



# Analyse marché boursier

Gaëtan Corin Fabien Vidor  
Lou Doussiet





# Sommaire

- 01. Problématique
- 02. Architecture
- 03. Données boursières
- 04. Actualités boursières
- 05. Futures évolutions





# Problématique

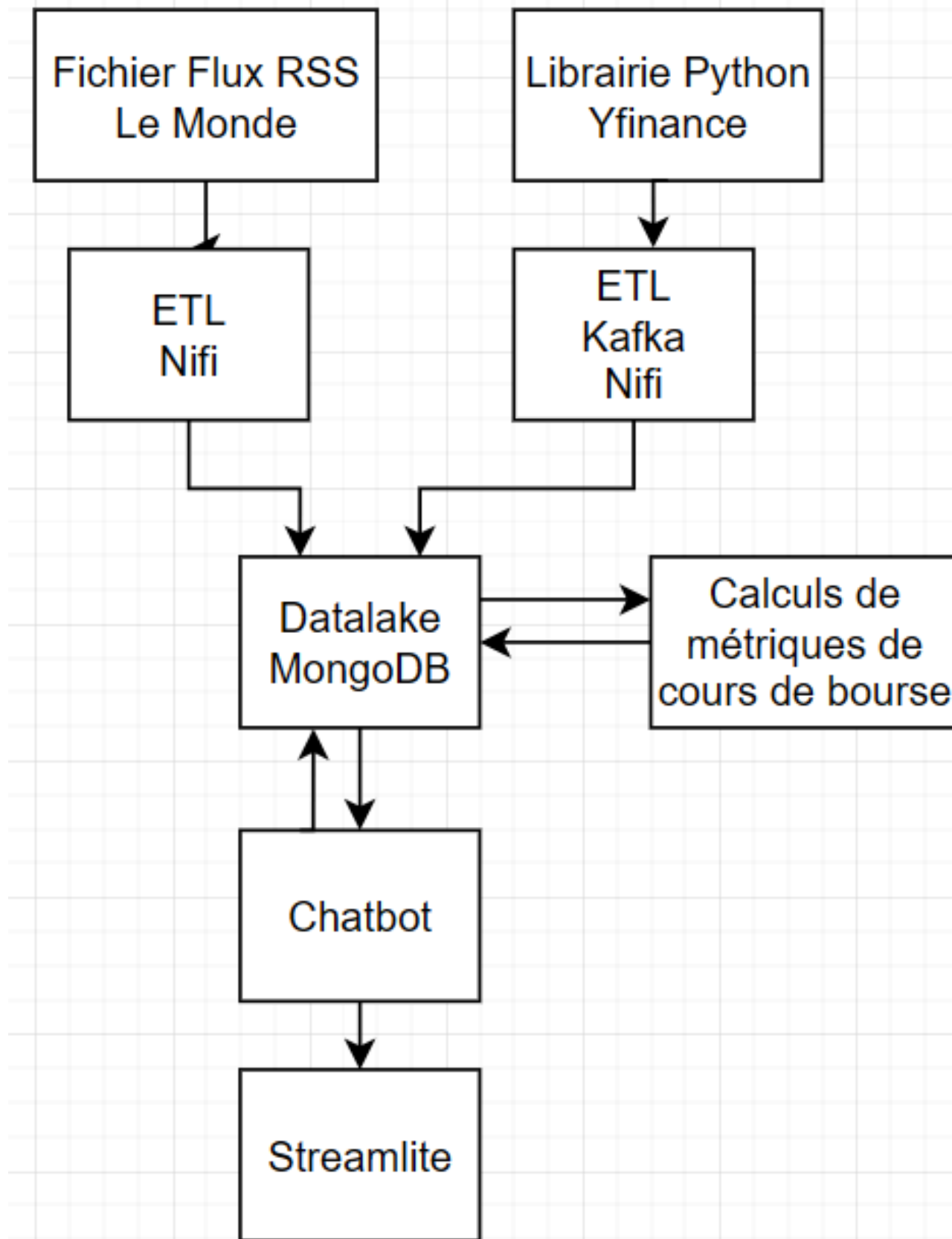
Comment créer un LLM qui analyse les indices boursiers en fonction d'un ensemble d'éléments comme l'évolution de son cours, les actualités de l'entreprise ou encore les événements mondiaux liés?



