

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



PHẠM THỊ XUYẾN

**HỆ THỐNG CẢNH BÁO BẤT THƯỜNG
CHO CẦU DÂY VĂNG**

LUẬN VĂN THẠC SỸ KỸ THUẬT

(Theo định hướng ứng dụng)

TP. HỒ CHÍ MINH – NĂM 2022

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



PHẠM THỊ XUYẾN

**HỆ THỐNG CẢNH BÁO BẤT THƯỜNG
CHO CẦU DÂY VĂNG**

CHUYÊN NGÀNH: HỆ THỐNG THÔNG TIN

MÃ SỐ: 8.48.01.04

LUẬN VĂN THẠC SỸ KỸ THUẬT

(Theo định hướng ứng dụng)

NGƯỜI HƯỚNG DẪN KHOA HỌC:

TS. HUỖNH TRỌNG THƯA

TP. HỒ CHÍ MINH – NĂM 2022

LỜI CAM ĐOAN

Tôi cam đoan rằng luận văn thạc sĩ hệ thống thông tin: “**Hệ thống cảnh báo bất thường cho cầu dây văng**” là công trình do tôi nghiên cứu tổng hợp và thực hiện.

Tôi cam đoan các số liệu, kết quả nêu trong luận văn là trung thực và chưa từng được ai công bố trong bất kỳ công trình nào khác.

Không có sản phẩm/nghiên cứu nào của người khác được sử dụng trong luận văn này mà không được trích dẫn theo đúng quy định.

TP. HCM, ngày 16 tháng 07 năm 2022

Học viên thực hiện luận văn

Phạm Thị Xuyên

LỜI CẢM ƠN

Trước tiên Tôi xin cảm ơn Thầy - **TS. HUỖNH TRỌNG THỬA** - Giảng viên hướng dẫn trực tiếp của tôi. Cảm ơn Thầy rất nhiều vì những góp ý và sự chỉ dẫn của Thầy đã giúp tôi hoàn thiện luận văn một cách dễ dàng nhất. Thầy luôn động viên cho phép tôi nêu quan điểm đồng thời đưa ra những nhận xét, dẫn dắt tôi đi đúng hướng trong suốt thời gian nghiên cứu, thực hiện đề tài luận văn thạc sĩ.

Tôi cũng xin trân trọng cảm ơn các thầy cô giáo Khoa Đào tạo Sau đại học - Học viện Công nghệ Bưu chính viễn thông Cơ sở Thành phố Hồ Chí Minh đã truyền đạt những kiến thức chuyên môn sâu rộng trong ngành cho tôi, có được nền tảng kiến thức hỗ trợ rất lớn cho tôi trong quá trình nghiên cứu, hoàn thiện luận văn.

Tôi cũng muốn bày tỏ sự biết ơn đến các Lãnh đạo trong công ty cổ phần đầu tư xây dựng Phú Mỹ đã cho phép Tôi được sử dụng dữ liệu trực tiếp về các thông số của cầu Phú Mỹ.

Cuối cùng, tôi xin gửi lời cảm ơn chân thành đến gia đình, bạn bè và các anh chị em trong lớp cũng như anh chị em công tác cùng lĩnh vực vì đã luôn đồng hành, tạo điều kiện, hỗ trợ và khuyến khích tôi cố gắng trong suốt những năm tháng học tập cũng như quá trình nghiên cứu viết luận văn này.

Xin chân thành cảm ơn!

TP.HCM, ngày 16 tháng 07 năm 2022

Học viên thực hiện luận văn

Phạm Thị Xuyên

MỤC LỤC

LỜI CAM ĐOAN	i
LỜI CẢM ƠN	ii
MỤC LỤC.....	iii
DANH SÁCH HÌNH VẼ.....	v
DANH SÁCH BẢNG	viii
DANH MỤC CÁC THUẬT NGỮ, CHỮ VIẾT TẮT	ix
MỞ ĐẦU	1
1. Lý do chọn đề tài	1
4. Đối tượng và phạm vi nghiên cứu	4
5. Phương pháp nghiên cứu	5
6. Cấu trúc luận văn	5
CHƯƠNG 1: NGHIÊN CỨU TỔNG QUAN	6
1.1. Tổng quan về Big Data	6
1.2. Các công nghệ Big Data	7
1.3. Cơ sở phân tích dữ liệu quan trắc cầu.....	7
1.4. Yêu cầu báo cáo số liệu cho hệ thống.....	8
CHƯƠNG 2: PHÂN TÍCH THIẾT KẾ	9
2.1 Mô tả vấn đề	9
2.2. Kiến trúc Big Data	9
2.2.1. <i>Lớp thu thập dữ liệu</i>	11

2.2.2. <i>Lớp xử lý dữ liệu</i>	14
2.2.3. <i>Lớp lưu trữ dữ liệu</i>	16
2.3. Đánh giá và lựa chọn mô hình phát triển hệ thống.....	19
CHƯƠNG 3: TRIỂN KHAI ỨNG DỤNG	21
3.1. Giới thiệu tổng quan về cầu dây văng Phú Mỹ	21
3.2. Thực trạng khai thác và bảo trì cầu dây văng Phú Mỹ	22
3.3. Thực tế triển khai ứng dụng.....	25
3.3.1. <i>Triển khai hệ thống thu thập, xử lý và lưu trữ dữ liệu từ các cảm biến</i>	25
3.3.2. <i>Phân tích và mô hình hóa dữ liệu</i>	31
3.3.3. <i>Quản lý khai thác và bảo trì cầu dây văng Phú Mỹ dựa trên kết quả phân tích</i>	39
CHƯƠNG 4: ĐÁNH GIÁ KẾT QUẢ THỬ NGHIỆM	40
4.1. Mô tả về quá trình truyền dữ liệu phát dữ liệu	40
4.2. Mô tả hệ thống lưu trữ, xử lý và quản trị dữ liệu	41
4.3. Phân tích và xuất trình báo cáo.....	42
4.3.1 <i>Theo dõi dữ liệu</i>	42
4.3.2. <i>Cảnh báo</i>	46
4.3.3. <i>Báo cáo</i>	48
4.4. Nhận xét kết quả thử nghiệm ứng dụng.....	49
KẾT LUẬN	51
DANH MỤC TÀI LIỆU THAM KHẢO	52

DANH SÁCH HÌNH VẼ

Hình 2.1: Kiến trúc Big Data tổng quan	10
Hình 2.2: Cấu trúc cơ bản của Kafka	11
Hình 2.3: Các thành phần chính của Kafka	12
Hình 2.4: So sánh Spark Structured Streaming với các công cụ truyền phát khác.....	14
Hình 2.5: Luồng dữ liệu dưới dạng bảng không giới hạn.....	15
Hình 2.6: Mô hình lập trình cho Structured Streaming.....	16
Hình 2.7: Minh họa Time-series Data.....	17
Hình 2.8: Bảng xếp hạng 10 cơ sở dữ liệu chuỗi thời gian hàng đầu thế giới theo DB-Engines	18
Hình 2.9: Mô hình xử lý tổng quan.....	19
Hình 2.10: Mô hình xử lý chi tiết của hệ thống	20
Hình 3.1: Đồ hình vị trí cầu Phú Mỹ.....	21
Hình 3.2: Kích thước hình học chính cầu dây văng Phú Mỹ.....	22
Hình 3.3: Kích thước mặt cắt ngang dầm cầu Phú Mỹ.....	22
Hình 3.4: Sơ đồ chi tiết hệ truyền dẫn	23
Hình 3.5: Dữ liệu nguồn.....	24
Hình 3.6: Chi tiết định dạng tên và cấu trúc lưu trữ các file dữ liệu.....	24

Hình 3.7: Các cột dữ liệu trong file csv thu được từ cRIO 9031	26
Hình 3.8: Các cột dữ liệu trong file csv thu được từ cRIO 9065	27
Hình 2.9: Sơ đồ luồng dữ liệu vào Kafka	28
Hình 3.10: Các bản ghi trong mỗi file dữ liệu thu được sau mỗi phút	28
Hình 3.11: Các thư mục lưu trữ dữ liệu	29
Hình 3.12: Sơ đồ thu thập dữ liệu từ hệ hệ quan trắc cầu	30
Hình 3.13: Mã nguồn lấy dữ liệu từ Kafka	30
Hình 3.14: Mã nguồn ghi dữ liệu vào Kafka	31
Hình 3.15: Mô hình cơ sở dữ liệu	31
Hình 3.16: Kết nối Grafana với InfluxDB	32
Hình 3.17: Truy vấn dữ liệu và trực quan hóa dữ liệu.....	33
Hình 3.18: Mã code biến đổi FFT xác định tần số dao động riêng của cáp văng.....	35
Hình 3.19: Cài đặt tần số và ngưỡng cảnh báo	36
Hình 3.20: Cài đặt kênh nhận cảnh báo	37
Hình 3.2: Cấu hình SMTP để gửi mail trong Grafana	37
Hình 3.22: Lựa chọn mốc thời gian cho các bảng báo cáo	38
Hình 3.23: Mã nguồn lấy và tính toán các giá trị min, max, mean từ cơ sở dữ liệu.....	38
Hình 4.1: Các file dữ liệu chuyển sang folder finished sau khi truyền dữ liệu kết thúc	40

Hình 4.2: Bảng điều khiển khi dữ liệu truyền phát trực tiếp bị gián đoạn.....	41
Hình 4.3: Cơ sở dữ liệu thực tế được ghi vào InfluxDB.....	41
Hình 4.4: Đồ thị theo dõi lực căng dây cáp 2215.....	42
Hình 4.5: Bảng điều khiển theo dõi các thông tin thời tiết	43
Hình 4.6: Bảng điều khiển theo dõi các thông tin dao động của 4 cáp theo các phương cùng các giá trị min, max, mean tương ứng.....	43
Hình 4.7: Bảng điều khiển theo dõi độ nghiêng đỉnh tháp và chân tháp	44
Hình 4.8: Bảng điều khiển theo dõi các thông tin dao động dầm theo các phương cùng các giá trị min, max, mean tương ứng.....	44
Hình 4.9: Dashboard của hệ thống giám sát cầu dây văng đã triển khai	45
Hình 4.10: Đồ thị hiển thị tốc độ gió bình thường.....	46
Hình 4.11: Đồ thị hiển thị tốc độ gió vượt ngưỡng cảnh báo được cài đặt.....	47
Hình 4.12: Cảnh báo gửi về email khi giá trị vượt ngưỡng cảnh báo đã cài đặt	47
Hình 4.13: Báo cáo kết quả thực tế theo Tuần.....	48
Hình 4.14: Báo cáo kết quả thực tế theo Tháng.....	48
Hình 4.15: Báo cáo kết quả tháng 3 năm 2022	49

DANH SÁCH BẢNG

Bảng 3.1: Danh sách các cảm biến.....	23
Bảng 3.2: Chi tiết về các trường dữ liệu của các file csv thu được từ cRIO 9031	26
Bảng 3.3: Chi tiết về các trường dữ liệu của các file csv thu được từ cRIO 9065	27
Bảng 3.4: Đặc trưng cấp.....	34
Bảng 4.1: Giá trị cảnh báo.....	46

DANH MỤC CÁC THUẬT NGỮ, CHỮ VIẾT TẮT

STT	Chữ viết tắt	Tiếng Anh	Tiếng Việt
1	Big Data	-	Dữ liệu lớn
2	URL	Uniform Resource Locator	Đường dẫn hay địa chỉ dùng để tham chiếu tới các tài nguyên trên mạng internet
3	FFT	Fast Fourier Transform	Biến đổi nhanh Fourier
4	HDFS	Hadoop Distributed File System	Hệ thống tập tin phân tán của Hadoop
5	DAU	Data Acquisition Unit	Hệ thống thu thập dữ liệu
6	SHM	Structural Health Monitoring	Hệ thống quan trắc cầu
7	TSDB	Time Series Database	Cơ sở dữ liệu chuỗi thời gian
8	IoT	Internet of Things	Internet Vạn Vật
9	SQL	Structured Query Language	Ngôn ngữ truy vấn có cấu trúc

MỞ ĐẦU

1. Lý do chọn đề tài

Với sự phát triển vượt bậc của khoa học kỹ thuật, số lượng cầu dây văng trên thế giới được xây dựng ngày càng nhiều. Do vai trò đặc biệt quan trọng của chúng, cộng với giá thành xây dựng rất đắt đỏ, các cầu này cần thiết phải được duy tu, bảo dưỡng thường xuyên và kịp thời. Quá trình này thường gặp nhiều khó khăn do nhiều lý do như vị trí xây dựng khắc nghiệt, số chi tiết nhiều, cấu tạo phức tạp, chi phí lớn... [1]. Chính vì vậy, nhằm đáp ứng nhu cầu thực tiễn của công tác duy tu, bảo dưỡng cầu, nhất là cầu nhịp lớn, hệ thống quan trắc cầu (Structural Health Monitoring – SHM) đã từng bước được áp dụng tại Việt Nam, đặc biệt sau khi Bộ GTVT ban hành Thông tư 52/2013/TT-BGTVT quy định về quản lý, khai thác và bảo trì công trình đường bộ, trong đó yêu cầu các công trình cầu có nhịp dài hơn 150m hoặc trụ cầu cao hơn 50m phải được lắp đặt hệ thống quan trắc [2].

Hệ thống quan trắc cầu dây văng là một hệ thống phức tạp, được tích hợp nhiều công nghệ tiên tiến nhằm quan trắc, giám sát trạng thái công trình liên tục trong thời gian thực [3]. Hiện nay, tại Việt Nam đã và đang xây dựng rất nhiều cầu lớn và hệ thống quan trắc giám sát sửa khỏe cầu đã được lắp đặt phổ biến trên các cây cầu lớn như: Cầu Cần Thơ, cầu Bãi Cháy, cầu Bình, cầu Phú Mỹ... nhưng phần lớn là hệ thống thiết bị và phần mềm do các công ty nước ngoài cung cấp, vì thế khi các sự cố xảy ra đối với hệ thống này rất khó có thể tự can thiệp và khắc phục [4]. Bên cạnh đó, việc quản lý, phân tích, xử lý và lưu trữ dữ liệu, thông số thu thập được từ hệ thống quan trắc trong quá trình khai thác ở nhiều công trình còn lúng túng, chưa hiệu quả và số lượng chuyên gia còn hạn chế [5].

Vấn đề đặt ra là làm sao triển khai được một hệ thống không những có thể quản lý khối lượng lớn dữ liệu thu được mỗi ngày một cách hiệu quả, dễ mở rộng, khả năng chịu lỗi cao mà còn có thể phân tích và đưa ra cảnh báo bất thường ghi nhận được về tình trạng kỹ thuật cầu, đảm bảo khai thác cầu an toàn giảm chi phí vận hành, bảo dưỡng cầu là một việc hết sức cấp thiết. Từ những lý do trên, tôi xin lựa chọn đề tài nghiên cứu “**Hệ thống cảnh báo bất thường cho cầu dây văng**”.

2. Tổng quan vấn đề nghiên cứu

Trên thế giới đã có nhiều công trình nghiên cứu ứng dụng công nghệ Big Data vào việc quản lý và phân tích dữ liệu cho hệ thống SHM. Seongwoon Jeong et al. [6, 7] đã đề xuất mô hình quản lý và phân tích dữ liệu lớn cho việc giám sát cầu dựa trên dữ liệu từ hệ thống quan trắc. Mô hình này bao gồm các máy ảo đám mây, hệ thống cơ sở dữ liệu phân tán như Cassandra, NoSQL cùng với Spark cluster và máy chủ. Kết quả cho thấy khung đề xuất không chỉ quản lý dữ liệu lớn và phức tạp được thu thập từ việc giám sát cầu, mà còn đối phó với các nhu cầu thay đổi về xử lý dữ liệu.

S. Gunner et al. [8] triển khai hệ thống quản lý dữ liệu mã nguồn mở tích hợp với hệ thống quan trắc công trình cầu sử dụng mạng cảm biến không dây cho cầu treo Clifton, Vương Quốc Anh. Hệ thống cho phép thu thập, lưu trữ, truy xuất, xử lý và trực quan hóa dữ liệu về việc sử dụng cầu dựa trên lưu lượng tải trên cầu một cách linh hoạt với phần lớn phần mềm là miễn phí như: Kafka giúp đơn giản hóa việc di chuyển dữ liệu, cơ sở dữ liệu chuỗi thời gian InfluxDB cung cấp khả năng lưu trữ lâu dài cho dữ liệu và giao diện trực quan hóa dữ liệu dựa trên Grafana.

Andrew Scullion [9] đã khai thác công nghệ dữ liệu lớn mã nguồn mở để xử lý dữ liệu của hệ thống SHM và giải thích cách một số công nghệ mã nguồn mở mới nhất như Python, Redis, Apache, PostgreSQL, WebGL, JQuery, Hadoop ecosystem... đã cho phép chuyển đổi dữ liệu thành thông tin và có thể đưa ra quyết định hiệu quả trong việc quản lý cầu.

Có nhiều nghiên cứu áp dụng các công nghệ khác nhau để thu thập, quản lý lưu trữ và phân tích dữ liệu thu được cho hệ thống SHM. Tuy nhiên, phạm vi đề tài luận văn chỉ tập trung áp dụng các công nghệ Big Data trên hệ thống cảnh báo bất thường cho cầu dây văng từ dữ liệu thu được của hệ thống SHM. Việc ứng dụng công nghệ Big Data mã nguồn mở không những tối ưu về mặt chi phí mà còn có khả năng mở rộng, khả năng chịu lỗi và độ tin cậy cao.

3. Mục tiêu nghiên cứu

Mục tiêu tổng quát

- Xây dựng hệ thống cảnh báo bất thường cho cầu dây văng dựa trên dữ liệu thô thu được từ các cảm biến của hệ quan trắc cấu trúc công trình cầu bằng công nghệ Big Data.

