

PONCET Paul
ROUVEURE Kylian
EDIN Alexis

SAÉ 5.02

Migration de données vers où depuis un environnement

Sous la direction de M.AGOUN

BUT Science des données
3ème année - IUT Lumière
Année universitaire 2023/2024

SOMMAIRE

1. Introduction	3
2. Données	4
3. Migration	5
4. Requête et analyse	9
5. Conclusion	13

INTRODUCTION

Dans ce dossier, nous avons pour but d'intégrer et d'exploiter des données dans un système NoSQL orienté graphe. Cela permet de gérer efficacement la masse de données et d'éviter la lourdeur des jointures.

L'étude de cas porte sur les événements sportifs en ville et le commanditaire est l'institut IGN. Nous devons réaliser une comparaison temporelle (avant/après) et spatiale pour une étude d'impact à la suite d'un événement sportif.

Pour le projet, nous avons choisi comme événement sportif l'Euro de football 2016 qui a eu lieu en France. L'événement a eu lieu du 10 juin 2016 au 10 juillet 2016 et s'est réparti dans 10 stades différents. Nous allons donc pouvoir analyser l'impact qu'a pu avoir cet événement.

Nous avons ainsi réfléchi à une problématique pour le projet :

Quel est l'impact de l'Euro 2016 sur les stades participants ainsi que sur leurs accessibilités ?

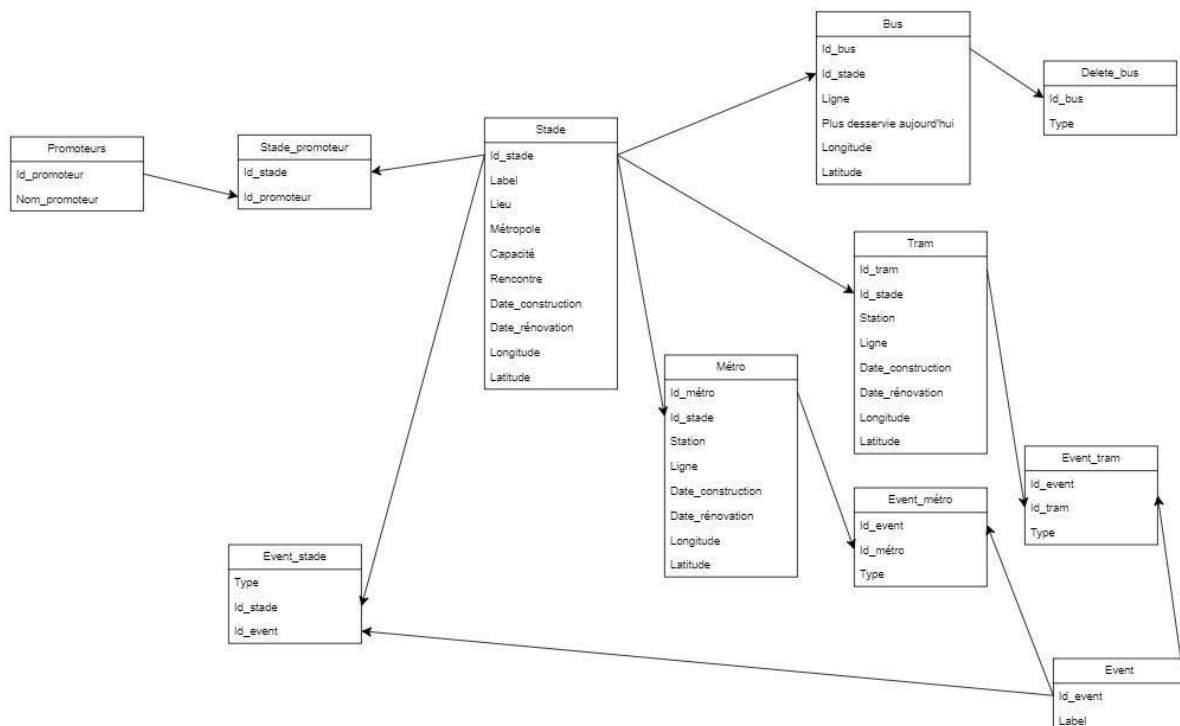
À travers cette problématique, nous allons pouvoir voir si certain stade ont pu subir des travaux ou autres modifications ? Est-ce que des lignes de métro ou de tramway ont été ajoutées à la suite de l'annonce de l'Euro ? Quelles installations ont été mises en place pour l'événement ? Nous allons essayer d'y répondre.

