

A decorative graphic on the left side of the slide, consisting of a network of light blue lines and circles on a dark grey background, resembling a circuit board or data flow diagram.

# PROJET 12

OPENCLASSROOMS: FORMATION  
INGÉNIEUR DATA

A decorative graphic on the left side of the slide, consisting of a network of light blue lines and circles, resembling a circuit board or a data network, set against a dark blue background.

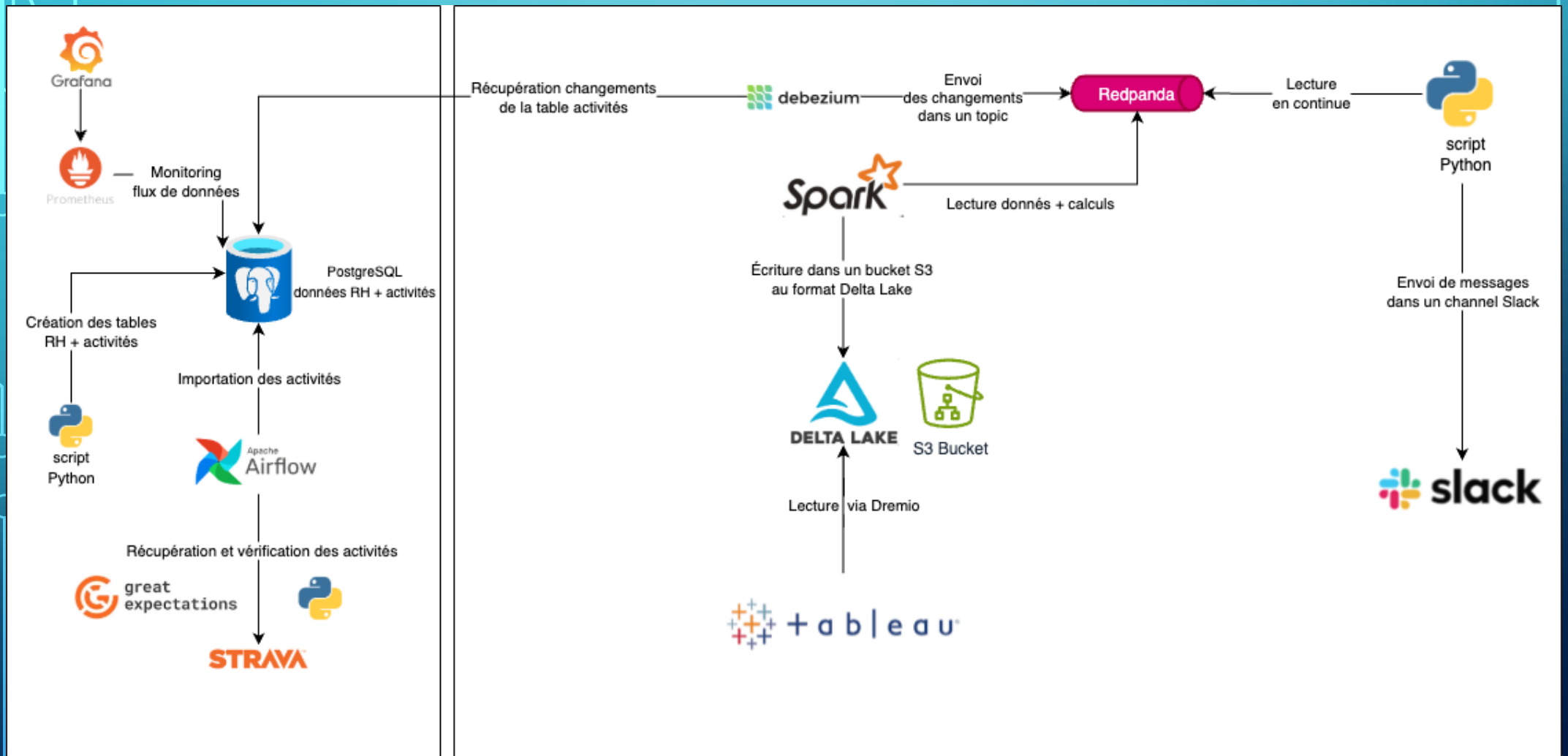
Contexte:

Data Engineer chez Sport Data Solution.

But du projet:

Développer une solution proposant des avantages à l'ensemble des collaborateurs s'ils ont une pratique du sport régulière.