Mục tiêu cụ thể

- Cung cấp một Dashboard cho giám sát cầu dây văng để: 1) Hiện thị kết quả dữ liệu quan trắc đã qua xử lý và các dữ liệu lịch sử bằng các dạng đồ thị khác nhau theo yêu cầu cụ thể của hệ thống trong thời gian thực, 2) Hiện thị các thông tin, thiết lập các mức cảnh báo và giám sát cần thiết cho cầu dây văng.
- Xác định tình trạng bất thường bằng cách sử dụng các ngưỡng cảnh báo cho phép đối với tốc độ gió, lực căng cáp, độ nghiêng, độ dao động dầm, dao động cáp qua đó so sánh đối chiếu với giá trị đã được xử lý từ dữ liệu thô để hiện thị thông tin cảnh báo.
- Gửi các thông tin cảnh báo bất thường ghi nhận được về tình trạng kỹ thuật cầu thông qua một số trình thông báo khác nhau bao gồm PagerDuty, SMS, email hoặc Slack như: 1) Cảnh báo về tốc độ gió báo động, 2) Lực căng cáp báo động, lực căng cáp nguy hiểm, 3) Độ nghiêng đỉnh tháp, độ nghiêng chân tháp, 4) Dao động dầm, dao động cáp, 5) Các thông tin về thời tiết bất lợi cho cầu.

- Đánh giá về tình trạng tổng quát cầu Phú Mỹ dựa vào tình trạng của tốc độ gió, lực căng cáp, độ nghiêng, độ dao động dầm, dao động cáp với ba mức độ cảnh báo: a) Mức 1: Bình thường, b) Mức 2: Cảnh báo, c) Mức 3: Nguy hiểm.

4. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu

- Biến đổi dữ liệu thô thu được từ các cảm biến của hệ quan trắc cấu trúc công trình cầu dây văng Phú Mỹ thành dữ liệu phục vụ cho mục đích nghiên cứu.
- Kỹ thuật phân tích dữ liệu thô như phương pháp tiếp cận miền tần số (biến đổi Fourier), lý thuyết về tính lực căng, góc nghiêng, dao động.
- Ngưỡng an toàn cho cầu dây văng dựa trên các ngưỡng cảnh báo đã được phân tích tính toán từ mô hình phân tích 3D của công trình và dựa trên cơ sở thực nghiệm.
- Các công nghệ Big Data để quản lý, phân tích dữ liệu như: Kafka, Apache Spark Structured Streaming, InfluxDB, Grafana.

Phạm vi nghiên cứu

- Sử dụng dữ liệu từ các cảm biến thu được thông qua hệ thống quan trắc của cầu dây văng Phú Mỹ như: Cảm biến đo nghiêng, gia tốc kế, cảm biến thời tiết, phong kế.
- Tập trung cảnh báo: 1) Tốc độ gió báo động, 2) Lực căng cáp báo động, lực căng cáp nguy hiểm, 3) Độ nghiêng đỉnh tháp, độ nghiêng chân tháp, 4) Dao động dầm, dao động cáp, 5) Các thông tin về thời tiết bất lợi cho cầu. 6) Tình hình tổng quát của cầu dây văng Phú Mỹ.
- Đánh giá tình hình tổng quát cầu ở các mức độ: Bình thường, cảnh báo, nguy hiểm.

5. Phương pháp nghiên cứu

Đề tài này sử dụng phương pháp nghiên cứu lý thuyết kết hợp với xây dựng ứng dụng thực nghiệm:

- Thu thập các tài liệu có liên quan tới đề tài, các thông số và ngưỡng cho phép về các bộ phận của cầu dây văng Phú Mỹ.
- Ứng dụng các công nghệ Big Data như Kafka, Apache Spark Structured Streaming, InfluxDB, Grafana để phát triển hệ thống thực nghiệm.
- Tiến hành đánh giá kết quả thực nghiệm, đưa ra hướng phát triển mở rộng của đề tài để đáp ứng những nhu cầu triển khai thực tế.

6. Cấu trúc luận văn

Ngoài phần mở đầu, mục lục, kết luận và kiến nghị, danh mục hình vẽ, danh mục bảng biểu, tài liệu tham khảo, phụ lục, phần chính của luận văn gồm 4 chương như sau:

Chương 1: NGHIÊN CỨU TỔNG QUAN

Chương 2: PHÂN TÍCH THIẾT KẾ

Chương 3: TRIỂN KHAI ỨNG DỤNG

Chương 4: ĐÁNH GIÁ KẾT QUẢ THỬ NGHIỆM

CHƯƠNG 1: NGHIÊN CỨU TỔNG QUAN

1.1. Tổng quan về Big Data

Dữ liệu lớn (Big Data) là thuật ngữ dùng để mô tả các bộ dữ liệu có kích thước rất lớn, tốc độ phát triển nhanh và khó thu thập, lưu trữ, quản lý và phân tích với các công cụ hay ứng dụng truyền thống. Khi công nghệ tiến bộ theo thời gian, kích thước của các dữ liệu cũng sẽ tăng lên và Big Data được hiểu luôn có kích thước lớn hơn rất nhiều. Theo một phân tích gần đây của IDC cho thấy việc tạo và tiêu thụ dữ liệu toàn cầu ngày càng tăng, dự báo rằng vào năm 2025, tổng dung lượng dữ liệu toàn cầu (global datasphere) sẽ tăng lên tới 175 Zettabytes so với mức 33 Zettabytes được tạo ra trong năm 2018 [10]. Xét trên quan điểm toàn diện hơn, Big Data là “một hiện tượng văn hóa, công nghệ và học thuật dựa trên sự tương tác của: (1) Công nghệ tối đa hóa sức mạnh tính toán và độ chính xác thuật toán để thu thập, phân tích, liên kết, và so sánh các tập dữ liệu lớn. (2) Phân tích tạo ra trên dữ liệu lớn để xác định mô hình giải thích các hiện tượng kinh tế, xã hội, kỹ thuật và pháp lý” [11].

Có rất nhiều khía cạnh để tìm hiểu và ứng dụng lên quan tới Big Data như: Ứng dụng Big Data giúp doanh nghiệp tìm hiểu được giá trị thông tin thực sự nằm sau dữ liệu sẵn có, đặc biệt trong một xã hội toàn cầu hóa và thế giới số như hiện nay. Từ các dữ liệu lớn thu được qua các cảm biến theo dõi máy móc, cơ sở hạ tầng như các thiết bị thông gió, cầu hoặc động cơ máy bay. Dữ liệu này có thể được sử dụng cho các phân tích, để sửa chữa hoặc thay thế trước khi chúng bị hỏng... Tuy nhiên, phạm vi đề tài luận văn chỉ tập trung áp dụng các công nghệ Big Data trên hệ thống cảnh báo bất thường cho cầu dây văng từ dữ liệu thu được của hệ thống SHM.

1.2. Các công nghệ Big Data

Công nghệ Dữ liệu lớn có thể được định nghĩa là một Tiện ích-Phần mềm được thiết kế để phân tích, xử lý và trích xuất thông tin từ một bộ dữ liệu cực kỳ phức tạp và lớn mà phần mềm xử lý dữ liệu truyền thống khó có thể xử lý được. Các công nghệ Big Data hàng đầu được chia thành 4 lĩnh vực được phân loại như sau:

- Lưu trữ dữ liệu: Hadoop, MongoDB, InfluxDB, Rainstor...
- Khai thác dữ liệu: Presto, Elasticsearch, Apache Spark...
- Phân tích dữ liệu: Apache Kafka, Splunk, KNIME, Apache Spark...
- Trực quan hóa dữ liệu: Tableau, Plotly, Grafana ...

1.3. Cơ sở phân tích dữ liệu quan trắc cầu

Việc thu nhận dữ liệu được thực hiện một cách liên tục theo thời gian thực, khi có những sự kiện đặc biệt, một số cảm biến riêng biệt sẽ đo những số liệu liên quan của các sự kiện này. Tất cả các dữ liệu này đều được lưu trữ và phân tích dựa theo tính chất của chúng, khác với giá trị thống kê thô trước đó. Chẳng hạn như đối với thông số dao động thì cần đi đến thông tin về tần số dao động để phân tích so sánh.

Để các cơ sở dữ liệu phản ánh một cách chính xác trạng thái kết cấu và sức khỏe của công trình cầu thì việc thu thập các dữ liệu cần phải được thực hiện một cách đồng bộ. Phần mềm quản lý dữ liệu sẽ được thiết kế cho các trạng thái, mức độ làm việc khác nhau của công trình liên quan đến tần suất thu thập đồng bộ tất cả các đại lượng vật lý từ các cảm biến và các thiết bị quan trắc.

- Mức 1 – Hoạt động bình thường: đồng bộ thu thập tất cả các dữ liệu cứ mỗi 1 phút cho các cảm biến, gia tốc và thiết bị đo nghiêng, đo độ dao động.
- Mức 2 – Cảnh báo: Thông báo cho người chịu trách nhiệm quản lý các kịch bản có thể xảy ra.

- Mức 3 – Nguy hiểm: Thông báo cho người chịu trách nhiệm về nguy cơ hư hại của (công trình ví dụ như yêu cầu cấm cầu) [12].

Để phục vụ tốt cho công tác duy tu bảo dưỡng, hệ thống không chỉ cho phép thiết lập các mức cảnh báo và giám sát cần thiết cho công trình mà còn thực hiện đánh giá các mức cảnh báo một cách tự động thông qua các giá trị nhận được bởi tính toán phân tích dữ liệu.

1.4. Yêu cầu báo cáo số liệu cho hệ thống

Việc báo cáo thống kê phải mang tính chất định kỳ bao gồm:

- Báo cáo hàng ngày,
- Báo cáo hàng tháng,
- Báo cáo năm

Trong đó các kết quả phải phân tích đánh giá đến thông số yêu cầu, so sánh đánh giá theo thời gian quan trắc.

CHƯƠNG 2: PHÂN TÍCH THIẾT KẾ

Trong chương này trình bày một số công nghệ Big Data, so sánh đánh giá ưu nhược điểm của các công nghệ này để lựa chọn mô hình phù hợp nhất cho yêu cầu xây dựng hệ thống cảnh báo bất thường cho cầu dây văng.

2.1 Mô tả vấn đề

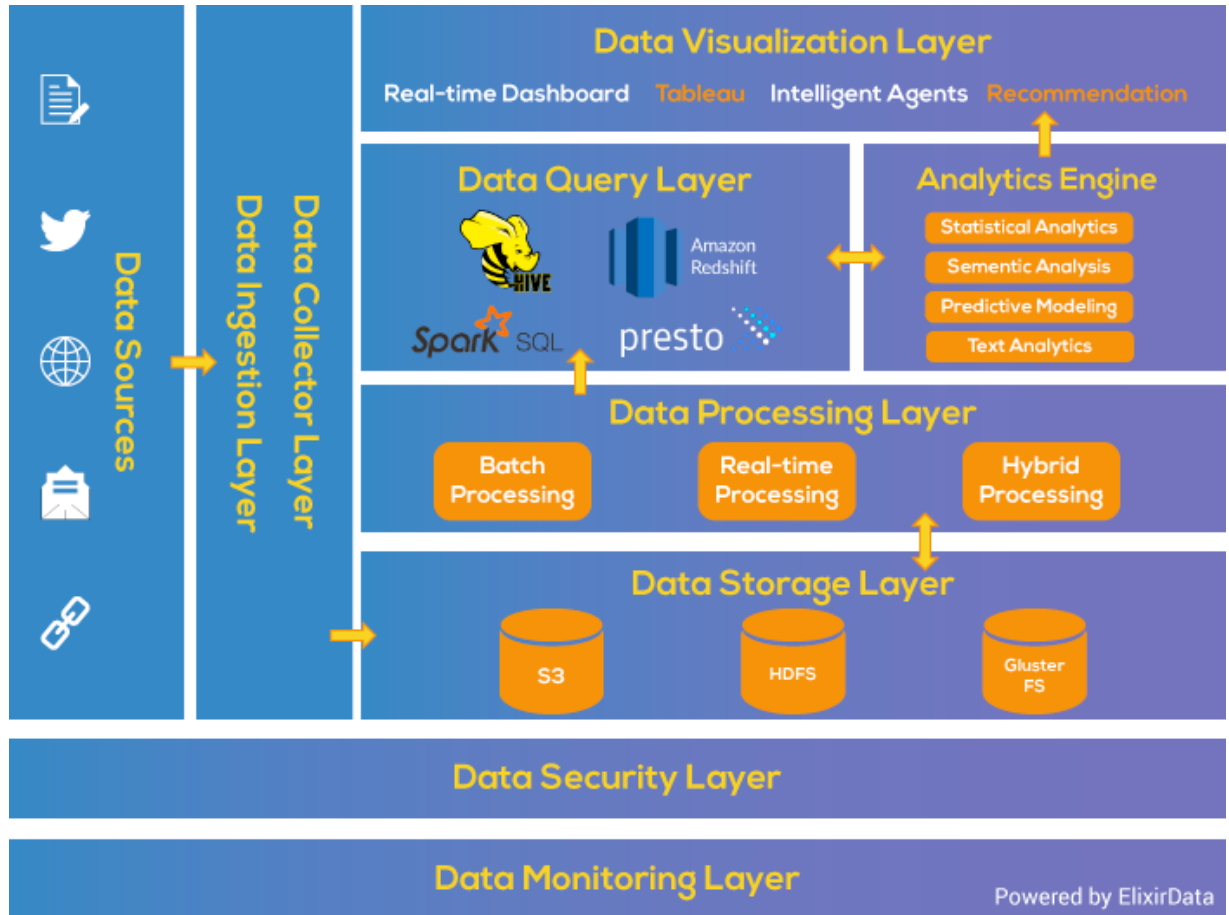
Có nhiều nghiên cứu áp dụng các công nghệ khác nhau để thu thập, quản lý lưu trữ và phân tích dữ liệu thu được cho hệ thống SHM. Seongwoon Jeong et al. [6, 7] đã sử dụng các máy ảo đám mây, hệ thống cơ sở dữ liệu phân tán như Cassandra, NoSQL cùng với Spark cluster và máy chủ để xây dựng hệ thống giám sát các cây cầu dọc theo hành lang I-275 nằm ở bang Michigan. S. Gunner et al. [8] thì sử dụng Kafka, InfluxDB và Grafana cho hệ thống quan trắc cầu treo Clifton, Vương Quốc Anh. Andrew Scullion [9] đã phân tích một số công nghệ mã nguồn mở mới nhất như Python, Redis, Apache, PostgreSQL, WebGL, JQuery, Hadoop ecosystem... để xử lý dữ liệu của hệ thống SHM.

Thông qua việc đánh giá nguồn dữ liệu thu được từ hệ quan trắc là kiểu dữ liệu chuỗi thời gian và khối lượng dữ liệu thu được là rất lớn do nó được tạo ra mỗi phút liên tục trong thời gian thực. Do đó việc áp dụng các công nghệ Big Data mã nguồn mở trên hệ thống cảnh báo bất thường cho cầu dây văng từ dữ liệu thu được của hệ thống SHM không những tối ưu về mặt chi phí mà còn có khả năng mở rộng, khả năng chịu lỗi và độ tin cậy cao.

2.2. Kiến trúc Big Data

Kiến trúc Big Data bao gồm một khối logic, do đó để khám phá những lợi ích của kiến trúc đòi hỏi phải hiểu rõ về các lớp khác nhau và chức năng của chúng. Hình 2.1

dưới đây mô tả một trong những phương pháp hay dùng của kiến trúc Big Data đó là kiến trúc phân lớp.



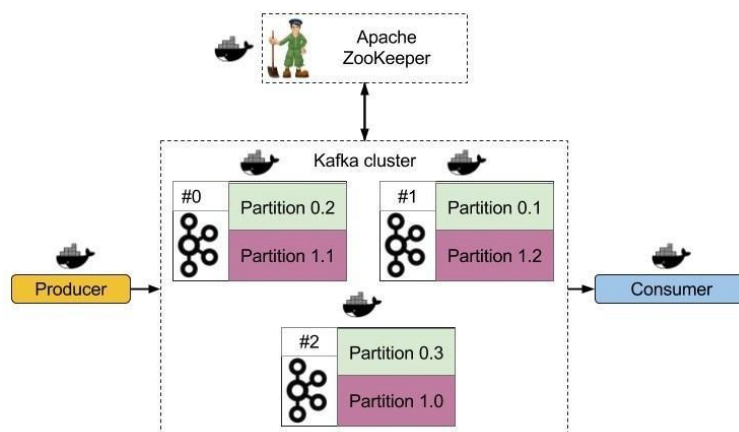
Hình 2.1: Kiến trúc Big Data tổng quan

Kiến trúc phân lớp được chia thành các lớp khác nhau trong đó mỗi lớp thực hiện một chức năng cụ thể. Kiến trúc này bao gồm các lớp kiến trúc chính như sau: Lớp nguồn dữ liệu, lớp thu thập và truyền dữ liệu, lớp xử lý dữ liệu, lớp lưu trữ dữ liệu, lớp truy vấn dữ liệu, lớp trực quan hóa dữ liệu. Kiến trúc phân lớp giúp xác định các thành phần cần thiết để xác định giải pháp dữ liệu lớn cho bài luận văn này.

2.2.1. Lớp thu thập dữ liệu

Lớp này là bước đầu tiên để dữ liệu từ các nguồn khác nhau bắt đầu hành trình của nó. Dữ liệu ở đây được đánh giá, phân loại và được vận chuyển đến các lớp còn lại của đường ống dữ liệu. Với yêu cầu nguồn dữ liệu đầu vào là nguồn dữ liệu từ các cảm biến truyền tới liên tục mỗi phút trong thời gian thực, vậy làm thế nào triển khai các kết nối truyền dữ liệu từ nơi nó được tạo ra đến nhiều nơi khác nhau để xử lý và đảm bảo rằng dữ liệu vào sẽ không bị mất trước khi nó được chuyển tới các lớp khác. Để làm được điều đó thì “messaging” là một công nghệ quan trọng để kết nối dữ liệu trong môi trường này.

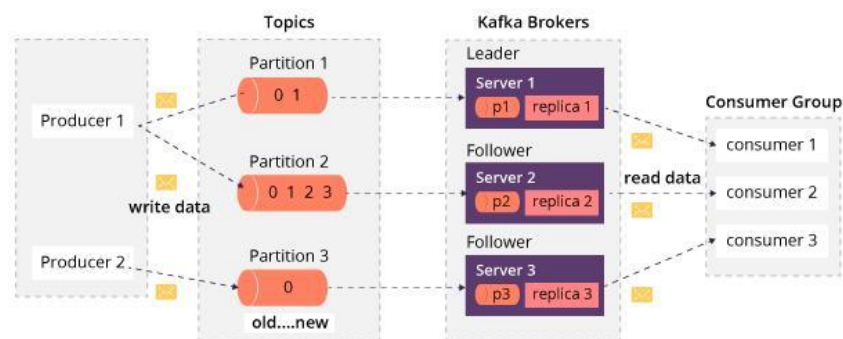
Messaging là một thuật ngữ bao gồm một số mô hình khác nhau dựa trên các dữ liệu được chuyển từ người gửi sang người nhận. Hai mô hình chính trong công nghệ này là “message queue” và “publish-subscribe messaging”. Khác với message queue – mỗi tin nhắn sẽ được gửi cho một topic hay channel và xử lý bởi chính xác một consumer thì publish-subscribe messaging cho phép nhiều consumer nhận từng message trong một topic. Hơn nữa publish-subscribe messaging đảm bảo rằng mỗi consumer nhận được message trong một topic theo thứ tự chính xác mà hệ thống message nhận được.