Nous avons coupé le travail en deux parties, la première création de la base SQL et la migration sur NoSQL par Kylian et la création du dossier et les analyses par Paul. Le plan de ce dossier sera les sources et l'intégration des données, puis notre méthodologie sur la migration du modèle SQL sur Neo4J et enfin les requêtes et les analyses.

DONNÉES

Comme dit précédemment, nous travaillons sur l'impact de l'Euro 2016 sur les stades et leurs accessibilités. Nous n'avons pas trouvé de jeux de données déjà prêts à être utilisés, qui contenait toutes les données nécessaires à notre analyse. Il a donc fallu créer nos fichiers CSV.

Voici notre MLD afin d'avoir un premier aperçu de nos données :



Nous avons ainsi un total de 11 tables que nous allons détailler.

Pour commencer, nous avons la table Stade composée d'un ID par stade ainsi que du nom, le lieu, la métropole, la capacité, le nombre de rencontres, les dates de construction et rénovation ainsi que les données géographiques. Nous avons une table promoteur avec les noms des promoteurs des stades. Nous avons dû créer une table pivot Stade_promoteur, car certains stades ont plusieurs promoteurs et il n'était donc pas possible de les ajouter directement à la table Stade.

Nous avons les tables Métro et Tram qui sont sensiblement les mêmes, avec leur ID, l'ID du stade qu'ils desservent, le nom de la station, la ligne, les dates de construction et rénovations et les données géographiques. La table bus est-elle différente. On retrouve les 2 ID ainsi que la ligne avec les données géographiques, cependant nous avons également

la possibilité de savoir si cette ligne est encore desservie ou non aujourd'hui. Nous avons décidé de prendre uniquement les lignes ajoutées pour l'Euro afin de savoir si elles ont, ou non, été gardées suite à l'événement. Cela signifie qu'il peut y avoir des lignes de bus qui sont non présentes dans nos données car elles étaient déjà en place et n'ont pas été impactées par l'Euro. Cela peut nous permettre de savoir si cette dernière a par exemple été mise en place seulement pour l'événement.

Nous avons la table événement qui est composée du nom de l'événement, en l'occurrence l'Euro 2016. Elle est reliée aux trois tables Event_tram, Event_metro et Event_stade. Ces trois tables sont similaires avec l'ID, l'ID_stade ainsi que le Type qui permet de savoir si ce fut construit ou rénové pour l'événement.

Pour avoir toutes ces données et créer nos fichiers CSV, nous avons mené plusieurs recherches. Pour ce qui est des stades, nous avons trouvé nos données sur Wikipédia avec la page des stades, mais également la page de l'Euro 2016. Pour les transports, nous avons pris comme source France Bleu, qui proposait le guide d'accès des stades de l'Euro (Guide d'accès [Nom du stade]). Pour connaître les transports aujourd'hui, nous avons utilisé Moovit. Concernant les stations, nous sommes allés sur les pages Wikipédia des stations concernées.

Nous avons ensuite unifié les articles afin de créer nos différents fichiers CSV qui correspondent aux tables présentées précédemment.

