

Centre de Calcul
de l'Institut National de Physique Nucléaire
et de Physique des Particules

Informatique Décisionnelle au CC-IN2P3

Cathala Dominique & Evesque Cécile

- Introduction
- Le décisionnel : contexte
- La data visualisation : exemples
- Les données : extraction-traitement
- L'architecture mise en place
- Conclusion

Le décisionnel au Centre de Calcul de l'IN2P3

Des données volumineuses et une infrastructure complexe :



The screenshot shows the CCIN2P3 website homepage. At the top, there is a navigation menu with the following items: QUI SOMMES-NOUS, NOS SERVICES, DOMAINES D'APPLICATION, ESPACE UTILISATEURS, ACTUALITÉS, and CONTACT. Below the menu is a large hero image of a server room with the text "Centre de Calcul de l'IN2P3" and "Calcul scientifique et traitement de données". A "En savoir plus" button is located below the hero image. Below the hero image are three service icons: "Traitement de données" (gear icon), "Outils d'aide à la production scientifique" (document icon), and "Hébergement" (server rack icon). At the bottom, there is a section titled "Une infrastructure performante" with four statistics: 52000 Calculs en parallèle, 1700 mètres carrés, 340 Pétaoctets, and 80 Expériences.

Le décisionnel au Centre de Calcul de l'IN2P3

Des données volumineuses et une infrastructure complexe :

Comment donner du sens aux données pour répondre aux besoins des services et utilisateurs



The screenshot shows the homepage of the CCIN2P3 website. At the top, there is a navigation menu with links: QUI SOMMES-NOUS, NOS SERVICES, DOMAINES D'APPLICATION, ESPACE UTILISATEURS, ACTUALITÉS, and CONTACT. Below the menu is a large hero image of a server room with the text "Centre de Calcul de l'IN2P3" and "Calcul scientifique et traitement de données". A button "En savoir plus" is visible. Below the hero image are three service icons: "Traitement de données" (gear icon), "Outils d'aide à la production scientifique" (document icon), and "Hébergement" (server rack icon). At the bottom, a section titled "Une infrastructure performante" displays four statistics: 52000 Calculs en parallèle, 1700 mètres carrés, 340 Pétaoctets, and 80 Expériences.

Statistique	Valeur
Calculs en parallèle	52000
mètres carrés	1700
Pétaoctets	340
Expériences	80

Le décisionnel au Centre de Calcul de l'IN2P3

Des données volumineuses et une infrastructure complexe :

Comment donner du sens aux données pour répondre aux besoins des services et utilisateurs

- ❑ Démarche initiée en 2010
- ❑ Basée sur MS SQL Server & Tableau Software
- ❑ Pour les décideurs et les utilisateurs



The screenshot shows the homepage of the CCIN2P3 website. At the top, there is a navigation menu with links: QUI SOMMES-NOUS, NOS SERVICES, DOMAINES D'APPLICATION, ESPACE UTILISATEURS, ACTUALITÉS, and CONTACT. Below the menu is a large banner image of a server room with the text "Centre de Calcul de l'IN2P3" and "Calcul scientifique et traitement de données". A "En savoir plus" button is located at the bottom of the banner. Below the banner are three service icons: "Traitement de données" (gear icon), "Outils d'aide à la production scientifique" (document icon), and "Hébergement" (server rack icon). At the bottom, there is a section titled "Une infrastructure performante" with four statistics: 52000 Calculs en parallèle, 1700 mètres carrés, 340 Pétaoctets, and 80 Expériences.

Le décisionnel au Centre de Calcul de l'IN2P3

Des données volumineuses et une infrastructure complexe :

Comment donner du sens aux données pour répondre aux besoins des services et utilisateurs

- ❑ Démarche initiée en 2010
- ❑ Basée sur MS SQL Server & Tableau Software
- ❑ Pour les décideurs et les utilisateurs

Le décisionnel, c'est toute une histoire !

Storytelling « Raconte moi les données »

DataViz « Dessine moi les données »



Une infrastructure performante			
52000	1700	340	80
Calculs en parallèle	mètres carrés	Pétaoctets	Expériences

Fournir une solution permettant
de s'assurer que le CC répond correctement
aux besoins des expériences

Fournir une solution permettant
de s'assurer que le CC répond correctement
aux besoins des expériences

- Analyser nos données métiers complexes
- Identifier des métriques et KPIs (Key Performance Indicators)
- Interpréter et suivre nos données et activités

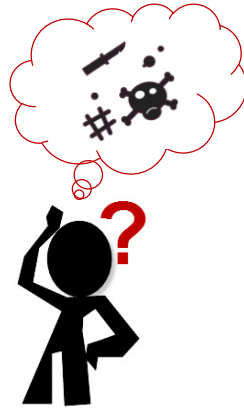
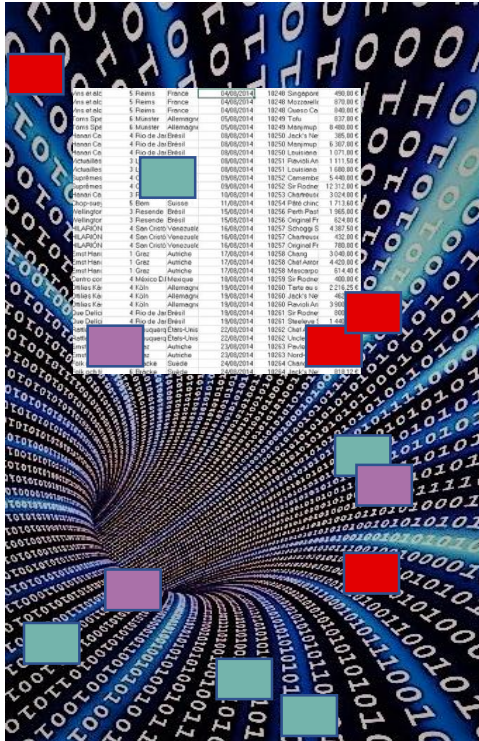
Fournir une solution permettant
de s'assurer que le CC répond correctement
aux besoins des expériences

- Analyser nos données métiers complexes
- Identifier des métriques et KPIs (Key Performance Indicators)
- Interpréter et suivre nos données et activités

