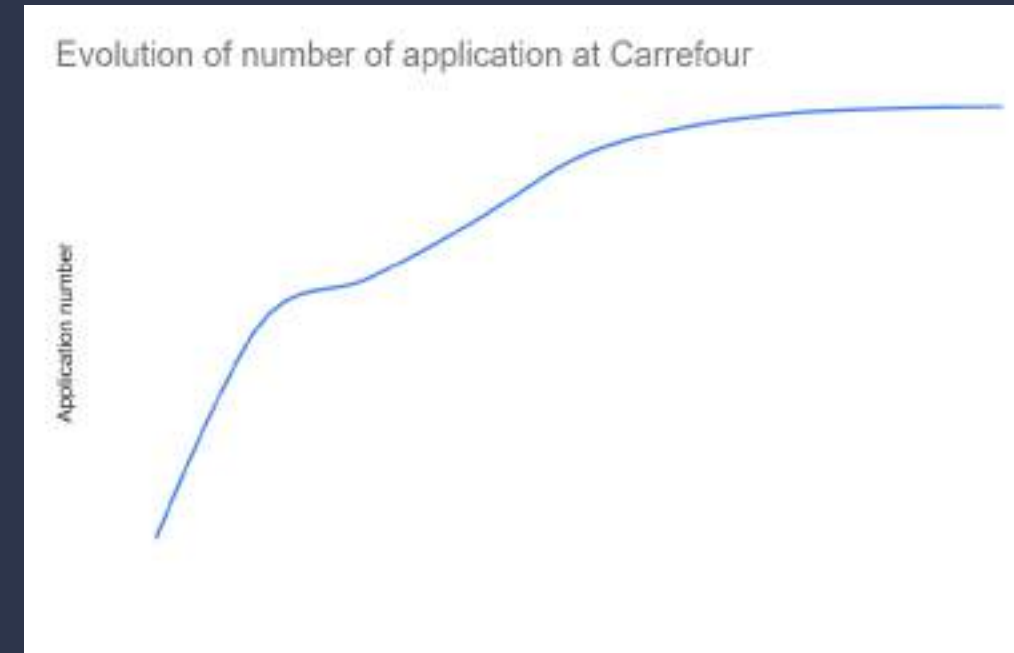
Campagne de tests de
charge en non
production

Audit et
accompagnement
webperf

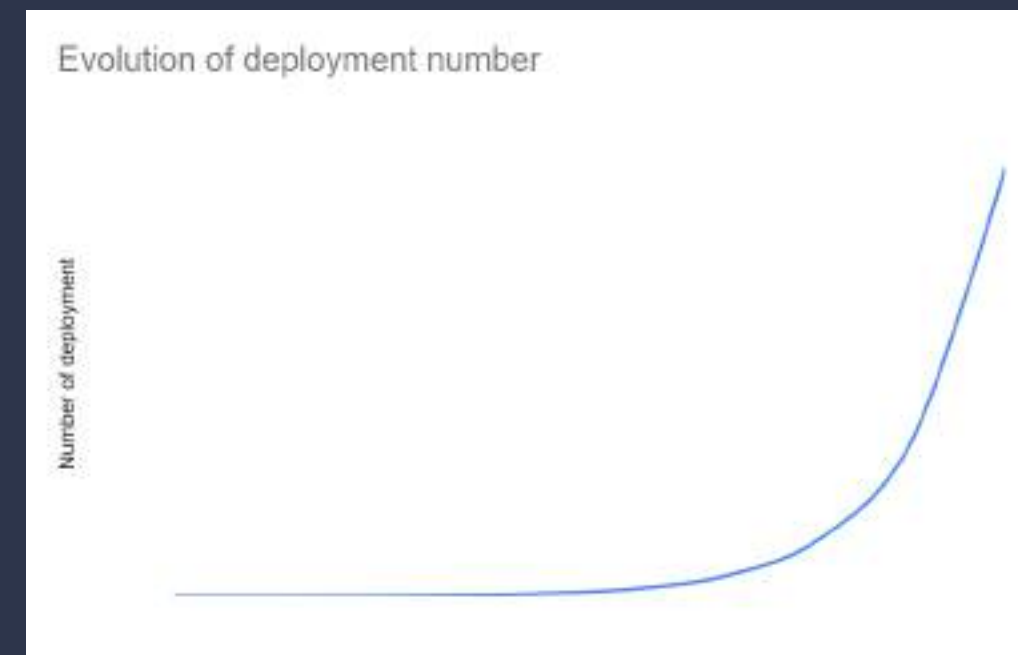
Le contexte de la performance chez Carrefour



De plus en plus de déploiements



De plus en plus d'applications (+1000 applications)





Besoin de transformer l'équipe performance

Plusieurs problématiques à résoudre

Augmenter la productivité



Permettre à nos clients
d'être autonomes



Et plus précisément ?

JOURNÉE
FRANÇAISE
DES TESTS
LOGICIELS



Être indépendant des outils

JOURNÉE
FRANÇAISE
DES TESTS
LOGICIELS



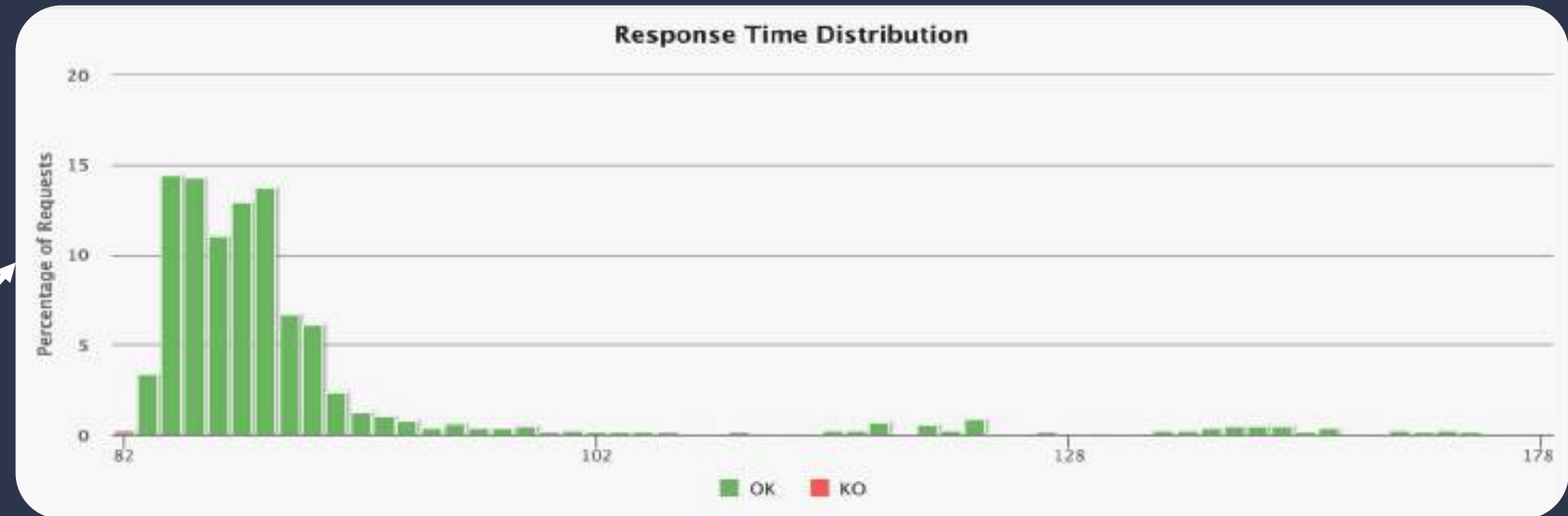
Beaucoup d'outils et donc besoin d'une forte expertise

JOURNÉE
FRANÇAISE
DES TESTS
LOGICIELS



Des choix par défaut et pas toujours pertinents. Des données agrégées difficiles à exploiter

Graphiques choisis pour nous et pas forcément optimaux (ce qui peut fausser nos analyses)



Des choix par défaut, pas toujours pertinents. Des données agrégées ne sont pas assez fiables

Des statistiques sur les requêtes OK et KO
Et donc des statistiques sur des choses différentes

