

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



Lê Dương Phong

**NGHIÊN CỨU PHÁT TRIỂN NỀN TẢNG TÍCH
HỢP PHÂN TÍCH DỮ LIỆU DÒNG**

Chuyên ngành: Hệ thống thông tin

Mã số: 8.48.01.04

TÓM TẮT LUẬN VĂN THẠC SĨ

TP.HCM - NĂM 2023

Luận văn được hoàn thành tại:
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: **PGS.TS. THOẠI NAM**

Phản biện 1: -----

Phản biện 2: -----

Luận văn sẽ được bảo vệ trước Hội đồng chấm luận
văn thạc sĩ tại Học viện Công nghệ Bưu chính Viễn
thông

Vào lúc: giờ ngày tháng năm 2023.

Có thể tìm hiểu luận văn tại:

- Thư viện của Học viện Công nghệ Bưu chính
Viễn thông.

PHẦN MỞ ĐẦU

1. Lý do chọn đề tài

Hiện nay, theo xu hướng xây dựng đô thị thông minh tại Việt Nam cũng như trên thế giới, hệ thống camera giám sát an ninh, giao thông, hỗ trợ du lịch là một thành phần cấu thành không thể thiếu luôn được ưu tiên khi lựa chọn đầu tư triển khai. Việc lắp đặt camera giám sát an ninh ở khu dân cư, các nút giao thông, các điểm du lịch với mục đích chính là phục vụ hiệu quả công tác phòng, chống các loại tội phạm về trật tự xã hội, bảo đảm an ninh trật tự trên địa bàn, góp phần giảm thiểu tai nạn giao thông, ùn tắc giao thông. Bên cạnh hệ thống camera giám sát an ninh, hệ thống camera còn tích hợp các công nghệ thông minh để hỗ trợ trong việc nhận diện biển số xe, nhận diện khuôn mặt, đo đếm lưu lượng phương tiện giao thông tại các điểm cửa ngõ của tỉnh/thành phố; hỗ trợ phát hiện, theo dõi các xe nghi ngờ, lưu trữ và trích xuất dữ liệu phục vụ công tác điều tra của các cơ quan quản lý nhà nước, v.v.

Đối với các hệ thống giám sát đặc biệt là hệ thống giám sát giao thông hiện đại ngày nay, số lượng dữ liệu

được sinh ra ngày càng tăng do các hệ thống này được kết nối vô số cảm biến. Các cảm biến này có thể được lắp đặt trên các phương tiện giao thông di chuyển trên đường (thiết bị giám sát hành trình) hay là các hệ thống camera giám sát trên đường, bảng báo điện tử, thiết bị di động, v.v. Để đối phó với dữ liệu phức tạp, các hệ thống giám sát cần phải được xây dựng trên hạ tầng dữ liệu hiện đại, có khả năng lưu trữ, xử lý cũng như truy vấn khối lượng lớn dữ liệu.

Vì vậy, việc nghiên cứu phát triển nền tảng tích hợp phân tích dữ liệu dòng trong thời gian thực ở thời điểm hiện tại là rất cần thiết, đáp ứng nhu cầu xây dựng đô thị thông minh của các địa phương. Đó cũng chính là động lực để thực hiện luận văn này.

2. Mục tiêu và nhiệm vụ nghiên cứu

Mục tiêu của luận văn hướng đến là hiện thực vận dụng các giải pháp cho bài toán lưu trữ dữ liệu đo đếm phương tiện giao thông qua các công việc như sau:

- Nghiên cứu các giải pháp lưu trữ dữ liệu;
- Đề xuất giải pháp lưu trữ cho hệ thống tích hợp lưu trữ dữ liệu giao thông;

- Hiện thực triển khai thực tế giải pháp lưu trữ dữ liệu lớn cho dữ liệu đo đếm phương tiện giao thông song song với việc đánh giá hiệu năng;
- Hiện thực mô hình dự báo ngắn hạn lưu lượng giao thông sử dụng Support Vector Regression.

3. Phạm vi nghiên cứu

- Tìm hiểu kiến trúc Data Lakehouse;
- Tìm hiểu công nghệ lưu trữ dữ liệu lớn;
- Tìm hiểu giải thuật Support Vector Regression;
- Xây dựng kiến trúc triển khai thí điểm giải pháp trên thực tế;
- Đánh giá thực nghiệm dựa trên dữ liệu thực.