➔ Aider à la prise de décision

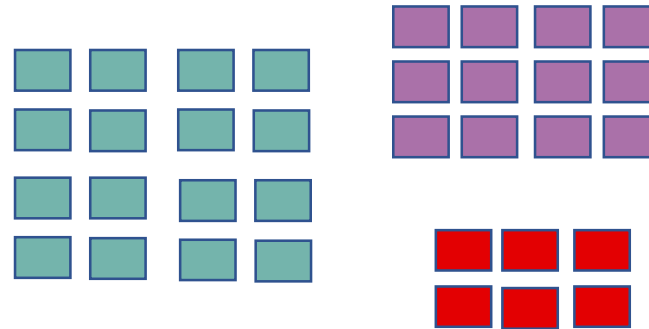
Sources de données

- Complexité
- Hétérogénéité

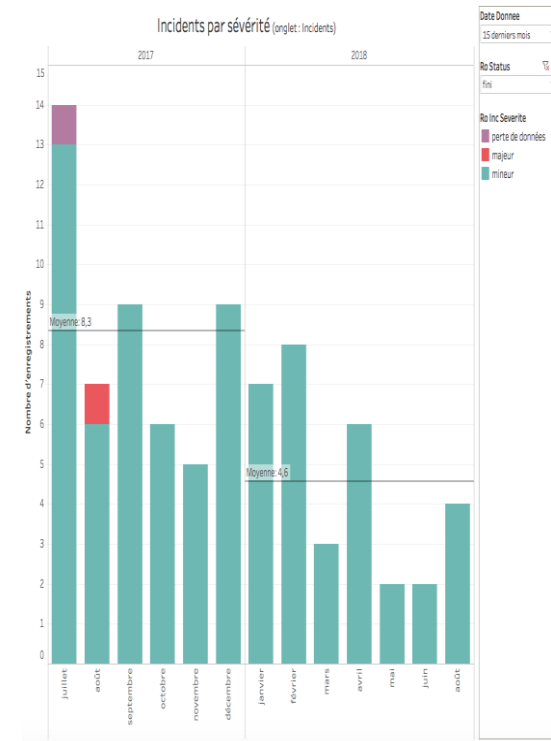


Data Warehouse

- Structuré
- Centralisé
- Actualisé



- ➔ Vues dynamiques
- ➔ Analyse visuelle
- ➔ Interprétation



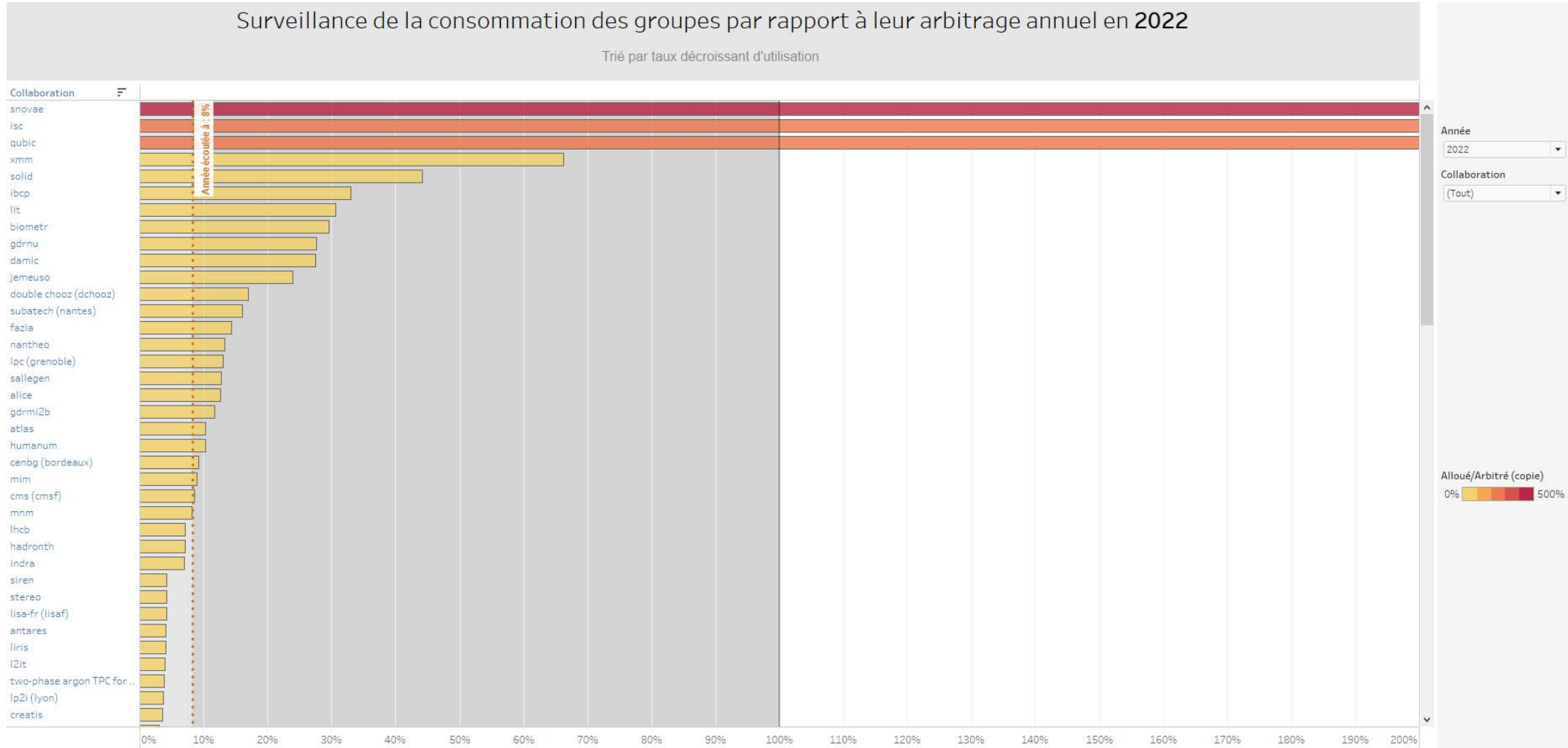
- **Traitement visuel rapide**
 - 13 millisecondes pour traiter une image
 - Les visuels traités 60000x plus vite / textes

- **Bonnes pratiques**
 - Un visuel compris par tous
 - Facilité et rapidité d'interprétation