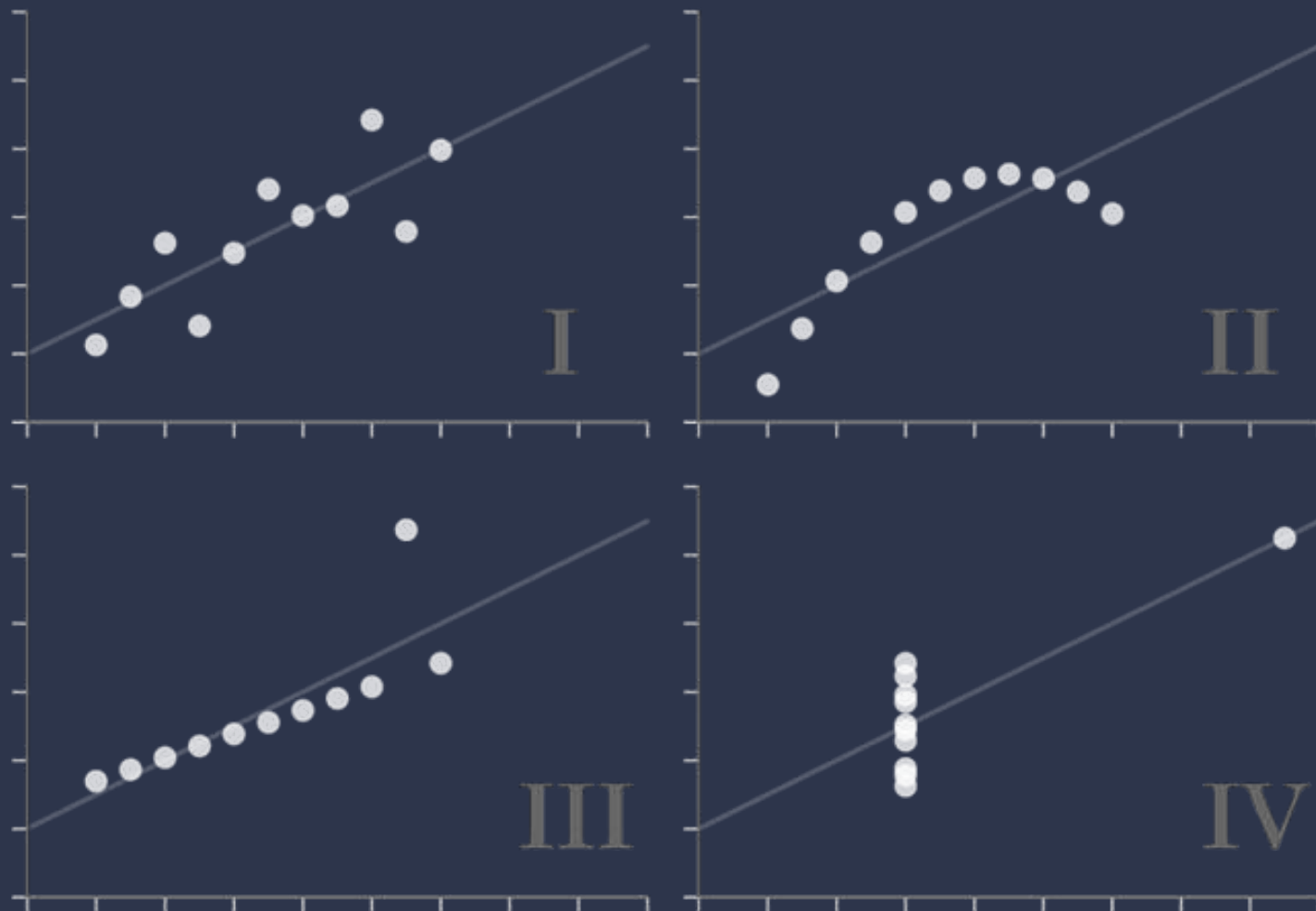
Requests ^	Executions					Response Time (ms)							
	Total ⇅	OK ⇅	KO ⇅	% KO ⇅	Cnt/s ⇅	Min ⇅	50th pct ⇅	75th pct ⇅	95th pct ⇅	99th pct ⇅	Max ⇅	Mean ⇅	Std Dev ⇅
Global Information	73	70	3	4%	1.825	2	3586	4036	4806	5183	5497	3109	1423
▶ get starter infos	2	2	0	0%	0.05	8518	10175	11004	11666	11799	11832	10175	1657
▶ get rand...ll infos	10	9	1	10%	0.25	12218	15008	17443	19780	20392	20545	15528	2649
get pikachu info	2	2	0	0%	0.05	1636	1924	2068	2183	2206	2212	1924	288
get oran berry info	5	3	2	40%	0.125	4	3634	4126	4522	4601	4621	2478	2044
get cher...rry info	5	5	0	0%	0.125	3315	4084	4489	4495	4496	4496	4084	431
get pech...rry info	5	5	0	0%	0.125	1389	3161	3515	3572	3583	3586	2937	801

Des statistiques pour toute la durée du test (y compris la montée en charge et la descente).
Donc des statistiques sur des charges différentes

Plus généralement des données agrégées peuvent masquer des informations

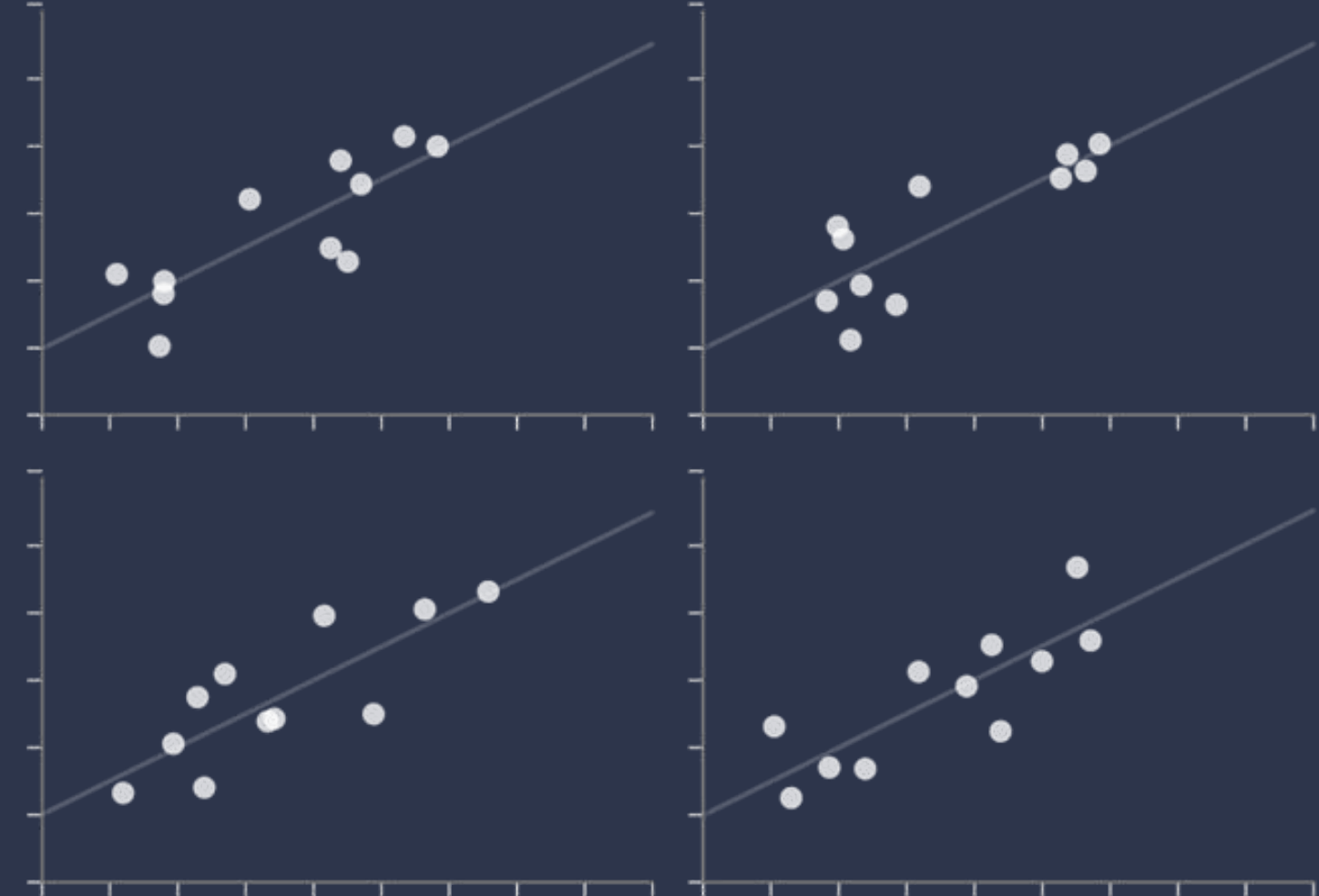
✓ Anscombe's Quartet

Each dataset has the same summary statistics (mean, standard deviation, correlation), and the datasets are *clearly different*, and *visually distinct*.



✗ Unstructured Quartet

Each dataset here also has the same summary statistics. However, they are not *clearly different* or *visually distinct*.