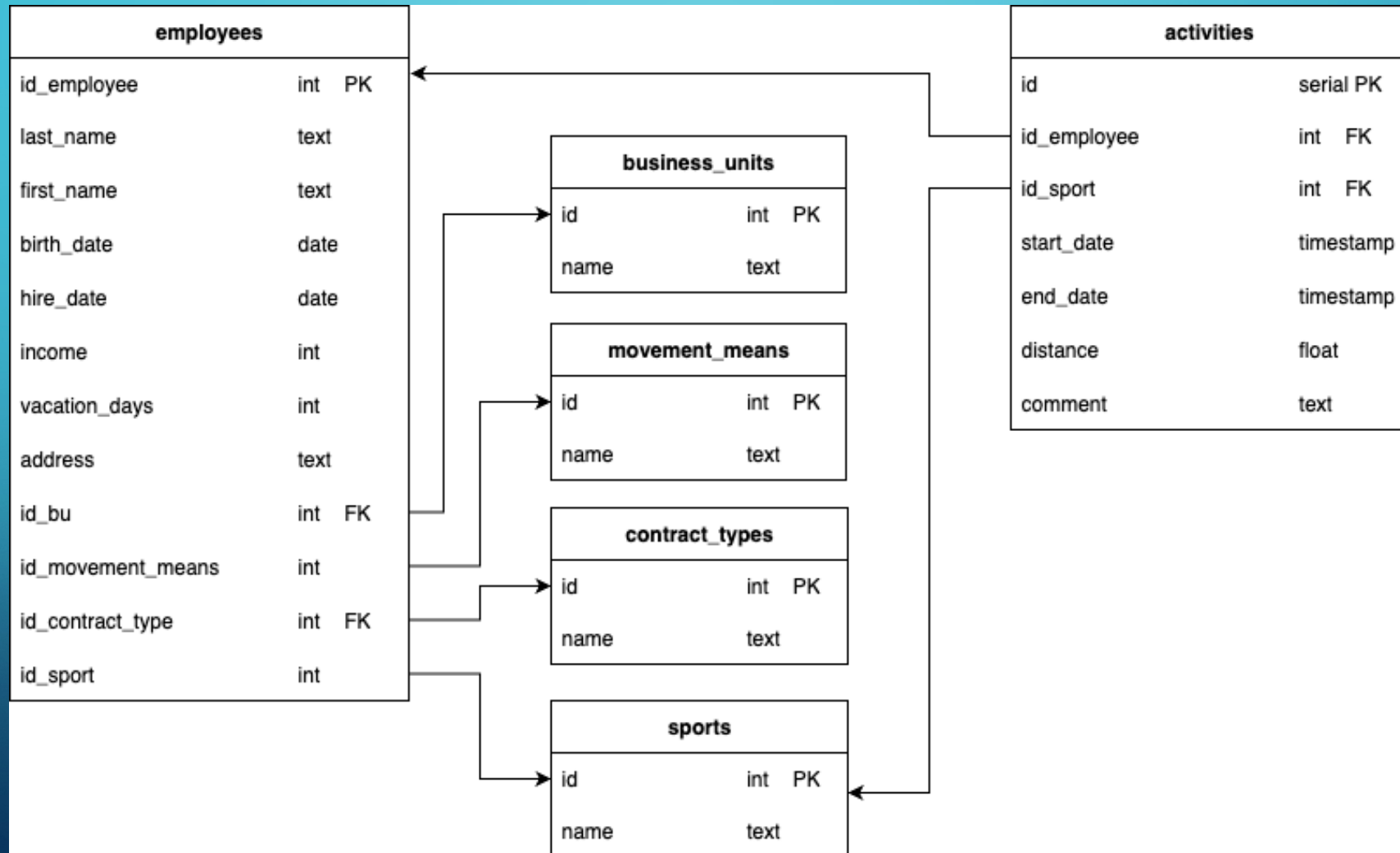
# ARCHITECTURE



## RÉPERTOIRE GITHUB

<https://github.com/nau81000/rewards>

# SCHÉMAS DE DONNÉES



## SERVICES DOCKER

Nom	Description
sport-data-db	Base PostgreSQL gérant les données de l'entreprise
init-sport-data-db	Script d'initialisation de la base de données
redpanda	Plateforme de streaming de données Redpanda
redpanda-console	Console du service Redpanda
debezium	Capture des changements de la base de la table activités et alimentant un topic Redpanda
debezium-connector-init	Initialisation du connecteur Debezium
slack-consumer	Service consommant le topic Redpanda et envoyant des messages sur un channel Slack
airflow-webserver	Serveur Web Airflow
airflow-scheduler	Scheduler Airflow
airflow-triggerer	Triggerer Airflow
airflow-init	Initialisation d'Airflow