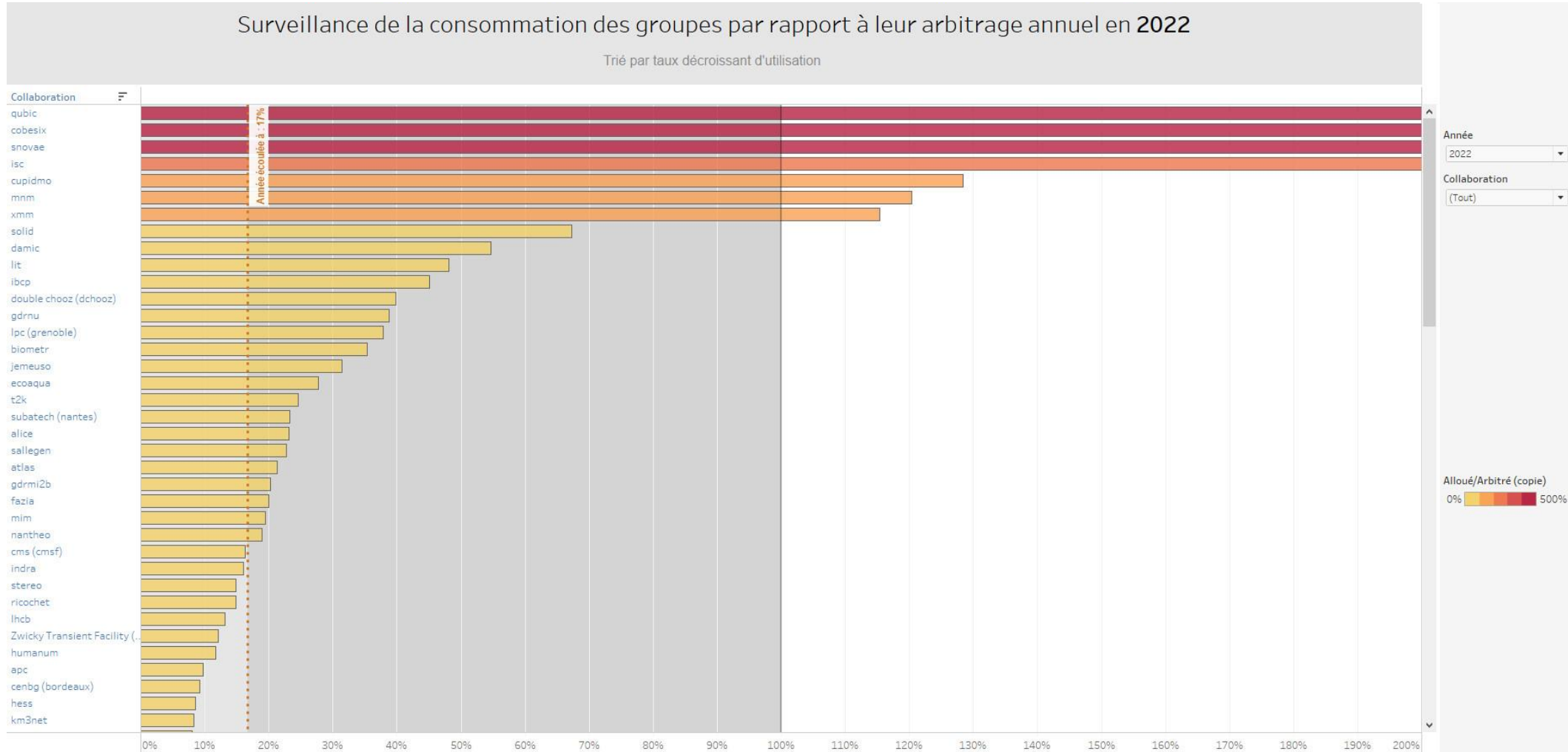
- **Bonne compréhension du besoin**
 - Accompagner & Comprendre
 - Identifier & Analyser

CF : <https://www.lebigdata.fr/pourquoi-datavis-cerveau>

Surveillance des dépassements de consommation



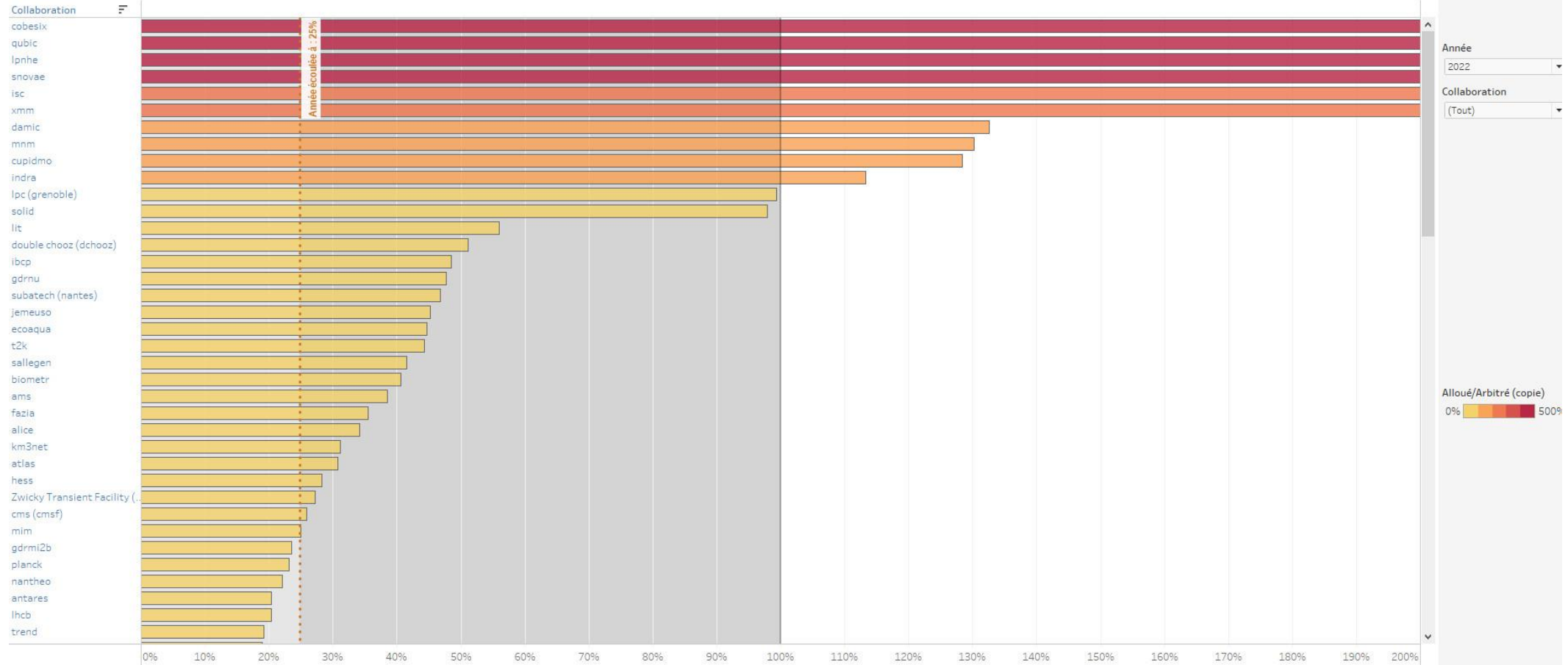
Surveillance des dépassements de consommation