Hình 2.2: Cấu trúc cơ bản của Kafka

Trong tất cả các công nghệ sử dụng publish-subscribe messaging thì Kafka [13] là một mô hình có tốc độ nhanh, khả năng mở rộng, độ tin cậy và đảm bảo tính chịu lỗi cao. Ngoài ra Kafka có thể truyền một lượng lớn dữ liệu trong thời gian thực, vì có độ tin cậy và tính chịu lỗi cao nên trong trường hợp bên nhận chưa nhận, dữ liệu vẫn được lưu trữ sao lưu trên một hàng đợi và cả trên ổ đĩa để đảm bảo an toàn.

Các thành phần chính của Kafka được minh họa trong Hình 2.3.



Hình 2.3: Các thành phần chính của Kafka

- *Topic:* Dữ liệu truyền trong Kafka được phân chia theo topic, khi cần truyền dữ liệu cho các ứng dụng khác nhau thì sẽ tạo ra các topic khác nhau.
- *Partition:* Đây là nơi dữ liệu cho một topic được lưu trữ. Một topic có thể có một hay nhiều partition. Trên mỗi partition thì dữ liệu lưu trữ cố định và được gán cho một ID gọi là offset. Trong một Kafka cluster thì một partition có thể replicate (sao chép) ra nhiều bản. Trong đó có một bản leader chịu trách nhiệm đọc ghi dữ liệu và các bản còn lại gọi là follower. Khi bản leader bị lỗi thì sẽ có một bản follower lên làm leader thay thế. Nếu muốn dùng nhiều consumer đọc song song dữ liệu của một topic thì topic đó cần phải có nhiều partition.
- *Replicas:* là một bản backup của partition, chỉ được dùng để lưu trữ, tránh mất dữ liệu và nó không được dùng để đọc ghi.

- *Broker*: là một Kafka server. Nếu chạy đồng thời nhiều Kafka server thì sẽ gọi đó là Cluster. Mỗi Broker (Kafka server) sẽ là một node trong cluster đó. Như vậy khi nói đến Broker, node thì sẽ hiểu đó là Kafka server. Chịu trách nhiệm điều tiết thông điệp input output. Partition sẽ nằm bên trong Broker. Một Broker có thể có không, hoặc một hoặc nhiều partition đặt bên trong đó.
- *Producer*: Kafka lưu, phân loại message theo topic, sử dụng producer để publish message vào các topic. Dữ liệu được gửi đến partition của topic lưu trữ trên Broker.
- *Consumer*: Kafka sử dụng consumer để subscribe vào topic, các consumer được định danh bằng các group name. Nhiều consumer có thể cùng đọc một topic.
- *Leader*: Mỗi partition sẽ có một Broker chịu trách nhiệm chính trong việc đọc và ghi dữ liệu vào partition đó. Broker đó được gọi là Leader.
- *Follower*: Đây là một con replicas của partition. Khi có một Broker làm Leader thì các Broker còn lại trong Cluster sẽ làm replicas của partition, nó sẽ làm Follower cho partition đó. Nó sẽ theo dõi Leader, nếu Leader bị chết thì Zookeeper sẽ chịu trách nhiệm bầu ra một trong số các Follower lên làm Leader thay thế. Follower hoạt động giống như một con Consumer, nhiệm vụ của nó là lấy dữ liệu từ Leader về lưu cho riêng nó.
- *Zookeeper*: Được sử dụng để quản lý điều phối các Kafka Broker. Zookeeper được sử dụng chủ yếu cho việc thông báo cho producer và consumer biết thông tin của một broker mới thêm vào hoặc thông tin của một Kafka broker lỗi trong hệ thống Kafka. Sau mỗi thông báo của Zookeeper về việc broker thêm vào hay broker bị lỗi thì producer và consumer sẽ quyết định và bắt đầu phối hợp với một broker khác.

Việc truyền tải dữ liệu giữa các hệ thống khác với Kafka broker được thực hiện thông qua giao thức TCP. Việc phát triển, phân tích dữ liệu có thể thực hiện trên nhiều nền tảng ngôn ngữ lập trình khác nhau.

2.2.2. Lớp xử lý dữ liệu

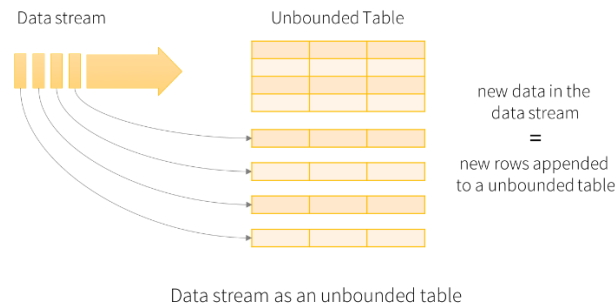
Các công nghệ hỗ trợ xử lý dữ liệu thời gian thực được đánh giá cao và sử dụng trong các mô hình ở trên là Apache Spark [14] cụ thể là Apache Spark Streaming và Apache Spark Structured Streaming. Spark Structured Streaming nhận dữ liệu từ các nguồn streaming data, cụ thể ở trong các mô hình trên là Kafka. Về mặt khái niệm, Structured Streaming coi dữ liệu stream là một bảng (table) không giới hạn và dữ liệu liên tục được thêm vào từng dòng của bảng này. Do đó mô hình xử lý Stream của Spark cũng tương tự với mô hình xử lý batch (coi dữ liệu trong mỗi batch là khoảng 1 giây, 5 giây...) và ta có thể sử dụng các truy vấn stream trong Spark tương tự như các truy vấn trong xử lý batch. Structured Streaming là phiên bản mới, xử lý stream dựa trên Dataset/DataFrame và có nhiều tính năng cải thiện hơn so với các công cụ truyền phát khác như Hình 2.4.

Property	Structured Streaming	Spark Streaming	Apache Storm	Apache Flink	Kafka Streams	Google Dataflow
Streaming API	incrementalize batch queries	integrates with batch	separate from batch	separate from batch	separate from batch	integrates with batch
Prefix Integrity Guarantee	✓	✓	✗	✗	✗	✗
Internal Processing	exactly once	exactly once	at least once	exactly once	at least once	exactly once
Transactional Sources/Sinks	✓	some	some	some	✗	✗
Interactive Queries	✓	✓	✗	✗	✗	✗
Joins with Static Data	✓	✓	✗	✗	✗	✗

Hình 2.4: So sánh Spark Structured Streaming với các công cụ truyền phát khác

(Nguồn tham khảo [14])

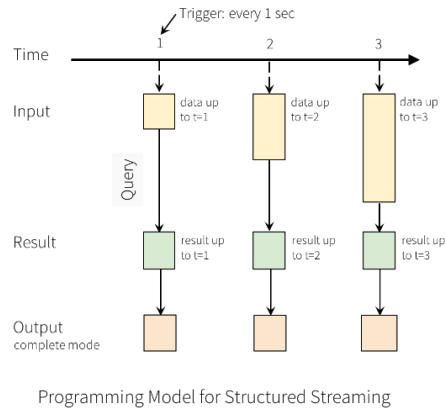
Mô hình lập trình Spark Structured Streaming coi dòng dữ liệu đầu vào là “Input Table”. Mỗi mục dữ liệu tới trên luồng được thêm vào trong dòng của “Input Table” này như Hình 2.5.



Hình 2.5: Luồng dữ liệu dưới dạng bảng không giới hạn

Mỗi truy vấn ở đầu vào sẽ tạo ra “Result Table”. Mỗi khoảng thời gian kích hoạt (giả sử, cứ sau một giây), các hàng mới sẽ được thêm vào “Input Table”, cuối cùng sẽ cập nhật “Result Table”. Bất cứ khi nào “Result Table” được cập nhật, các hàng kết quả đã thay đổi sẽ được viết vào một hệ thống bên ngoài, chẳng hạn như S3, HDFS hoặc cơ sở dữ liệu như InfluxDB. Với mục đích này, Structured Streaming cung cấp ba chế độ đầu ra. Mô hình lập trình cho Structured Streaming được mô phỏng trong Hình 2.6.

- *Complete Mode*: Toàn bộ bảng kết quả được cập nhật sẽ được ghi vào bộ nhớ ngoài.
- *Append Mode*: Chỉ các hàng mới được thêm vào “Result Table” vì lần kích hoạt cuối cùng sẽ được ghi vào bộ nhớ ngoài. Điều này chỉ áp dụng cho các truy vấn mà các hàng hiện có trong Bảng Kết quả sẽ không thay đổi.
- *Update Mode*: Chỉ những hàng đã được cập nhật trong Bảng kết quả kể từ lần kích hoạt cuối cùng sẽ được ghi vào bộ nhớ ngoài. Chế độ này khác với “Complete Mode” ở chỗ chế độ này chỉ xuất ra các hàng đã thay đổi kể từ lần kích hoạt cuối cùng.



Hình 2.6: Mô hình lập trình cho Structured Streaming

Mô hình này rất khác với các công cụ xử lý phát trực tuyến khác. Nhiều hệ thống phát trực tuyến yêu cầu người dùng duy trì trạng thái. Trong mô hình này, Spark chịu trách nhiệm cập nhật “Result Table” khi có dữ liệu mới, do đó giúp người dùng không phải suy đoán về nó.

2.2.3. Lớp lưu trữ dữ liệu

Trong đề tài luận văn sử dụng dữ liệu thu được từ các cảm biến và là kiểu dữ liệu chuỗi thời gian, trước khi đi vào chi tiết về lớp lưu trữ ta sẽ tìm hiểu về kiểu dữ liệu Time-series Data.

a. Time-series Data

Time-series Data, còn được gọi là dữ liệu được đóng dấu thời gian, là một chuỗi các điểm dữ liệu được lập chỉ mục theo thứ tự thời gian. Các điểm dữ liệu này thường bao gồm các phép đo liên tiếp được thực hiện từ cùng một nguồn trong một khoảng thời gian và được sử dụng để theo dõi sự thay đổi theo thời gian. Hình 2.7 Minh họa một tập dữ liệu theo chuỗi thời gian từ các cảm biến thu được trên cầu dây văng Phú Mỹ, dữ liệu được cập nhật sau mỗi 02 millisecond và liên tục theo thời gian.

time	G11ACC01Y	G11ACC01Z	P06INC01	P07INC01
2020-04-23T00:00:28.19Z	0.016108	0.093451	0.012595	0.011712
2020-04-23T00:00:28.21Z	0.016155	0.093382	0.012596	0.01171
2020-04-23T00:00:28.23Z	0.016157	0.093673	0.012594	0.011709
2020-04-23T00:00:28.25Z	0.016209	0.093928	0.012595	0.011711
2020-04-23T00:00:28.27Z	0.016249	0.093999	0.012594	0.01171
2020-04-23T00:00:28.29Z	0.016135	0.093898	0.012594	0.011709
2020-04-23T00:00:28.31Z	0.016171	0.093822	0.012593	0.011711
2020-04-23T00:00:28.33Z	0.016072	0.093738	0.012594	0.01171

Hình 2.7: Minh họa Time-series Data

b. Ứng dụng của Time-series Data

Dữ liệu chuỗi thời gian được sử dụng trong nhiều ngữ cảnh khác nhau, trong đó phổ biến nhất là:

- *Phân tích chuỗi thời gian*: Phân tích chuỗi thời gian được sử dụng để tìm hiểu cách một biến số nhất định thay đổi theo thời gian. Ví dụ: Phân tích về độ nghiêng của đỉnh tháp cầu dây văng Phú Mỹ theo thời gian, có thể được coi là một nhiệm vụ Phân tích chuỗi thời gian.
- *Phân tích hồi quy*: Phân tích hồi quy có thể được sử dụng để kiểm tra xem những thay đổi liên quan đến một biến cụ thể có thể gây ra sự thay đổi như thế nào đối với các biến khác trong cùng một khoảng thời gian. Ví dụ: Phân tích về các giá trị thu được từ các cảm biến trên cầu Phú Mỹ, cụ thể thông số về lực căng cáp của cầu theo thời gian ảnh hưởng như thế nào tới thông số của các vị trí khác như độ nghiêng của

cầu hoặc kết cấu của cầu trong một khoảng thời gian, có thể được coi là nhiệm vụ Phân tích hồi quy.

- *Dự báo chuỗi thời gian*: Dự báo chuỗi thời gian sử dụng thông tin liên quan đến các giá trị lịch sử và các mẫu liên quan để dự đoán hoạt động trong tương lai. Ví dụ: Dự báo thời tiết, Dự báo đứt cáp, v.v.

Từ các phân tích về kiểu dữ liệu chuỗi thời gian ở trên, để lưu trữ một lượng lớn dữ liệu thời gian thực, cần một kho lưu trữ dữ liệu hỗ trợ ghi nhanh và có khả năng mở rộng. Một vài cơ sở dữ liệu chuỗi thời gian (Time Series Database - TSDB) và thứ tự xếp hạng của chúng được hiển thị trong Hình 2.8 của DB-Engines. DB-Engines là một trang web tích hợp, xếp hạng cơ sở dữ liệu dựa trên mức độ phổ biến của công cụ tìm kiếm, tham khảo trên mạng xã hội, các bài đăng công việc và tần suất thảo luận kỹ thuật. Tính tới tháng 9 năm 2020, InfluxDB [15] là một cơ sở dữ liệu chuỗi thời gian được xếp hạng số 1 trong danh sách này. Hình 2.8 cũng thể hiện sự thay đổi theo thời gian của các cơ sở dữ liệu nhưng có thể thấy rằng, vào tháng 9 năm 2020, InfluxDB cũng là TSDB hàng đầu và vẫn giữ được sự ổn định trong thứ hạng của nó [16].

Rank			DBMS	Database Model	Score		
Sep 2020	Aug 2020	Sep 2019			Sep 2020	Aug 2020	Sep 2019
1.	1.	1.	InfluxDB	Time Series	23.34	+0.46	+4.69
2.	2.	2.	Kdb+	Time Series, Multi-model	7.43	+0.63	+1.93
3.	3.	3.	Prometheus	Time Series	5.69	+0.33	+2.24
4.	4.	4.	Graphite	Time Series	4.31	+0.20	+0.99
5.	5.	5.	RRDtool	Time Series	3.05	+0.03	+0.48
6.	6.	8.	TimescaleDB	Time Series, Multi-model	2.73	+0.19	+1.33
7.	8.	6.	OpenTSDB	Time Series	2.30	+0.17	+0.45
8.	7.	7.	Druid	Multi-model	2.29	+0.11	+0.56
9.	9.	11.	FaunaDB	Multi-model	1.86	+0.15	+1.43
10.	11.	10.	GridDB	Time Series, Multi-model	0.75	+0.09	+0.26

Hình 2.8: Bảng xếp hạng 10 cơ sở dữ liệu chuỗi thời gian hàng đầu thế giới theo DB-Engines

(Nguồn tham khảo [17])

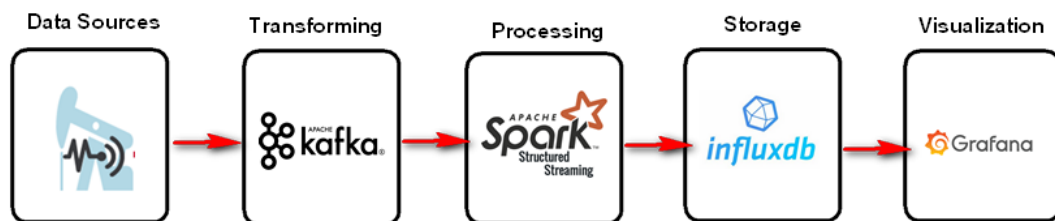
2.2.4. Lớp trực quan hóa dữ liệu

Kibana và Grafana đều là nền tảng phân tích và trực quan hóa mã nguồn mở, sử dụng trong việc phân tích các dữ liệu thu thập được và hiển thị một cách trực quan dữ liệu thu thập được ở nhiều dạng khác nhau. Tuy nhiên về cốt lõi chúng đều được sử dụng cho các loại dữ liệu và các ứng dụng khác nhau: Grafana [18] cùng với cơ sở dữ liệu chuỗi thời gian thực như Graphite hoặc InfluxDB là sự kết hợp được sử dụng để phân tích số liệu. Trong khi đó, Kibana là một phần của Elasticsearch, được sử dụng để khám phá dữ liệu nhật ký.

Ngoài các chức năng biểu diễn dữ liệu với các bảng điều khiển mạnh mẽ thì Grafana cũng hỗ trợ cài đặt và gửi cảnh báo qua một số trình thông báo khác nhau bao gồm PagerDuty, SMS, email hoặc Slack. Trong khi đó Kibana không có khả năng cảnh báo vượt trội, để thêm cảnh báo cho Kibana người dùng có thể triển khai thêm các gói tiện ích bổ sung khác như ElastAlert hoặc sử dụng X-Pack.

2.3. Đánh giá và lựa chọn mô hình phát triển hệ thống

Dựa vào kết quả đánh giá các công nghệ sử dụng trong kiến trúc Big Data ở trên cùng với việc đánh giá các mô hình tương tự đang được nghiên cứu và áp dụng trong thực tế được công bố gần đây, mô hình được lựa chọn để xây dựng trong bài luận văn này được mô tả như trong Hình 2.9: Mô hình xử lý tổng quan.