4. Kết cấu luận văn

Chương 1: Giới thiệu

Chương 2: Cơ sở lý thuyết

Chương 3: Bài toán và giải pháp cho hệ lưu trữ và truy vấn dữ liệu giao thông

Chương 4: Thực nghiệm và đánh giá kết quả

Chương 5: Kết luận

CHƯƠNG 2: CƠ SỞ LÝ THUYẾT

2.1. Apache Kafka

Apache Kafka là một hệ thống nhắn tin đăng ký / xuất bản mã nguồn mở, được thiết kế để xây dựng các đường ống dữ liệu trực tuyến thời gian thực lấy dữ liệu giữa nhiều hệ thống hoặc ứng dụng độc lập một cách đáng tin cậy. Nó thường được mô tả là “nhật ký cam kết phân tán” hoặc gần đây hơn là “nền tảng phát trực tuyến phân phối”. Hệ thống tệp hoặc nhật ký cam kết cơ sở dữ liệu được thiết kế để cung cấp một bản ghi lâu dài về tất cả các giao dịch để chúng có thể được phát lại để dễ dàng xây dựng trạng thái của hệ thống. Tương tự, dữ liệu trong Kafka được lưu trữ lâu dài, theo thứ tự và có thể được đọc một cách xác định. Ngoài ra, dữ liệu có thể được phân phối trong hệ thống để cung cấp các biện pháp bảo vệ bổ sung chống lại các lỗi, cũng như không có cơ hội đáng kể để mở rộng hiệu suất.

2.2. Apache Spark

Apache Spark là một framework xử lý dữ liệu giúp nhanh chóng thực hiện các tác vụ xử lý trên các tập dữ liệu

rất lớn. Spark giúp phân phối các tác vụ xử lý dữ liệu trên nhiều máy tính hoặc cùng với các công cụ tính toán phân tán khác. Spark cũng giúp giảm gánh nặng cho các nhà phát triển với các công cụ đơn giản hoá việc phân bổ tài nguyên cho lưu trữ và tính toán song song, phân tán trên nhiều node.

2.3. Tình hình nghiên cứu trong nước

Trong nhiều năm gần đây, yêu cầu phát triển đô thị thông minh được lãnh đạo nhiều tỉnh/thành phố rất quan tâm. Nhiều dự án thí điểm về đô thị thông minh được xây dựng và triển khai như tại Bình Phước, Bắc Kạn, Thanh Hoá, v.v. Tuy nhiên bài toán đô thị thông minh là bài toán khó, nên các dự án thí điểm này chủ yếu tập trung vào một số nội dung chính như:

- Chính quyền điện tử
- Xây dựng trung tâm IOC cho đô thị
- Xây dựng giải pháp CCTV giám sát đô thị

Riêng đối với bài toán giao thông, đa phần các dự án dừng ở mức giám sát và phát hiện các hành vi vi phạm bằng thủ công hoặc bán tự động để xử lý phạt nguội. Khi số lượng camera bùng nổ dẫn đến nhu cầu giám sát giao thông

tự động bằng phần mềm trí tuệ nhân tạo tăng cao. Tại Việt Nam có nhiều nhóm nghiên cứu, công ty xây dựng các giải pháp ứng dụng trí tuệ nhân tạo trong giám sát giao thông như Đại học Bách Khoa, Đại học quốc gia Thành phố Hồ Chí Minh, Viettel, VNPT v.v. Các ứng dụng này chủ yếu khai thác các mô hình trí tuệ nhân tạo để phân tích tự động hình ảnh camera. Đối với bài toán tích hợp và lưu trữ, bất kỳ một trong những đặc điểm trên của dữ liệu ITS đều có thể tạo ra thách thức đối với các hệ quản trị cơ sở dữ liệu truyền thống và một số đặc điểm là không thể xử lý được đối với các hệ thống lưu trữ dữ liệu truyền thống. Do đó, để đối phó với bài toán thu thập, tích hợp, lưu trữ dữ liệu giao thông cần khai thác tối đa sức mạnh của công nghệ trong đó phải kể đến các công nghệ về xử lý dữ liệu lớn. Vì vậy bài toán thu thập, tích hợp, lưu trữ dữ liệu giao thông là một bài toán khó. Để xây dựng được giải pháp giải quyết bài toán khó này đòi hỏi phải khai thác được triệt để sức mạnh công nghệ tiên tiến đồng thời kết hợp được kiến thức sâu chuyên môn về giao thông vận tải, đặc biệt sự am hiểu về giao thông đặc thù của đô thị.