## SERVICES DOCKER

Nom	Description
airflow-cli	Client Airflow
postgres	Base de données servant à Airflow
prometheus	Service de monitoring
grafana	Visualisation des données de monitoring
postgresql-exporter	Service exportant les données de Prometheus
strava-like-server	Service créant aléatoirement des données d'activités, utilisé par Airflow
spark-master	Master du service Spark
spark-worker	Worker du service Spark
spark-job	Initialisation de l'application Spark
dremio	Service faisant l'interface entre un bucket S3 en Delta Lake et Tableau pour la partie analytique
dremio-init	Service initialisant le service Dremio

# MESSAGERIE SLACK

← → ⌚

Rechercher dans sport\_data\_prj\_12

**sport\_data\_prj\_12** ⌵ ⚙️ 📝

🕒 3 jours restants dans le cadre de l'essai ➤

🗨️ Fils de discussion

📞 Appels d'équipe

▼ Canaux

**# activities**

# general

🔒 sport-data

+ Ajouter des canaux

▼ Messages directs

**N** NICOLAS 81 vous

+ Inviter des personnes

▼ Applications

👤 python messenger

+ Ajouter des applications

**# activities**

🗨️ Messages ➤ 🖼️ Ajouter un canevas ➤

Bravo Emmendene Torres! Vous avez fait 2 heures de Runing le dimanche 17 août 2025

Bravo Gégore Guillot! Vous avez fait 2 heures de Runing le dimanche 17 août 2025

Bravo Julie Lambert! Vous avez fait 2 heures de Randonnée le dimanche 17 août 2025

Bravo Roger Techer! Vous avez fait 2 heures de Randonnée le dimanche 17 août 2025

Bravo Michel Normand! Vous avez fait 2 heures de Runing le dimanche 17 août 2025

Bravo Michelle Dumont! Vous avez fait 2 heures de Tennis le dimanche 7 septembre 2025

Bravo Eugénie Foucher! Vous avez fait 2 heures de Runing le dimanche 7 septembre 2025

Bravo Claudine Joubert! Vous avez fait 2 heures de Football le dimanche 7 septembre 2025

Bravo Charles Roussel! Vous avez fait 2 heures de Tennis le dimanche 7 septembre 2025

Bravo Jules Schmitt! Vous avez fait 2 heures de Natation le dimanche 7 septembre 2025

Bravo Yves Pons! Vous avez fait 2 heures de Runing le dimanche 7 septembre 2025

Bravo Arthur Pendragon! Vous avez fait 2 heures de Judo le dimanche 7 septembre 2025

Bravo Astrid Mendes! Vous avez fait 2 heures de Randonnée le dimanche 7 septembre 2025

