



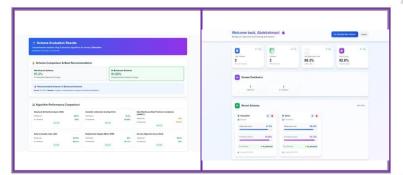


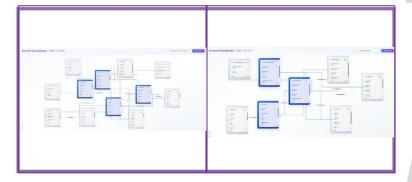
## كلية الحاسبات والمعلومات جامعة عين شمس

المعـرض العلمـى بالكليـة الفنية العسـكرية - يوليو ٢٠٢٥

# **DataForge**







#### أوجه الاستفادة من الدراسة

- تسريع عملية التصميم: توفير ما يصل إلى 90 % من الوقت مقارنة بالتصميم
  - تقليل الأخطاء : زيادة موثوقية المخازن وتقليل إعادة العمل.
  - دعم قطاعات متعددة :قابلية التطبيق في المالية، والرعاية الصحية، والتجزئة.
- تحسين اتخاذ القرار :تمكين المحللين من تعديل المخطط وتحديثه في الوقت الحقيقي

#### الهدف من المشروع

- تطوير أداة آلية قائمة على الذكاء الاصطناعي لتوليد مخططات مستودعات البيانات من SQL DDL بسرعة وكفاءة.
  - تقليل الجهد اليدوي والأخطاء البشرية في تصميم المخازن.
- دعم تحليلات البيانات الفورية عبر واجهة تفاعلية تسمح بالتحرير الفوري والتحقق

#### الدراســـة النظريــــة

- التحليل الهجين لـ :SQL دمج التحليل المبني على التعابير النمطية (Regex) مع شجرة البنية المجردة (AST) لاستخراج الجداول والأعمدة والعلاقات.
- تصنيف المخططات بالقاعدة العرافية :استخدام كثافة المفاتيح الأجنبية ونسب عددية وأعتاب التكرار لتصنيف الجداول إلى نجوم أو سنوفليك (star/snowflake)
- تحسين مدفوع بالذكاء الاصطناعي : تطبيق نموذج BERT مُدرّب على مخططات JSON لتحديد مجالات البيانات (domain detection) واقتراح

#### التصميـــمات

الهيكلية العامة للنظام:

واجهة أمامية تفاعلية مبنية على ReactFlow وReact لعرض المخططات وتحريرها. خادم تطبيقات ( (Django يستقبل SQL ويحوّله إلى JSONحدانقياً.

خدمات AWS ECSلتشغيل التحليل والتصنيف والتحسين.

تخزين المخرجات في خدمة تخزين الكائنات (.(Object Storage

سىر عمل :DataForge

SQL Input  $\to$  2. Hybrid Parsing (Parsed Output)  $\to$  3. Heuristic Classification (Classified Output)  $\to$  4. Al Optimization (Optimized Output)  $\to$  5. Interactive Visualization

### أهم نتائج الدراسـة

النتيجة	المقياس
96 %	دقة تحليل (Parsing)
92 %	دقة تحديد المجال ((Domain
< 5 ثواني	زمن توليد المخطط (100 جدول)
80 %	خفض الأخطاء
4.2/5	رضا المستخدم (على مقياس ١-٥)