Surveillance des dépassements de consommation

Surveillance de la consommation des groupes par rapport à leur arbitrage annuel en 2022

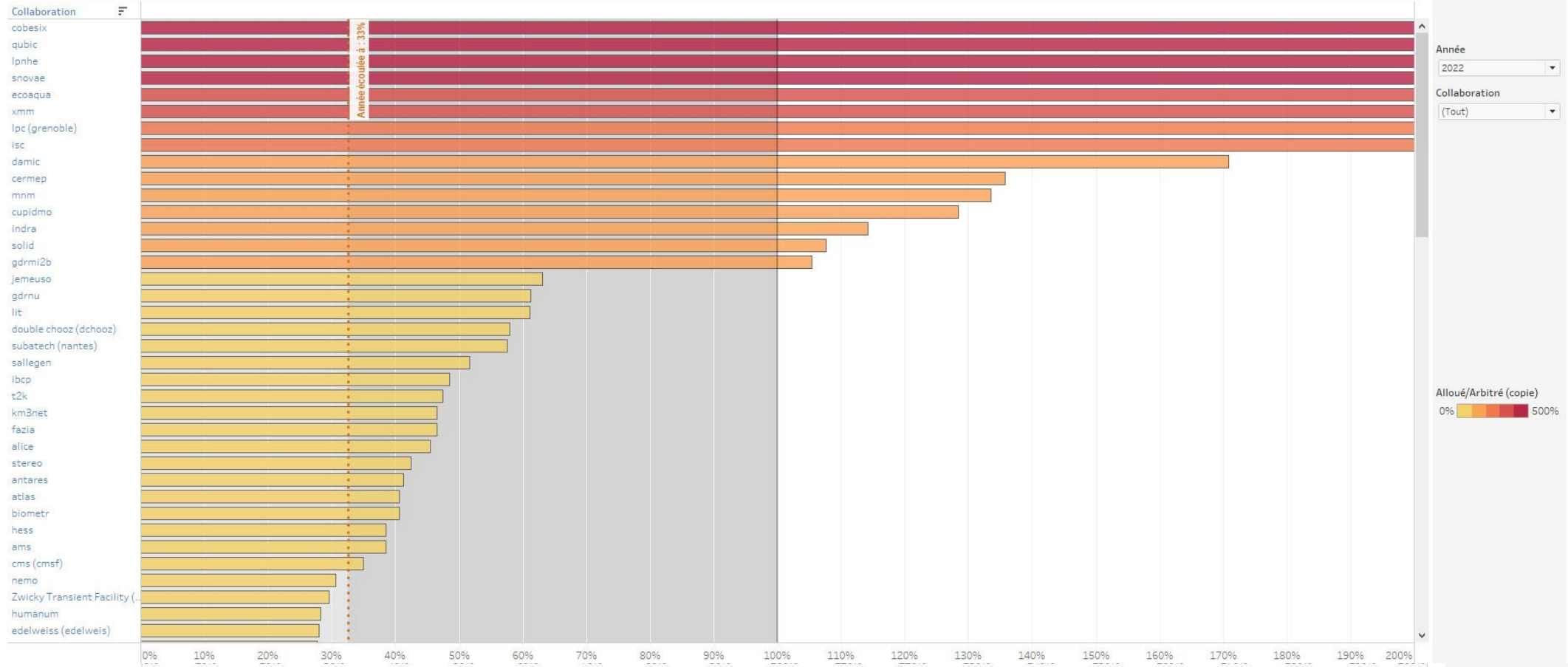
Trié par taux décroissant d'utilisation



Surveillance des dépassements de consommation

Surveillance de la consommation des groupes par rapport à leur arbitrage annuel en 2022