MIGRATION

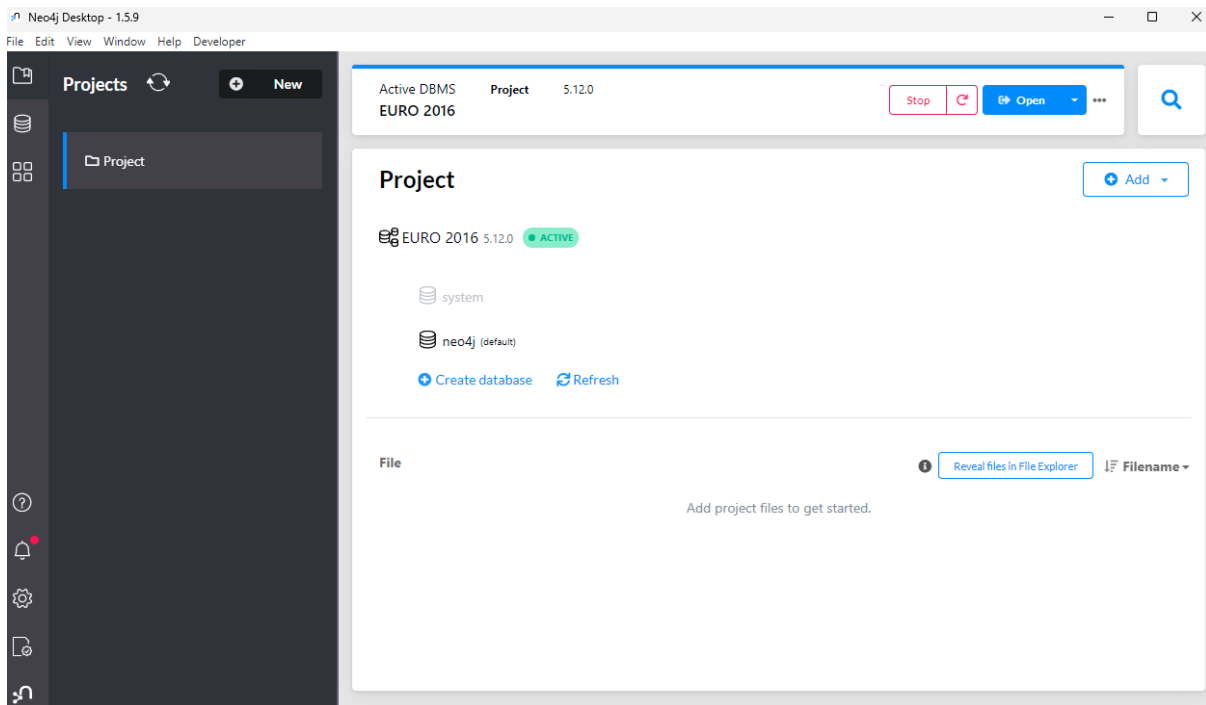
1. Création de la base SQLite

Pour créer la base de données SQLite, nous avons dû, au préalable, réaliser le MLD présenté dans la partie préparation des données. Cela nous a permis de ne pas nous égarer dans la création de la base. Pour la créer, nous avons d'abord utilisé le logiciel looping avant d'utiliser draw.io pour des raisons d'accessibilité.

Par la suite, nous avons dû importer les données présentes dans nos fichiers CSV dans la base de données SQLite.

2. Création de la base Neo4J

Pour commencer la création de la base Neo4j, après avoir installé ce système, nous avons créé un projet et une base de données DBMS. Nous l'avons nommée Euro 2016.



À partir de là, nous allons utiliser la bdd par défaut neo4j. Nous avons créé un user admin (pseudo : sae & mdp : azerty01) pour pouvoir insérer des données à partir d'un script python afin d'automatiser le processus de création de base orienté graphe.

3. Script python

Nous avons choisi de faire une migration avec python. Le but est d'automatiser le processus et d'avoir une évolutivité des bases puisque que si la base SQL accueille de nouvelles données, nous aurons juste à lancer le script pour mettre à jour la base Neo4J.

Pour commencer, on a dû installer les librairies pour faire le process.

```
> pip install py2neo  
> pip install neo4j
```

Voici un exemple de code pour les nœuds :

