



École Nationale Supérieure d'Informatique et d'Analyse des Systèmes
Centre d'Études Doctorales en Sciences des Technologies de l'Information et de l'Ingénieur

AVIS DE SOUTENANCE DE THÈSE DE DOCTORAT

Madame Chaimae ASAAD

Soutiendra publiquement sa thèse de Doctorat en Informatique

Le Vendredi 26 Septembre à 10h00 au Grand Amphi à l'ENSIAS de Rabat

Leveraging Artificial Intelligence for NoSQL Database Modeling, Querying, and Optimization

Président :

Pr. Mohammed BOULMALF, PES, Université Internationale de Rabat

Directeur de thèse :

Pr. Karim BAÏNA, PES, ENSIAS, Université Mohammed V de Rabat

Co-Encadrant de thèse :

Pr. Mounir GHOGHO, PES, Université Mohammed VI Polytechnique, Rabat

Rapporteurs :

Pr. Khalid BELHAJJAME, MCF-HDR, Université Paris Dauphine, France

Pr. Boutaina HDIOUD, MCH, ENSIAS, Université Mohammed V de Rabat

Pr. Youssef GAHI, MCH, ENSA, Université Ibn Tofail, Kénitra

Examineur :

Pr. Mahmoud EL HAMLAOUI, MCH, ENSIAS, Université Mohammed V de Rabat

Résumé: Cette thèse propose un cadre unifié et extensible pour la modélisation et l'interrogation des systèmes non relationnels, reposant sur des hypergraphes typés à propriétés, des transformations à base de graphes et des pipelines assistés par l'intelligence artificielle. La première contribution introduit une méthodologie de conception de schémas de données, à la fois conceptuelle et logique, adaptée aux systèmes de données hybrides. Cette approche offre une manière de modélisation rigoureuse pour représenter les entités, les relations multiples, les rôles et les hiérarchies. Le modèle est implémenté via un pipeline de transformation et encodage conceptuel vers logique, permettant des transformations canoniques vers des schémas physiques de données pour les bases de données de type document, graphe et clé-valeur. La deuxième contribution majeure est un cadre de méta-requêtage : un pipeline de transformation multi-étapes et indépendant des langages, permettant l'interopérabilité sémantique entre des systèmes NoSQL hétérogènes. À travers trois étapes de transformation, les requêtes sont analysées dans un format logique structuré, représentées sous forme de graphes acycliques dirigés (DAG) puis traduites en requêtes syntaxiquement correctes et sémantiquement équivalentes pour les bases de données cibles. Des grands modèles de langage (LLM) sont intégrés pour améliorer l'analyse, la transformation et l'optimisation. La troisième contribution propose un cadre d'évaluation de la qualité, indépendant des bases de données, visant à examiner des critères tels que la flexibilité, la complétude et la correction. Ce cadre sert de guide pour la modélisation et aborde la dimension, largement négligée, de la qualité dans les modèles de données NoSQL. La quatrième et dernière contribution de ce travail est la proposition d'une approche collaborative et inclusive de développement NoSQL. En intégrant la modélisation participative, la modélisation visuelle, la modélisation et l'interrogation en langage naturel, et en proposant l'intelligence artificielle comme collaborateur au sein de l'écosystème, nous présentons une intégration complète, innovante et optimisée de l'IA tout au long du cycle de vie des bases de données NoSQL et non relationnelles.

Des études de cas approfondies, notamment dans le domaine des maladies respiratoires, de la recherche sur les épidémies et de l'analyse bibliographique, démontrent l'utilité, l'adaptabilité et les capacités intelligentes de l'approche proposée. Ce travail jette ainsi les bases d'une conception unifiée des schémas de données NoSQL, de l'ingénierie et de l'interrogation inter-plateformes à l'ère des données hybrides et de l'IA générative.

Mots-clés: Bases de Données NoSQL, Conception de Schémas, Évaluation de la qualité, Grands Modèles de Langage, Hypergraphes à propriétés, Intelligence artificielle, Méta-requête, Modélisation Assistée par l'IA.

Abstract: This dissertation proposes a unified and extensible framework for modeling and querying NoSQL systems grounded in typed property hypergraphs, graph-based transformations, and AI-assisted pipelines. The first contribution introduces a conceptual and logical schema design methodology tailored to hybrid data systems, offering a principled way of representing entities, multi-way relationships, roles, and hierarchies. This model is

implemented through a conceptual-to-logical mapping pipeline, with canonical transformations to physical schemas in document stores, graph databases, and key-value systems. The second major contribution is a meta-query based framework: a multi-stage, language-agnostic transformation pipeline that enables semantic interoperability across heterogeneous NoSQL systems. Through three transformational stages, queries are parsed into a structured logical format, represented as directed acyclic graphs (DAGs), and translated into syntactically correct, semantically equivalent queries for target databases. The framework integrates large language models (LLMs) to enhance parsing, transformation, and optimization. The third major contribution is a database-agnostic quality characterization framework, aiming to dissect various criteria such as flexibility, completeness, and correctness, act as a modeling guideline, and contribute to the largely overlooked dimension of quality in NoSQL data models. The fourth and culminating contribution of this work is the proposal of a collaborative and stakeholder-inclusive approach to NoSQL development. By integrating collaborative and participatory modeling, visual modeling, natural language-based modeling and querying, and proposing Artificial Intelligence as a collaborator in the ecosystem, we present a comprehensive, novel, and end-to-end integration of AI in an optimized NoSQL data lifecycle.

Extensive case studies, including applications in asthma and respiratory diseases, epidemic research, and bibliographic literature analysis, demonstrate the utility, adaptability, and intelligent components of the proposed approach. This work lays the foundation for unified NoSQL schema design, engineering, and cross-platform querying in the age of hybrid data and generative AI.

Keywords: AI-assisted modeling, Artificial Intelligence, Large Language Models, Meta-query, NoSQL Databases, Property Hypergraphs, Quality Assessment, Schema Design.