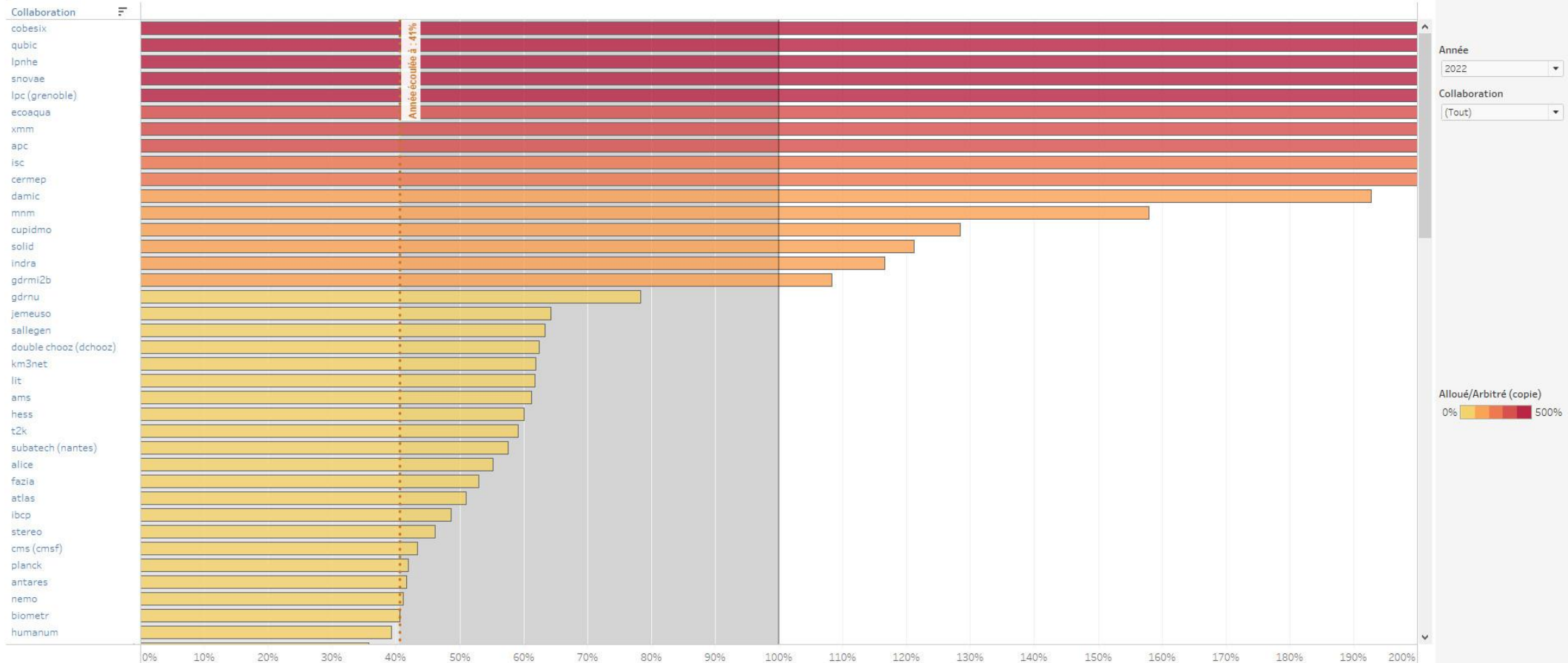
Trié par taux décroissant d'utilisation



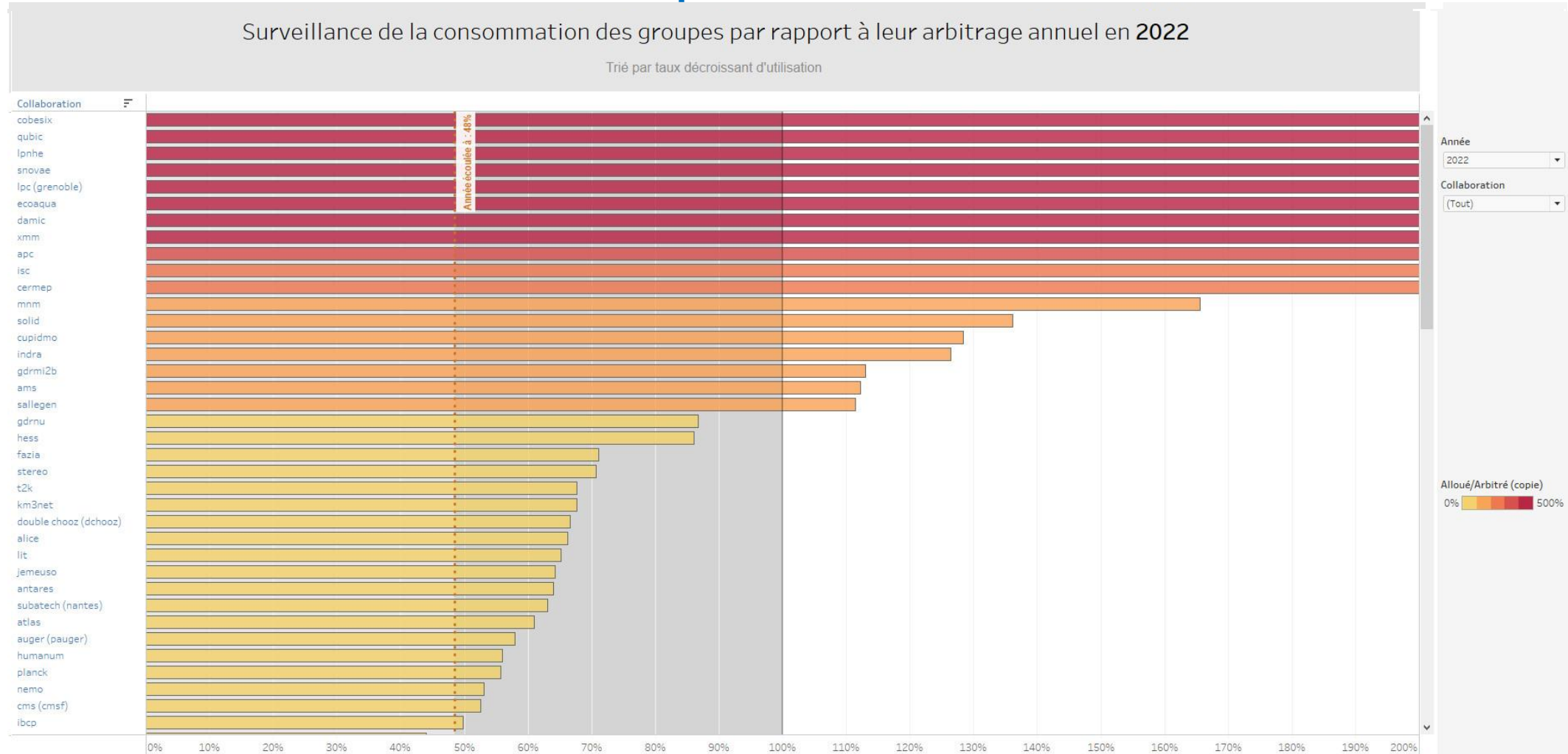
Surveillance des dépassements de consommation

Surveillance de la consommation des groupes par rapport à leur arbitrage annuel en 2022

Trié par taux décroissant d'utilisation



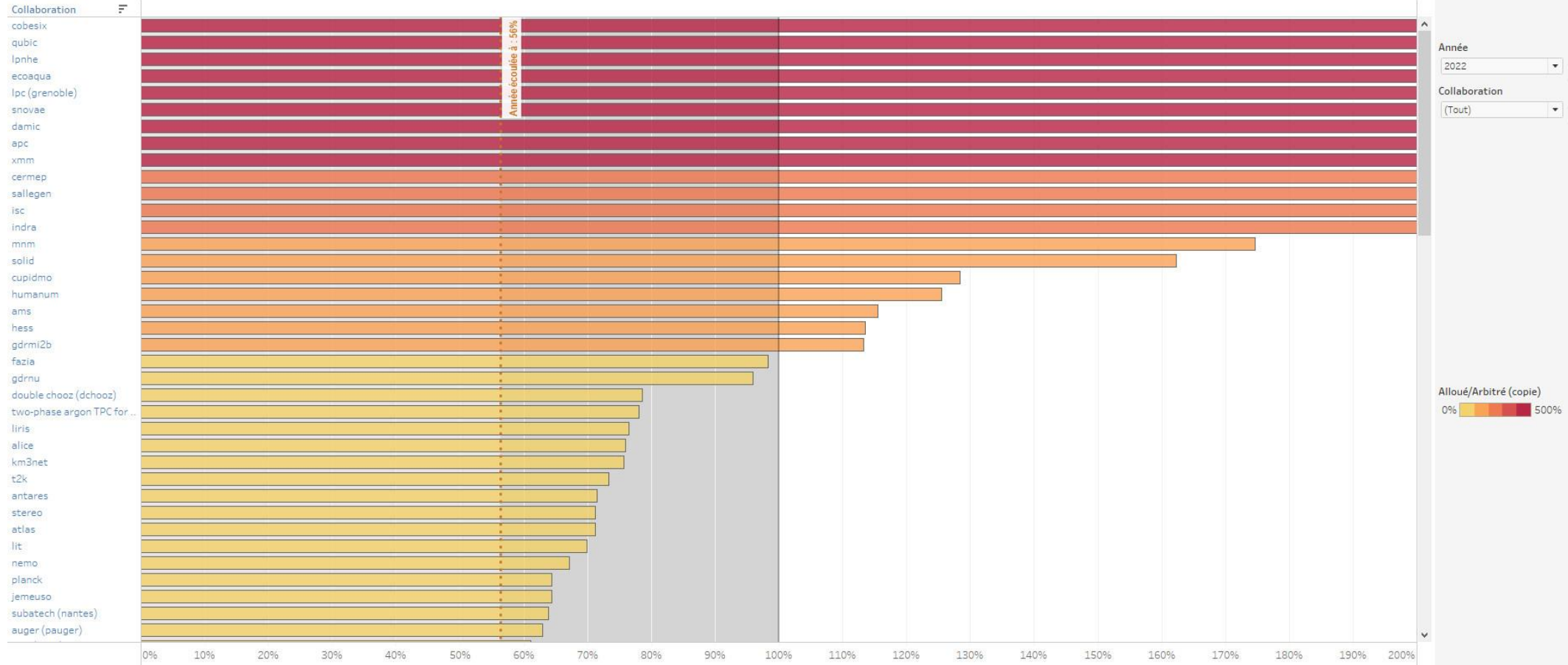
Surveillance des dépassements de consommation