Accélérer l'analyse des résultats de tests

JOURNÉE
FRANÇAISE
DES TESTS
LOGICIELS



Le Shift left uniquement pour tester la charge sans automatiser l'analyse est problématique

Group	Health	Job	Success Percentage	Status	Last P
[Group Name]		Line	50.0%	STILL FAILING	2023-11-21T0
		Fi	100%	SUCCESS	2023-11-21T0
[Group Name]		LL	100%	SUCCESS	2023-11-20T1
		D	100%	SUCCESS	2023-11-20T0
		SL	100%	SUCCESS	2023-11-21T0
[Group Name]		b	100%	SUCCESS	2023-11-21T0
[Group Name]		[Job Name]	100%	SUCCESS	2023-11-21T0
		A	100%	SUCCESS	2023-11-21T0
[Group Name]		[Job Name]	95.45%	STILL FAILING	2023-11-21T0
		[Job Name]	[Success %]	[Status]	[Last P]

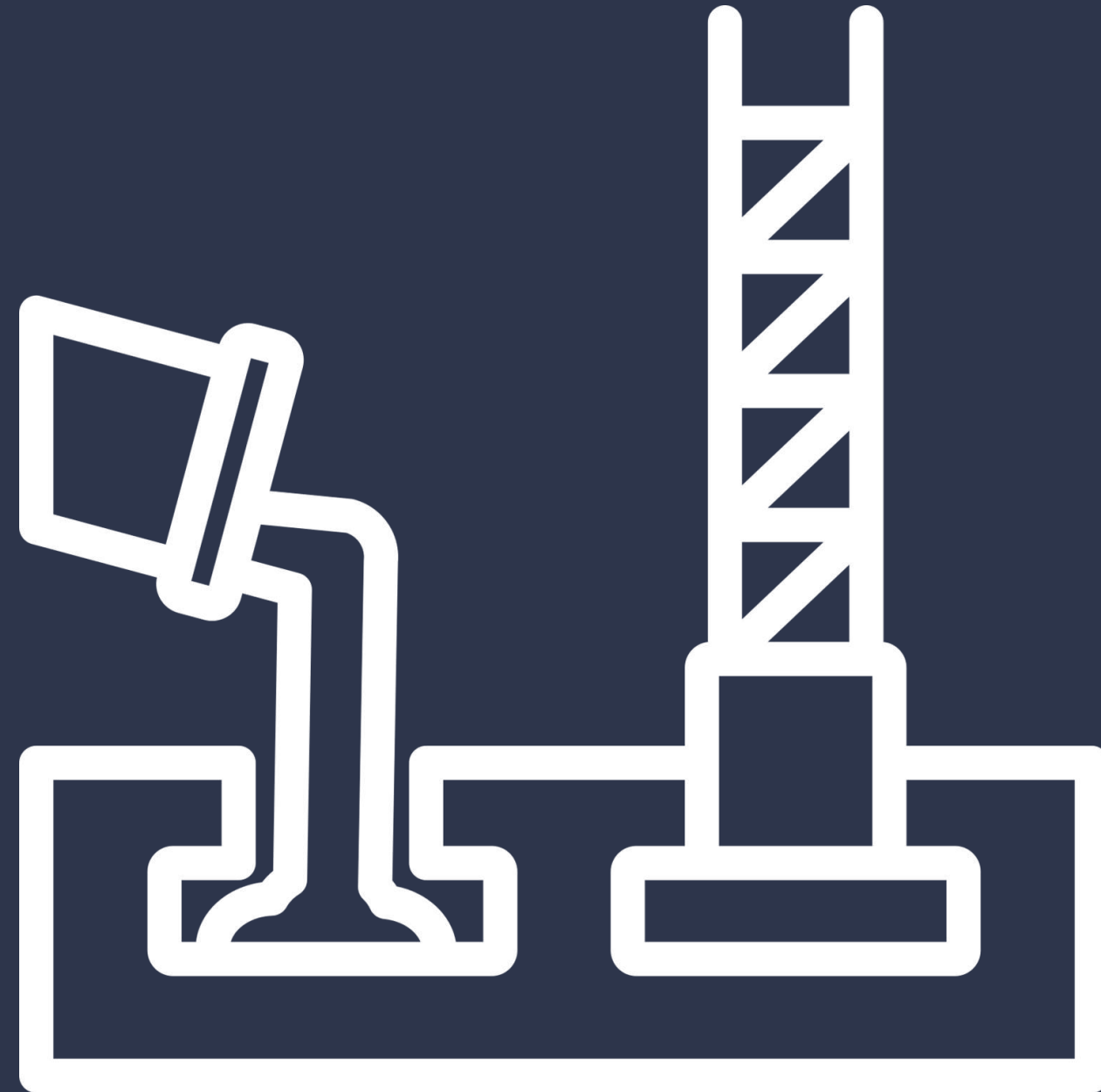
Comment résoudre ces problèmes

JOURNÉE
FRANÇAISE
DES TESTS
LOGICIELS



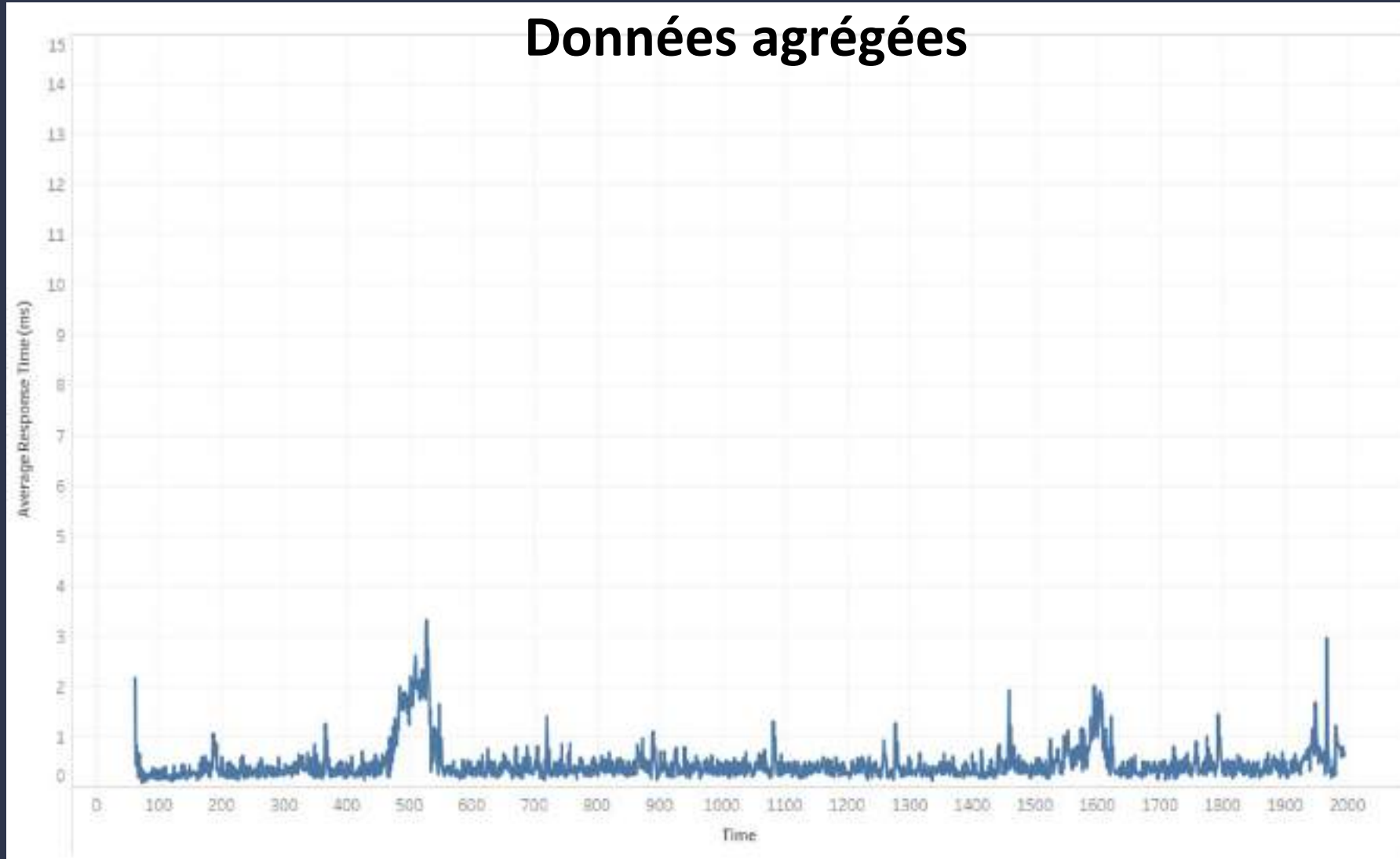
