

Projet Homework/ Dr.Djerbi

Système Autonome de Diagnostic et d'Optimisation de Performance de Base de Données (SADOP)

Ce projet se concentre sur l'idée d'utiliser l'IA pour **améliorer la performance et automatiser l'administration** d'une base de données MySQL, un domaine en forte demande d'innovation.

1. Objectifs Principaux du Projet

1. Conception et Administration MySQL :

- Mise en place d'une **base de données MySQL de grande taille** (simulée ou réelle, ex: données de transactions, logs de serveurs, etc.).
- Maîtrise de l'**administration** (monitoring des performances, logs, etc.) et de l'**optimisation** (indexation, schémas, requêtes complexes).

2. Intégration et Traitement des Données pour l'IA :

- Extraction des données de performance (temps d'exécution des requêtes, usage CPU/RAM, plans d'exécution, etc.) depuis MySQL pour alimenter les modèles d'IA.

3. Développement de Modèles d'IA Innovants :

- Création de modèles pour le diagnostic **prédictif** et l'optimisation.
- Implémentation d'une interface en langage naturel pour l'administration.

2. Composantes Techniques et IA

Domaine	Composante/Modèle d'IA	Implication MySQL	Difficulté (Note)
Machine Learning (ML) / Deep Learning (DL)	Moteur de Diagnostic et de Prédiction de Performance : Classification/Régression pour prédire les requêtes lentes (slow_query) ou les goulots d'étranglement futurs .	Analyse des logs de requêtes, des statistiques internes de l'InnoDB, des métriques du serveur.	Élevée (4 Pts)
Deep Learning (DL) / Apprentissage par Renforcement (RL)	Agent d'Optimisation d'Index : Un agent RL qui apprend à recommander des index ou des changements de schéma en fonction des performances observées, sans intervention humaine .	Exécution des requêtes EXPLAIN, analyse de l'impact des index. C'est le cœur innovant du projet.	Très Elevée (6 Pts)
IA Générative (LLM / LangChain)	Interface d'Administration en Langage Naturel : Utilisation d'un LLM (local ou via API) orchestré par LangChain (ou similaire) pour traduire une demande en langage naturel (ex: "Trouve la requête la plus lente hier soir") en une requête SQL optimisée et/ou une commande d'administration MySQL.	Génération de SQL, exécution de commandes SHOW Variables, etc.	Élevée (4 Pts)

3. Répartition des Tâches (une proposition pour un groupe de six personnes) :

- **Personnes 1 & 2 : Ingénieurs Base de Données (DBA) & Administration (Focus MySQL)**
 - Conception du schéma MySQL, création du jeu de données massif.
 - Mise en place du monitoring, configuration des **slow_query_logs**, extraction des métriques.
 - Création des scripts d'**injection de charge** pour générer des données de performance variées (simuler des pics de trafic, des requêtes inefficaces).
- **Personnes 3 & 4 : Data Scientists & ML/DL (Focus Modèles Prédictifs)**
 - Prétraitement et *Feature Engineering* des données de performance MySQL.
 - Développement du **Moteur de Diagnostic** (ML/DL) pour prédire les requêtes qui deviendront lentes.
 - Évaluation des modèles et interprétation des résultats.
- **Personnes 5 & 6 : Chercheurs IA & Développement (Focus RL/IA Générationne)**
 - Développement de l'**Agent RL (le défi le plus ardu)** pour l'optimisation des index/schémas.
 - Mise en place de l'interface utilisateur/API et intégration de l'**IA Générationne** pour l'interface en langage naturel.
 - Connexion du système complet : l'Agent RL et le Moteur DL alimentent les recommandations pour l'IA Générationne.

4. Potentiel Académique et d'Innovation

Ce projet est très important, car (1) il combine le concept de l'Administration et Programmation des BD (votre module en M2) avec les nouveaux concepts de l'IA (ML/DL, **Apprentissage par Renforcement (RL)**, **LangChain**, **IA Générationne et LLMs..**), mais aussi (2) il nécessite de lier les actions prises dans la BD (création d'index, modification de schémas) avec l'**impact sur les performances** (mesuré par le ML).

- **Production Scientifique** : Le travail sur l'**Agent RL pour l'optimisation d'index** est un sujet de recherche. Il est innovant de laisser un modèle apprendre par lui-même les meilleures stratégies d'indexation.
- **Innovation Technique** : L'approche **IA Générationne** pour interagir avec une base de données en langage naturel est une application très actuelle du **Large Language Model** au domaine de l'administration système.

Le livrable final devrait être un **rappor académique détaillé (30% de la note du 15)**, un **code source bien documenté** et une **démonstration live** du SADOP fonctionnant sur leur base de données MySQL (70%).

Faites voir l'Annexe A pour plus de détail technique.