

الجمهورية الشعبية الديمقراطية الجزائرية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
المدرسة العليا للإعلام الآلي 080 ماي 5491. بسيدي بلعباس
École Supérieure en Informatique
-08 Mai 1945- Sidi Bel Abbès



ENGINEERING REPORT

To obtain the diploma of **Engineer**
Field: **Computer Science**
Specialty: **Intelligence Artificielle et Sciences de Données (IASD)**

Theme

**Querying Probabilistic OLAP data cubes with
hierarchies: models and algorithms**

Presented by:
TALBI Youcef
BENDOU Ahmed El Mahdi

Submission Date: **June, 2026**
In front of the jury composed of:

Prof. Nabil Keskes
Dr. Belkacem Khaldi
Prof. Alfredo Cuzzocrea
Dr. Mohamed Salih Bendella

President
Supervisor
Co-Supervisor
Examiner

Academic Year : 2025/2026

Abstract

This engineering report presents the end-to-end design, implementation, and evaluation of an OLAP pipeline that produces a fully allocated, uncertainty-aware multidimensional cube in a relational database. The contribution is a software architecture that implements eight allocation policies across four algorithm families, a Medallion-architecture ETL pipeline ingesting real ERA5 weather data into a SQL Server star schema, and a dual Power BI dashboard that compares naive OLAP output against the uncertainty-tagged allocated cube.

Keywords: OLAP, uncertain data, imprecise data, allocation policy, expectation-maximization, Dirichlet mixture, ETL pipeline, SQL Server Integration Services, Power BI, data warehousing.

Résumé

L'OLAP classique suppose que chaque fait est un point précis dans un cube multidimensionnel, mais les données réelles contiennent de l'incertitude dans les mesures et de l'imprécision dans les dimensions, nécessitant une allocation fondée des faits de région imprécis aux cellules feuilles. Les fondements théoriques de cette allocation ont été établis, mais leur réalisation en pipeline de production a reçu peu d'attention. Ce rapport d'ingénieur comble cet écart en présentant la conception et l'évaluation de bout en bout d'un pipeline OLAP implémentant huit politiques d'allocation, ingérant des données ERA5 via un pipeline ETL à architecture médaillon, et produisant un cube étiqueté d'incertitude exploitable par Power BI.

Mots-clés: OLAP, données incertaines, données imprécises, politique d'allocation, espérance-maximisation, mélange de Dirichlet, pipeline ETL, SQL Server Integration Services, Power BI, entreposage de données.

ملخص

تفترض نماذج OLAP التقليدية أن كل حقيقة تمثل نقطة دقيقة داخل مكعب متعدد الأبعاد، في حين أن البيانات الفعلية تنطوي على عدم يقين في المقاييس وعدم دقة في قيم الأبعاد، وهو ما يستوجب تخصيصاً قائماً على مبادئ منهجية لحقائق المناطق غير الدقيقة إلى الخلايا الطرفية. وقد تم تأسيس الأسس النظرية لهذا التخصيص، غير أن تجسيدها على هيئة خط معالجة إنتاجي لم يحظَ إلا باهتمام محدود. يسدّ هذا التقرير الهنسي تلك الثغرة، إذ يعرض التصميم والتقييم الشاملين من البداية إلى النهاية لخط معالجة OLAP إنتاجي ينفذ ثنائي سياسات تخصيص ضمن أربع عائلات خوارزمية، ويستوعب بيانات ERA5 الفعلية عبر خط معالجة ETL مُصمَّم وفق معمارية الميدالية (Medallion Architecture)، لينتج في النهاية مكعباً موسوماً بعدم اليقين قابلاً للاستخدام المباشر في Power BI.

الكلمات المفتاحية:

OLAP، بيانات غير مؤكدة (uncertain data)، بيانات غير دقيقة (imprecise data)، سياسة التخصيص (allocation policy)، خوارزمية التوقع والتعظيم (expectation-maximization)، خليط دركلييه (Dirichlet mixture)، خط معالجة ETL، خوارزمية التوقع والتعظيم والتعظيم (expectation-maximization)، خليط دركلييه (Dirichlet mixture)، خط معالجة ETL، Power BI، SQL Server Integration Services، مستودعات البيانات.