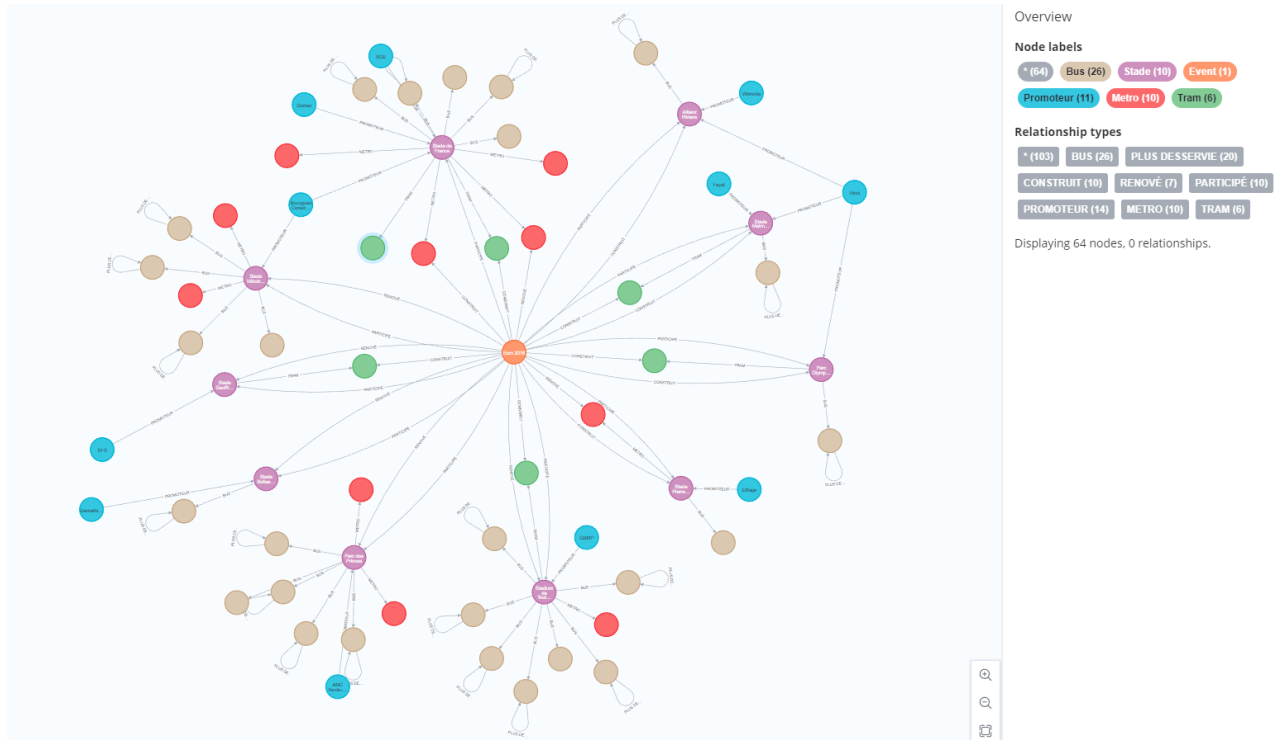
```
cursor.execute("SELECT * FROM stades") // Récupérer toute la table  
stades = cursor.fetchall()  
for stade in stades: // boucle pour chaque ligne  
    stade_node = Node("Stade", id=stade[0], label=stade[1],  
        lieu=stade[2], metropole=stade[3],  
        capacite=stade[4], rencontre=stade[5],  
        construction=stade[6], renovation=stade[7],  
        lon=stade[8], lat=stade[9])  
graph.create(stade_node) // Insérer un noeud dans Neo4J
```

Voici un exemple de code pour les relations :

```
cursor.execute("SELECT * FROM event_stades") // Récupérer toute la table  
event_stades = cursor.fetchall()  
for rel_type, id_event, id_stade in event_stades: // boucle pour chaque ligne  
    stadium=graph.nodes.match("Stade", id=id_stade).first()// recup l'id stade  
    event =graph.nodes.match("Event", id=id_event).first()// recup l'id evenement  
  
    if stadium and event:  
        relationship = Relationship(event, rel_type.upper(), stadium)  
        graph.create(relationship) // Insérer une relation dans Neo4J
```

4. Résultat

Après avoir lancé le script, voici le résultat de la base :