# Architecture

L'architecture du projet se compose de 4 composants distincts pouvant fonctionner avec le datalake de manière indépendantes en lien avec le Datalake

- Le composant composé du fichier flux RSS le monde et de l'ETL Nifi. Ce composant permet de récupérer les actualités boursières journalières
- Le composant composé de la librairie python Yfinance, et de l'ETL constitué de Kafka et de Nifi. Ce composant permet de récupérer les données boursières journalières
- Le composant de calcul de métriques de cours de bourse. Ce composant sert à enrichir les données afin de rajouter de la pertinences aux analyses.
- Le composant de chatbot et de Streamlite, qui permet à l'utilisateur de comprendre les demandes de l'utilisateur en langage naturel et de lire les données du Datalake de manière structurer



# Architecture

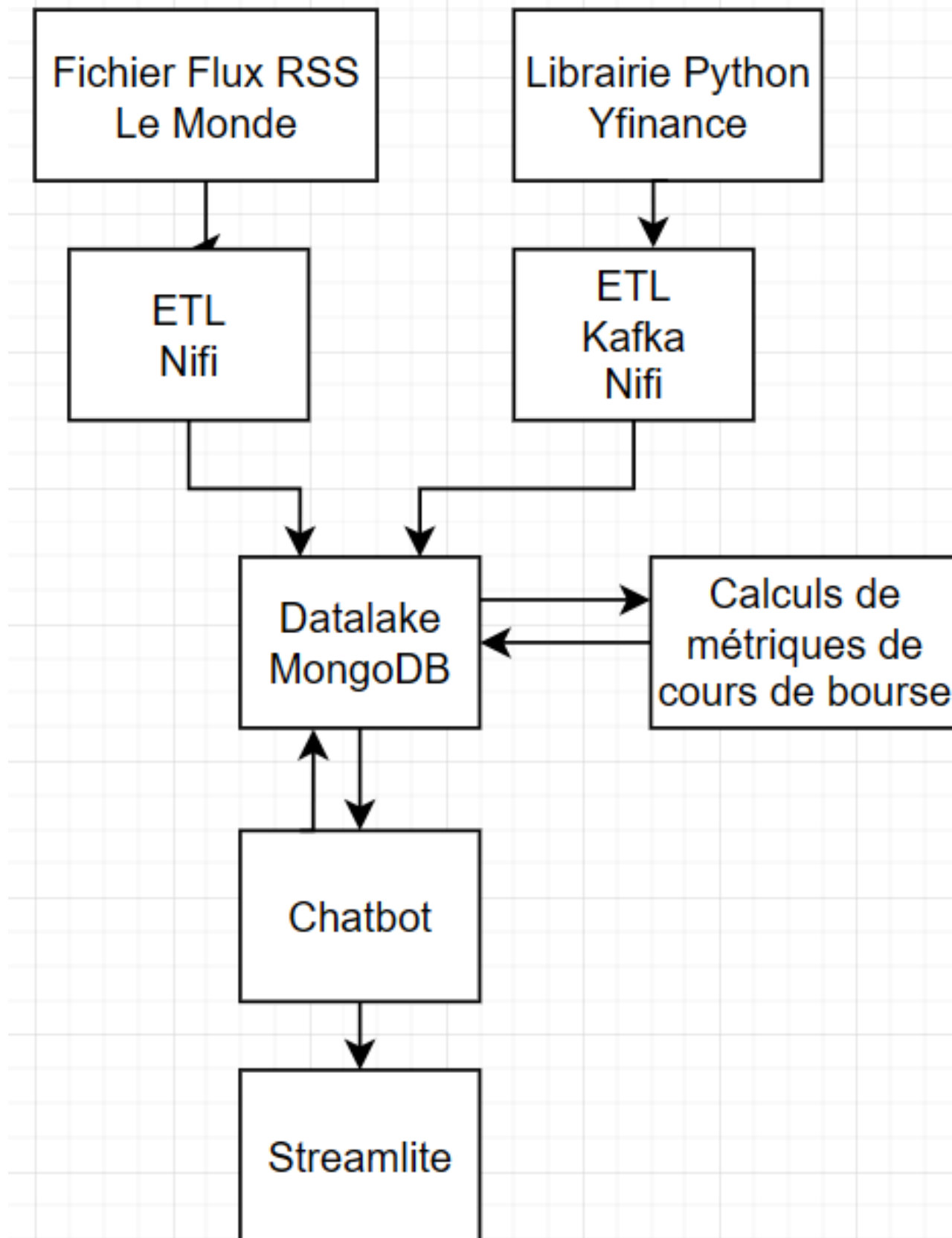
L'architecture a été décidée sous format de modules séparés afin d'améliorer la résilience et la disponibilité de la solution, que cela soit en terme d'implémentation de nouvelles données, aussi bien que de la lecture d'utilisateur par le chatbot.