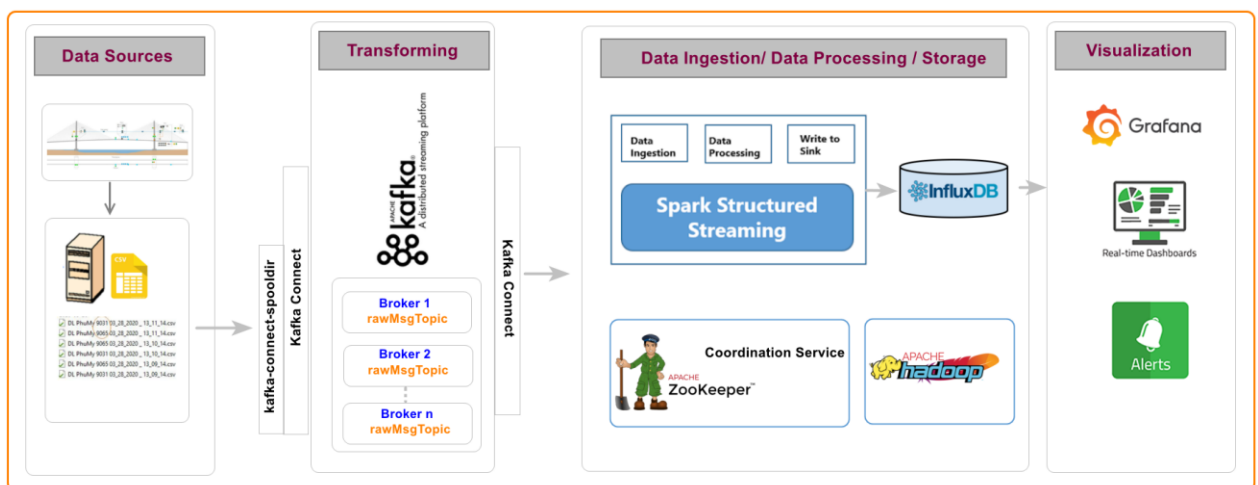
Hình 2.9: Mô hình xử lý tổng quan

Các cảm biến được đặt tại cầu Phú Mỹ tạo ra dữ liệu là các file .csv phát trực tuyến và được lưu trữ trong thư mục nguồn. Thư mục lưu trữ dữ liệu được giám sát bởi

thư viện **Kafka Connect SpoolDir** dựa trên **Kafka Connect** để đọc và ghi dữ liệu từ các file .csv vào Kafka và từ Kafka ra hệ thống bên ngoài. **Kafka Connect SpoolDir** theo dõi thư mục nguồn được cấu hình thông qua “*input.path*” và đọc dữ liệu từ các tập tin mới được đưa vào thư mục. Khi một tập tin đã đọc xong nó sẽ được chuyển sang thư mục được cấu hình thông qua “*finished.path*”. Dữ liệu từ Kafka được đẩy ra và được xử lý bởi Spark Structured Streaming sau đó sẽ lưu trữ trong InfluxDB để sử dụng các công cụ phân tích và báo cáo khác nhau.

Dữ liệu về tất cả các sự kiện đều được lưu trữ trong InfluxDB trong thời gian thực, có nghĩa là khi dữ liệu của một topic trong Kafka được thêm mới, ngay lập tức nó sẽ được lưu vào hệ thống và truyền đến Spark Structured Streaming xử lý để lưu trữ vào InfluxDB. Số liệu tóm tắt, tổng hợp và báo cáo hay các cảnh báo cũng được thể hiện dựa vào các thay đổi của các cảm biến diễn ra trong thời gian thực và được trực quan hóa thông qua Grafana.

Từ mô hình xử lý tổng quan Hình 2.9 kết hợp với dữ liệu thực tế của cầu dây văng cũng như lý thuyết tổng quan về các lớp dữ liệu trong mục 2.1, 2.2, 2.3 ta có mô hình chi tiết của hệ thống như sau:



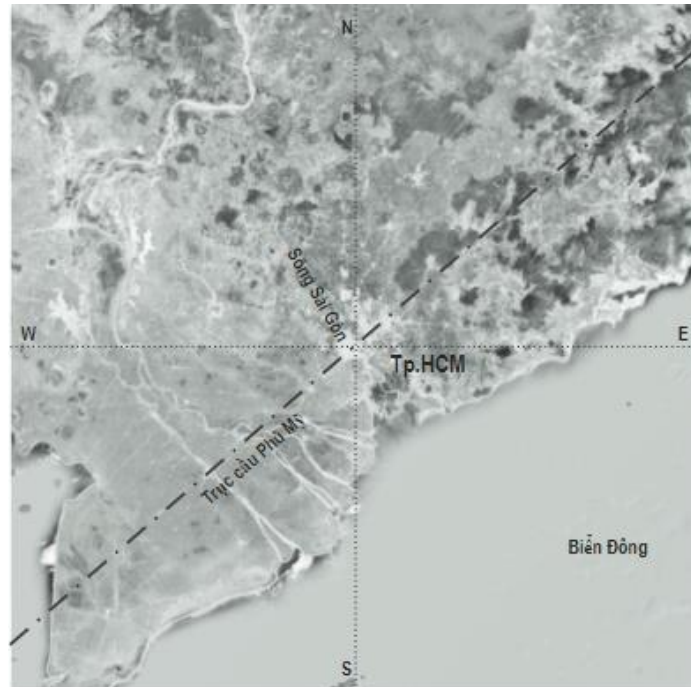
Hình 2.10: Mô hình chi tiết của hệ thống

CHƯƠNG 3: TRIỂN KHAI ỨNG DỤNG

3.1. Giới thiệu tổng quan về cầu dây văng Phú Mỹ

- Thông tin chung:

- Đối tượng vượt: sông Sài Gòn, nối quận 7 và quận 2, thành phố Hồ Chí Minh
- Lý trình: nằm trên đường vành đai ngoài thành phố Hồ Chí Minh
- Thời điểm xây dựng: tháng 9/2005 đến tháng 9/2009 [19]

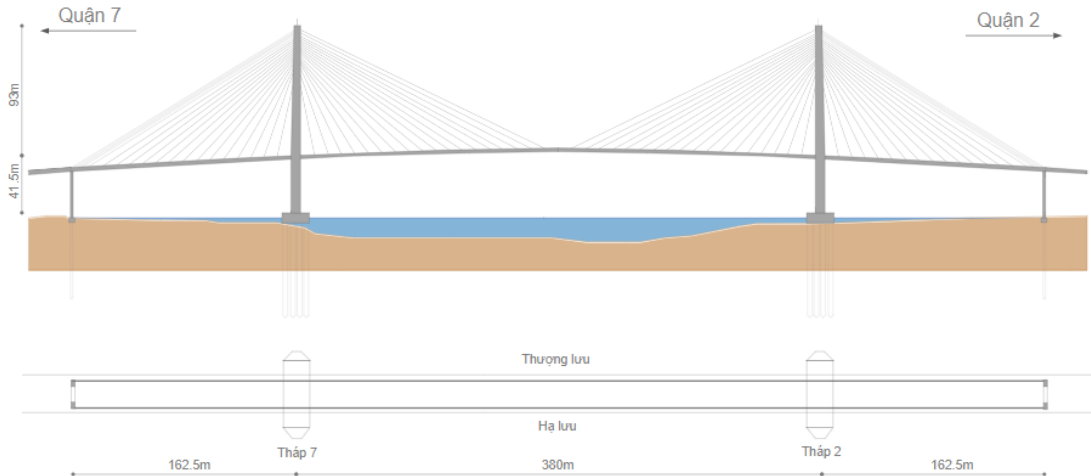


Hình 3.1: Đồ hình vị trí cầu Phú Mỹ

(Nguồn công ty cổ phần đầu tư xây dựng Phú Mỹ)

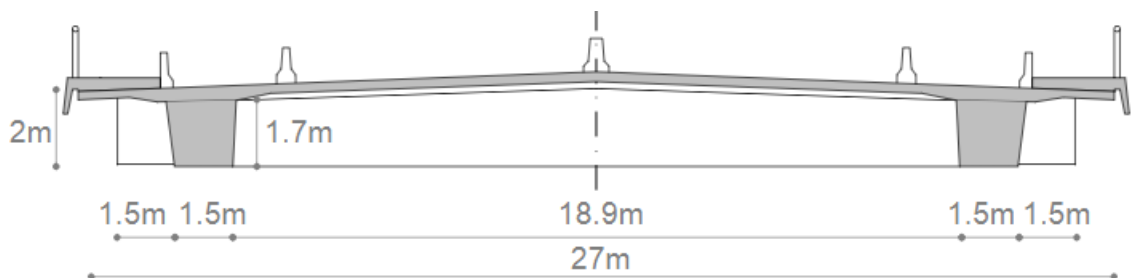
- Kết cấu chính:
 - Cầu hai mặt phẳng dây văng, nhịp chính: 162,5+380+162,5m
 - Móng trụ tháp được đặt trên hệ cọc khoan nhồi đường kính 2,05m sâu 75m

- Trụ tháp cầu dạng chữ H cao 134,5m tính từ bệ móng
- Hệ dầm cầu - bản mặt cầu ở nhịp chính gồm 2 dầm dọc chủ liên kết với nhau bằng bản mặt cầu và các dầm ngang
- Hình học chính của cầu dây văng Phú Mỹ:



Hình 3.2: Kích thước hình học chính cầu dây văng Phú Mỹ

(Nguồn công ty cổ phần đầu tư xây dựng Phú Mỹ)



Hình 3.3: Kích thước mặt cắt ngang dầm cầu Phú Mỹ

(Nguồn công ty cổ phần đầu tư xây dựng Phú Mỹ)

3.2. Thực trạng khai thác và bảo trì cầu dây văng Phú Mỹ

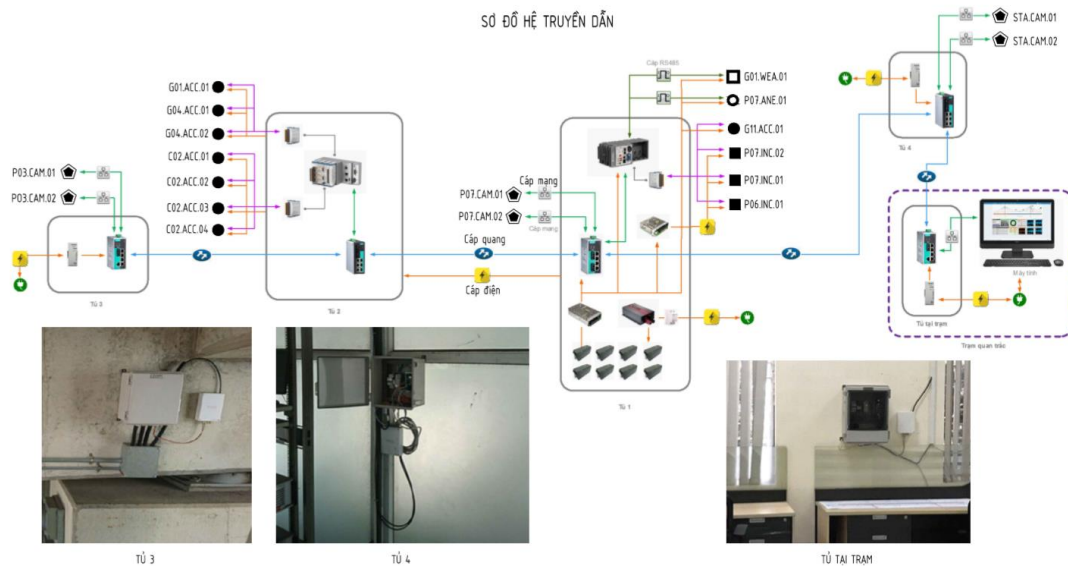
Cầu dây văng Phú Mỹ được trang bị hệ thống quan trắc để thu thập dữ liệu hỗ trợ cho công tác vận hành. Hệ quan trắc của cầu bao gồm các thiết bị đo, sensor, các trạm

GPS, các camera (Bảng 3.1) cho phép thu thập các số liệu về tình trạng làm việc của kết cấu thân cầu, các số liệu chuyển vị, số liệu về điều kiện khí hậu (tốc độ gió, nhiệt độ, lượng mưa) và hiện trạng giao thông trên cầu...

Bảng 3.1: Danh sách các cảm biến

(Nguồn công ty cổ phần đầu tư xây dựng Phú Mỹ)

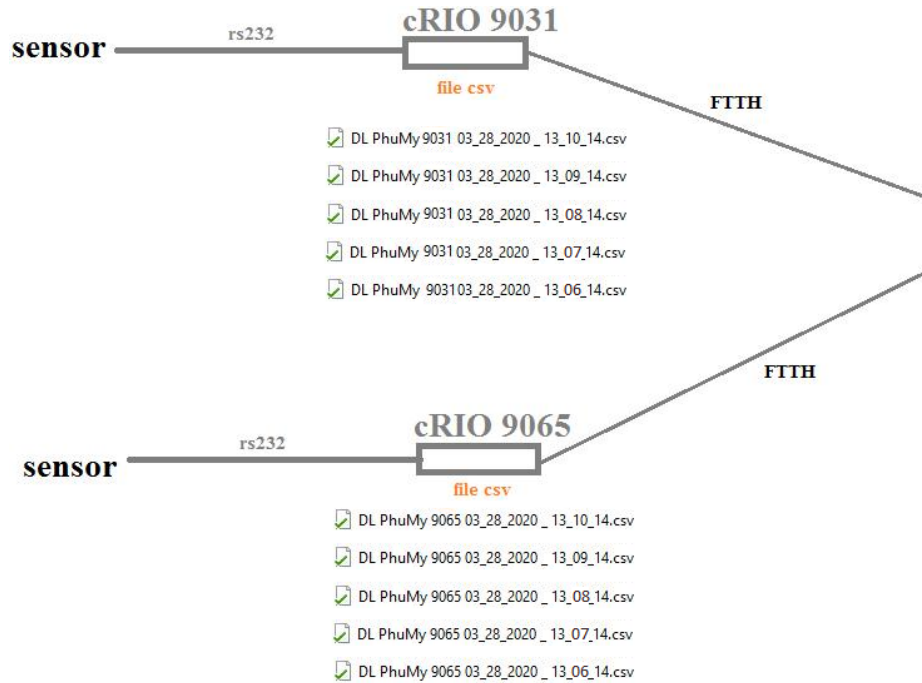
STT.	THIẾT BỊ	KÝ HIỆU	SỐ LƯỢNG	MÔ TẢ VỊ TRÍ
A				
QUAN TRẮC KẾT CẤU				
1	GIA TỐC KẾ	ACC ●	8	4 THIẾT BỊ TRÊN DẦM CẦU 4 THIẾT BỊ TRÊN CÁP VĂNG
2	ĐO NGHIÊNG	INC ■	3	3 THIẾT BỊ TRÊN THÁP 2
B				
QUAN TRẮC KHÍ TƯỢNG				
1	CẢM BIẾN THỜI TIẾT	WEA □	1	1 THIẾT BỊ TRÊN DẦM CẦU
2	PHONG KẾ	ANE ○	1	1 THIẾT BỊ TRÊN THÁP CẦU



Hình 3.4: Sơ đồ chi tiết hệ truyền dẫn

(Nguồn công ty cổ phần đầu tư xây dựng Phú Mỹ)

Dữ liệu thu được từ mười sáu cảm biến được liệt kê chi tiết ở Bảng 3.1. Mỗi một phút hai máy cRIO 9031 và cRIO 9065 sẽ trả về một file định dạng csv liệt kê các thông số mà các cảm biến thu được.



Hình 3.5: Dữ liệu nguồn

DL PhuMy 9031 03_28_2020_13_11_14.csv	28/03/2020 1:12 CH	CSV
DL PhuMy 9065 03_28_2020_13_11_14.csv	28/03/2020 1:12 CH	CSV
DL PhuMy 9065 03_28_2020_13_10_14.csv	28/03/2020 1:11 CH	CSV
DL PhuMy 9031 03_28_2020_13_10_14.csv	28/03/2020 1:11 CH	CSV
DL PhuMy 9065 03_28_2020_13_09_14.csv	28/03/2020 1:10 CH	CSV
DL PhuMy 9031 03_28_2020_13_09_14.csv	28/03/2020 1:10 CH	CSV

Hình 3.6: Chi tiết định dạng tên và cấu trúc lưu trữ các file dữ liệu

Mặc dù được quản lý bởi Tổng cục Đường bộ VN nhưng hệ thống quan trắc chỉ mới dừng lại ở việc thu thập dữ liệu, các dữ liệu thô sau khi được thu thập sẽ được gửi tới các cơ quan chuyên gia ở nước ngoài (hội quan trắc cầu thế giới) để phân tích đánh

giá chuyên sâu. Từ thực tế trên ta thấy được việc xử lý dữ liệu của cầu dây văng còn gặp nhiều khó khăn do thiếu các chuyên gia lẫn phần mềm chuyên dụng, dẫn đến việc không tận dụng được các trang thiết bị hiện có [4].

3.3. Thực tế triển khai ứng dụng

Phần này sẽ đi vào cài đặt hệ thống, lập trình và cấu hình hệ thống để thu thập, xử lý, lưu trữ và mô hình hóa dữ liệu thu được từ các cảm biến. Thiết lập các cảnh báo theo ngưỡng thông số cảnh báo từ các đơn vị chuyên môn trong đơn vị quản lý, khai thác và bảo trì cầu dây văng Phú Mỹ.