REQUÊTES & ANALYSES

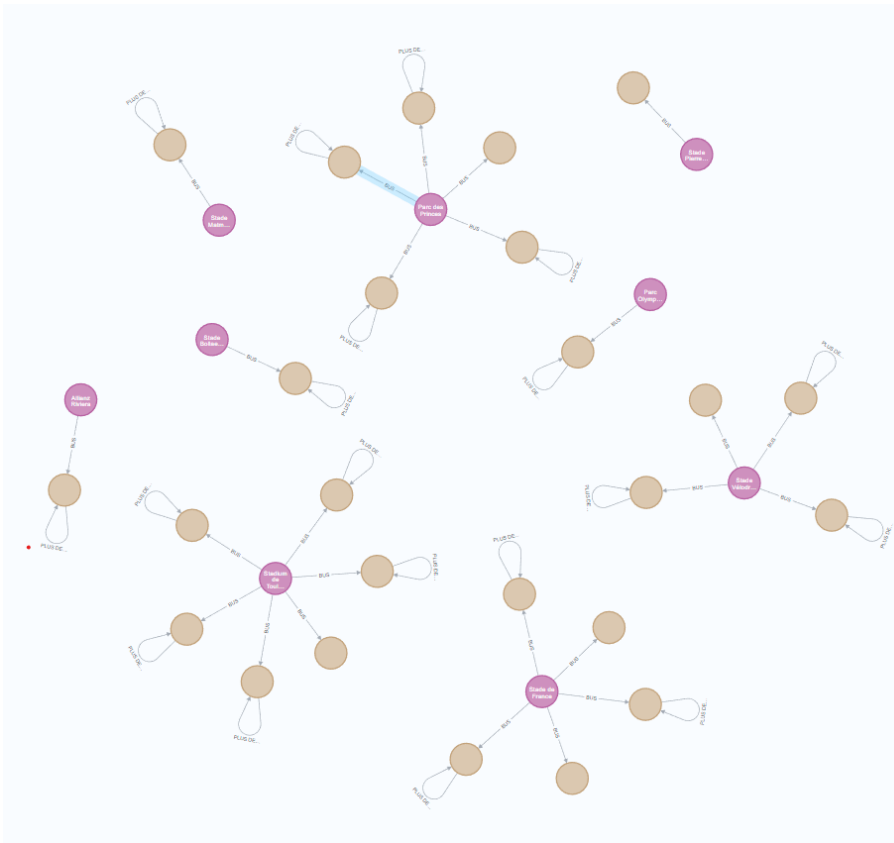
Requête 1 : Quelles sont les construction et rénovations pour l'Euro 2016 :

Requête	MATCH p=()-[r:CONSTRUIT RENOVÉ]->() RETURN p
Résultat	

Analyse :

A travers ce graphique, nous souhaitons montrer l'impact global de l'Euro sur les stades et leur moyen d'accès. Nous pouvons voir que sur les 10 stades participant à l'euro il y en a 9 qui ont subi une rénovation ou qui ont été construits. Nous avons trois lignes de métro rénovées ou construites ainsi que 4 lignes de tramway. Les nouveaux stades comme le Parc OL ou le Matmut Atlantic, n'ont pas simplement été construits pour l'événement, mais il y a également eu l'ajout d'un accès en tramway. Cela n'a pas été le cas pour l'Allianz Riviera dû au fait d'une localisation assez isolée par rapport à la ville.

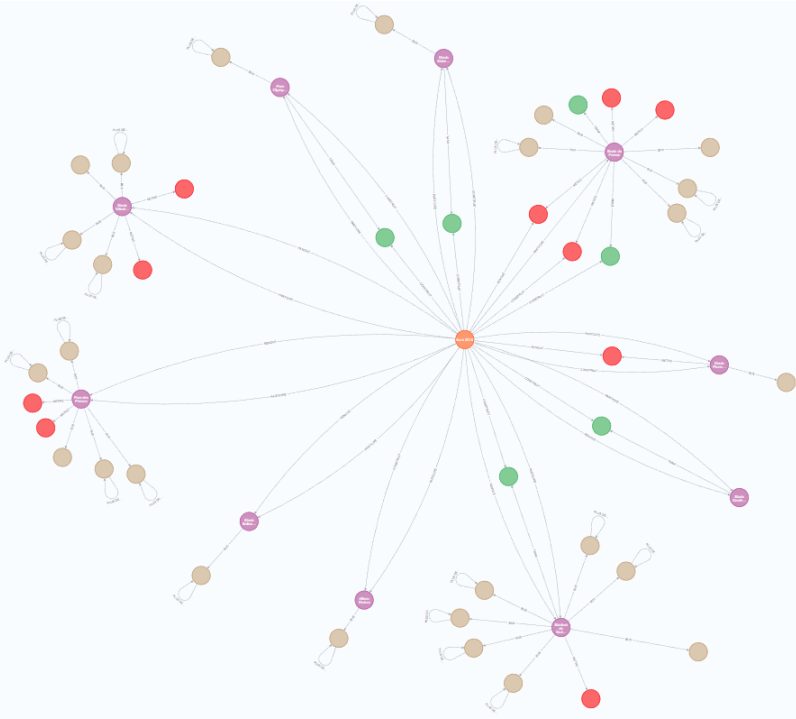
Requête 2 : Parmi les lignes de bus mises en place pour l'événement, est-ce qu'il y en a certaines qui ont été supprimées suite à l'Euro ?

