



# كلية الحاسبات والمعلومات جامعة عين شمس



المعرض العلمي بالكلية الفنية العسكرية - يوليو ٢٠٢٥

## DataForge

### الهدف من المشروع

- تطوير أداة 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 بالتحليل الفوري والتحقق من المتطلبات.

### الدراسة النظرية

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

### التصميمات

- الهيكلية العامة للنظام:
- واجهة أمامية تفاعلية مبنية على React و ReactFlow لعرض المخططات وتحريرها.
- خادم تطبيقات (Django) يستقبل SQL ويحوّله إلى JSON لاحقاً.
- خدمات AWS ECS لتشغيل التحليل والتصنيف والتحسين.
- تخزين المخرجات في خدمة تخزين الكائنات (Object Storage).

سير عمل DataForge:

SQL Input → 2. Hybrid Parsing (Parsed Output) → 3. Heuristic Classification (Classified Output) → 4. AI Optimization (Optimized Output) → 5. Interactive Visualization

### أهم نتائج الدراسة

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

### أوجه الاستفادة من الدراسة

- تسريع عملية التصميم : توفير ما يصل إلى 90 % من الوقت مقارنة بالتصميم اليدوي.
- 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 بسهولة.

بالتعاون مع

أكاديمية البحث العلمي والتكنولوجيا



تحت إشراف  
الكلية الفنية العسكرية  
مركز الاتصال للبحوث والتطوير