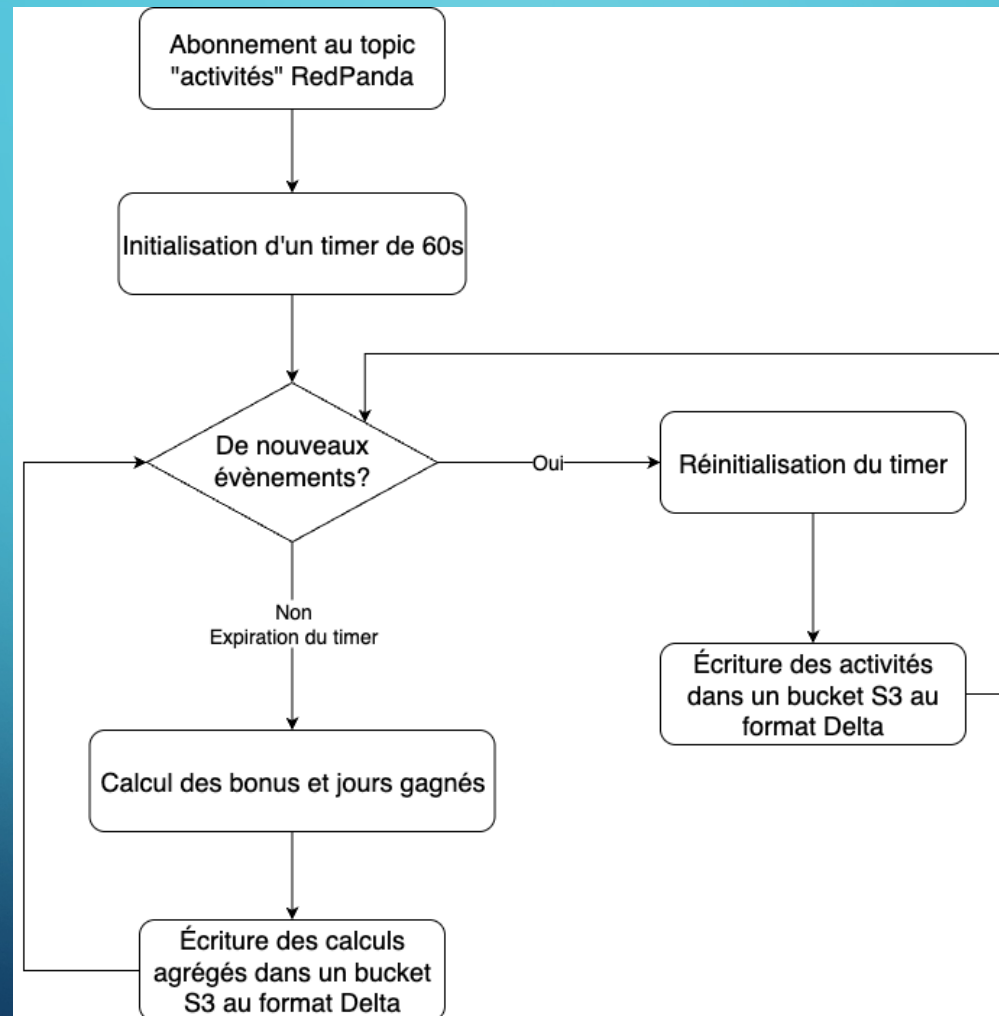
Bravo Francis Morin! Vous avez fait 2 heures de Runing le dimanche 7 septembre 2025

Bravo Capucine Payet! Vous avez fait 2 heures de Équitation le dimanche 7 septembre 2025

↑ 552 nouveaux messages ✕



# SYNOPSIS APPLICATION SPARK



# BUCKET S3

The screenshot shows the AWS S3 console interface. At the top, there's a search bar and a navigation breadcrumb: **Amazon S3** > **Buckets** > **nau-sport-data** > **delta/**. On the left, a sidebar lists various S3 features like 'General purpose buckets', 'Directory buckets', etc. The main area is titled 'delta/' and has two tabs: 'Objects' (selected) and 'Properties'. Under the 'Objects' tab, there's a section 'Objects (2)' with a refresh icon and buttons for 'Copy S3 URI', 'Copy URL', and 'Download'. Below this is a search bar 'Find objects by prefix' and a table listing the objects.

<input type="checkbox"/>	Name	Type	Last modified
<input type="checkbox"/>	<a href="#">activities/</a>	Folder	-
<input type="checkbox"/>	<a href="#">rewards/</a>	Folder	-

# INTERFACE DREMIO

The screenshot displays the Dremio web interface in a browser window. The address bar shows the URL `127.0.0.1:9047/source/sport-data/folder/nau-sport-data/delta`. The interface is divided into a left sidebar and a main content area.

**Left Sidebar:**

- Search bar: "Search Spaces and Datasets"
- Navigation menu with icons for Datasets, Spaces, Sources, and Object Storage.
- Current view: "sport-data.nau-sport-data.delta"