Requête	<code>MATCH p=()-[r:BUS]->() RETURN p</code>
Résultat	

Analyse :

Ici, nous pouvons voir que la majorité des lignes de bus ont été supprimées suite à l'événement. Il en reste généralement une ou deux. Nous pouvons voir que certains stade n'avaient qu'une seule ligne de bus qui a été supprimée par la suite. Pour des stades comme Bollaert et l'Allianz Riviera, il y a eu un gain d'accessibilité pendant l'Euro, mais cela est revenu à la normale après l'événement. Pour ces stades, il n'y a donc pas eu d'amélioration des transports après l'Euro, contrairement au stade de Geoffroy Guichard qui a eu la construction du Tram pour l'Euro.

Requête 3 : Quels sont les stades les mieux desservies durant l'Euro ?

Requête	<code>MATCH p=()-[r:BUS TRAM METRO CONSTRUIT]->() RETURN p</code>
Résultat	

Analyse :

Avec ce graph, nous souhaitons voir quels étaient les stades avec la meilleure accessibilité mise en place lors de l'Euro. Nous pouvons voir que le Stade de France ainsi que le Vélodrome et le parc des princes sont les 3 stades avec la meilleure accessibilité durant l'Euro. Nous pouvons tout de même voir qu'après l'Euro, le stade de France reste le stade qui a gardé le plus des transports qui en place pour l'événement.

Requête 4 : Quels sont les promoteurs qui ont le plus participé à l'événement ?

Requête	<code>MATCH p=()-[r:PROMOTEUR]->() RETURN p</code>
Résultat	<pre> graph TD SPS((SPS)) -- PROMOTEUR --> SG1((Stade Geoffroy...)) Fayat((Fayat)) -- PROMOTEUR --> SM1((Stade M...)) GBMP((GBMP)) -- PROMOTEUR --> ST1((Stadium de Traut...)) Vinci((Vinci)) -- PROMOTEUR --> SM2((Stade M...)) Vinci -- PROMOTEUR --> PV1((Parc Olymp...)) Vinci -- PROMOTEUR --> SV1((Stade Vélod...)) AllianzRiviera((Allianz Riviera)) -- PROMOTEUR --> SM2 AllianzRiviera -- PROMOTEUR --> SV1 Wilmotte((Wilmotte)) -- PROMOTEUR --> SM2 Wilmotte -- PROMOTEUR --> SV1 SougouardComet((Sougouard Comet...)) -- PROMOTEUR --> SV1 SougouardComet -- PROMOTEUR --> SF1((Stade de France)) Dumez((Dumez)) -- PROMOTEUR --> SF1 SGE((SGE)) -- PROMOTEUR --> SF1 ParcPrinces((Parc des Princes)) -- PROMOTEUR --> AMC1((AMC...)) StadePereire((Stade Pierre...)) -- PROMOTEUR --> Eflage((Eflage)) </pre>

Analyse :

Ce graph nous permet de voir quels sont les promoteurs ayant eu le plus d'impact dans les rénovations et construction des stades. Le promoteur qui sort du lot et Vinci qui a été promoteur de 3 stades au total. Il y a quand même un nombre assez important de promoteurs.

CONCLUSION

En conclusion, notre analyse de l'impact de l'Euro 2016 sur les stades et leurs accessibilités a révélé des transformations significatives dans l'infrastructure urbaine et les réseaux de transport.

Notre étude a mis en évidence que la majorité des stades participants à l'Euro 2016 ont bénéficié de rénovations ou de constructions, non seulement pour améliorer les infrastructures sportives, mais aussi pour optimiser les moyens d'accès via les transports en commun. L'ajout et la rénovation de lignes de métro et de tramway soulignent l'effort pour augmenter l'accessibilité des spectateurs. Cependant, notre analyse a également révélé que beaucoup des lignes de bus spécialement mises en place pour l'événement ont été supprimées suite pour être remplacées par des lignes permanentes. Ce choix est sûrement dû au manque de temps pour développer des lignes de bus permanentes.

En synthèse, l'Euro 2016 a engendré des avancées importantes dans la préparation et l'amélioration des infrastructures sportives et de transport. Toutefois, les impacts varient, certains bénéficiant d'améliorations durables tandis que d'autres voient des avantages plus éphémères. Cette dualité souligne la nécessité d'une planification et d'un investissement réfléchis pour que de grands événements sportifs puissent laisser un héritage positif et durable aux villes hôtes et à leurs habitants. Avec cet événement, nous pouvons affirmer qu'il a eu un impact très positif pour la France, ses divertissements et ses transports.