2.4. Cơ sở lý luận

Hệ thống giao thông là một dạng hệ thống phức tạp được cấu thành từ nhiều thành phần từ phần cứng thiết bị, đường truyền viễn thông, hạ tầng máy chủ, phần mềm ứng dụng... Các hệ thống ITS trong quá trình vận hành sinh ra một lượng dữ liệu khổng lồ. Đặc điểm của dữ liệu này có thể được mô tả gói gọn trong 5 tính chất “5V” của dữ liệu lớn: (1) Volume – dung lượng dữ liệu, Variety - đa dạng, (3) Velocity – tốc độ, (4) Veracity – tính xác thực, và (5) Value - giá trị.

Do đó, để đối phó với bài toán thu thập, tích hợp, lưu trữ dữ liệu giao thông cần khai thác tối đa sức mạnh của công nghệ trong đó phải kể đến các công nghệ về lưu trữ và xử lý dữ liệu lớn.

2.5. Lý thuyết về các kiến trúc và thuật ngữ

2.5.1. Data Warehouse

Đây là nơi lưu trữ dữ liệu bằng thiết bị điện tử của một tổ chức, doanh nghiệp, nhằm hỗ trợ việc phân tích dữ liệu và lập báo cáo. Về cơ bản, có thể hiểu, Data Warehouse là một tập hợp dữ liệu hoặc thông tin có chung một chủ đề,

được tổng hợp từ nhiều nguồn khác nhau trong nhiều mốc thời gian.

Đối với kiến trúc Data Warehouse, dữ liệu có cấu trúc từ các database thông qua quá trình trích xuất, biến đổi sẽ được lưu vào “nhà kho”. Dữ liệu trong nhà kho này sẽ được sử dụng để xuất báo cáo, trực quan cho người sử dụng. Mục đích của kiến trúc Data Warehouse là phục vụ các yêu cầu phân tích, hoặc khai phá cụ thể được gọi là chủ đề.

2.5.2. Data Lake

Trong một nhánh tương tự như Data Warehouse, các kho dữ liệu cục bộ (data mart) nổi lên với một mục đích sử dụng cụ thể hoặc được phân loại theo một chất lượng nhất định. Data mart đã thành công hơn vì việc sử dụng dữ liệu được hiểu rõ hơn, và kết quả có thể được cung cấp nhanh hơn. Tuy nhiên, tính chất ngăn cách của các Data Mart đã khiến chúng trở nên ít hữu ích hơn đối với các bài toán có lượng dữ liệu khổng lồ, và cần sử dụng dữ liệu đó một cách đa chức năng

Vì lý do này, Data Lake đã phát triển do khả năng đáp ứng nhu cầu dữ liệu ở mọi quy mô. Chúng có thể tăng tốc mọi thứ, làm cho dữ liệu dễ sử dụng hơn cho các nhu

cầu chưa được xác định trước đó. Sự xuất hiện của điện toán đám mây quy mô lớn với sức mạnh tính toán khổng lồ và khả năng lưu trữ gần như vô hạn đã khiến phương pháp tiếp cận hồ dữ liệu này trở nên khả thi.

2.5.3. Data Lakehouse

Xu hướng hiện tại vẫn đang xây dựng mô hình dữ liệu hai tầng Lake + Warehouse. Những công nghệ đầu tiên được phát triển đó là các định dạng lưu trữ như Parquet và ORC cùng với những hỗ trợ của chúng, cho phép người dùng có thể truy vấn dữ liệu trong Data Lake với cùng một cơ chế SQL như trong Data Warehouse. Tuy nhiên chúng vẫn không làm cho dữ liệu trong Data Lake dễ quản lý hơn, và chúng cũng không thể giải quyết được sự thiếu linh động, phức tạp và những thách thức dữ liệu đến từ các phân tích bậc cao. Thứ hai là, đã có những đầu tư cho các cơ chế truy vấn có thể truy vấn trực tiếp từ Data Lake như Spark SQL, Trino, Hive hay AWS Athena. Tuy nhiên, những cơ chế này không giải quyết các vấn đề của Data Lake: thiếu hụt những tính năng quản lý dữ liệu như ACID Transaction hay những phương thức truy cập dữ liệu như indexing để đạt hiệu năng của Data Warehouse.