En effet, grâce à cette architecture, en cas d'échec de l'un des modules, les autres modules continuent à fonctionner.

L'utilisation de Nifi lors de l'implémentation de nouvelles données permet de réaliser des reprises automatique en cas d'échec de traitement. Cela permet de sauvegarder les données non traitées afin de ne pas avoir de perte.

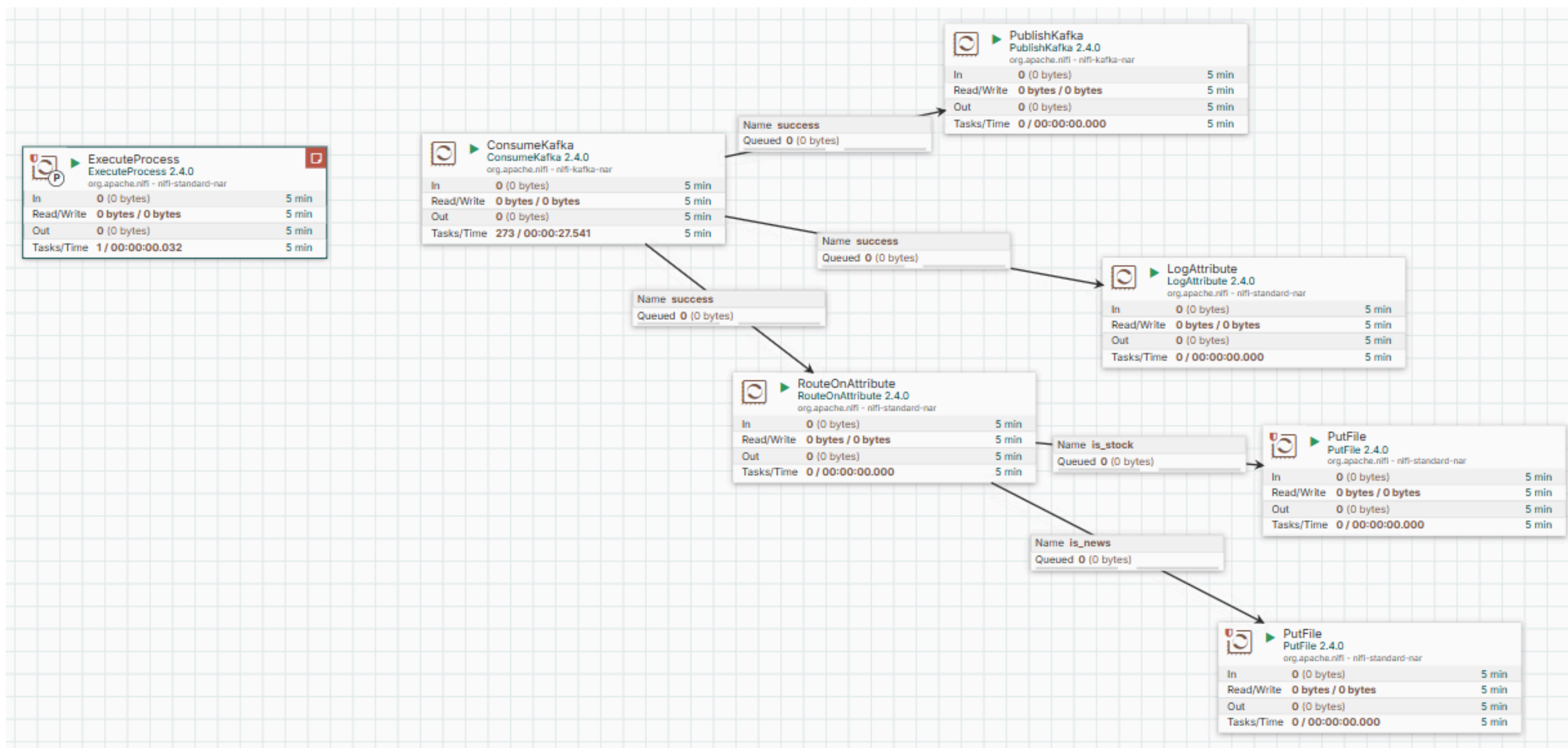
Le clustering est disponible sur Nifi. Il sera donc possible lors d'une scalabilité futur de conserver les outils actuels.