3.3.1. Triển khai hệ thống thu thập, xử lý và lưu trữ dữ liệu từ các cảm biến

a. Cài đặt hệ thống

Để xây dựng hệ thống cảnh báo bất thường cho cầu dây văng trong đề tài này, ta sử dụng ngôn ngữ lập trình, thư viện và các phần mềm chính như sau:

- Ngôn lập trình: Python, Scala;
- Cơ sở dữ liệu: InfluxDB v2.1.1;
- Thư viện đọc và truyền dữ liệu: Kafka v2.3.1, Spark v3.1.3;
- Thư viện quan hóa dữ liệu: Grafana v8.5.1;

Ngoài ra còn có các thư viện mở rộng phục vụ cho việc thu thập, xử lý dữ liệu như: kafka-connect-spooldir [20];

b. Nguồn dữ liệu

Dữ liệu nguồn là các file .csv không giới hạn đến liên tục mỗi phút theo thời gian thực. Chi tiết các cột dữ liệu thu được trong mỗi file .csv như cấu trúc minh họa bên dưới:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ThisRecordStartTime=00:00:14 03/27/2020													
2	Time (ms)	G11ACC01	G11ACC01	P06INC01	P07INC01	P07INC02	G01WEA0	G01WEA0	G01WEA0	G01WEA0	G01WEA0	P07ANE01	P07ANE01WD	
3	1585267214086.0000	0.016316	0.08882	0.012601	0.011695	0.01168	0	31.2	27.7	78	1005	3.72	81	
4	1585267214106.0000	0.016352	0.089298	0.0126	0.011694	0.011681	0	31.2	27.7	78	1005	3.72	81	
5	1585267214126.0000	0.016295	0.089856	0.0126	0.011694	0.011681	0	31.2	27.7	78	1005	3.72	81	
6	1585267214146.0000	0.016149	0.09025	0.0126	0.011693	0.01168	0	31.2	27.7	78	1005	3.72	81	
7	1585267214166.0000	0.016129	0.090275	0.012599	0.011692	0.011681	0	31.2	27.7	78	1005	3.72	81	
8	1585267214186.0000	0.016276	0.090406	0.0126	0.011692	0.01168	0	31.2	27.7	78	1005	3.72	81	
9	1585267214206.0000	0.016149	0.090693	0.0126	0.011693	0.01168	0	31.2	27.7	78	1005	3.72	81	
10	1585267214226.0000	0.016082	0.090849	0.012604	0.011717	0.011679	0	31.2	27.7	78	1005	3.72	81	
11	1585267214246.0000	0.016198	0.090974	0.012605	0.011717	0.011679	0	31.2	27.7	78	1005	3.72	81	
12	1585267214266.0000	0.016159	0.090803	0.012604	0.011718	0.011701	0	31.2	27.7	78	1005	3.72	81	
13	1585267214286.0000	0.016112	0.090953	0.012605	0.011716	0.0117	0	31.2	27.7	78	1005	3.72	81	
14	1585267214306.0000	0.016166	0.09106	0.012604	0.011717	0.0117	0	31.2	27.7	78	1005	3.72	81	
15	1585267214326.0000	0.016177	0.091037	0.012605	0.011718	0.0117	0	31.2	27.7	78	1005	3.72	81	
16	1585267214346.0000	0.016096	0.091048	0.012604	0.011717	0.011699	0	31.2	27.7	78	1005	3.72	81	
17	1585267214366.0000	0.016239	0.091097	0.012605	0.011717	0.011699	0	31.2	27.7	78	1005	3.72	81	
18	1585267214386.0000	0.016184	0.0909	0.012604	0.011683	0.011699	0	31.2	27.7	78	1005	3.72	81	
19	1585267214406.0000	0.016179	0.090697	0.01262	0.011684	0.011693	0	31.2	27.7	78	1005	3.72	81	
20	1585267214426.0000	0.016325	0.090666	0.01262	0.011683	0.011693	0	31.2	27.7	78	1005	3.72	81	

Hình 3.7: Các cột dữ liệu trong file csv thu được từ cRIO 9031

Chi tiết của các cột dữ liệu như bảng sau:

Bảng 3.2: Chi tiết về các trường dữ liệu của các file csv thu được từ cRIO 9031

Trường dữ liệu	Giá trị đo
G11ACC01Y;G11ACC01Z;	Gia tốc kế - đo dao động dầm theo phương Y và Phương Z
G01WEA01AT	Nhiệt độ không khí
G01WEA01AP	Áp suất không khí
G01WEA01AH	Độ ẩm không khí
P07ANE01WS	Vận tốc gió
P07INC01; P07INC02	Độ nghiêng chân tháp
times	Thời gian dưới dạng mã Unix của milliseconds

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	ThisRecordStartTime=21:23:14 03/24/2020														
2	Time (ms)	G01ACC01Y	G01ACC01Z	G04ACC01Y	G04ACC01Z	G04ACC02Y	G04ACC02Z	C02ACC01Y	C02ACC01Z	C02ACC02Y	C02ACC02Z	C02ACC03Y	C02ACC03Z	C02ACC04Y	C02ACC04Z
3	1585084994659.0000	-0.005009	0.678991	-0.046332	1.042375	0.011392	1.044076	-0.00353	0.945749	0.045367	0.935846	0.035896	-2.09878	-0.04895	0.930488
4	1585084994679.0000	-0.00499	0.679028	-0.046315	1.042659	0.011347	1.043983	-0.0034	0.944868	0.045232	0.936456	0.035819	-2.09867	-0.04883	0.929832
5	1585084994699.0000	-0.004952	0.678966	-0.046235	1.042948	0.011323	1.043878	-0.00275	0.944326	0.046082	0.938024	0.035522	-2.09907	-0.04823	0.928395
6	1585084994719.0000	-0.004988	0.679008	-0.046123	1.0433	0.011325	1.043856	-0.00271	0.94467	0.046528	0.937986	0.035498	-2.09903	-0.04801	0.927903
7	1585084994739.0000	-0.004971	0.678966	-0.046309	1.043497	0.011321	1.044018	-0.0026	0.944985	0.04657	0.937315	0.035538	-2.09895	-0.04785	0.927591
8	1585084994759.0000	-0.004967	0.678985	-0.04628	1.043777	0.011255	1.044462	-0.00284	0.945115	0.046624	0.936159	0.035796	-2.09943	-0.04775	0.928147
9	1585084994779.0000	-0.004979	0.67905	-0.046165	1.043637	0.011257	1.044661	-0.00298	0.944994	0.046596	0.935493	0.035892	-2.10039	-0.0477	0.928601
10	1585084994799.0000	-0.004907	0.679029	-0.0463	1.043471	0.011136	1.044743	-0.00286	0.944725	0.04648	0.934933	0.035853	-2.10142	-0.04781	0.928926
11	1585084994819.0000	-0.00492	0.679028	-0.046213	1.043245	0.011211	1.045117	-0.00223	0.944266	0.046763	0.933891	0.035902	-2.10317	-0.04829	0.928893
12	1585084994839.0000	-0.004925	0.679006	-0.046247	1.043137	0.01114	1.045262	-0.00207	0.944565	0.04677	0.933414	0.035911	-2.10358	-0.04849	0.928947
13	1585084994859.0000	-0.004954	0.678999	-0.046267	1.043031	0.011209	1.045213	-0.00195	0.945151	0.046732	0.932988	0.035724	-2.1037	-0.04866	0.928975
14	1585084994879.0000	-0.004922	0.679038	-0.046221	1.042629	0.01108	1.04519	-0.00202	0.945798	0.046511	0.93276	0.035773	-2.10345	-0.04862	0.929216
15	1585084994899.0000	-0.004795	0.678991	-0.046232	1.041739	0.011114	1.045658	-0.00247	0.947033	0.045996	0.932692	0.035609	-2.10227	-0.04857	0.92965
16	1585084994919.0000	-0.00478	0.678968	-0.04617	1.041456	0.011123	1.04593	-0.00251	0.947441	0.04582	0.93309	0.035982	-2.10163	-0.0488	0.929196
17	1585084994939.0000	-0.0048	0.679023	-0.046257	1.041253	0.011235	1.045983	-0.00254	0.947311	0.045529	0.933688	0.036285	-2.10125	-0.04902	0.928736

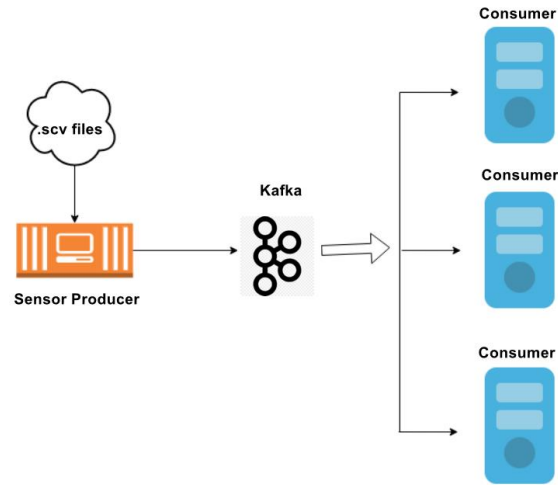
Hình 3.8: Các cột dữ liệu trong file csv thu được từ cRIO 9065

Bảng 3.3: Chi tiết về các trường dữ liệu của các file csv thu được từ cRIO 9065

Trường dữ liệu	Giá trị đo
G01ACC01Y; G01ACC01Z	Giá trị dao động dầm theo phương Y,Z
G04ACC01Y; G04ACC01Z	Giá trị dao động dầm theo phương Y,Z
G04ACC02Y; G04ACC02Z	Giá trị dao động dầm theo phương Y,Z
C02ACC01Y; C02ACC01Z	Giá trị dao động cấp 2218 theo phương Y,Z
C02ACC02Y; C02ACC02Z	Giá trị dao động cấp 2217 theo phương Y,Z
C02ACC03Y; C02ACC03Z	Giá trị dao động cấp 2216 theo phương Y,Z
C02ACC04Y; C02ACC04Z	Giá trị dao động cấp 2215 theo phương Y,Z
times	Thời gian dưới dạng mã Unix của milliseconds

b. Truyền dữ liệu từ các file csv vào Kafka

Như đã phân tích ở trong phần tổng quan về Kafka, Kafka là một nền tảng truyền dữ liệu phân tán (distributed streaming platform). Dữ liệu nguồn sẽ được thu thập và truyền dẫn thông qua hệ thống Kafka.



Hình 3.9: Sơ đồ luồng dữ liệu vào Kafka

Dữ liệu được sử dụng trong luận văn là dữ liệu dạng *Time-series*, được hệ thống thu thập dữ liệu (DAU - Data Acquisition Unit) của hệ quan trắc trả về liên tục mỗi phút sẽ thu được 2 files dữ liệu dưới định dạng .csv như hình 3.10. Trong mỗi file dữ liệu có 3000 bản ghi được ghi cách nhau 02 millisecond.

1	Time (ms)	G11ACC01Y	G11ACC01Z	P06INC01	P07INC01	P07INC02	G01WEAO'	G01WEAO'	G01WEAO'	G01WEAO'	G01WEAO'	P07ANE01	P07ANE01WD
2978	1585267273606	0.016071	0.094773	0.012577	0.01171	0.011707	0	31.2	27.7	78	1005	1.7	83.9
2979	1585267273626	0.015922	0.094934	0.012578	0.011712	0.011682	0	31.2	27.7	78	1005	1.7	83.9
2980	1585267273646	0.015976	0.095042	0.012576	0.011665	0.011682	0	31.2	27.7	78	1005	1.7	83.9
2981	1585267273666	0.015984	0.095045	0.012577	0.011666	0.011681	0	31.2	27.7	78	1005	1.7	83.9
2982	1585267273686	0.016002	0.09513	0.012577	0.011666	0.011682	0	31.2	27.7	78	1005	1.7	83.9
2983	1585267273706	0.015977	0.095188	0.012577	0.011665	0.011683	0	31.2	27.7	78	1005	1.7	83.9
2984	1585267273726	0.01608	0.095509	0.012578	0.011667	0.011684	0	31.2	27.7	78	1005	1.7	83.9
2985	1585267273746	0.016044	0.095534	0.012618	0.011666	0.011684	0	31.2	27.7	78	1005	1.7	83.9
2986	1585267273766	0.016134	0.095596	0.012618	0.011667	0.011684	0	31.2	27.7	78	1005	1.7	83.9
2987	1585267273786	0.016224	0.095679	0.012619	0.011701	0.011674	0	31.2	27.7	78	1005	1.7	83.9
2988	1585267273806	0.016067	0.095814	0.012618	0.011701	0.011674	0	31.2	27.7	78	1005	1.7	83.9
2989	1585267273826	0.016157	0.095846	0.012618	0.011701	0.011674	0	31.2	27.7	78	1005	1.7	83.9
2990	1585267273846	0.01619	0.095804	0.012619	0.011701	0.011674	0	31.2	27.7	78	1005	1.7	83.9
2991	1585267273866	0.016209	0.095655	0.012619	0.0117	0.011674	0	31.2	27.7	78	1005	1.7	83.9
2992	1585267273886	0.016129	0.095595	0.012618	0.0117	0.011674	0	31.2	27.7	78	1005	1.7	83.9
2993	1585267273906	0.016181	0.095535	0.012618	0.0117	0.011674	0	31.2	27.7	78	1005	1.7	83.9
2994	1585267273926	0.016211	0.095494	0.012618	0.0117	0.011673	0	31.2	27.7	78	1005	1.7	83.9
2995	1585267273946	0.01611	0.095255	0.012618	0.011699	0.011723	0	31.2	27.7	78	1005	1.7	83.9
2996	1585267273966	0.016179	0.095245	0.012618	0.011712	0.011723	0	31.2	27.7	78	1005	1.7	83.9
2997	1585267273986	0.016114	0.095304	0.012618	0.011711	0.011723	0	31.2	27.7	78	1005	1.7	83.9
2998	1585267274006	0.016144	0.095323	0.012617	0.011712	0.011723	0	31.2	27.7	78	1005	1.7	83.9
2999	1585267274026	0.016168	0.095364	0.012618	0.011711	0.011723	0	31.2	27.7	78	1005	1.7	83.9
3000	1585267274046	0.01622	0.095295	0.012618	0.011711	0.011723	0	31.2	27.7	78	1005	1.72	60.9

Hình 3.10: Các bản ghi trong mỗi file dữ liệu thu được sau mỗi phút

Để truyền dữ liệu một cách liên tục từ các file CSV trong thư mục dữ liệu vào Kafka, ta sẽ sử dụng plugin **kafka-connect-spooldir** [20]. Nhằm phục vụ cho việc đọc, và lấy data, để quản lý nguồn dữ liệu đã được xử lý ta sẽ tạo ba thư mục như sau:

- Một thư mục chứa các file dữ liệu gốc (*.\source*)
- Một thư mục chứa các file dữ liệu đã được xử lý (*.\finished*)
- Và một thư mục chứa bất kỳ các file lỗi (*.\error*)

```
PS D:\hethongcanhbaoluanvan> ls .\Data\source\

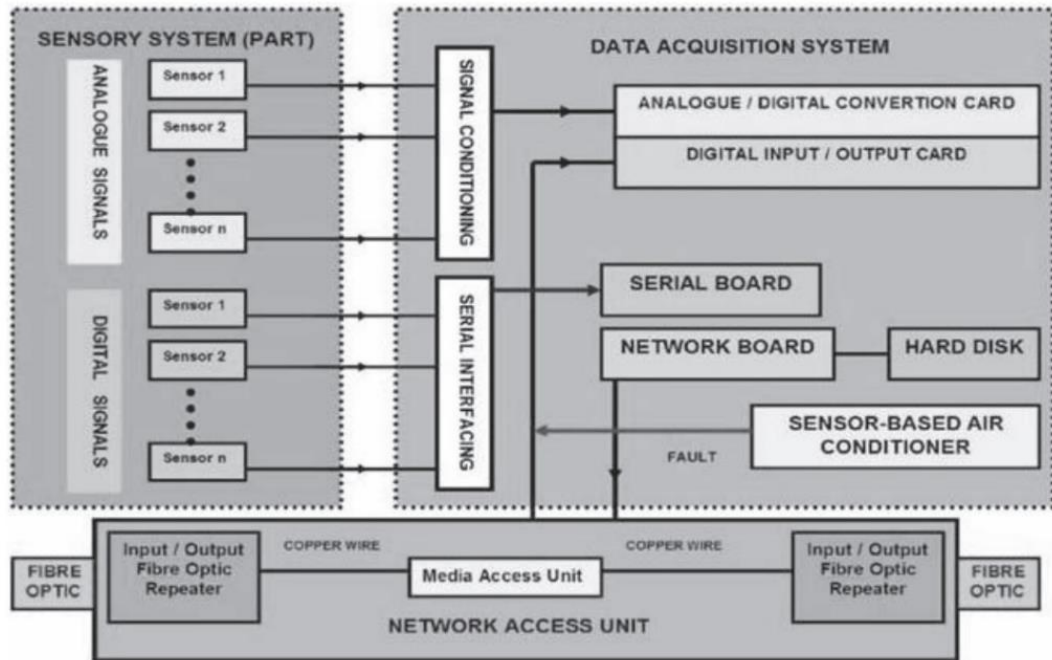
Directory: D:\hethongcanhbaoluanvan\Data\source

Mode                LastWriteTime         Length Name
----                -
-a-----          3/27/2020  12:02 AM         408774 DL PhuMy 9031 03_27_2020 _ 00_00_14.csv
-a-----          3/27/2020  12:03 AM         408724 DL PhuMy 9031 03_27_2020 _ 00_01_14.csv
-a-----          3/27/2020  12:04 AM         409099 DL PhuMy 9031 03_27_2020 _ 00_02_14.csv
-a-----          3/27/2020  12:05 AM         408849 DL PhuMy 9031 03_27_2020 _ 00_03_14.csv
-a-----          3/27/2020  12:06 AM         409076 DL PhuMy 9031 03_27_2020 _ 00_04_14.csv
-a-----          3/27/2020  12:01 AM         456190 DL PhuMy 9065 03_27_2020 _ 00_00_14.csv
-a-----          3/27/2020  12:02 AM         456190 DL PhuMy 9065 03_27_2020 _ 00_01_14.csv
-a-----          3/27/2020  12:03 AM         456190 DL PhuMy 9065 03_27_2020 _ 00_02_14.csv
-a-----          3/27/2020  12:05 AM         456190 DL PhuMy 9065 03_27_2020 _ 00_04_14.csv
-a-----          3/27/2020  12:06 AM         456190 DL PhuMy 9065 03_27_2020 _ 00_05_14.csv
```

Hình 3.11: Các thư mục lưu trữ dữ liệu

b. Lấy dữ liệu từ Kafka thông qua Spark structured Streaming vào database

Dữ liệu từ các file .csv là dữ liệu thu được từ hệ quan trắc cầu, dữ liệu đã được xử lý nhiều thông qua hệ thống thu thập data của hệ quan trắc (DAU) như Hình 3.12. Các cảm biến trên cầu sẽ thu thập tín hiệu đo và truyền tín hiệu tới bộ thu nhận dữ liệu DAU. Các bộ thu nhận dữ liệu DAU này thu nhận các tín hiệu từ các cảm biến, số hóa các tín hiệu analog và truyền các dữ liệu vào phòng trung tâm. Việc bố trí tối ưu các bộ thu nhận dữ liệu đã được xem xét kỹ để khử nhiễu/tổn thất khi truyền và để bảo vệ các bộ DAU khỏi nhiễu nhằm đảm bảo chất lượng và tính chính xác của số liệu thu nhận.



Hình 3.12: Sơ đồ thu thập dữ liệu từ hệ hệ quan trắc cầu

(Nguồn tham khảo [5])

Spark Structured Streaming nhận dữ liệu từ các nguồn streaming data, cụ thể ở đây là Kafka. Để ghi dữ liệu vào InfluxDB, trước hết ta phải tạo một **Class InfluxDBWriter**. Class này có hai methods là – **switch_database** và **write_rows**.