Data Lakehouse là một hệ thống quản lý dữ liệu dựa trên nền tảng lưu trữ chi phí thấp và các cơ chế truy cập dữ liệu mà trong đó các tính năng quản trị dữ liệu truyền thống từ các DBMS và các tính năng như ACID transaction, indexing, tối ưu truy vấn... được hiện thực. Data Lakehouse kết hợp những tính năng chủ chốt của cả Data Lake và Data Warehouse: lưu trữ chi phí thấp trên các định dạng mở và tính năng quản lý và tối ưu dữ liệu. Cần lưu ý rằng một hệ thống Lakehouse cần có khả năng triển khai trên môi trường đám mây, đó là lưu trữ phân tán và xử lý song song.

Kiến trúc Lakehouse được kì vọng sẽ giải quyết các vấn đề lớn mà mô hình hai lớp Data Lake và Data Warehouse, cho phép người sử dụng có một địa điểm tập trung duy nhất cho công việc phân tích dữ liệu, học máy cũng như xuất báo cáo. Đồng thời vẫn đảm bảo tốc độ nhanh chóng trong các truy vấn.

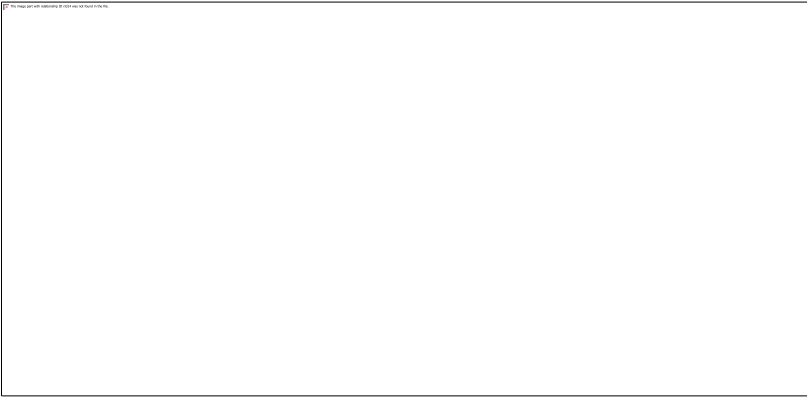
Bảng 2.1: So sánh giữa Data Warehouse, Data Lake và Data Lakehouse

	Data Warehouse	Data Lake	Data Lakehouse
Loại dữ liệu lưu trữ	Dữ liệu có cấu trúc	Dữ liệu bán cấu trúc và phi cấu trúc	Cả dữ liệu cấu trúc và phi cấu trúc
Tác vụ hỗ trợ	Sinh báo cáo	Học máy và phân tích dữ liệu	Phục vụ được cả hai tác vụ
Chi phí lưu trữ	Cao	Thấp	Thấp
ACID	Hỗ trợ	Không hỗ trợ	Hỗ trợ mức đọc ghi dữ liệu

CHƯƠNG 3: BÀI TOÁN HỆ LƯU TRỮ VÀ TRUY VẤN DỮ LIỆU GIAO THÔNG

3.1. Mô tả bài toán

Do tính chất 5V của dữ liệu giao thông đã phân tích ở Chương 1 phần Cơ sở lý luận, hệ thống giám sát giao thông có lượng dữ liệu lớn, cụ thể là dữ liệu đo đếm phương tiện, cần được lưu trữ trên hạ tầng có khả năng lưu trữ, xử lý và truy vấn khối lượng lớn dữ liệu.



Hình 3.1: Hệ thống đo đếm phương tiện giao thông

Trong lược đồ Hình 2.1, hệ thống phân tích trí tuệ nhân tạo xác định số lượng phương tiện đã được triển khai bao gồm hai chức năng:

- (1) *Xác định số lượng phương tiện* tự động thông qua phân tích trực tiếp các luồng camera giao thông phát hiện ra số lượng các loại phương tiện đang di chuyển ngang qua điểm giao thông được giám sát. Hệ thống trả về số lượng phương tiện theo từng loại.
- (2) *Đọc biểu số phương tiện* thông qua phân tích trực tiếp các luồng camera phát hiện các biểu số phương tiện đang lưu thông qua các nút giao thông. Kết quả trả về là thông tin biểu số các phương tiện.
- (3) *Dự báo số lượng phương tiện trong ngắn hạn* sử dụng phương pháp học máy truyền thống – cụ thể là mô hình Support Vector Regression.