Avoir des bonnes bases

JOURNÉE
FRANÇAISE
DES TESTS
LOGICIELS

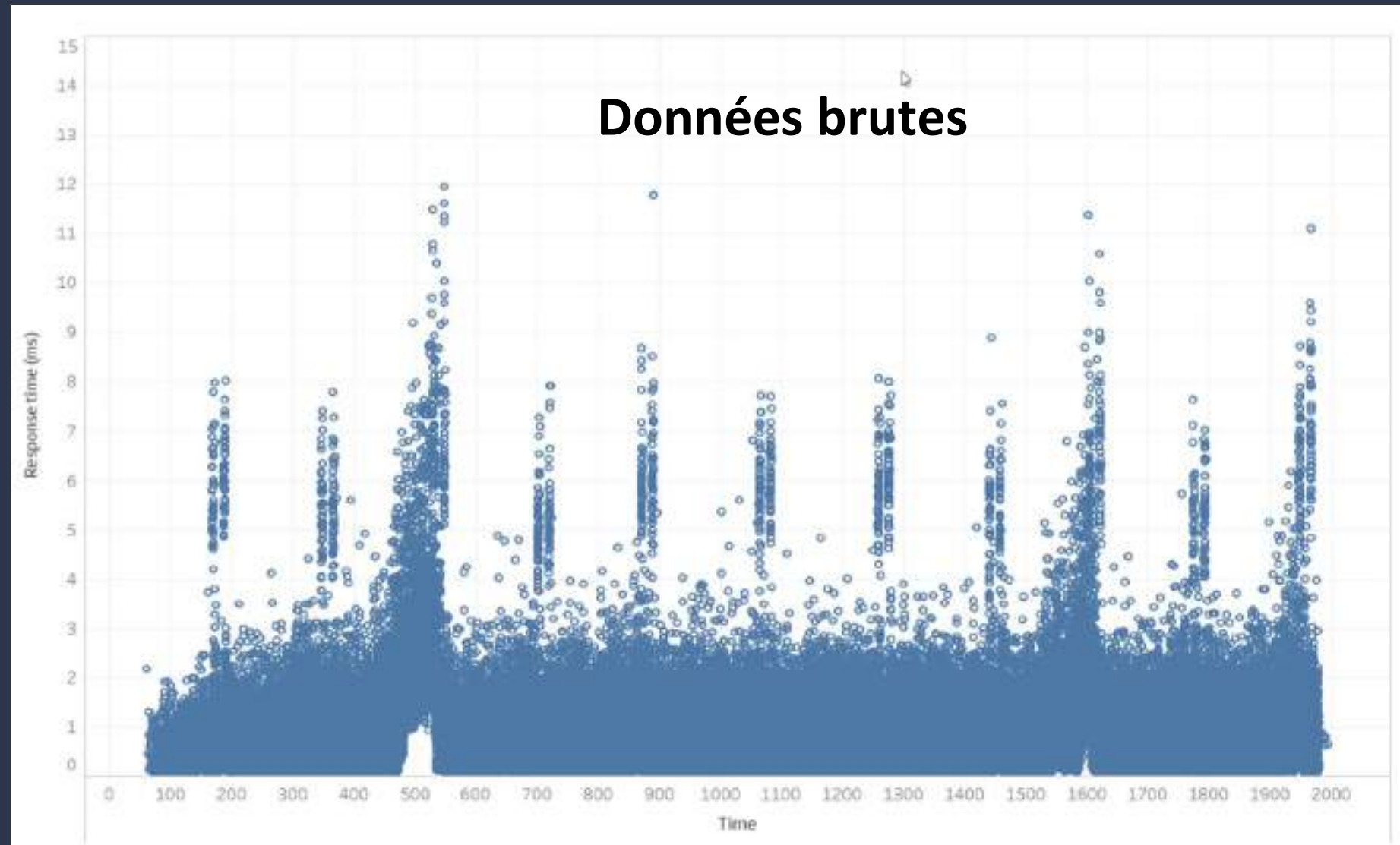


Travailler avec les données brutes

Données agrégées

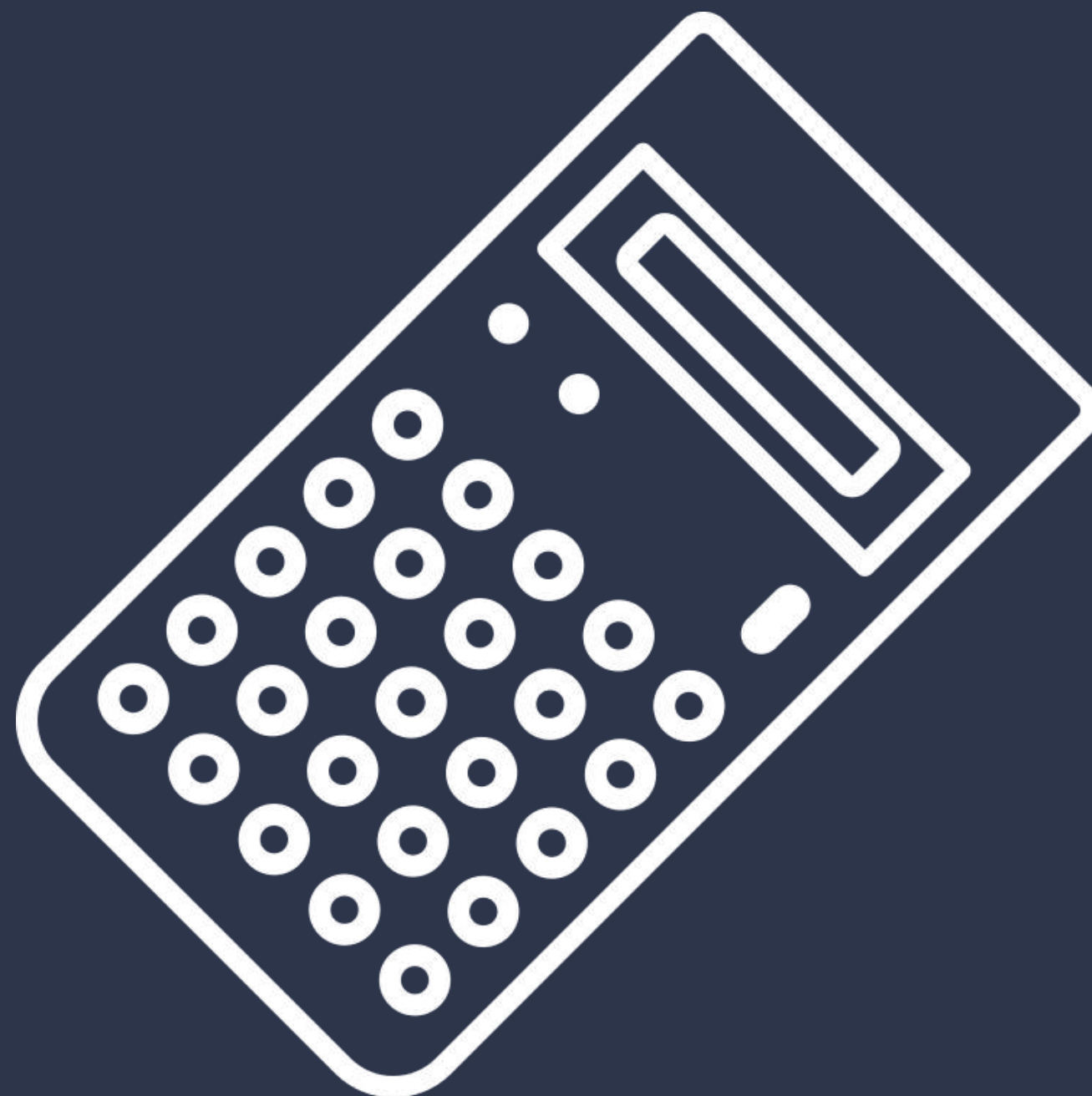


Données brutes



Maîtriser le calcul des statistiques

JOURNÉE
FRANÇAISE
DES TESTS
LOGICIELS



Exemple : séparer automatiquement les données en succès et en échec

Statistique sur l'ensemble des données

	Success	Mean	95pct
TR_0X_*****	All	1083	1504



Statistique par statut

	Success	Mean	95pct
TR_0X_*****	no	4139	4426
	yes	817	1185

Le modèle Fail-Fast préconise une panne de service rapide et une réponse rapide aux entités appelantes, ce qui est essentiel pour des systèmes de mise à l'échelle efficaces.

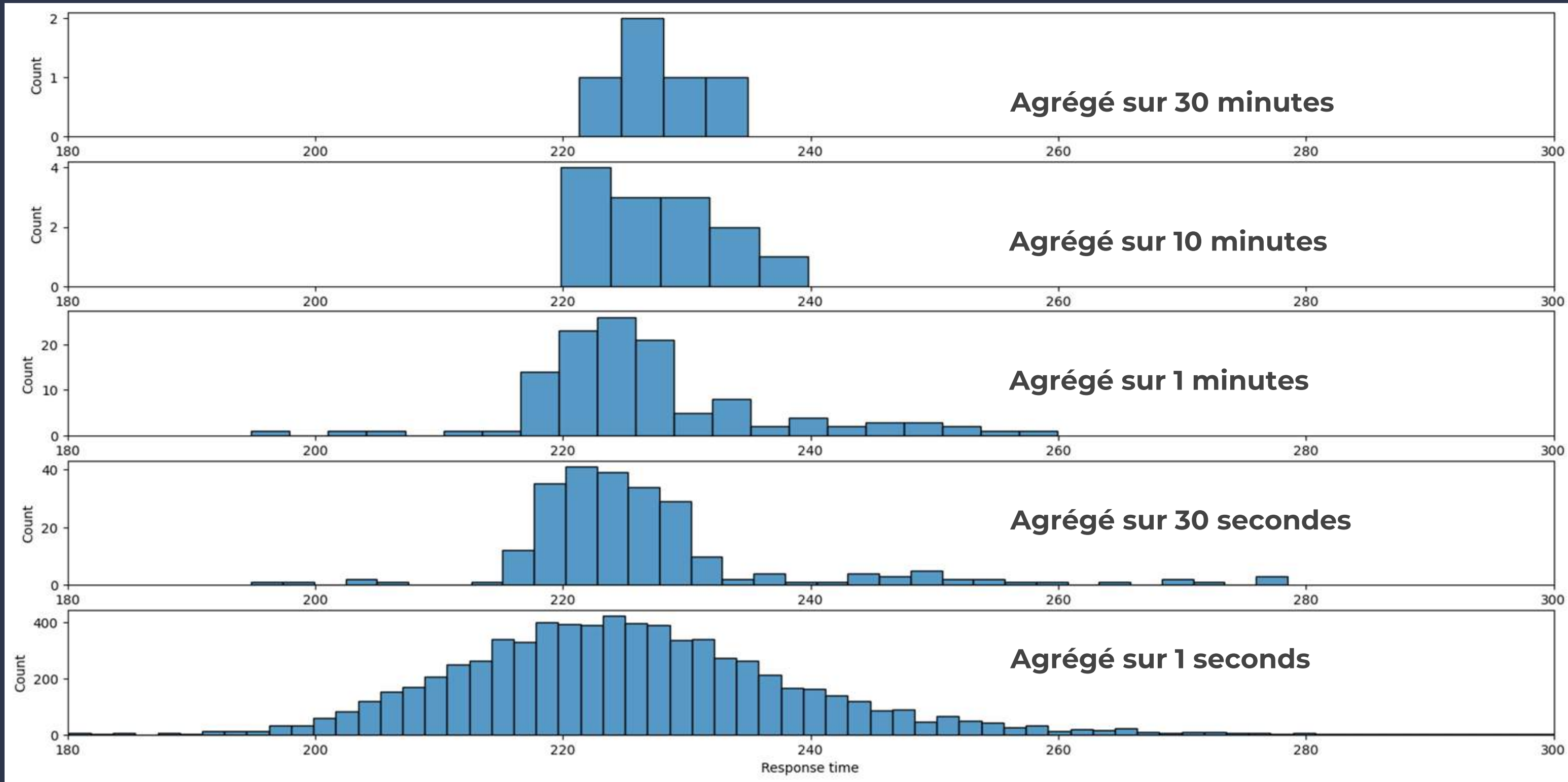
Ici, nous pouvons voir :