Bên cạnh đó, ta cũng tạo một Kafka consumer để lấy dữ liệu từ Kafka. Hàm **value_deserializer** được sử dụng để chuyển các bản ghi đến thành dạng JSON để có thể phân tích cú pháp một cách dễ dàng hơn:

```

1 def main():
2     config = read_yaml(path)
3     logging.info("Starting writing to InfluxDb")
4
5     consumer = KafkaConsumer(
6         config["kafka_sink"]["topic"],
7         bootstrap_servers=config["kafka_sink"]["bootstrap_servers"],
8         enable_auto_commit=True,
9         value_deserializer=lambda x: json.loads(x.decode('utf-8')))
10
11     db = InfluxDb(config["database"]["host"], config["database"]["database"], config["database"]["port"])
12
13     for message in consumer:
14         message = message.value
15         db.write_rows(message)

```

Hình 3.13: Mã nguồn lấy dữ liệu từ Kafka

Để ghi dữ liệu vào Kafka, ta cần làm ngược lại những gì đã làm khi đọc dữ liệu từ Kafka. Đoạn mã sau chuyển đổi dữ liệu thành dạng JSON và bắt đầu stream.

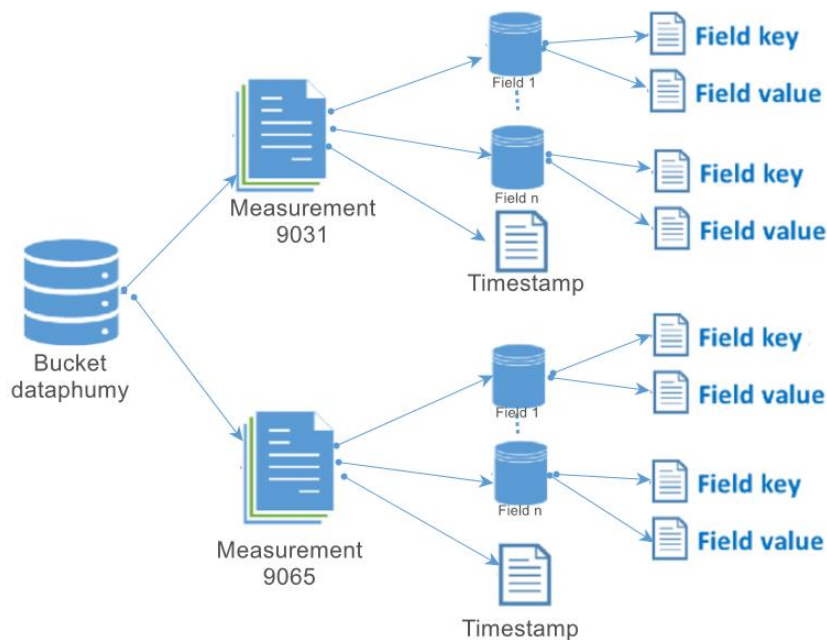
```

1 df = df.withColumn("value", to_json(struct(df.columns)))
2
3 query = df.writeStream \
4   .format("kafka") \
5   .option("kafka.bootstrap.servers", ".join(config["kafka_sink"]["bootstrap_servers"])) \
6   .option("checkpointLocation", config["spark_checkpoint"] + "/kafka/") \
7   .option("topic", config["kafka_sink"]["topic"]) \
8   .start()

```

Hình 3.14: Mã nguồn ghi dữ liệu vào Kafka

Từ cấu trúc dữ liệu của data thu thập được, ta thiết kế cấu trúc database để lưu dữ liệu vào InfluxDB như sau:



Hình 3.15: Mô hình cơ sở dữ liệu

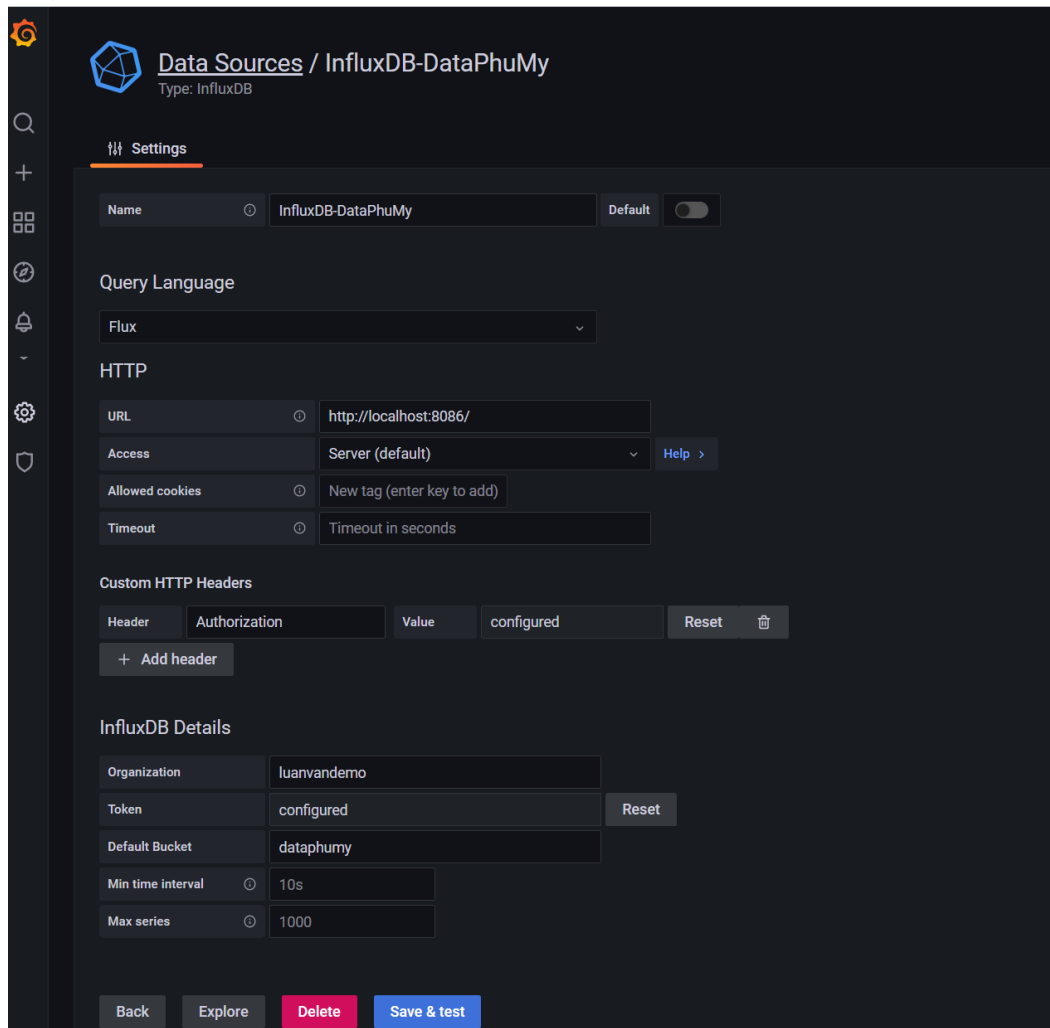
3.3.2. Phân tích và mô hình hóa dữ liệu

a. Cấu hình để Grafana đọc dữ liệu từ InfluxDB

Để phân tích và mô hình hóa dữ liệu, ta kết nối Grafana tới InfluxDB nơi lưu trữ thông tin dữ liệu của cảm biến đã hoàn thành ở phần 3.3.1.

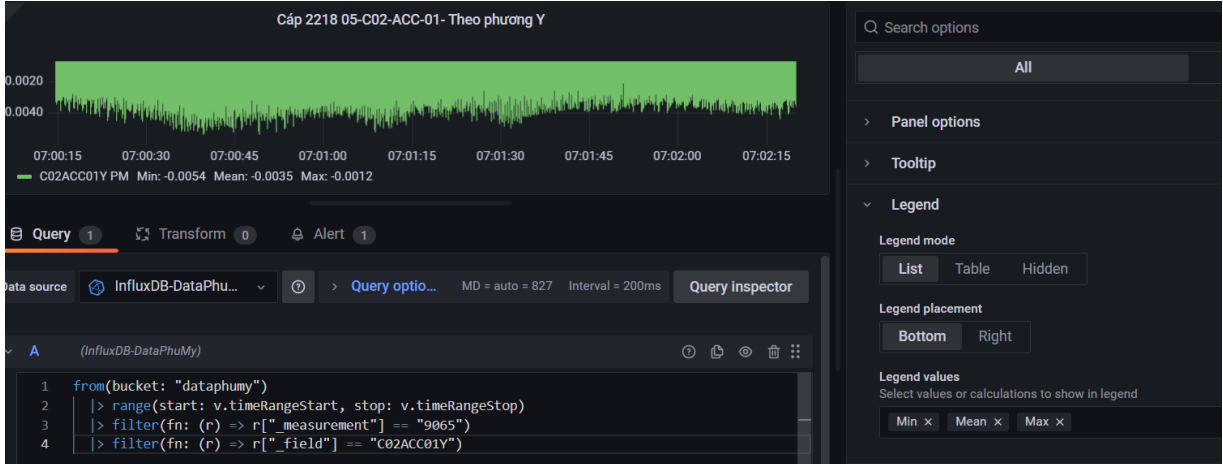
Để kết nối Grafana với InfluxDB ta sử dụng các thông số sau:

- URL: là địa chỉ của máy có chứa InfluxDB, như trong thực nghiệm của luận văn là *http://localhost:8086/*
- Database: thông tin database nơi lưu trữ data của hệ thống



Hình 3.16: Kết nối Grafana với Influxdb

Sau khi hoàn thành kết nối Grafana với InfluxDB ta tiến hành tạo các bảng điều khiển để phân tích, theo dõi quan sát và cảnh báo cho hệ thống.



Hình 3.17: Truy vấn dữ liệu và trực quan hóa dữ liệu

Tương tự như phần truy vấn trên, ta đổ dữ liệu và lựa chọn đồ thị để mô hình hóa dữ liệu lên bảng điều khiển cho các thông số sau dựa vào giá trị cảm biến tương ứng. Phần mô hình hóa dữ liệu thành các biểu đồ sẽ được bộ phận công nghệ cấu hình. Sau khi hoàn thành cấu hình bộ phận công nghệ sẽ bàn giao các bảng điều khiển cho bộ phận vận hành theo dõi, giám sát để phục vụ cho công việc vận hành và bảo dưỡng cầu. Trong mô hình xây dựng của bài luận văn này sử dụng Grafana để biểu diễn dữ liệu về các vấn đề sau:

- Biểu diễn độ nghiêng đỉnh tháp dựa vào dữ liệu của cảm biến Đo Nghiêng trên đỉnh tháp.
- Biểu diễn độ nghiêng chân tháp dựa vào dữ liệu của cảm biến Đo Nghiêng dưới chân tháp.
- Biểu diễn dao động dầm dựa vào dữ liệu của Gia Tốc Kế trên dầm cầu.
- Biểu diễn dao động cáp dựa vào dữ liệu của Gia Tốc Kế trên cáp văng.
- Biểu diễn thời tiết dựa vào số liệu của Cảm Biến Thời Tiết trên dầm cầu.
- Biểu diễn tọa độ dựa vào số liệu của Phong Kế trên tháp cầu.
- Biểu diễn tình hình hoạt động của hệ thống và đưa ra cảnh báo về các thiết bị hư hỏng, mất tín hiệu.

b. Tính lực căng cáp

Để mô hình hóa dữ liệu của lực căng cáp và cảnh báo về ngưỡng lực căng cáp, ta tiến hành tính lực căng cáp trước khi hiển thị và báo cáo. Lực căng cáp được tính toán dựa trên các thông số của cáp văng như bảng 3.4. Đặc trưng cáp và tần số dao động riêng thông qua phương pháp biến đổi nhanh Fourier (FFT) từ dữ liệu thu được ở miền thời gian của gia tốc kế. Trước tiên ta tìm hiểu về các thông số đặc trưng của cáp nơi lắp đặt các gia tốc kế để đo dao động cáp như ở Bảng 3.4: Đặc trưng cáp.

Bảng 3.4: Đặc trưng cáp

(Nguồn công ty cổ phần đầu tư xây dựng Phú Mỹ)

Tên cáp (*)	Số lượng bện cáp	Khối lượng đơn vị (kg/m)	Tiết diện mặt cắt ngang cáp (mm ²)	Đường kính ống bao ngoài (mm)	Hệ số rỗng	Mô men quán tính (cm ⁴)	Mô đun đàn (GPa)	Chiều dài (m)	Góc nghiêng với (độ)	Lực kéo đứt (KN)
2218	70	82.6	10500	341	0.115	7631	195	202.04	24.9	18585
2217	58	68.4	8700	326	0.104	5765		192.40	25.8	15399
2216	54	63.7	8100	284	0.128	4080		182.81	27.1	14337
2215	51	60.2	7650	284	0.121	3854		173.28	28.4	13541

(*) Cáp phía hạ lưu

Tính lực căng trong dây văng từ kết quả đo dao động

$$T = 4mL^2 \left(\frac{f_n}{n} \right)^2 - \frac{EJ\pi^2}{L^2} n^2 \quad (3.1)$$

Trong đó:

- T : là lực căng
- m : là trọng lượng đơn vị theo chiều dài của cáp văng.
- L : là khoảng cách giữa hai đầu cố định của dây văng.
- f_n : Tần số dao động riêng thứ n của cáp văng.
- EJ : Độ cứng chống uốn. $EJ = (B \cdot H^3)/12$, với B là chiều rộng còn H là chiều cao tiết diện cáp.

Xác định tần số dao động riêng của cáp văng

Để xác định tần số dao động riêng của cáp văng ta sử dụng phương pháp biến đổi nhanh Fourier (FFT). Phương pháp FFT dùng thuật toán để biến đổi chuỗi dữ liệu từ miền thời gian sang miền tần số. Công thức cơ bản của phương pháp là:

$$X_{\kappa} = \sum_{j=0}^{N-1} x_j e^{-\frac{2\pi i}{N} \kappa j}, \quad \kappa = 0, 1, 2, \dots, N-1 \quad (3.2)$$

trong đó: x_j - biến dữ liệu trong miền thời gian và x_{κ} - biến dữ liệu miền tần số, $N=2^n$ với n là số nguyên; $t_j=j\Delta t$ với $j = 0$ đến $N-1$; $i = \sqrt{-1}$ là số phức.

```
from scipy.fftpack import fft

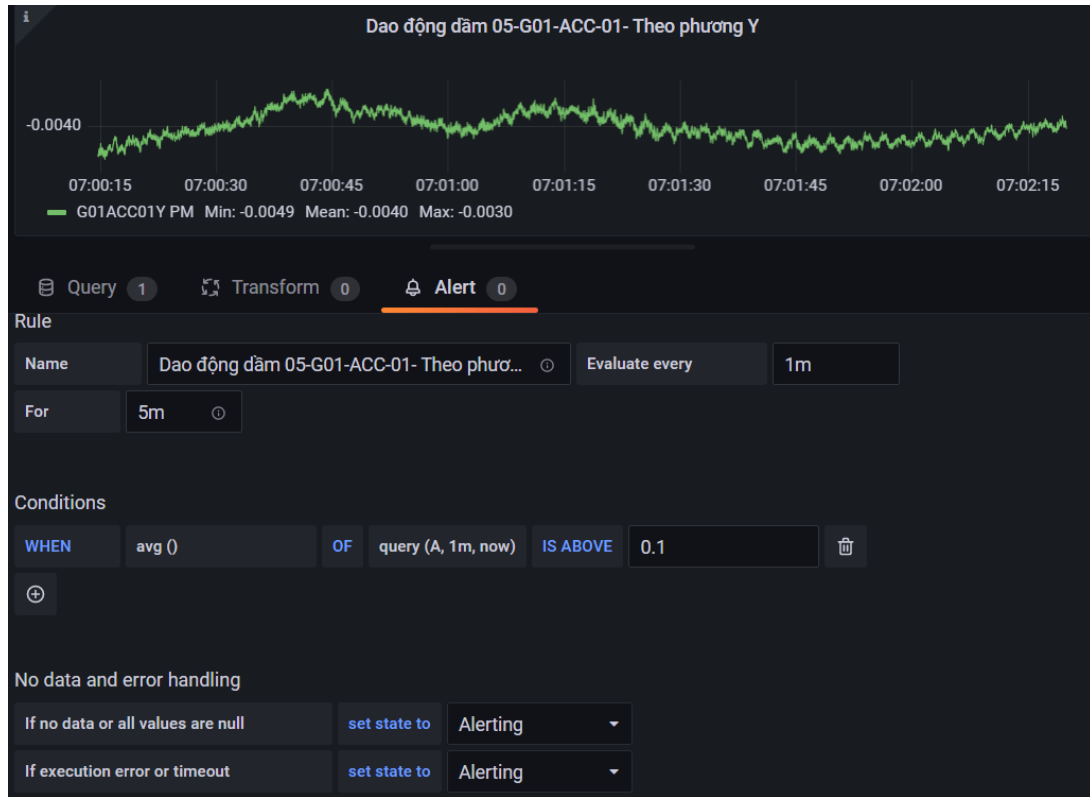
def get_fft_values(y_values, T, N, f_s):
    f_values = np.linspace(0.0, 1.0/(2.0*T), N//2)
    fft_values_ = fft(y_values)
    fft_values = 2.0/N * np.abs(fft_values_[0:N//2])
    return f_values, fft_values
```

Hình 3.18: Mã code biến đổi FFT xác định tần số dao động riêng của cáp văng

c. Cấu hình cảnh báo

Bên cạnh việc xây dựng các màn hình theo dõi, hệ thống còn cấu hình cảnh báo khi giá trị vượt qua ngưỡng hoặc mất tín hiệu (Hình 3.19) và gửi thông tin cảnh báo qua các kênh như slack, email...(Hình 3.20).