Surveillance des dépassements de consommation

Surveillance de la consommation des groupes par rapport à leur arbitrage annuel en 2022

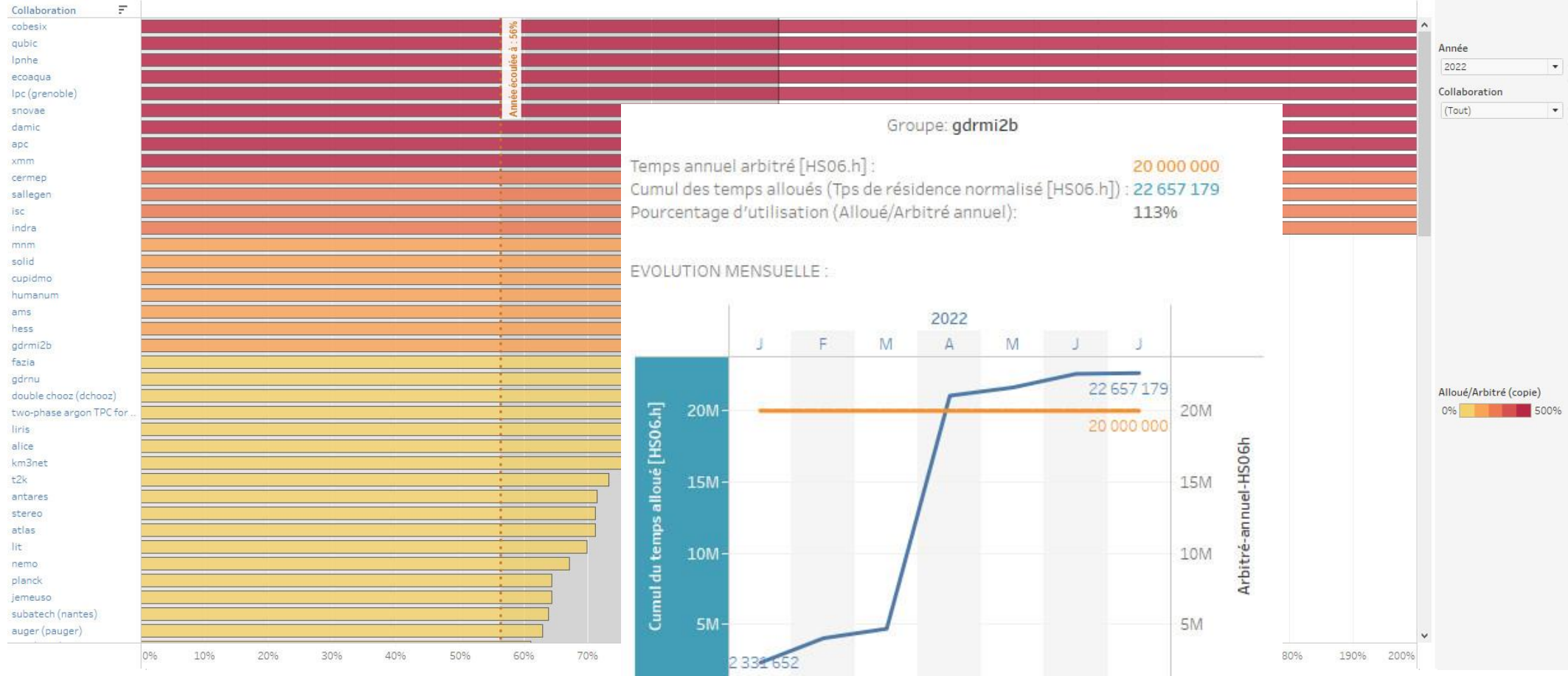
Trié par taux décroissant d'utilisation



Surveillance des dépassements de consommation

Surveillance de la consommation des groupes par rapport à leur arbitrage annuel en 2022

Trié par taux décroissant d'utilisation



Nos DataViz : localisation des serveurs

Villeurbanne2

2 442 machines dans cette salle

Dont 2 442 répondent aux filtres que vous avez sélectionnés

Données mises à jour le 16/09/2022 05:01:14



Filtres disponibles :

- Type
 - (Tout)
 - PHYSICAL
 - SWITCH
 - VM
- Rack
 - (Valeurs multiples)
- Service
 - (Tout)
- Realm
 - (Tout)
- PuppetRole
 - (Tout)
- PuppetInstance
 - (Tout)
- Status
 - (Tout)
- MachinePlatformName
 - (Tout)
- OSRelease
 - (Tout)
- Machine
 -
- SerialNumber
 -
- No Commande
 -
- Date Fin Garantie
 - 01/01/2001 28/07/2028

Nos DataViz : localisation des serveurs

Villeurbanne2

2 442 machines dans cette salle

Dont 441 répondent aux filtres que vous avez sélectionnés

Données mises à jour le 16/09/2022 05:01:14



Filtres disponibles :