- **De mauvais temps de réponse pour les requêtes en échec**
- **De meilleurs temps de réponse pour les requêtes en succès**

Maîtriser l'affichage des statistiques



Pouvoir choisir les paramètres



Ajouter des statistiques

JOURNÉE
FRANÇAISE
DES TESTS
LOGICIELS

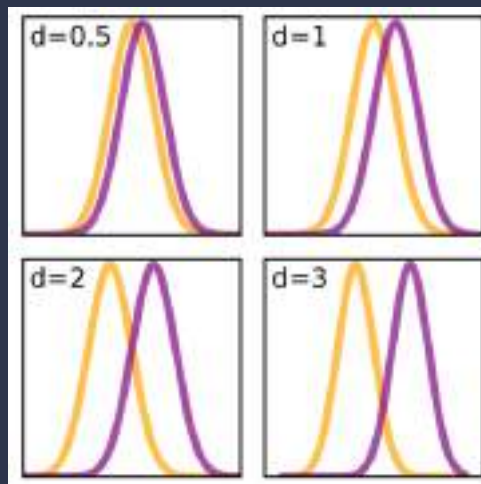


Exemple : d de Cohen pour prioriser les analyses lors de la comparaison de 2 tests

Element	Median (delta with previous test)
TR02_***	-187.0 (-8.03%)
TR03_***	3.0 (0.24%)
TR04_***	6.0 (1.29%)
TR05_BIS_***	9.0 (1.91%)
TR05_***	0.0 (0.0%)
TR06_BIS_***	-6.0 (-2.37%)
TR06_***	-8.0 (-1.08%)
TR07_BIS_***	-7.0 (-0.98%)
TR07_***	-109.0 (-2.47%)
TR07_TER_***	-2.0 (-4.55%)
TR08_***	-10.0 (-0.46%)
TR09_***	1.0 (0.48%)
TR10_***	-5.0 (-1.02%)
TR_0X_***	-7.0 (-2.97%)
TR_0X_TER_***	0.0 (0.0%)

Avons nous des régressions ?

Effect size	d
Very small	0.01
Small	0.20
Medium	0.50
Large	0.80
Very large	1.20
Huge	2.0



d de Cohen :

Element	Cohen's d
TR02_***	-0.58
TR03_***	-0.01
TR04_***	0.12
TR05_BIS_***	0.04
TR05_***	-0.01
TR06_BIS_***	-0.05
TR06_***	-0.02
TR07_BIS_***	-0.10
TR07_***	-0.09
TR07_TER_***	-0.17
TR08_***	-0.04
TR09_***	0.04
TR10_***	-0.02
TR_0X_***	-0.07
TR_0X_TER_***	0.01



Analyses en
priorité la
transaction TR02

La taille d'effet est une mesure de la force de l'effet observé d'une variable sur une autre et est fréquemment utilisée pour comparer deux distributions. Quand ces distributions sont normales un indicateur populaire est le **d de Cohen**. Malheureusement ce dernier ne donne pas de bons résultats lorsqu'il est appliqué à des distributions non-normales

