

# SPRINT 2

---

Management de Projets (Scrum)

# ÉQUIPE

**Nadir BOUDJERIDA**  
Scrum Master



**Jean-Louis HU**  
Data Analyst



**Thomas DAHROUJ**  
Product Owner



# GLOSSAIRE

## 1. Table owa\_click.csv

Cette table enregistre les **clics des utilisateurs** sur le site. Elle inclut des informations telles que :

- **Identifiants** : `id`, `visitor_id`, `session_id`, `site_id`, etc.
- **Informations temporelles** : `timestamp`, `year`, `month`, `day`, `hour`, `minute`, `second`, etc.
- **Informations sur l'utilisateur** : `is_new_visitor`, `is_repeat_visitor`, `language`.
- **Données du clic** : `click_x`, `click_y` ( coordonnées du clic), `page_width`, `page_height`.
- **Contexte du clic** : `target_id`, `target_url`, `document_id` (éléments cliqués).
- **Informations sur le DOM** : `dom_element_name`, `dom_element_id`, `dom_element_text` (éléments du document HTML cliqués).
- **Campagnes et sources** : `campaign_id`, `ad_id`, `source_id`.

Cette table est idéale pour analyser le comportement des utilisateurs, notamment où les clics sont effectués.

## 2. Table owa\_request.csv

Elle capture toutes les **requêtes envoyées par les utilisateurs** sur le site :

- **Identifiants** : `visitor_id`, `session_id`, `site_id`.
- **Informations temporelles** : `timestamp`, `year`, `month`, `day`.
- **Adresse IP et localisation** : `ip_address`, `location_id`.
- **Source et campagnes marketing** : `source_id`, `campaign_id`, `medium`.
- **Données spécifiques** : `document_id`, `inbound_visitor_id`.

Cette table permet de suivre les pages consultées et l'origine du trafic.

## 3. Table owa\_session.csv

Cette table contient les informations sur les **sessions des utilisateurs** :

- **Identifiants** : `id`, `visitor_id`, `site_id`.
- **Durée et localisation** : `session_duration`, `ip_address`.
- **Informations temporelles** : `timestamp`, `year`, `month`, `day`.

Elle est utile pour analyser la durée des sessions et le comportement des utilisateurs.