Type

- (Tout)
- PHYSICAL
- VM

Rack

(Tout)

Service

(Aucun)

slurm

slurm

(Tout)

PuppetRole

(Tout)

PuppetInstance

(Tout)

Status

(Tout)

MachinePlatformName

(Tout)

OSRelease

(Tout)

Machine

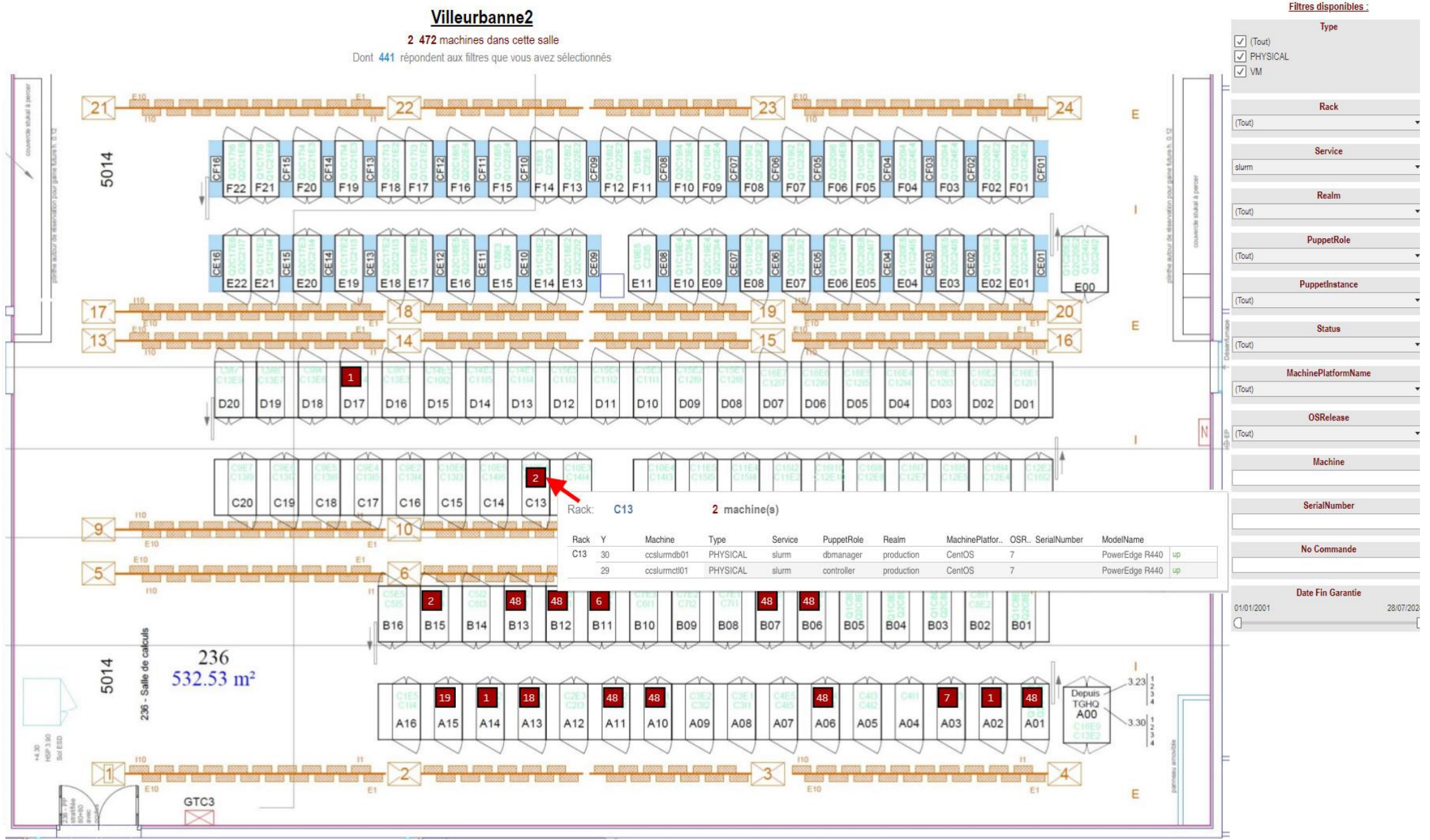
SerialNumber

No Commande

Date Fin Garantie

01/01/2001 28/07/2028

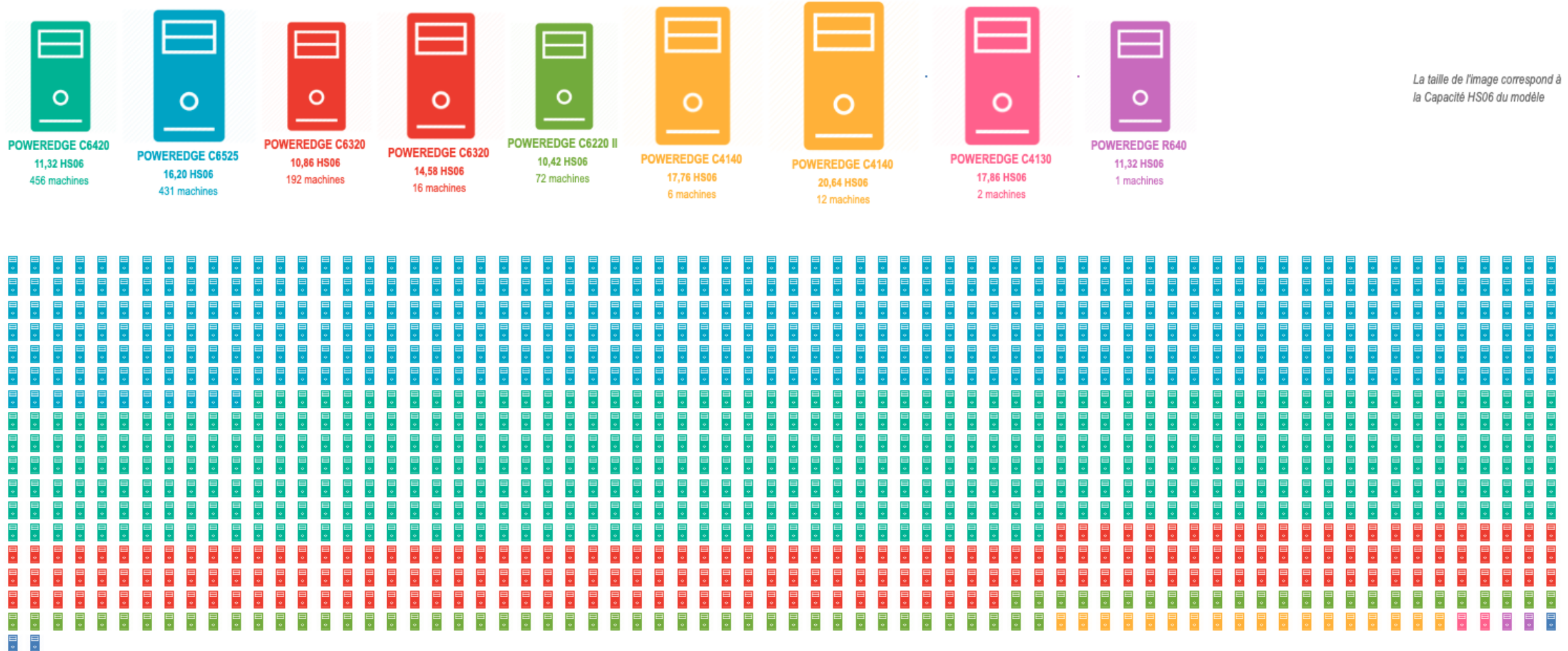
Nos DataViz : localisation des serveurs



Nos DataViz : parc des machines de calcul

24 août 2022

Modèles et capacité HS06 des 1 192 machines



Une grande hétérogénéité au CC

- De nombreux services dont 2 cœurs de métier :
 - le stockage
 - le calcul
- Diversité des outils
 - Pour le stockage : hpss, dCache, xRootD, iRods, Gpfs, Sps, Bases de données
 - Pour le calcul : HTCondor, Slurm, Grid Engine, BQS
- Diversité des sources de données
 - Fichiers plats : json, xml, csv, texte