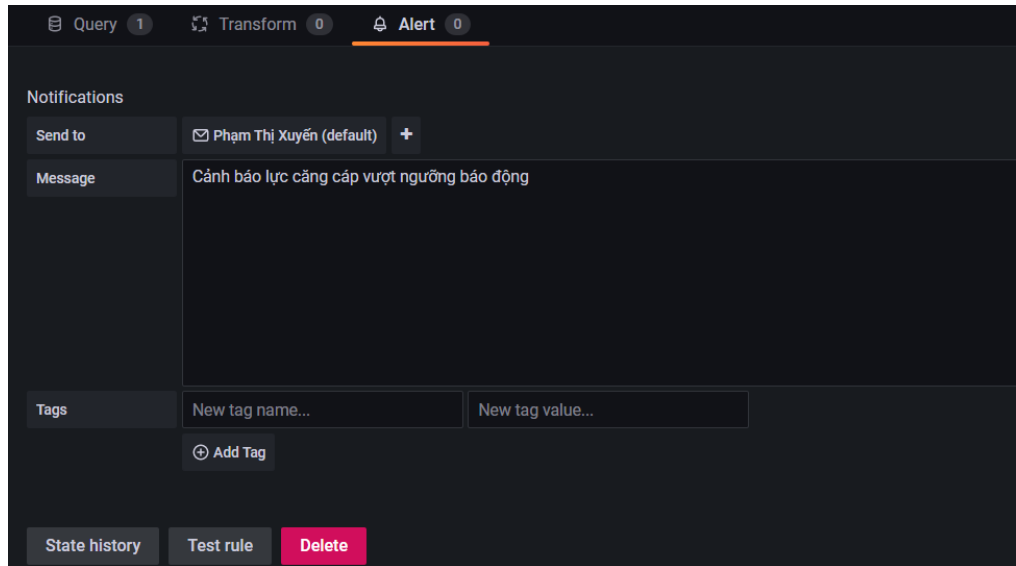
Các quy tắc cảnh báo sẽ được đánh giá dựa vào các điều kiện cần đáp ứng để cảnh báo thay đổi trạng thái và kích hoạt thông báo. Các quy tắc cảnh báo này có thể được thêm mới hoặc chỉnh sửa hoặc xóa thông qua tab cảnh báo ở các bảng biểu đồ. Dưới đây là các trường cần cấu hình cho mỗi cảnh báo:



Hình 3.19: Cài đặt tần số và ngưỡng cảnh báo

- *Name:* Nhập tên mô tả. Tên sẽ được hiển thị trong danh sách các cảnh báo trên bảng điều khiển.
- *Evaluate every:* Chỉ định tần suất người lập lịch sẽ đánh giá quy tắc cảnh báo. Đây được gọi là khoảng thời gian đánh giá.
- *For:* Chỉ định thời gian truy vấn cần vi phạm các ngưỡng đã định cấu hình trước khi kích hoạt thông báo cảnh báo.

Cài đặt kênh nhận cảnh báo



Hình 3.20: Cài đặt kênh nhận cảnh báo

- *Send to*: Chọn một kênh thông báo cảnh báo nếu bạn đã thiết lập một kênh.
- *Message*: Nhập tin nhắn văn bản sẽ được gửi trên kênh thông báo. Một số trình thông báo hỗ trợ chuyển đổi văn bản sang HTML hoặc các định dạng khác.
- *Tags*: Chỉ định danh sách các thẻ (khóa / giá trị) được đưa vào thông báo.

Trong giới hạn của luận văn này tôi sử dụng gửi cảnh báo thông qua email, để gửi thông tin cảnh báo vượt ngưỡng báo động được cài đặt.

Để mail có thể gửi đi được thông qua phương thức SMTP ta phải cấu hình các thông tin trong file grafana.ini như sau:

```

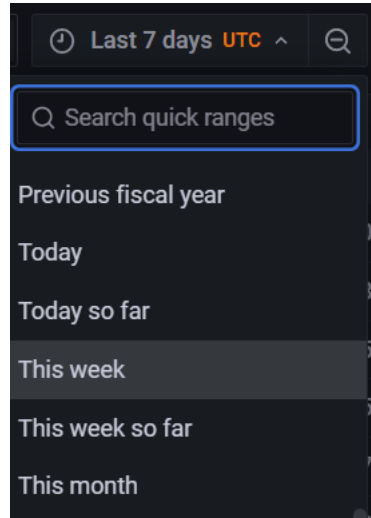
1  ##### SMTP / Emailing #####
2  [smtp]
3  enabled = true
4  host = localhost:25
5  user =
6  # If the password contains # or ; you have to wrap it with triple quotes. Ex ""#password;""
7  password =
8  cert_file =
9  key_file =
10 skip_verify = false
11 from_address = admin@grafana.localhost
12 from_name = Grafana
13 ehlo_identity =

```

Hình 3.21: Cấu hình SMTP để gửi mail trong Grafana

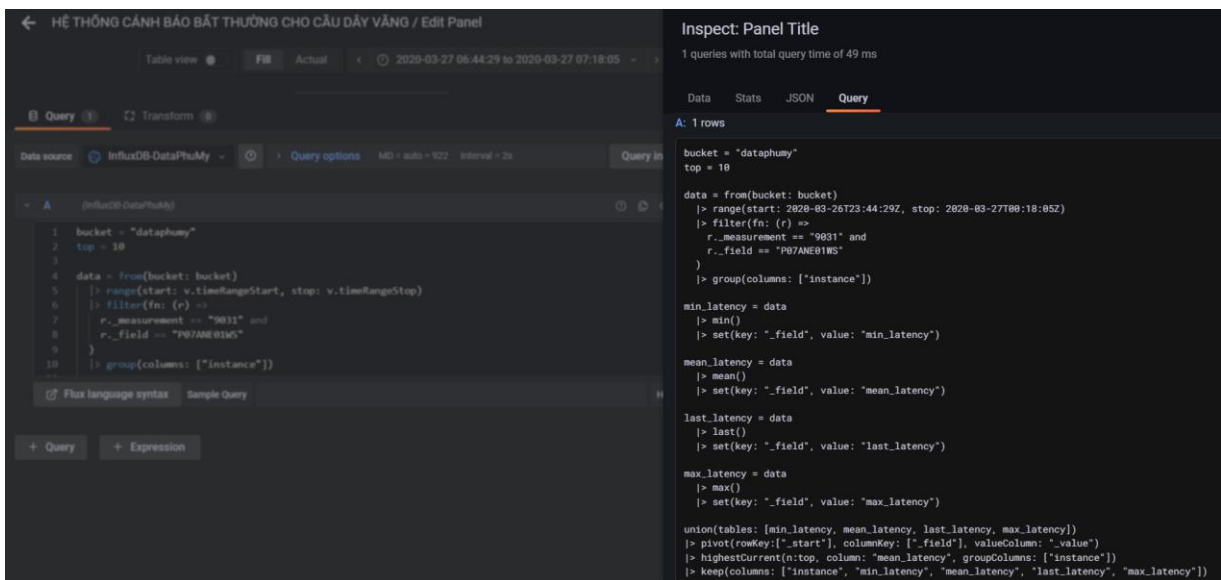
d. Cấu hình báo cáo

Sử dụng panel dạng table để hiển thị các báo cáo theo thời gian quy định như Ngày, Tuần, hoặc Tháng....



Hình 3.22: Lựa chọn mốc thời gian cho các bảng báo cáo

Sử dụng các hàm *union* and *pivot* để tổng hợp các giá trị lớn nhất, nhỏ nhất hoặc các giá trị trung bình của dữ liệu theo thời gian.



Hình 3.23: Mã nguồn lấy và tính toán các giá trị min, max, mean từ cơ sở dữ liệu

3.3.3. Quản lý khai thác và bảo trì cầu dây văng Phú Mỹ dựa trên kết quả phân tích

Dựa trên Dashboard được xây dựng bằng Grafana, các đơn vị vận hành cầu sẽ theo dõi các thông số trên bảng điều khiển, khi có một thông số bất kỳ vượt ngưỡng báo động, hệ thống sẽ gửi email cảnh báo tới email đã được cài đặt thông qua ứng dụng Grafana, các đơn vị vận hành sẽ tiến hành các biện pháp nghiệp vụ để đảm bảo an toàn cho cầu và cho các phương tiện đang và sắp lưu thông trên cầu nắm được.

CHƯƠNG 4: ĐÁNH GIÁ KẾT QUẢ THỬ NGHIỆM

Chương này sẽ mô tả chi tiết về quá trình truyền phát dữ liệu từ các cảm biến. Sau đó sẽ mô tả quá trình xử lý, lưu trữ, quản lý các dữ liệu đã thu thập được. Dựa vào các yêu cầu cần theo dõi như độ nghiêng, dao động dầm, dao động cáp... để phân tích và xuất trình báo cáo trực quan. Cuối cùng là một số nhận xét về kết quả thử nghiệm.

4.1. Mô tả về quá trình truyền dữ liệu phát dữ liệu

Toàn bộ dữ liệu được thu thập trong thời gian chạy thử nghiệm đã được đọc, truyền và import vào cơ sở dữ liệu InfluxDB. Như đã trình bày ở mục 3.3.1 b, plugin **kafka-connect-spooldir** sẽ giám sát thư mục nguồn (`.\Data\source`), khi có bất kỳ file dữ liệu mới nào được đưa vào thư mục nguồn, thì hệ thống sẽ đọc, truyền dữ liệu đó tới Kafka ngay lập tức. Điều này đảm bảo được dữ liệu luôn được ghi vào cơ sở dữ liệu và phát dữ liệu trong thời gian thực. Các file dữ liệu sau khi được xử lý xong được chuyển qua folder *finished*.

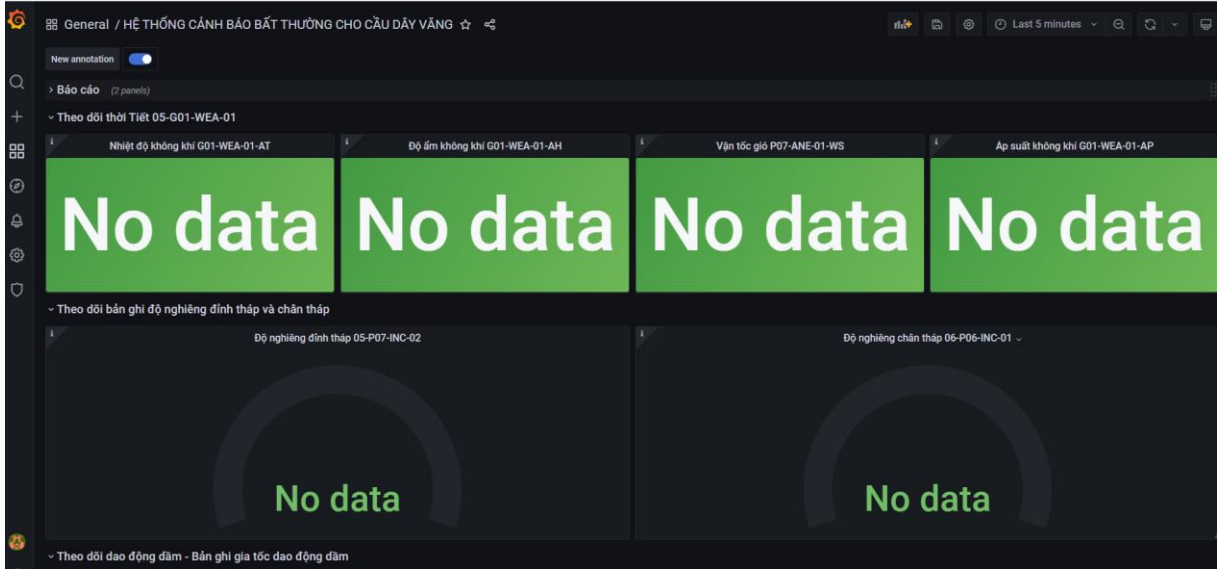
```
Directory: D:\hethongcanhbaoluanvan\Data\finished
```

Mode	LastWriteTime	Length	Name
-a----	3/27/2020 12:02 AM	408774	DL PhuMy 9031 03_27_2020 _ 00_00_14.csv
-a----	3/27/2020 12:03 AM	408724	DL PhuMy 9031 03_27_2020 _ 00_01_14.csv
-a----	3/27/2020 12:04 AM	409099	DL PhuMy 9031 03_27_2020 _ 00_02_14.csv
-a----	3/27/2020 12:05 AM	408849	DL PhuMy 9031 03_27_2020 _ 00_03_14.csv
-a----	3/27/2020 12:06 AM	409076	DL PhuMy 9031 03_27_2020 _ 00_04_14.csv
-a----	3/27/2020 12:01 AM	456190	DL PhuMy 9065 03_27_2020 _ 00_00_14.csv
-a----	3/27/2020 12:02 AM	456190	DL PhuMy 9065 03_27_2020 _ 00_01_14.csv
-a----	3/27/2020 12:03 AM	456190	DL PhuMy 9065 03_27_2020 _ 00_02_14.csv
-a----	3/27/2020 12:05 AM	456190	DL PhuMy 9065 03_27_2020 _ 00_04_14.csv
-a----	3/27/2020 12:06 AM	456190	DL PhuMy 9065 03_27_2020 _ 00_05_14.csv

Hình 4.1: Các file dữ liệu chuyển sang folder *finished* sau khi truyền dữ liệu kết thúc

Khi thư mục nguồn bị ngắt kết nối (không có dữ liệu được đưa vào) thì hệ thống truyền phát trực tiếp sẽ ghi nhận lại việc mất tín hiệu và hiển thị lên trên bảng điều khiển

là không có dữ liệu như Hình 4.2. Dựa vào đó bộ phận vận hành có thể đánh giá xem lại về cảm biến hoặc các trục trặc phát sinh ở bộ phận truyền dẫn dữ liệu liên quan.



Hình 4.2: Bảng điều khiển khi dữ liệu truyền phát trực tiếp bị gián đoạn

4.2. Mô tả hệ thống lưu trữ, xử lý và quản trị dữ liệu

Dữ liệu thực tế được ghi vào InfluxDB rất lớn và được ghi với thời gian cách nhau 02 millisecond.

time	9031.G01WEA01AH	9031.G01WEA01AP	9031.G01WEA01AT	9031.G01WEA01...	9031.G01WEA01...	9031.G11ACC01Y	9031.G11ACC01Z
03/27/2020 00:00:14.46	78.00	1005.00	27.70	31.20	0.00	0.02	0.09
03/27/2020 00:00:14.48	78.00	1005.00	27.70	31.20	0.00	0.02	0.09
03/27/2020 00:00:14.50	78.00	1005.00	27.70	31.20	0.00	0.02	0.09
03/27/2020 00:00:14.52	78.00	1005.00	27.70	31.20	0.00	0.02	0.09
03/27/2020 00:00:14.54	78.00	1005.00	27.70	31.20	0.00	0.02	0.09
03/27/2020 00:00:14.56	78.00	1005.00	27.70	31.20	0.00	0.02	0.09
03/27/2020 00:00:14.58	78.00	1005.00	27.70	31.20	0.00	0.02	0.09
03/27/2020 00:00:14.60	78.00	1005.00	27.70	31.20	0.00	0.02	0.09
03/27/2020 00:00:14.62	78.00	1005.00	27.70	31.20	0.00	0.02	0.09
03/27/2020 00:00:14.64	78.00	1005.00	27.70	31.20	0.00	0.02	0.09
03/27/2020 00:00:14.66	78.00	1005.00	27.70	31.20	0.00	0.02	0.09
03/27/2020 00:00:14.68	78.00	1005.00	27.70	31.20	0.00	0.02	0.09
03/27/2020 00:00:14.70	78.00	1005.00	27.70	31.20	0.00	0.02	0.09
03/27/2020 00:00:14.72	78.00	1005.00	27.70	31.20	0.00	0.02	0.09
03/27/2020 00:00:14.74	78.00	1005.00	27.70	31.20	0.00	0.02	0.09

Hình 4.3: Cơ sở dữ liệu thực tế được ghi vào InfluxDB

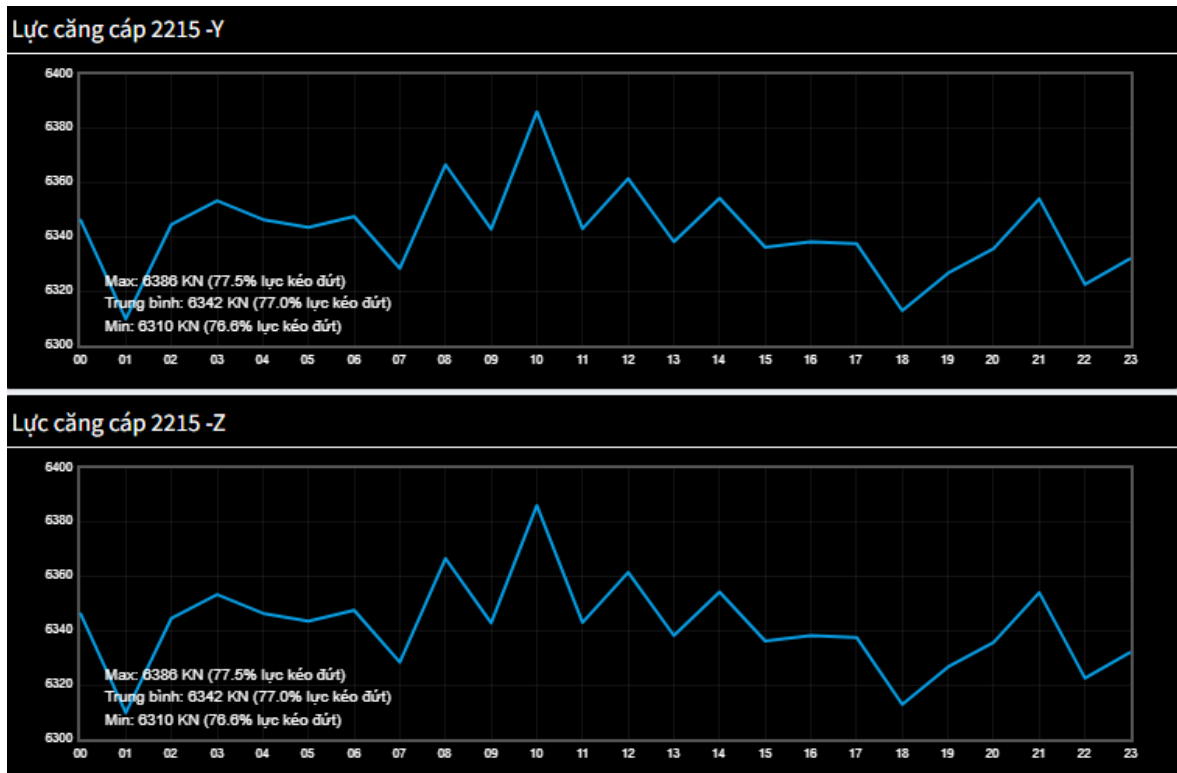
4.3. Phân tích và xuất trình báo cáo

4.3.1 Theo dõi dữ liệu

Phần này sẽ phân tích, xử lý các dữ liệu từ các cảm biến khác nhau và chuyển thành báo cáo trực quan để các phòng ban, bộ phận nắm bắt được kết quả qua đó lên kế hoạch quản lý, khai thác và bảo trì cầu dây văng Phú Mỹ theo kết quả phân tích.