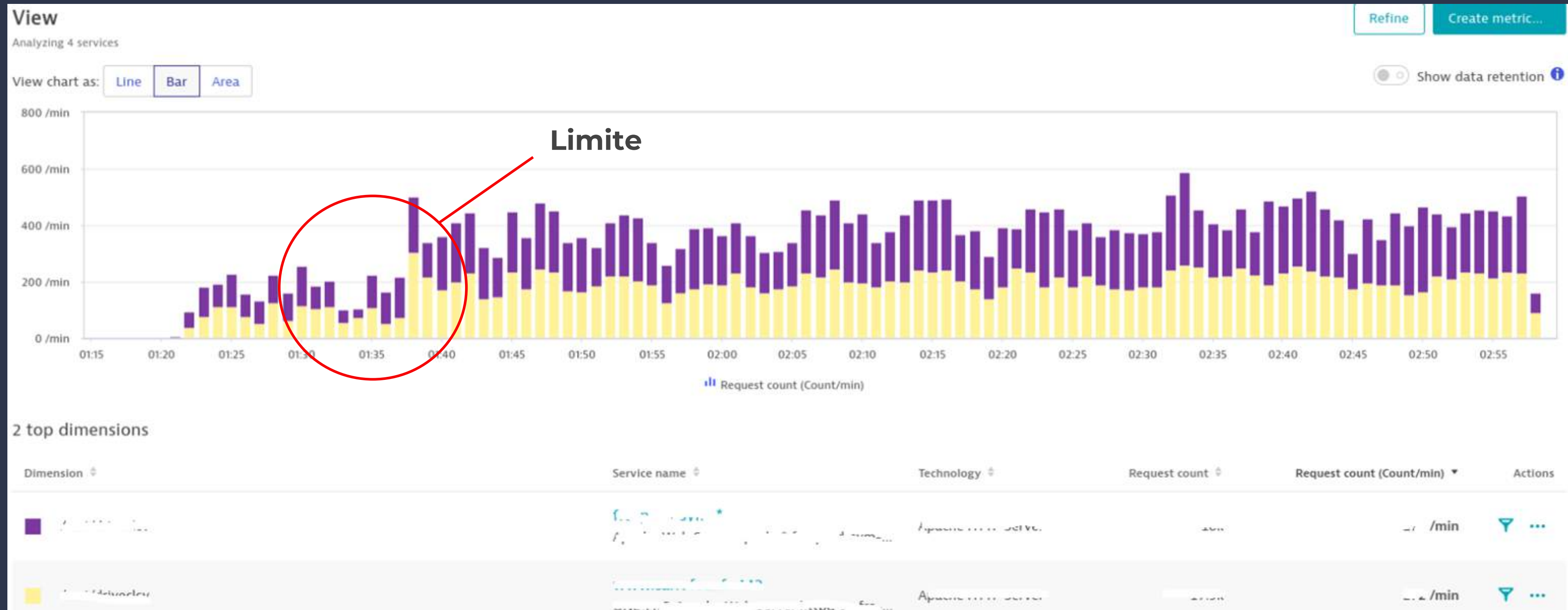
Ajouter des algorithmes

JOURNÉE
FRANÇAISE
DES TESTS
LOGICIELS



Exemple : trouver les limites d'une application avec du Change Point Detection

Aujourd'hui, nous vérifions manuellement la limite



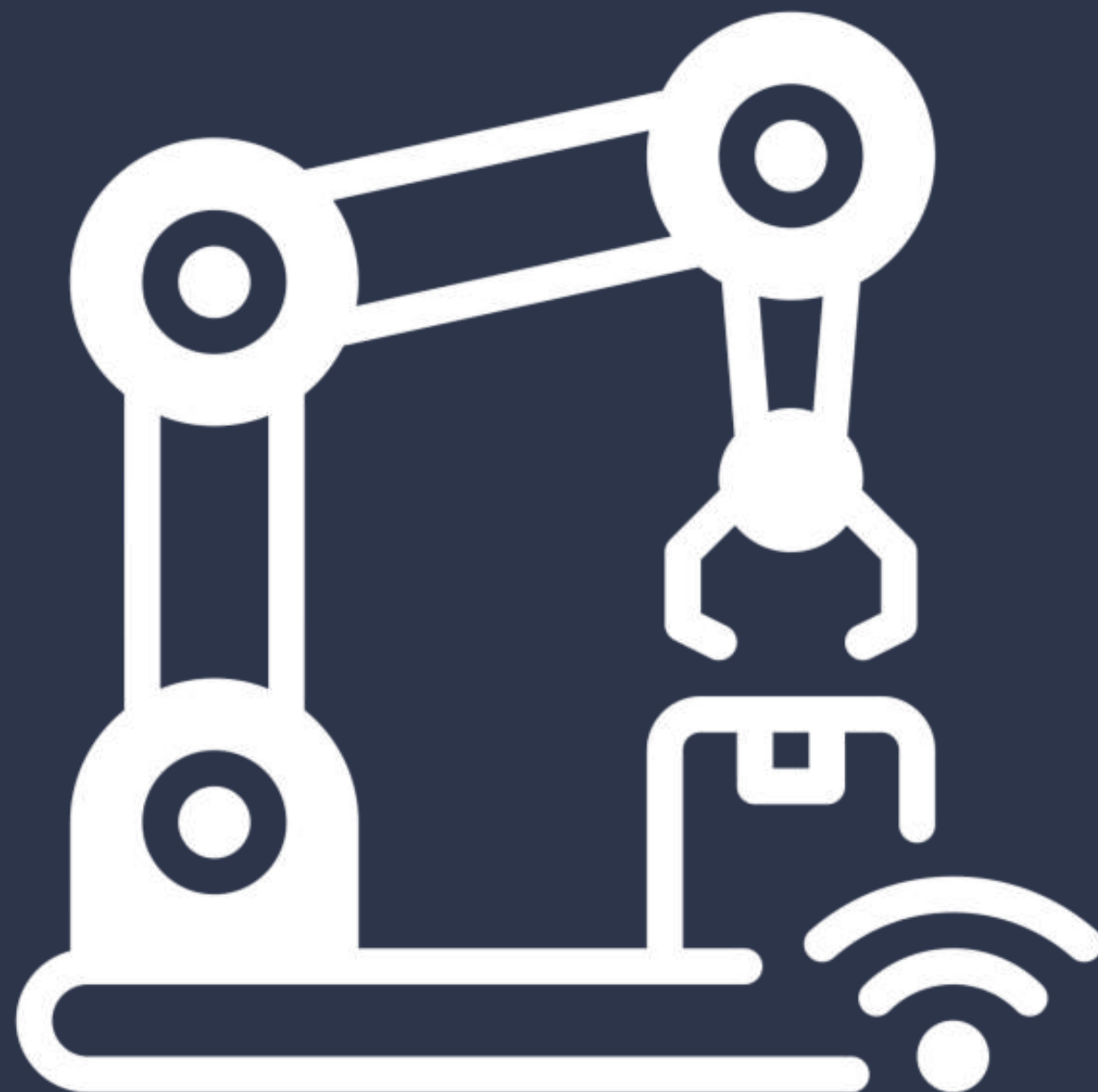
Notre solution

JOURNÉE
FRANÇAISE
DES TESTS
LOGICIELS



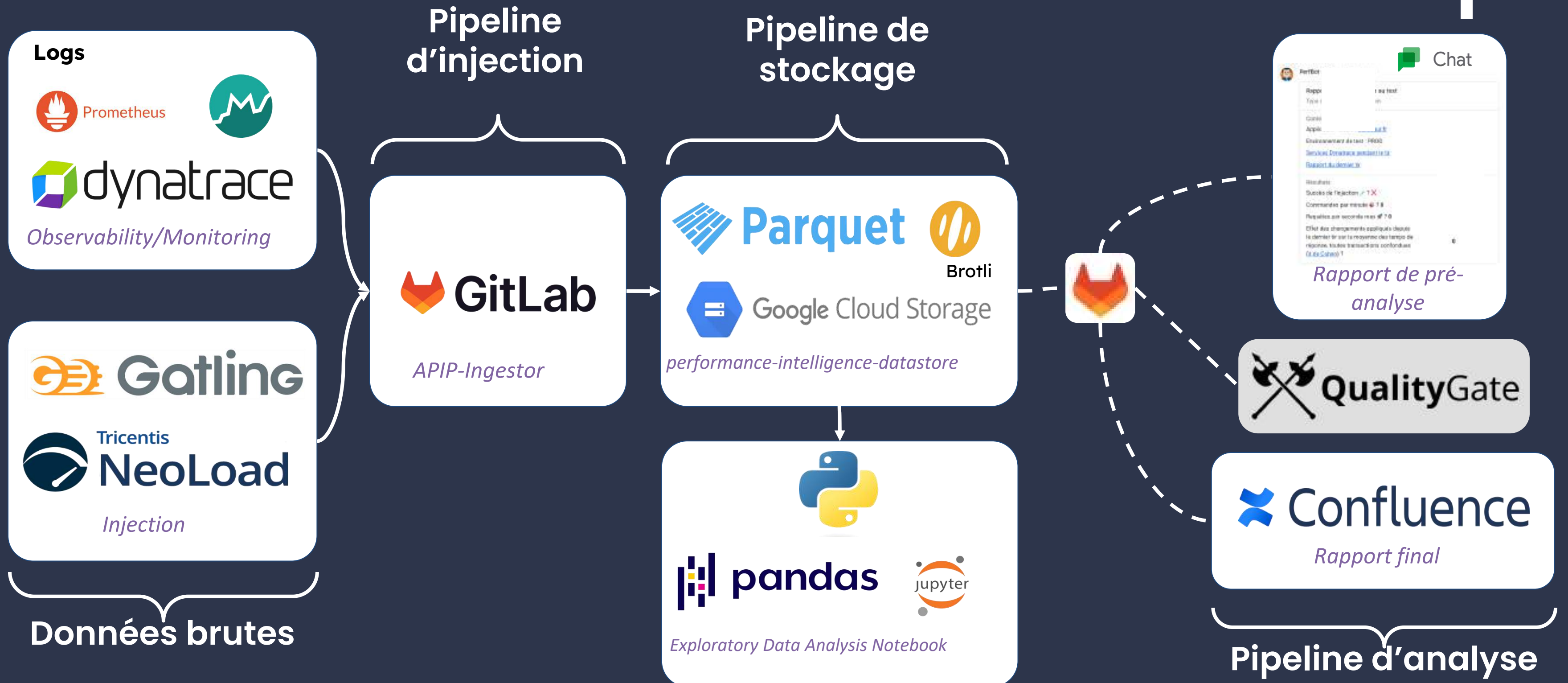
Automatiser le maximum de tâches

JOURNÉE
FRANÇAISE
DES TESTS
LOGICIELS



Utiliser les outils BI/AI/ML standards pour disposer d'un pipeline (ETL) permettant d'analyser les données brutes de nos tests de charge

JOURNÉE
FRANÇAISE
DES TESTS
LOGICIELS



Besoin d'une norme pour la supervision



**Tout le monde
examinera la même
métrique quel que soit
l'application**

Mise en place du tableau de bord RED

Rate (R) : Le nombre de requêtes par seconde
Errors (E) : Le nombre de requêtes en erreur
Duration (D) : Le temps de réponse de la requête

Element	Rate		Error		Duration			
	Count/s		Error rate		median		Perc95	
TR03_Chat_Serve		2.567		0.64		1217		1490
TR04_Chat_Serve		2.568		0.02		461		583
BIS_L_Chat_Serve		3.746		0.00		455		1042
TR05_Chat_Serve		15.427		0.00		1473		2050
BIS_L_Chat_Serve		3.650		0.00		241		436

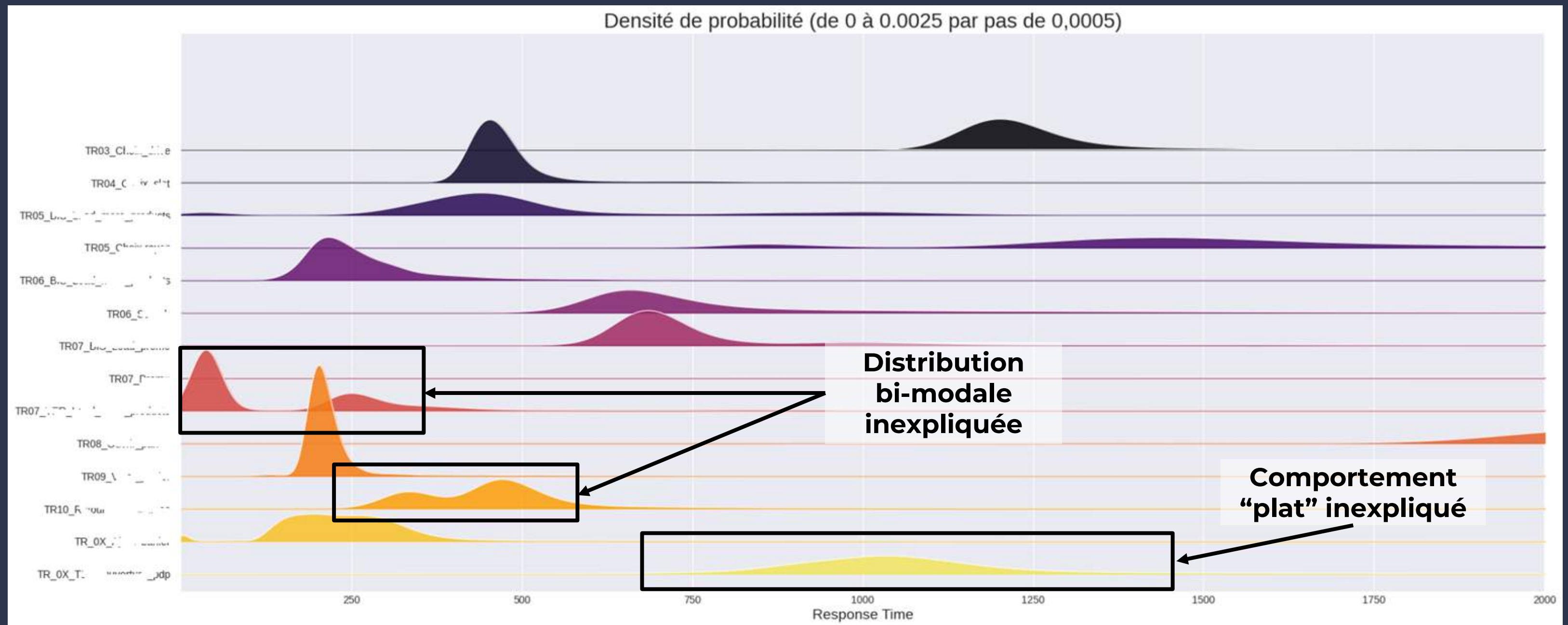
En un coup d'oeil, nous pouvons voir :