Các kênh phân phối dữ liệu cũng được triển khai đồng thời với hệ thống phân tích trí tuệ nhân tạo, phục vụ nhu cầu của các ứng dụng khác trong hệ sinh thái giám sát giao thông chung. Các bài toán chính giải quyết trong luận văn như sau:

1. Giải pháp lưu trữ và truy vấn dữ liệu giao thông đô thị, cụ thể là dữ liệu đo đếm phương tiện giao thông.
2. Kỹ thuật nâng cao hiệu năng truy vấn.

3. Ứng dụng học máy vào trong công tác dự báo dữ liệu lưu lượng giao thông.

3.2. Các vấn đề phân tích để giải quyết bài toán

3.2.1. Phân tích đặc trưng dữ liệu thực tế

Số lượng của từng loại phương tiện được gom lại thành một chuỗi (array) 6 thành phần bao gồm: xe máy, xe bus; xe hơi; xe ô tô, xe tải; xe container.

Biểu số của từng phương tiện sẽ đi kèm với loại phương tiện đó. Tốc độ dữ liệu sinh ra từ hệ thống CCTV là 74k record/ngày.

3.2.2. Phân tích yêu cầu lưu trữ

- *Khả năng mở rộng*: Hệ thống lưu trữ cần có khả năng mở rộng để đáp ứng dữ liệu tăng lên liên tục về dung lượng và số lượng vì lượng dữ liệu do hệ thống phân tích trí tuệ nhân tạo sinh ra tăng theo cấp số.;
- *Khả năng cập nhật nhanh*: hệ thống lưu trữ cần có khả năng cập nhật nhanh dữ liệu thời gian thực vì yêu cầu giám sát thực tiễn về giao thông đã kéo theo tốc độ tạo ra dữ liệu. Trong đó nhiều dữ liệu được thu thập liên tục, theo thời gian thực. Ví dụ như dữ

liệu camera sẽ được thu thập thời gian thực với tần suất có thể tính theo đơn vị giây;

- *Khả năng xử lý dữ liệu:* Để có thể xuất những báo cáo cho người dùng cuối, hệ thống lưu trữ cần đáp ứng khả năng chuẩn hóa dữ liệu, đưa về định dạng cấu trúc nhất định đúng với yêu cầu đặt ra;
- *Khả năng quản trị dữ liệu:* Hệ thống lưu trữ cần có khả năng quản lý dữ liệu, để dữ liệu giữa các tác vụ biến đổi ổn định và nhất quán, không gây ra tình trạng dư thừa, sai lệch dữ liệu. Nói cách khác, đó chính là tính năng ACID Transaction trong Data Warehouse.

3.2.3. Phân tích yêu cầu truy vấn

Giải pháp truy vấn về dữ liệu có nhiều yếu tố cần đánh giá. Trong luận văn này quan tâm đến giải pháp truy vấn dữ liệu đo đếm phương tiện giao thông nên hướng đến hai tham số sau:

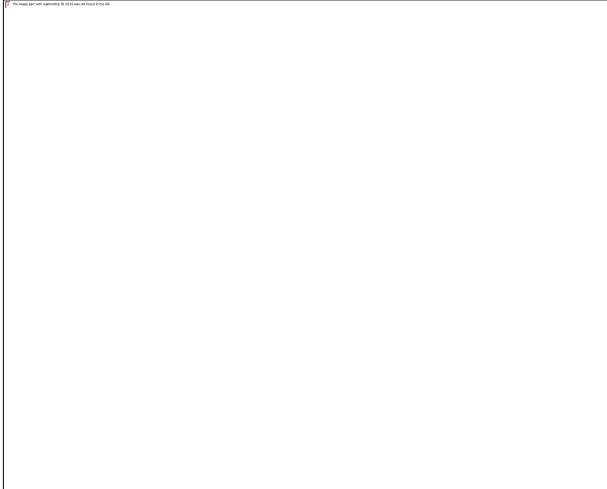
- *Thời gian đáp ứng:* Các báo cáo thời gian thực gặp vấn đề về độ trễ truy vấn dữ liệu;

- *Truy cập dễ dàng*: hệ thống lưu trữ sẽ cần hỗ trợ khả năng truy cập trực tiếp dữ liệu để phục vụ cho các công việc về học máy và phân tích dữ liệu cấp cao.

