Bradley Haskell

NVIDIA Data Engineering Insights

https://github.com/bradhaskell/sql-project

Project Proposal

Job Description

Data Engineering at NVIDIA focuses on building complex data platforms for advanced data analysis to enhance GPU server manufacturing. The role involves designing and maintaining ELT pipelines, ensuring data integrity, optimizing query performance and security through SQL.

This role directly aligns with my career goals in ISBA, providing an opportunity to work with large-scale datasets, optimize data processing systems, and apply SQL in real-world scenarios. Along with my strong interest in the tech industry, NVIDIA's ability to maintain technological developments, and sustain satisfied employees furthers my interest in this company and position. By the end of this project, I will have gained realistic experience in constructing ETL pipelines and querying company data in an industry and corporation I have a strong draw to.

Problem

The GPU market is dominated by NVIDIA and AMD. Comparing their GPU performance can offer strategic insights for consumers, manufacturers, and businesses making hardware decisions. The challenge lies in processing and interpreting benchmark and product data from both companies to reveal trends in performance, efficiency, and value.

Data Sources

1. API Data Sources
    a. NVIDIA Developer Portal: Offers tools and APIs for querying NVIDIA products and driver information
    b. Kaggle API: Useful for collecting benchmark results comparing NVIDIA to AMD on various titles and tasks
2. Web Scraping Sources
    a. TechPowerUp GPU Database: Contains detailed specs and benchmarks for both NVIDIA and AMD GPUs
    b. PassMark Benchmark: Provides comparative performance scores, prices, and popularity rankings for GPUs

These sources simulate real world engineering data pipelines at NVIDIA. They provide data for conducting SQL queries and present dashboards that benchmark NVIDIA and AMD GPUs.

Solution

To address this problem, I will collect GPU data using API and web scraping, store raw data into the relational database created from class examples, use SQL to clean, normalize and aggregate benchmark and price data, and create dashboards comparing NVIDIA and AMD GPUs across performance, cost, and efficiency. Once these steps have been completed, I will move into visualization and presentation, using applications such as Excel and PopSQL to create bar charts and line graphs displaying product performance.