- Rate/Error/Duration dans un format sparkline de série chronologique
- Des éléments avec des problèmes d'erreur/performance/débit
- Si une erreur affecte les temps de réponse
- Si le débit affecte les temps de réponse
- Etc.

Mise en place du tableau de bord de distribution

Graphique de distribution avec :

- Paramètre de granularité
- Paramètre du sous-ensemble de données
- Granularité automatique si nécessaire (algo *Kernel density estimation*)



Accélérer l'analyse

JOURNÉE
FRANÇAISE
DES TESTS
LOGICIELS

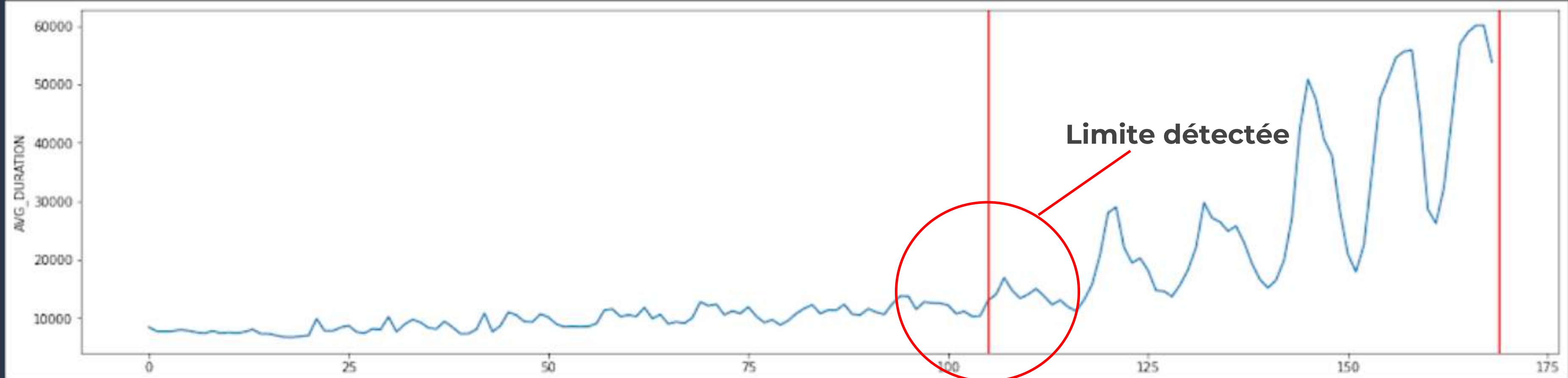


**Tout le monde examinera
la même métrique quel
que soit le service**

Mise en place d'algorithmes de détection de ruptures pour aider à trouver les limites de l'application

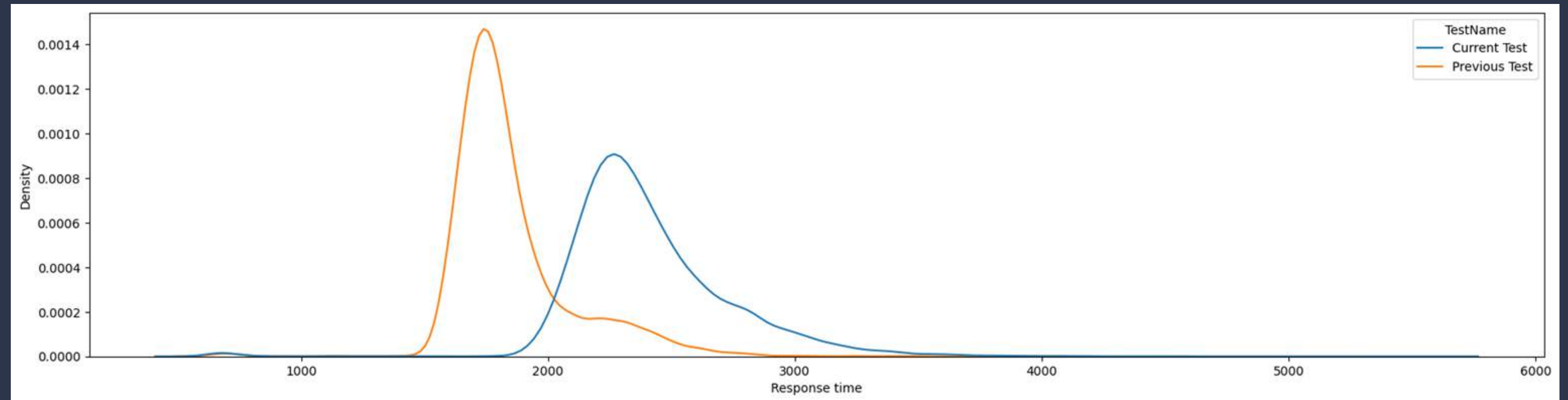
La détection de points de changements cherche à estimer le point où changent les propriétés statistiques d'une série d'observations

L'utilisation d'algorithmes de détection de points de changement comme le (PELT (Pruned Exact Linear Time) ou la segmentation binaire nous permet de détecter automatiquement les limites des applications sur plusieurs dimensions.



Comparer les distributions entre deux tests

Comment quantifier cette variation ?



Interprétation
visuelle

Indicateur
quantitatif

En théorie des probabilités et en théorie de l'information, la divergence de Kullback-Leibler (ou divergence K-L ou encore entropie relative) est une mesure de dissimilarité entre deux distributions de probabilités

0.12

Metric k8s sur l'auto scaling



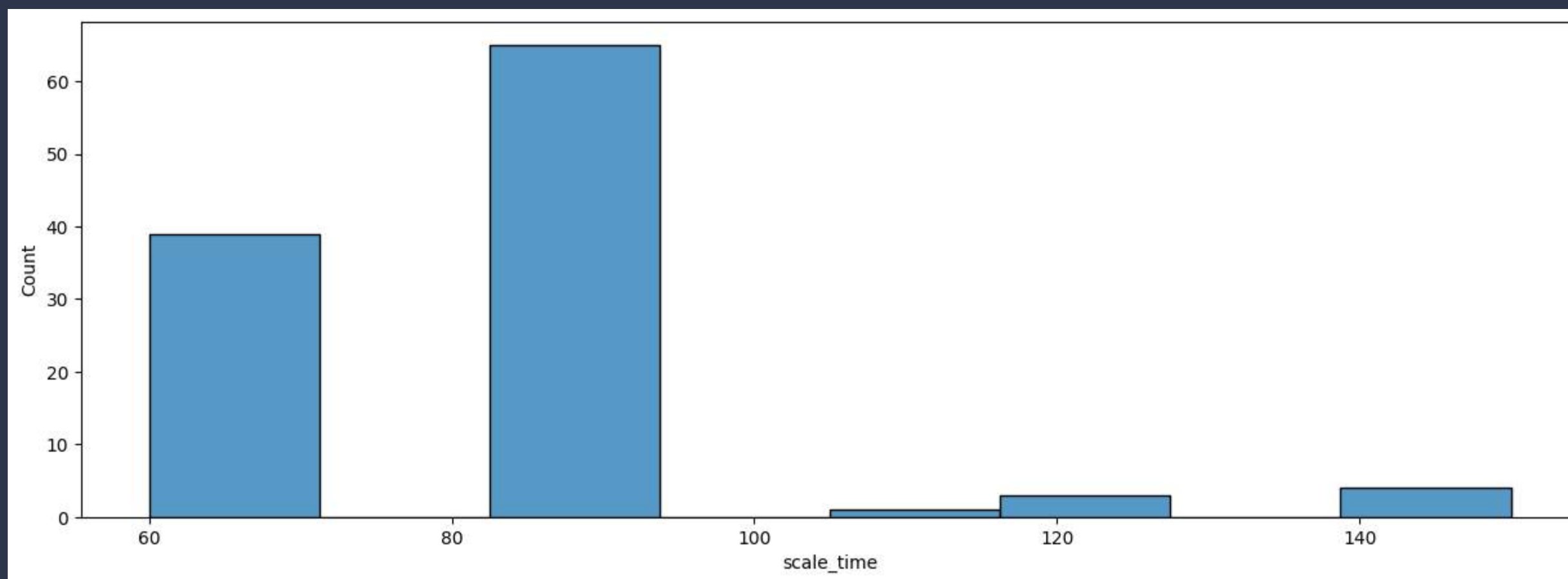
Prometheus

kube_pod_status_ready

Gauge

Describes whether the pod is ready to serve requests

```
kube_pod_status_ready{namespace="{namespace_name}", pod=~"{pod_name}", condition="false"}[{test_duration}:1s]
```



