

تحلیل داده با DuckDB

DuckDB for Data Analysis

آموزش تحلیل داده با SQL و DuckDB در محیط محلی؛ از اجرای پرس‌وجو روی فایل‌های CSV تا کار با JSON، Join و داده‌های راه دور، ساخت داشبورد و اجرای پروژه‌ی پایانی.

سطح دوره	مدت پیشنهادی	بخش و درس	شیوه‌ی برگزاری
مقدماتی تا متوسط	۴ روز (۳۲ ساعت)	۱۱ بخش، ۵۴ درس	حضوری یا آنلاین

— آنچه شرکت‌کنندگان می‌آموزند

- ◆ نصب و اجرای DuckDB از خط فرمان و پایتون روی هر سیستم‌عامل
- ◆ اجرای کوئری SQL مستقیماً روی فایل‌های CSV، JSON و Parquet، بدون بارگذاری و بدون سرور پایگاه داده
- ◆ Join، تجمیع و تحلیل کارایی کوئری‌ها با توابع پنجره‌ای و EXPLAIN
- ◆ ساخت اپلیکیشن‌ها و داشبوردهای داده‌ی محلی با Apache Superset و Streamlit

پیش‌نیازها

- برای بخش‌های ۱ تا ۶ نیازی به پایتون نیست؛ از بخش ۷ به بعد آشنایی مقدماتی با پایتون توصیه می‌شود
- نیازی به تجربه‌ی قبلی SQL یا پایگاه داده نیست؛ SQL لازم در طول دوره آموزش داده می‌شود
- آشنایی با خط فرمان و یک ویرایشگر کد
- Docker فقط برای دو درس از بخش ۱۰ لازم است

ابزارها و فناوری‌ها

Python SQL DuckDB
Apache Superset Streamlit
Docker

پروژه‌ی عملی

جریان کامل تحلیل داده با DuckDB برای فروشگاه Thistlewood Goods

۶ درس	۱ آشنایی با DuckDB	<p>۱. چرا DuckDB؟ تحلیل سریع با SQL جاسازی‌شده</p> <p>۲. مقایسه‌ی DuckDB با Postgres، SQLite و pandas</p> <p>۳. نصب DuckDB (خط فرمان و پایتون) و اتصال</p> <p>۴. اولین کوئری‌ها روی یک فایل واقعی</p> <p>۵. رابط کاربری DuckDB: بررسی بصری داده</p> <p>۶. پروژه‌ی راهنمادار: بررسی کامل یک فایل CSV</p>
۶ درس	۲ مبانی SQL به روش DuckDB	<p>۱. جدول‌ها، سطرها و انواع داده</p> <p>۲. فیلتر و مرتب‌سازی با WHERE و ORDER BY</p> <p>۳. تجمیع با GROUP BY</p> <p>۴. Join کردن جدول‌ها</p> <p>۵. CTE و زیرکوئری</p> <p>۶. پروژه‌ی راهنمادار: پاسخ به پرسش‌های کسب‌وکار با SQL</p>
۶ درس	۳ خواندن مستقیم فایل‌ها: CSV، Parquet و چندین فایل	<p>۱. کوئری روی CSV بدون بارگذاری</p> <p>۲. Parquet: قالب محبوب DuckDB</p> <p>۳. خواندن هم‌زمان چندین فایل با الگوهای Glob</p> <p>۴. تشخیص خودکار Schema و زمان تعیین دستی آن</p> <p>۵. خروجی‌گرفتن از نتایج با COPY</p> <p>۶. پروژه‌ی راهنمادار: اسکرپت دریافت داده از چندین فایل</p>
۵ درس	۴ ماندگاری داده: حافظه در برابر فایل پایگاه داده	<p>۱. پایگاه داده‌ی درون‌حافظه‌ای در برابر ماندگار</p> <p>۲. بارگذاری داده در جدول‌ها</p> <p>۳. View در برابر جدول</p> <p>۴. کار با چند پایگاه داده (ATTACH)</p> <p>۵. پروژه‌ی راهنمادار: ساخت پایگاه داده‌ی تحلیلی محلی</p>
۵ درس	۵ کار با JSON و داده‌های تودرتو	<p>۱. خواندن مستقیم فایل‌های JSON</p> <p>۲. کوئری روی Struct و List در SQL</p> <p>۳. تبدیل JSON تودرتو به جدول</p> <p>۴. ذخیره‌ی نتایج به صورت JSON</p> <p>۵. پروژه‌ی راهنمادار: ساختاردهی یک خروجی تودرتو</p>
۵ درس	۶ SQL پیشرفته برای تحلیل	<p>۱. توابع پنجره‌ای: ROW_NUMBER، RANK و سایر</p> <p>۲. جمع تجمعی و میانگین متحرک</p> <p>۳. PIVOT و UNPIVOT</p> <p>۴. توابع تاریخ و زمان</p> <p>۵. پروژه‌ی راهنمادار: گزارش تحلیل فروش</p>

۱. API پایتون: اتصال‌ها و Relationها
۲. کوئری مستقیم روی DataFrame های pandas و Polars
۳. دریافت نتایج به صورت pandas, Polars یا Arrow
۴. اسکریپت یا خط فرمان: کدام و کی؟
۵. پروژه‌ی راهنمادار: جریان پردازش پایتونی مبتنی بر DuckDB

۱. افزونه‌ی https: کوئری روی فایل‌ها از طریق HTTP و S3
۲. Parquet پارتیشن‌بندی‌شده و Predicate Pushdown
۳. خواندن مستقیم یک مجموعه داده‌ی عمومی از URL
۴. مدیریت داده‌های بزرگ‌تر از RAM
۵. پروژه‌ی راهنمادار: کوئری روی فایل راه دور بدون دانلود

۱. EXPLAIN و خواندن طرح اجرای کوئری
۲. راز سرعت DuckDB: اجرای ستونی و برداری
۳. ایندکس و ترتیب: چه زمانی واقعاً کمک می‌کنند
۴. تحلیل و رفع یک کوئری کند
۵. پروژه‌ی راهنمادار: افزایش سرعت یک گزارش واقعی

۱. از کوئری تا اپلیکیشن: اپلیکیشن داده‌ی محلی چیست
۲. داشبورد محلی با Streamlit و DuckDB
۳. تبدیل یک کوئری به ابزار خط فرمان قابل استفاده‌ی دوباره
۴. نصب Apache Superset با Docker و اتصال به DuckDB
۵. پروژه‌ی راهنمادار: ساخت یک داشبورد واقعی در Superset

۱. پروژه‌ی پایانی: جریان تحلیل داده با DuckDB برای Thistlewood Goods

درخواست برگزاری اختصاصی این دوره برای سازمان شما

ثبت درخواست این دوره

lotfinejad.ir/contact

سرفصل، مدت، سطح و تمرین‌های دوره بر اساس نیاز تیم شما تنظیم می‌شود و آموزش به زبان فارسی، به صورت حضوری یا آنلاین برگزار می‌شود. برای ثبت درخواست، فرم مربوط به همین دوره را تکمیل کنید.