 - Base de données : MySql, Oracle, PosgreSql

■ Outils :

- Propriétaires Microsoft : Outil de la suite SQL Microsoft
- OpenSource : Spark, SparkSQL, Airflow, Scala, Python

■ Processus ETL

- 3,5 millions de jobs traités par mois
- 10 000 extractions pour le stockage par mois
- 3 000 informations de machines par jour

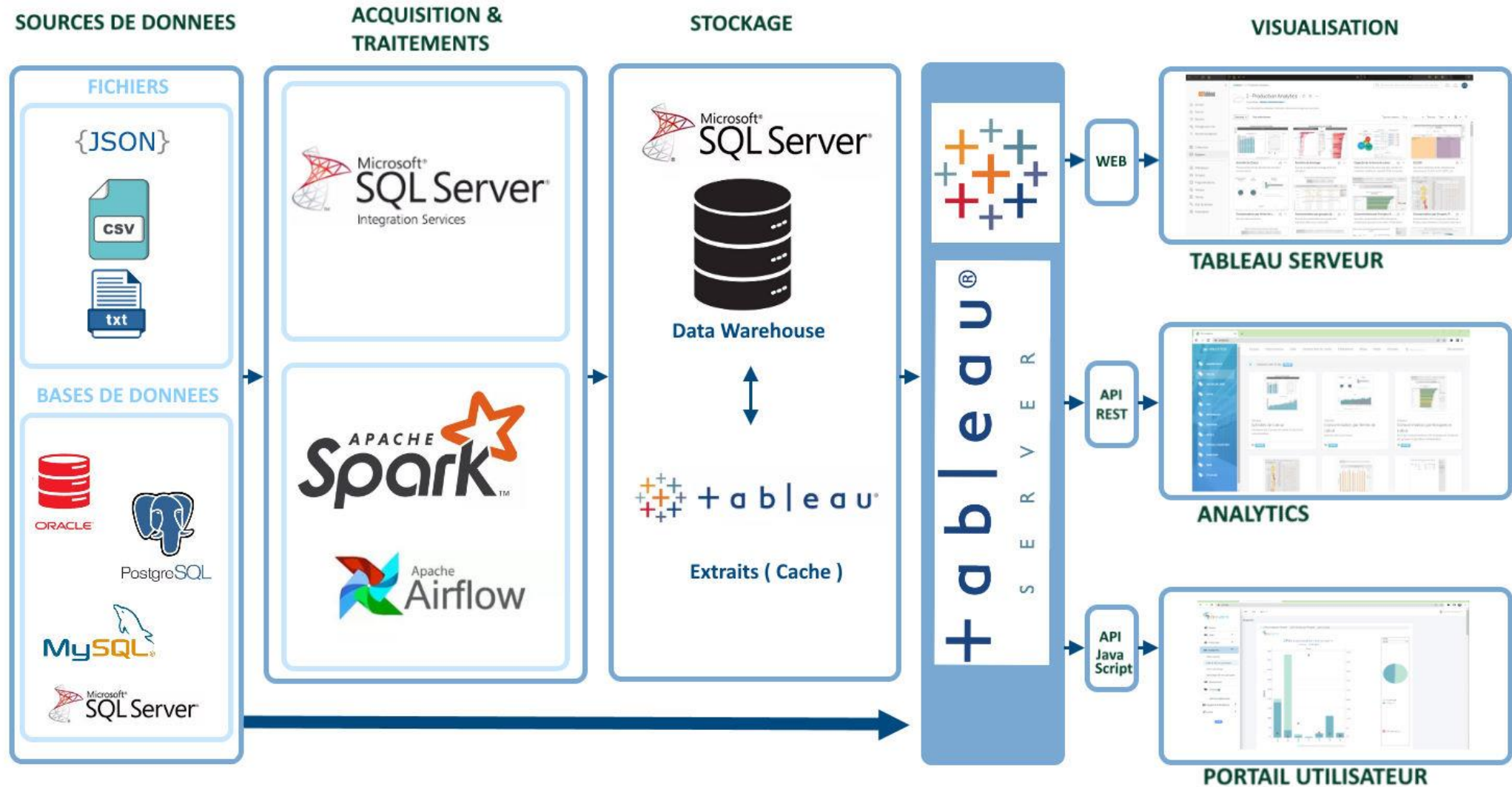
■ Extraction (Identification)

■ Traitement (Agrégation et Formatage)

■ Intégration (Datawarehouse, Vérification des données)

- 25 Go Datawarehouse

Architecture



Bilan rétrospectif du décisionnel au CC

Le décisionnel est aujourd'hui un référentiel incontournable pour le CC-IN2P3

- **80 tableaux de bord (400 visualisations)**
- **60 accès jour**

- **Nouveaux besoins régulièrement**
- **Evolution continue de notre SI**

- **Gain de temps sur le suivi de nos activités**

Merci pour votre écoute !



Des questions ?