La base de données MongoDB sera déployé avec un replicaset avec au moins 3 noeuds, afin d'obtenir une haute disponibilité.





# Données boursières





# Actualités boursières

Utilisation de flux RSS mis à disposition par les journaux

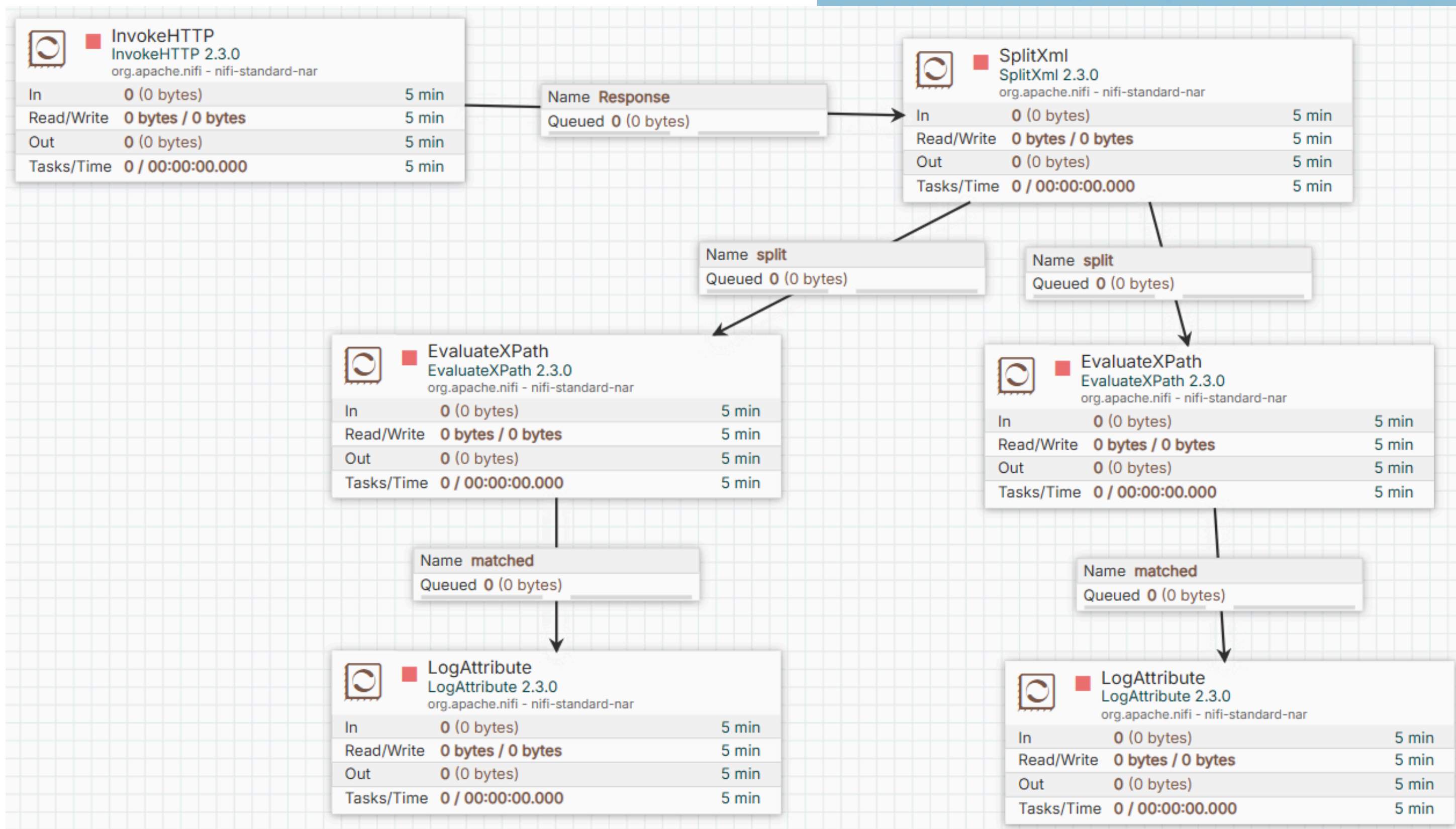
[https://www.lemonde.fr/economie/rss\\_full.xml](https://www.lemonde.fr/economie/rss_full.xml)

```
▼<item>
  ▼<title>
    <![CDATA[ Taxis et chauffeurs VTC prévoient de nouvelles manifestations à partir du 10 juin ]]>
  </title>
  <pubDate>Tue, 10 Jun 2025 12:20:44 +0200</pubDate>
  <updated>Tue, 10 Jun 2025 13:04:11 +0200</updated>
  ▼<description>
    <![CDATA[ Les taxis protestent contre la nouvelle convention régissant le transport des malades, tandis que les VTC dénoncent l'afflux de nouveaux chauffeurs sur les plateformes. ]]>
  </description>
  <guid isPermaLink="true">https://www.lemonde.fr/societe/article/2025/06/10/nouvelles-journees-de-mobilisations-pour-les-taxis-et-les-vtc-les-10-et-11-juin_6612032_3224.html</guid>
  <link>https://www.lemonde.fr/societe/article/2025/06/10/nouvelles-journees-de-mobilisations-pour-les-taxis-et-les-vtc-les-10-et-11-juin_6612032_3224.html</link>
  ▼<media:content width="644" height="322" url="https://img.lemde.fr/2025/05/19/1087/0/8192/4096/644/322/60/0/c6063f3_ftp-import-images-1-np4ynr2b7fxj-5947418-01-06.jpg">
    <media:description type="plain">Lors d'une manifestation de taxis, à Paris, le 19 mai 2025.</media:description>
    <media:credit scheme="urn:ebu">GEOFFROY VAN DER HASSELT / AFP</media:credit>
  </media:content>
</item>
```





