

Plateforme Analytique pour l'Université de Neuchâtel

Michael Stoffel

Description

Le projet a comme but la création d'un environnement qui permet à l'Université de Neuchâtel d'intégrer les techniques des sciences de donnée dans le processus de prise de décision.

Durant une première phase, ce développement se concentre sur les données académiques. Ces données sont prises du le système de gestion IS-Académia et stockées dans un data warehouse Oracle.

L'environnement à implémenter doit contenir des outils pour extraire les données, les transformer, les pré-traiter pour l'analyse, la modélisation et finalement la génération de connaissance actionnable.

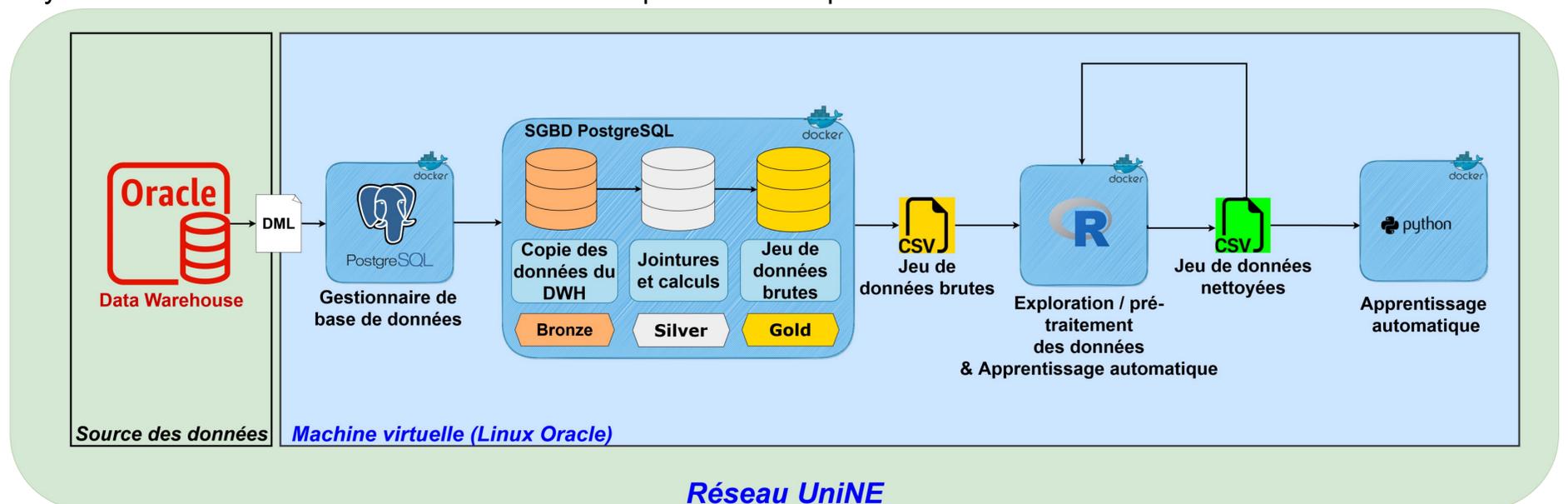
Du point de vue technologique, la plateforme devrait être, autant que possible, indépendant du système sous-jacent pour garantir un maximum de portabilité.

Pour illustrer le processus et pour le valider, une application typique doit être réalisée. Il s'agit d'analyser la durée d'étude et surtout le dépassement de la durée pour laquelle l'UniNE est subventionnée.

Déroulement

Le projet est organisé sous forme du développement d'un proof of concept (POC). Ce développement est structuré en trois étapes.

1. Choisir les outils et les technologies à utiliser, et définir une spécification du système complet.
2. L'implémentation de la plateforme en utilisant Docker pour garantir la portabilité du système, PostgreSQL pour réaliser les tâches de bases de données, et R et Python pour la préparation des données et la génération des modèles.
3. L'analyse de la durée d'étude est réalisée et ainsi le proof of concept est finalisé.



Résultats

Il y a trois principaux résultats sont obtenus dans le cadre de ce projet.

1. La conception de la plateforme. Nous avons pu montrer que Docker est parfaitement adapté à la création d'un environnement efficace et portable qui est indépendant du système sous-jacent. PostgreSQL, R et Python ont également complètement remplis leur mission.
2. La réalisation concrète du système. Le concept établi a pu être réalisé et facilement porté sur les différents systèmes exploitation et a tourné efficacement.
3. Cas d'utilisation. Nous avons pu montrer que l'environnement mis en place a permis de mettre l'application complète en place. Mais nous avons également dû constater que les données et certains algorithmes ont rencontré des limites. Des informations concrètes ont résulté, par exemple la forte dépendance de la durée d'étude en fonction de la faculté d'accueil de l'étudiant

Perspectives

La plateforme a certainement montré sa pertinence, mais elle a aussi mis en évidence que des efforts supplémentaires sont nécessaires pour qu'une telle plateforme puisse atteindre une maturité de production.

D'un côté, un certain nombre d'activités doivent être automatisées. Les informations de base doivent être produites systématiquement, p.ex. Sous forme de tableaux de bord. De l'autre côté la collecte de données doit être complétée pour satisfaire les tâches plus sophistiquées.