1. MỞ ĐẦU

## Tính cấp thiết của đề tài

Ngày nay, các công ty thu thập được rất nhiều dữ liệu từ các hành vi, tương tác của người dùng. Dữ liệu đã dần trở thành một tài sản quan trọng của các công ty này. Làm sao để tổ chức, quản lý và tận dụng nguồn dữ liệu này để hiểu người dùng, tăng lợi nhuận, và tạo được lợi thế cạnh tranh là mối quan tâm hàng đầu của các công ty. Điều này làm cho nhu cầu và công việc liên quan đến dữ liệu ở Việt Nam nói riêng và thế giới nói chung đang rất phát triển. Kể từ khi các doanh nghiệp bắt đầu nhận ra giá trị của dữ liệu, các kiến trúc mới bắt đầu ra đời như Data Lake, hay Data Warehouse, mỗi kiến trúc đều có ưu điểm và nhược điểm của riêng mỗi kiến trúc.

Vài năm trở lại đây, việc doanh nghiệp đang dịch chuyển dần cơ sở hạ tầng của họ lên các công nghệ điện toán đám mây dần mở đường cho sự bùng nổ của khối lượng dữ liệu. Vậy nên các kiến trúc cũ có thể không đáp ứng được một số nhu cầu của bản thân các doanh nghiệp nữa, dẫn tới các kiến trúc mới hơn ra đời nhằm giải quyết các bài toán đó. Một trong những kiến trúc khá nổi tiếng phải kể tới đó là kiến trúc Data Lakehouse, sự kế thừa các ưu điểm và khắc phục những yếu điểm của các kiến trúc đi trước.

Data Lakehouse là một trong những kiến trúc mới để tổ chức và xử lý dữ liệu lớn. tận dụng sự tuyệt vời từ các khả năng đáp ứng độ sẵn sàng cao của cộng nghệ điện toán đám mây và tính toán phân tán. Nhận thức được như cầu của thực tế nhóm quyết định chọn chủ đề tiểu luận chuyên ngành “Xây dựng Data Lakehouse để phân tích hoạt động kinh doanh nhà hàng” .

Ngày nay, các công ty thu thập được rất nhiều dữ liệu từ các hành vi, tương tác của người dùng. Dữ liệu đã dần trở thành một tài sản quan trọng của các công ty này. Làm sao để tổ chức, quản lý và tận dụng nguồn dữ liệu này để hiểu người dùng, tăng lợi nhuận, và tạo được lợi thế cạnh tranh là mối quan tâm hàng đầu của các công ty. Điều này làm cho nhu cầu và công việc liên quan đến dữ liệu ở Việt Nam nói riêng và thế giới nói chung đang rất phát triển. Kể từ khi các doanh nghiệp bắt đầu nhận ra giá trị của dữ liệu, các kiến trúc mới bắt đầu ra đời như Data Lake, hay Data Warehouse, mỗi kiến trúc đều có ưu điểm và nhược điểm của riêng mỗi kiến trúc.

Vài năm trở lại đây, việc doanh nghiệp đang dịch chuyển dần cơ sở hạ tầng của họ lên các công nghệ điện toán đám mây dần mở đường cho sự bùng nổ của khối lượng dữ liệu. Vậy nên các kiến trúc cũ có thể không đáp ứng được một số nhu cầu của bản thân các doanh nghiệp nữa, dẫn tới các kiến trúc mới hơn ra đời nhằm giải quyết các bài toán đó. Một trong những kiến trúc khá nổi tiếng phải kể tới đó là kiến trúc Data Lakehouse, sự kế thừa các ưu điểm và khắc phục những yếu điểm của các kiến trúc đi trước.

Data Lakehouse là một trong những kiến trúc mới để tổ chức và xử lý dữ liệu lớn. tận dụng sự tuyệt vời từ các khả năng đáp ứng độ sẵn sàng cao của cộng nghệ điện toán đám mây và tính toán phân tán. Nhận thức được như cầu của thực tế nhóm quyết định chọn chủ đề tiểu luận chuyên ngành “Xây dựng Data Lakehouse để phân tích hoạt động kinh doanh nhà hàng” .