# Flux Nifi







# Résultat de flux Nifi

Le flux nous permet de récupérer plusieurs attributs

- title
- updated
- description

✓ Formatted Original

```
1 <item xmlns:atom="http://www.w3.org/2005/Atom" xmlns:content="http://purl.org/rss/1.0/modules/content/" xmlns:media="http://search.yahoo.com/mrss/">
2   <title>En Chine, les villes face au gouffre financier de leurs métros</title>
3   <pubDate>Mon, 09 Jun 2025 16:28:10 +0200</pubDate>
4   <updated>Tue, 10 Jun 2025 10:19:19 +0200</updated>
5   <description>Après les lignes à grande vitesse, les lignes de métro sont en train de plomber les villes qu'elles desservent. La faute à des coûts de chantiers puis de fonctionnement colossaux, des recettes m:
6   <guid isPermaLink="true">https://www.lemonde.fr/economie/article/2025/06/09/en-chine-les-villes-face-au-gouffre-financier-de-leurs-metros_6611749_3234.html</guid>
7   <link>https://www.lemonde.fr/economie/article/2025/06/09/en-chine-les-villes-face-au-gouffre-financier-de-leurs-metros_6611749_3234.html</link>
8   <media:content width="644" height="322" url="https://img.lemde.fr/2025/06/09/842/0/6000/3000/644/322/60/0/b1a67a5_sirius-fs-upload-1-grznoxogsb04v-1749475762920-ibpremiumthree251894.jpg">
9     <media:description type="plain">La station de métro Gangxia Nord, à Shenzhen, en Chine, le 14 mai 2024.</media:description>
10    <media:credit scheme="urn:ebu">Markus Mainka/imageBROKER/MAXPPP</media:credit>
11  </media:content>
12 </item>
```



# Evolution Future

Mise en place de :

- Métriques des cours de Bourses
- Data Lake
- LLM
- Stream lit