Rapport de pré analyse 1/2

JOURNÉE
FRANÇAISE
DES TESTS
LOGICIELS

PerfBot App 1:12 AM
test is starting - duration: 20min - target 100 req/s - nb_injecteur 4

PerfBot App 1:45 AM
- status: SUCCESS

PerfBot App 2:06 PM
test is starting - duration: 60min - target 1700 req/s - nb_injecteur 6

PerfBot App 17 min
- status: FAILURE

Tir de performance -Lancement_manuel
Résultat de tir

Defined

Not_Defined

Informations du tir

22-02-2024 13:08:17 (GMT)

60 min

Statut **FAILURE**

Requêtes OK 5858271 req	Requêtes KO 729 req
Volumétrie 1513.95 req/s	95th percentile 198 ms

[View Gatling Report](#) [View Jenkins Logs](#)

test is starting - duration : 100s - target 10 req/s - <https://gitlab.com/>

PerfBot App 26 min

Tir de performance
Résultat de tir

Informations du tir

2024-05-06 14:34:52 (GMT)

113 s

Statut **FAILURE**

[Services Dynatrace pendant le tir](#)

Requêtes OK 4600 req	Requêtes KO 44 req	
VOLUMÉTRIE 41.05 req/s	MÉDIANE 66 ms	95E CENTILE 112 ms

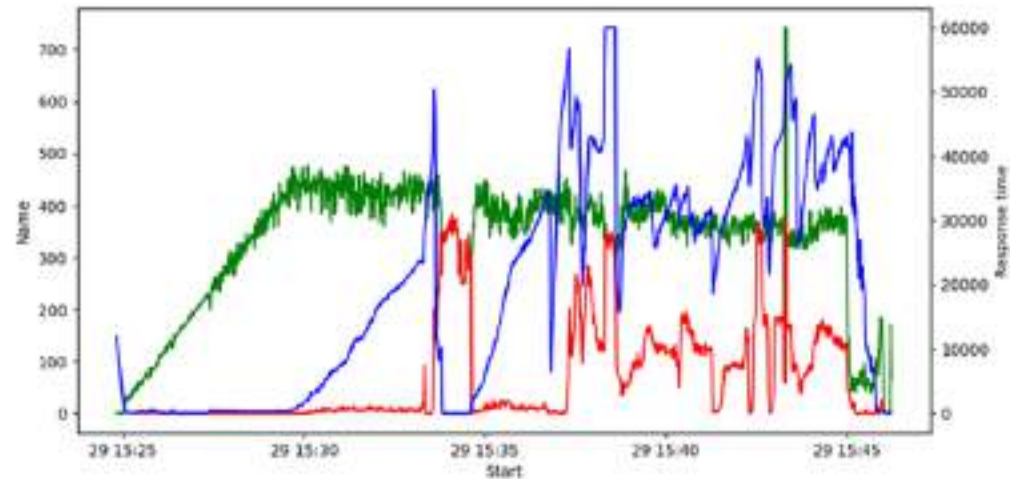
[Gatling Report](#) [Maven Execution Log](#)

Pre analysis report

Rapport de pré analyse 2/2

Pre analysis report for {{ ARIA

Rate, Error, Duration (RED) - Overview



RED - Sparkline by endpoint

Element	Rate		Error		Duration	
	Count/s		Error rate		median	Perc95
v2/slots/{slot_id}/baskets/{basket_id}	20.32		16.43		13667.3	60000.0
v2/slots/{slot_id_booked}/orders/{order_id}	14.51		13.81		8121.0	60000.0
/generate-perf-dataset	0.00		0.00		11921.0	11921.0
v2/slots/promove	3.38		25.79		25636.3	60000.0
v2/slots/{slot_id}	60.00		15.94		12687.5	60000.0
v2/slots/express	3.38		20.48		13013.3	60000.0
v2/slots	103.37		22.30		17365.3	60000.0
v2/service/{facility_service_id}/slots	94.27		24.01		17188.1	60000.0
v2/slots/{slot_id_booked}/baskets/{basket_id}/refresh	16.98		14.52		9832.0	60000.0
v2/slots/{slot_id_booked}	12.50		11.35		4764.0	60000.0

RED - Comparison with previous test

Endpoint/Method	Count	Count/s	Errors	Error rate	median	Perc95
v2/slots/{slot_id}/baskets/{basket_id}	-7440 (-22.16%)	-7.0 (-25.62%)	4104 (3994.28%)	16.34 (5160.05%)	11793.5 (629.82%)	48613.45 (426.94%)
v2/slots/{slot_id_booked}/orders/{order_id}	-14494 (-43.72%)	-12.47 (-46.22%)	2236 (642.53%)	12.6 (1219.25%)	6351.0 (358.80%)	48785.7 (435.17%)
/generate-perf-dataset	0 (0.0%)	0.0 (-4.44%)	0 (0.0%)	0.0 (0.0%)	-869.0 (-4.79%)	-869.0 (-4.79%)
v2/slots/promove	0 (0.0%)	-0.16 (-4.44%)	1122 (0.0%)	23.79 (0.0%)	14924.3 (139.33%)	36415.25 (154.4%)
v2/slots/{slot_id}	-21504 (-21.8%)	-20.29 (-25.27%)	12073 (5462.9%)	15.71 (7013.68%)	11184.5 (805.6%)	49630.0 (478.59%)
v2/slots/express	0 (0.0%)	-0.16 (-4.44%)	821 (1172.86%)	18.87 (1172.86%)	11883.0 (976.83%)	50160.42 (509.78%)
v2/slots	0 (0.0%)	-4.8 (-4.44%)	28944 (4206.98%)	21.78 (4206.98%)	14133.5 (437.13%)	46367.0 (340.11%)
v2/service/{facility_service_id}/slots	0 (0.0%)	-4.38 (-4.44%)	28444 (4349.24%)	23.47 (4349.24%)	13997.5 (438.66%)	46493.95 (344.25%)
v2/slots/{slot_id_booked}/baskets/{basket_id}/refresh	-11834 (-34.76%)	-10.26 (-37.66%)	2860 (922.58%)	13.39 (1467.53%)	8542.0 (862.17%)	30040.4 (302.43%)
v2/slots/{slot_id_booked}	-16730 (-51.0%)	-14.2 (-53.17%)	1654 (972.94%)	10.83 (2089.46%)	3356.0 (238.35%)	40543.3 (473.79%)

Overview
rapide des
métriques
RED

Tableau
sparkline RED

Tableau
comparatif tir
n-1

ROI chez Carrefour

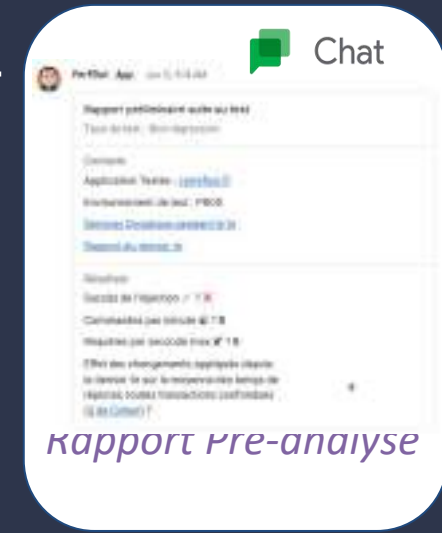
JOURNÉE
FRANÇAISE
DES TESTS
LOGICIELS



ROI : Diminution du temps d'analyse de nos tests de charge pour carrefour.fr

JOURNÉE
FRANÇAISE
DES TESTS
LOGICIELS

Détection de la limite + RED pour un **rapport 10min après le test**



Analyze
Performance
Intelligence Pipeline



Shift right: test au limite en production un fois par mois



Présentation des résultats + recommandations + solutions potentielles en **moins de 48h au lieu de 5 jours**



Analyse finalisée + Tickets JIRA



Coopérer avec les développeurs et les ops pour une analyse approfondie



Analyse continue par le **reste de l'équipe facilité par la norme de supervision**

Antonio
GOMES RODRIGUES

Raphael
DEBAN

Accélérer le Time To Market : Innovation et qualité grâce à
l'Analyse Performance Intelligence Pipeline

JOURNÉE
FRANÇAISE
DES TESTS
LOGICIELS



MERCI DE VOTRE ÉCOUTE
N'oubliez pas de voter



JOURNÉE
FRANÇAISE
DES TESTS
LOGICIELS

VOTEZ POUR LA
MEILLEURE
PRÉSENTATION