3.3. Đề xuất giải pháp cho hệ lưu trữ, truy vấn

3.3.1. Giải pháp công nghệ

Giải pháp Delta + HDFS



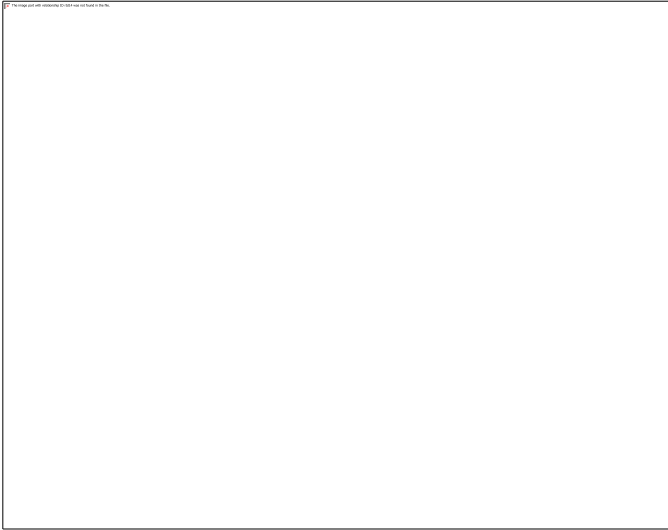
Hình 3.2: Giải pháp Delta + HDFS

Giải pháp Delta + MinIO



Hình 3.3: Giải pháp Delta + MinIO

Giải pháp Iceberg + MinIO + Trino



Hình 3.4: Giải pháp Iceberg + MinIO + Trino

3.3.2. Kỹ thuật tối ưu

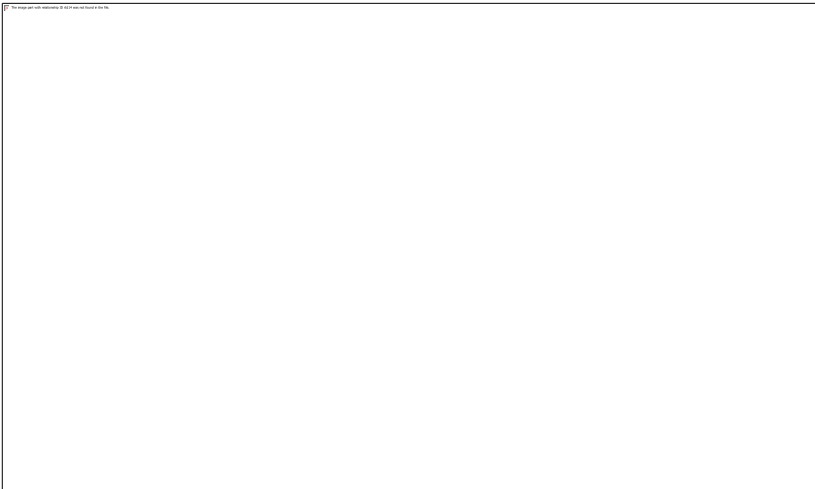
Mô hình dữ liệu tam cấp

Dữ liệu sẽ được lưu theo 3 tầng:

- Tầng dữ liệu thô hay còn gọi là Bronze là nơi lưu trữ những dữ liệu trích xuất nguyên bản từ nguồn thông tin mà không thông qua bất kỳ thay đổi nào.;
- Tầng dữ liệu sạch hay còn gọi là Silver là nơi tập trung những dữ liệu đã được làm sạch.;

- Tầng dữ liệu tổng hợp hay còn gọi là Gold là tầng lưu trữ những dữ liệu được thống kê biến đổi đặc trưng dành cho các bài toán cụ thể. Từ những dữ liệu ở tầng dữ liệu sạch, dữ liệu ở tầng này sẽ dùng cho mục đích trực quan và báo cáo, giúp cho người sử dụng có được lượng thông tin hữu ích để có thể ra các quyết định hoặc thiết lập các chiến lược.