**Main Content Area:**

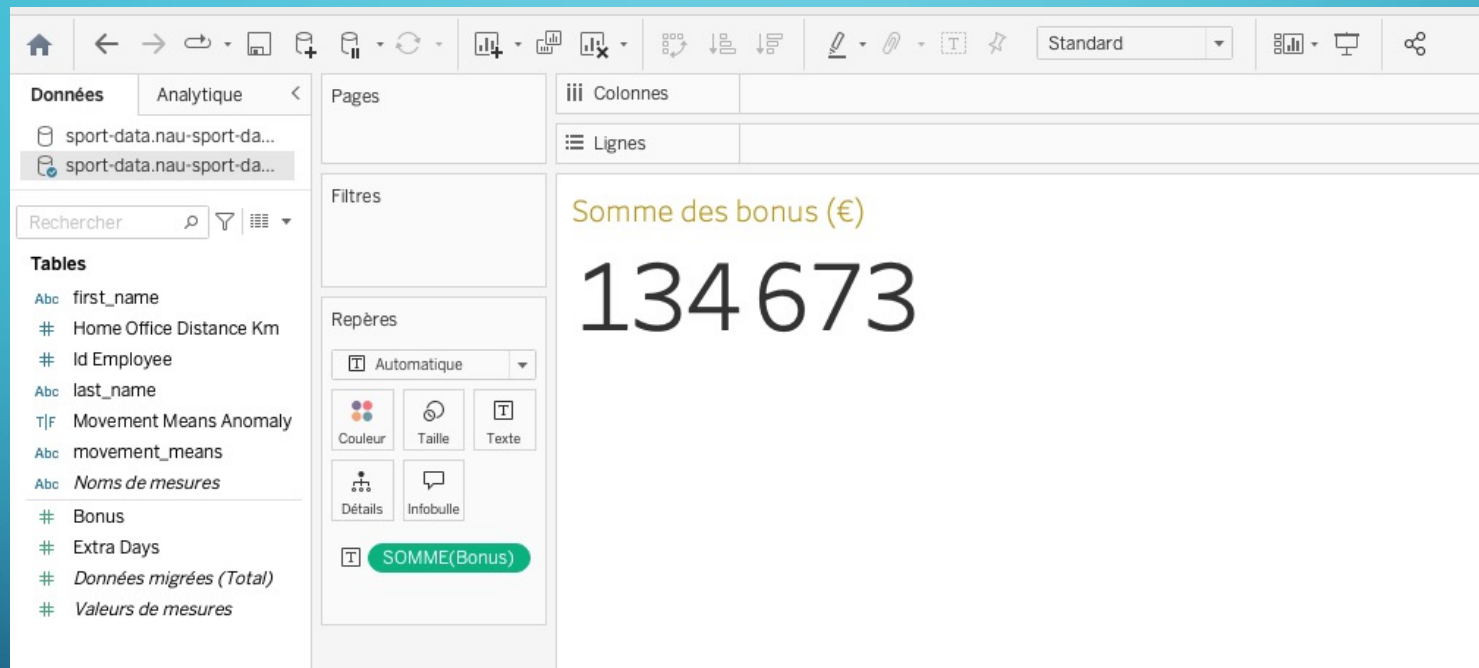
The main content area is divided into two panels. The left panel shows a tree view of the dataset structure:

- Datasets:** sport-data.nau-sport-data.delta
- Spaces (0):** nau (highlighted with a blue box)
- Sources:** (empty)
- Object Storage (1):** sport-data (2)

The right panel shows a list of datasets:

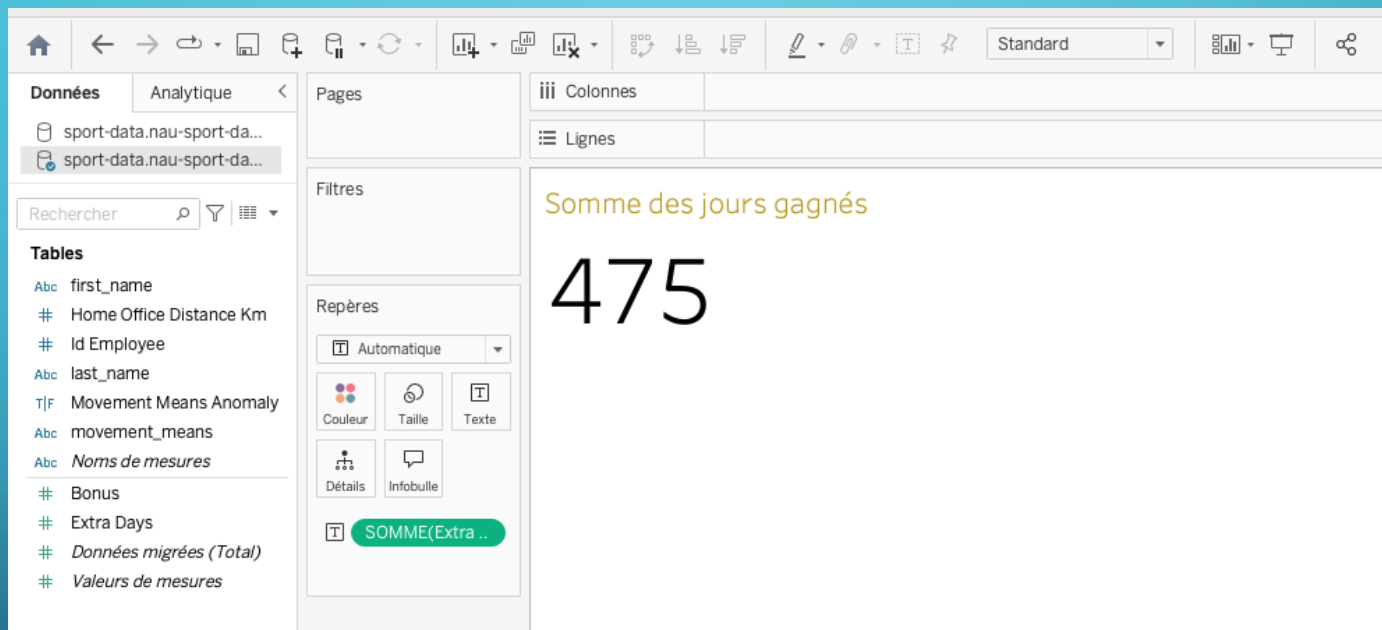
Name
activities
rewards

# TABLEAU : SOMME DES BONUS



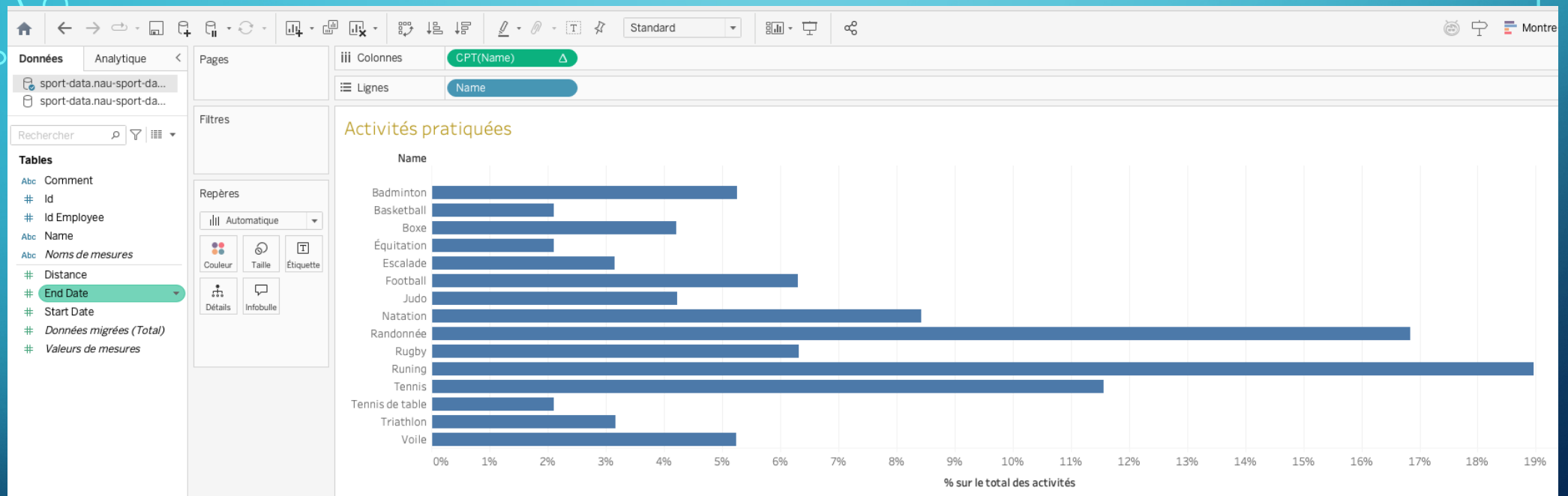


# TABLEAU : SOMME DES JOURS GAGNÉS





# TABLEAU : ACTIVITÉS PRATIQUÉES



# TABLEAU : INCOHÉRENCES

The screenshot shows the Tableau desktop application interface. The top navigation bar includes icons for home, back, forward, and various data manipulation tools. The left sidebar contains the 'Données' (Data) pane with a search bar and a list of tables. The main workspace displays a table titled 'Incohérences'.

**Données** Analytique < Pages

Colonne Colonne

Lignes

Movement Means An.. Id Employee first\_name last\_name Home Office Distance .. movement\_means

Rechercher

Filtres

Movement Means An..

Repères

Automatique

Couleur Taille Texte

Détails Infobulle

### Incohérences

Id Employee	first_name	last_name	Home Office Distance Km	movement_means	
100000	John	Doe	240.038	Marche/running	Abc

# GRAFANA: EXEMPLE METRICS