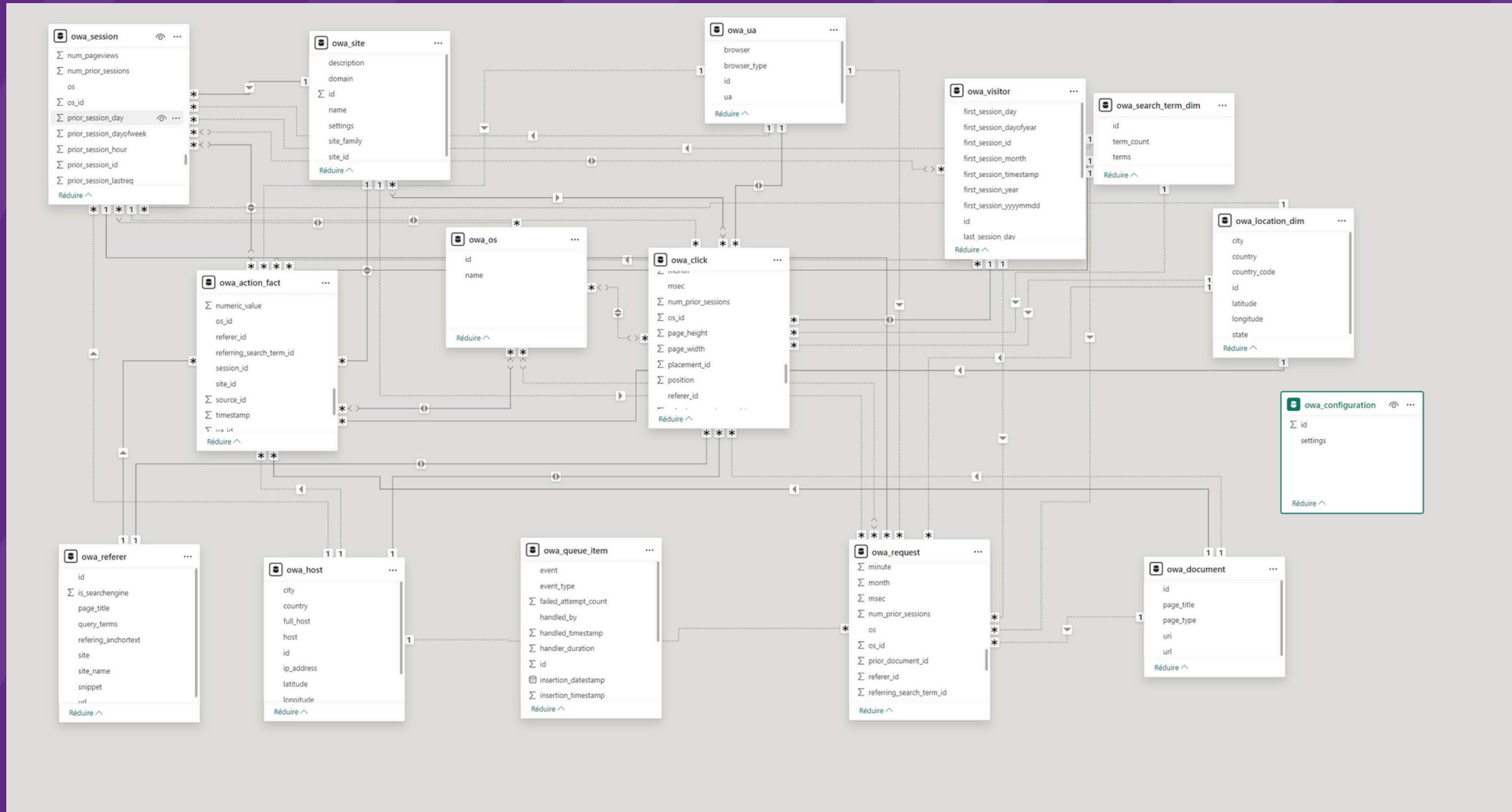
# DICTIONNAIRE

Nom de la table	Nom du champ	Type de données	Format de la donnée	Description ou définition du champ	Relation et dépendance avec d'autres champs
owa_ua.csv	browser	object	Texte	À définir ou spécifique à l'entité	Lié à owa_click.csv via la colonne id; Lié à owa_request.csv via la colonne id; Lié à owa_session.csv via la colonne id
	browser_type	object	Texte	À définir ou spécifique à l'entité	Lié à owa_click.csv via la colonne id; Lié à owa_request.csv via la colonne id; Lié à owa_session.csv via la colonne id
	id	int64	Numérique	Clé primaire de la table	Lié à owa_click.csv via la colonne id; Lié à owa_request.csv via la colonne id; Lié à owa_session.csv via la colonne id
	ua	object	Texte	À définir ou spécifique à l'entité	Lié à owa_click.csv via la colonne id; Lié à owa_request.csv via la colonne id; Lié à owa_session.csv via la colonne id

# LES 15 AUTRES TABLES SONT DANS LE DOSSIER

# CATALOGUE DE DONNÉES

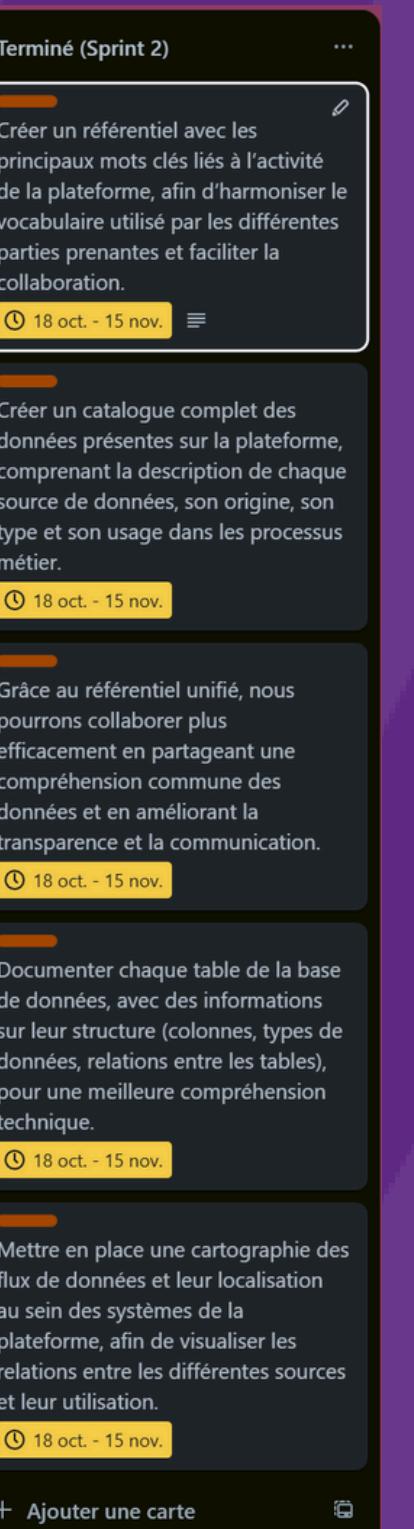
# CARTOGRAPHIE



# DIFFICULTÉ

- Compréhension des relations entre les tables
- Les notions de glossaire, dictionnaire et catalogue des données
- Trouver le bon outil pour visualiser les relations entre les tables

# SPRINT 2



# SPRINT 3

A FAIRE 4

Effectuez une analyse initiale des données pour identifier les problèmes, les lacunes, les incohérences et les erreurs du jeu de données.

SCRUM-26 

Définir des indicateurs de performances (KPI) pour surveiller en continu la qualité des données. 

SCRUM-27 

Contrôler la qualité des données (tests)

SCRUM-28 

Ajoutez des données manquantes ou améliorez la qualité des données existantes si nécessaire.

SCRUM-29 

+ Créer un ticket



**Paris  
School of  
Business**



**THOMAS DAHROUJ - JEAN-LOUIS HU - NADIR BOUDJERIDA**

# **LIEN BASE DE DONÉES :**

[https://docs.google.com/spreadsheets/d/1M99M4YwnjlwqlZeb  
IdBN26I3bvy6\\_dZgMhg0HHQ\\_Kio/edit?  
gid=2039841909#gid=2039841909](https://docs.google.com/spreadsheets/d/1M99M4YwnjlwqlZebIdBN26I3bvy6_dZgMhg0HHQ_Kio/edit?gid=2039841909#gid=2039841909)