Thiết kế lưu trữ và ETL cho dữ liệu đếm xe và biển số



Hình 3.5: Lưu đồ biến đổi dữ liệu đếm xe



Hình 3.6: Lưu đồ biến đổi dữ liệu đếm biến số

Kỹ thuật gom file và phân vùng dữ liệu

- *Tối ưu gom file* là ghi lại các file nhỏ thành một file dữ liệu lớn hơn một cách tối ưu. Table Format Delta Lake có thể giúp thực hiện quá trình này mà không làm thay đổi dữ liệu của file, điều này giúp cho việc gộp file có thể thực hiện đồng thời với ghi dữ liệu vào bảng
- *Phân vùng dữ liệu* là chia nhỏ một cơ sở dữ liệu lớn thành các phân vùng. Các phân vùng sẽ được truy xuất trực tiếp bằng các câu truy vấn SQL. Khi dữ liệu đã được phân vùng, các công việc chỉ cần thao tác trên các vùng nhỏ của dữ liệu, thay vì phải scan toàn bộ cơ sở dữ liệu.

CHƯƠNG 4: THỰC NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ

4.1. Mô hình triển khai



Hình 4.1: Mô hình kết nối camera

4.2. Kết quả và thực nghiệm đánh giá

4.2.1. Tóm tắt dữ liệu

- Số lượng bảng số xe trung bình theo ngày: 47.828
- Số lượng xe trung bình theo ngày: 36.645

4.2.2. Một số tính năng phân tích dữ liệu dòng giao thông

- Phân tích xe trong và ngoài tỉnh
- Phân tích lưu lượng xe

- Phân tích mật độ xe

4.2.3. Mô hình dự báo lưu lượng giao thông

Giải thuật Support Vector Regression Machines (SVR) là một phần mở rộng của Support-Vector Networks khi thay vì đưa ra các quyết định cho bài toán phân loại thì SVR được sử dụng để học và dự báo dữ liệu trong các bài toán hồi quy. SVR tận dụng được ưu điểm của SVM khi không phụ thuộc vào số chiều không gian của vector dữ liệu đầu vào và xử lý ở một chiều không gian đặc trưng cao hơn.

Bảng 4.1: Kiểm tra chất lượng dự báo

Tập dữ liệu	Số bước vào tương lai	R2
Validation	1	0.859
Test	1	0.851
Validation	5	0.822
Test	5	0.807

CHƯƠNG 5: KẾT LUẬN

5.1. Kết quả nghiên cứu của đề tài

Trong giải pháp giám sát giao thông đô thị, bài toán về đo đếm phương tiện giao thông là một bài toán cần thiết cho việc phân tích đánh giá mật độ giao thông. Lưu trữ được dữ liệu đo đếm phương tiện sẽ hỗ trợ nhiều báo cáo, phân tích, từ đó các nhà lãnh đạo có thể đưa ra những chính sách hợp lý để điều tiết phân luồng giao thông cũng như huy động lực lượng điều tiết tại các điểm nóng ùn tắc giao thông. Bài toán liên quan đến dữ liệu phương tiện giao thông đã được nghiên cứu và giải quyết trong luận văn này:

- Giải pháp lưu trữ và truy vấn dữ liệu đo đếm phương tiện giao thông với các kiến trúc đề xuất bao gồm: Delta + HDFS, Delta + MinIO, Iceberg + MinIO + Trino;
- Kỹ thuật nâng cao hiệu năng truy vấn: Gom file và Phân vùng dữ liệu.
- Ứng dụng học máy vào trong công tác dự báo dữ liệu lưu lượng giao thông.

Các giải pháp này cũng đã được triển khai đánh giá trên tập dữ liệu đo đếm và nhận diện biển số phương tiện giao thông thực với: 717.420 record dữ liệu biển số phương tiện và 549.675 record dữ liệu đo đếm số lượng phương tiện.

Kết quả nghiên cứu trong luận văn là cơ sở để triển khai hệ lưu trữ và truy vấn dữ liệu đo đếm phương tiện giao thông cho các đô thị.

5.2. Hạn chế luận văn

Số lượng mẫu dữ liệu còn hạn chế và thời gian nghiên cứu luận văn chưa đủ dài để đánh giá tính tối ưu nhất của giải pháp.

5.3. Hướng phát triển tiếp theo của đề tài nghiên cứu

Tiếp tục nghiên cứu triển khai đánh giá hiệu năng trên hệ thống phân tán nhiều máy chủ.

Cải tiến hiệu năng mô hình dự báo trong đó sẽ tiếp tục thu thập thêm dữ liệu mẫu và thử nghiệm dữ liệu mẫu này với các thuật giải học sâu như Deep Belief Network, Long-S (DBN), Long-Short Term Memory (LSTM), ...