Theo dõi lực căng dây cáp

Kết quả trực quan hóa của lực căng dây cáp được minh họa như hình sau:



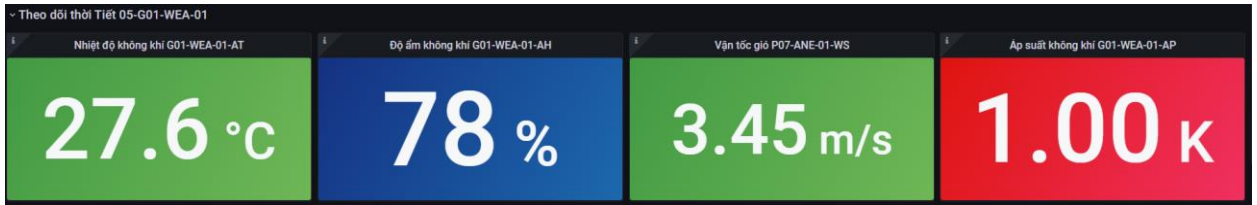
Hình 4.4: Đồ thị theo dõi lực căng dây cáp 2215

Đồ thị cho biết được lực căng ở thời điểm hiện tại với cả 4 cáp 2218, 2217, 2216, 2215 theo cả phương Y và Phương Z. Đối với mỗi thời điểm hệ thống đồng thời cho biết được lực căng trung bình (mean), lực căng nhỏ nhất ghi nhận được (min) và lực căng lớn

nhất đã ghi nhận được (max). Đồng thời hệ thống cũng hiện thị được phân trăm so với lực kéo đứt cáp tương ứng như ảnh trên.

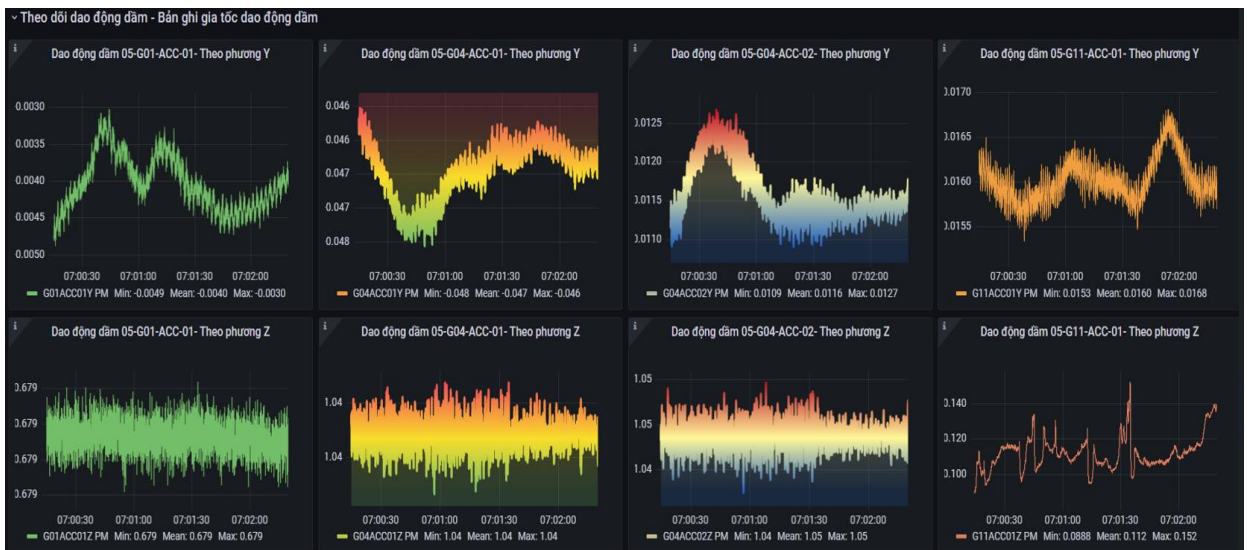
Theo dõi thời tiết

Theo dõi thời tiết bao gồm các thông số theo dõi như: Nhiệt độ không khí, độ ẩm không khí, vận tốc gió trên mặt cầu và áp suất không khí.



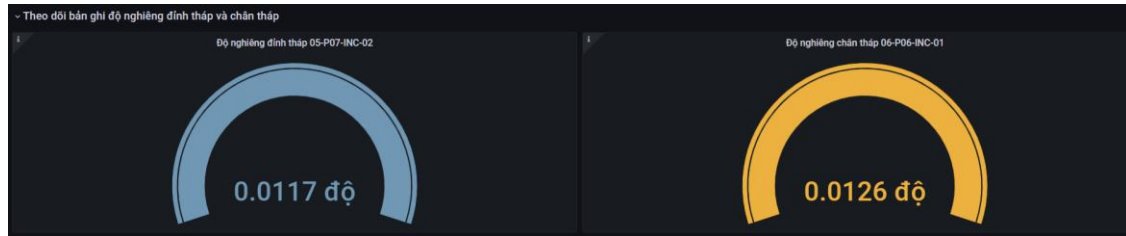
Hình 4.5: Bảng điều khiển theo dõi các thông tin thời tiết

Theo dõi dao động dầm



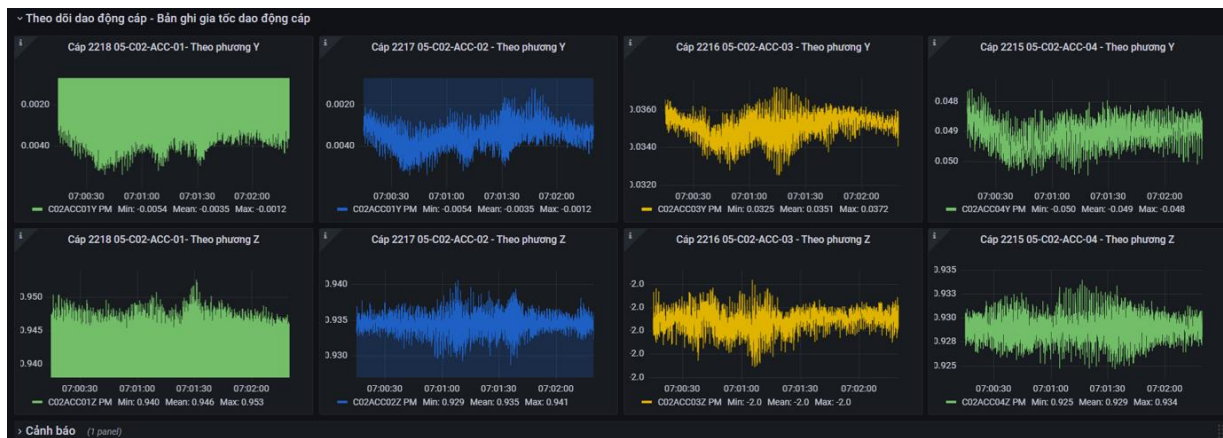
Hình 4.6: Bảng điều khiển theo dõi các thông tin dao động của 4 cáp theo các phương cùng các giá trị min, max, mean tương ứng

Theo dõi độ nghiêng



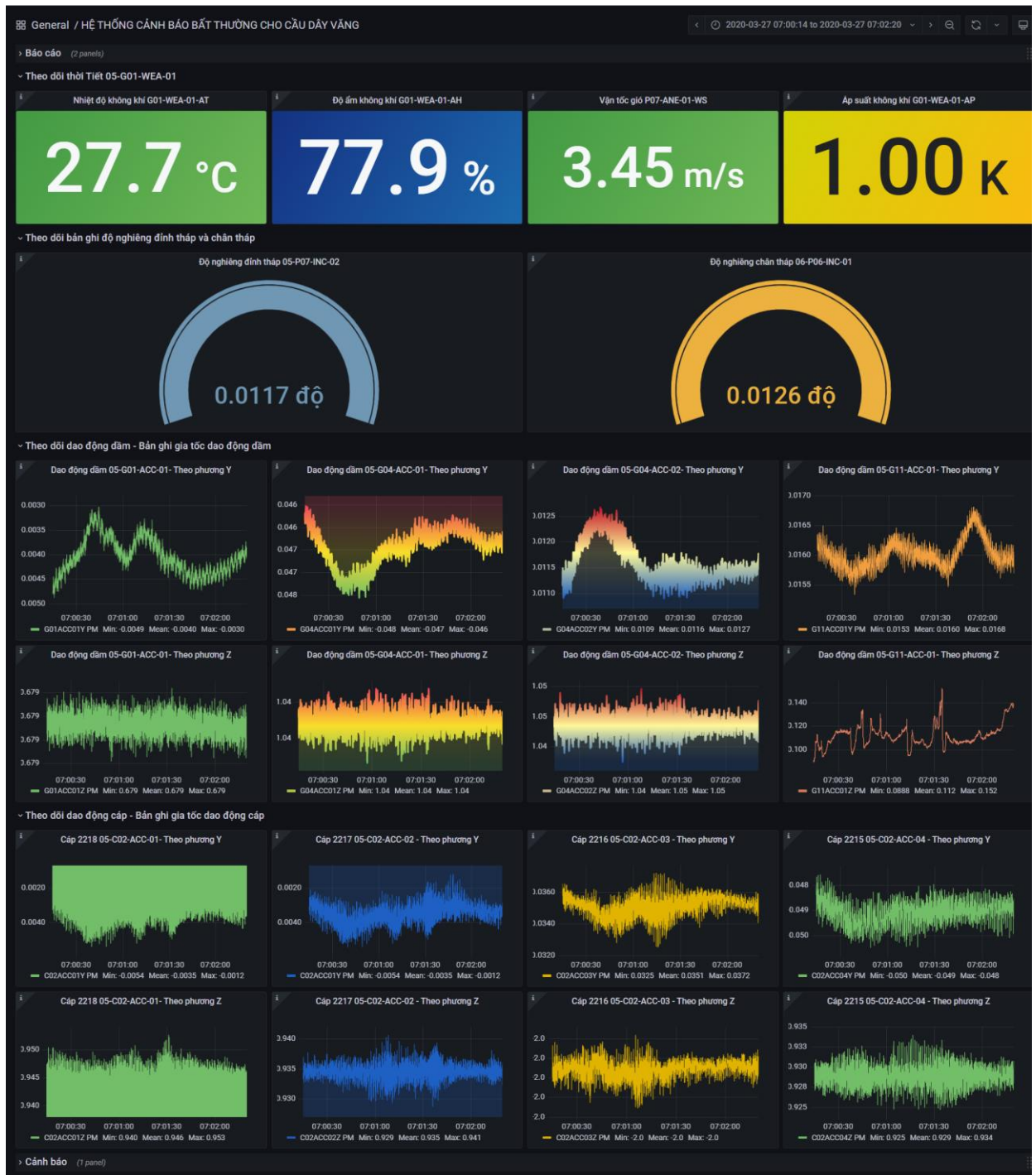
Hình 4.7: Bảng điều khiển theo dõi độ nghiêng đỉnh tháp và chân tháp

Theo dõi dao động cáp



Hình 4.8: Bảng điều khiển theo dõi các thông tin dao động dầm theo các phương cùng các giá trị min, max, mean tương ứng

Dashboard tổng quát của hệ thống giám sát cầu dây văng đã triển khai.



Hình 4.9: Dashboard của hệ thống giám sát cầu dây văng đã triển khai

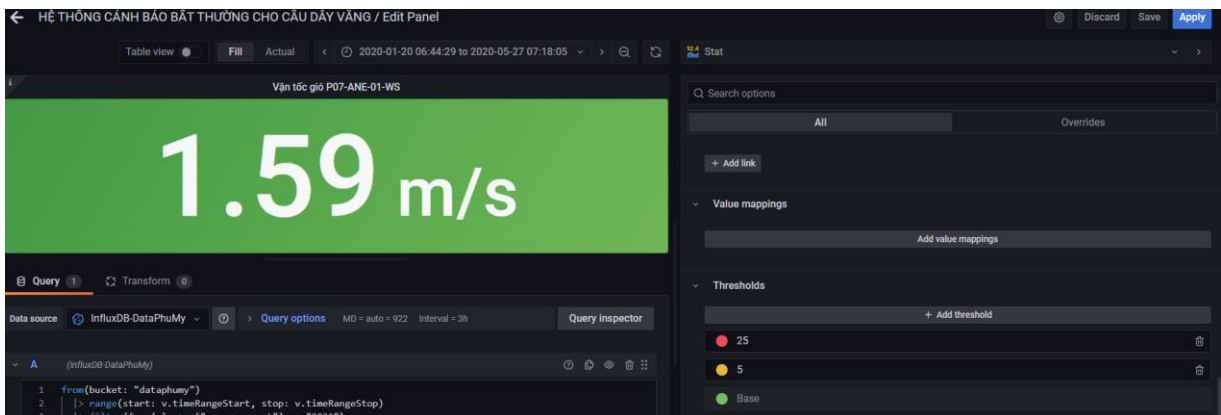
4.3.2. Cảnh báo

Sử dụng bảng giá trị cảnh báo (Bảng 4.1) để cài đặt các ngưỡng cảnh báo cho hệ thống, khi giá trị vượt ngưỡng thì hệ thống sẽ hiển thị màu sắc để phân biệt các giá trị bất thường đồng thời cũng sẽ gửi email đính kèm giá trị cảnh báo.

Bảng 4.1: Giá trị cảnh báo

(Nguồn công ty cổ phần đầu tư xây dựng Phú Mỹ)

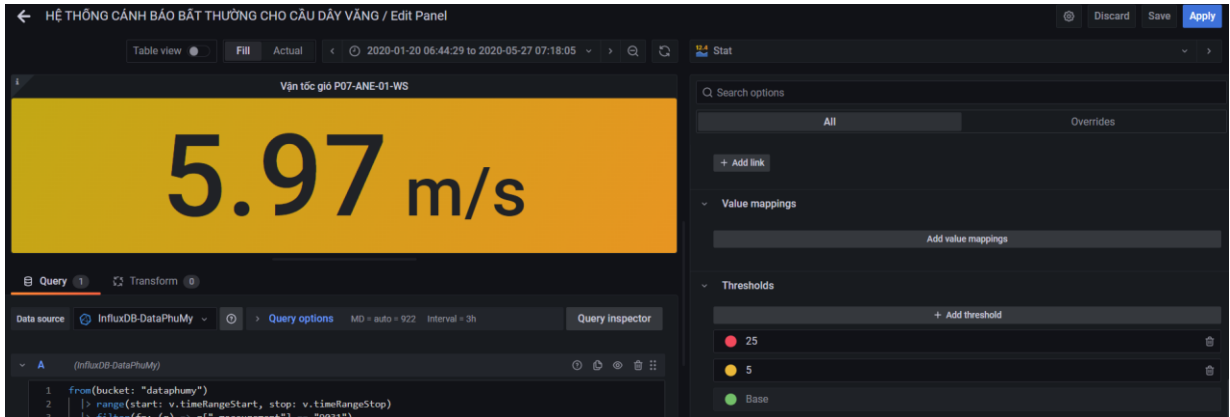
Loại cảnh báo	Giá trị
Ngưỡng vận tốc gió báo động (m/s)	25
Ngưỡng vận tốc gió nguy hiểm(m/s)	11
Cận trên ngưỡng lực căng cáp nguy hiểm (% so với lực kéo đứt cáp)	65
Cận dưới ngưỡng lực căng cáp nguy hiểm (% so với lực kéo đứt cáp)	15
Cận trên ngưỡng lực căng cáp báo động (% so với lực kéo đứt cáp)	75
Cận dưới ngưỡng lực căng cáp báo động (% so với lực kéo đứt cáp)	5



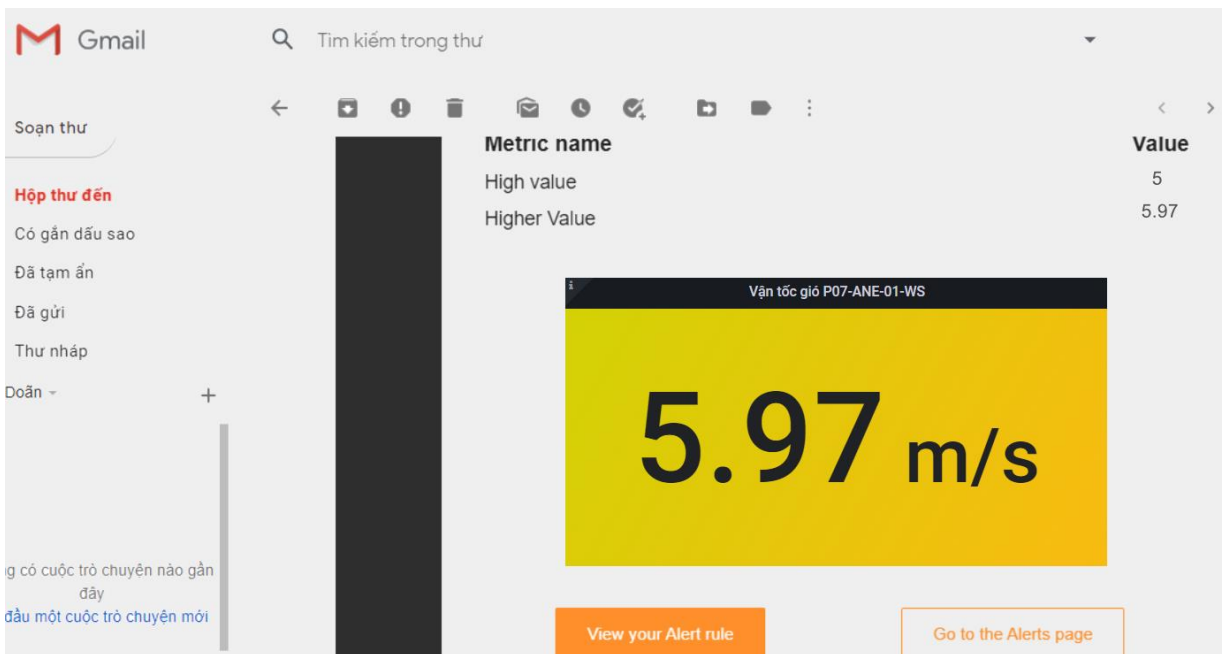
Hình 4.10: Đồ thị hiển thị tốc độ gió bình thường

Màu sắc biểu đồ thay đổi đối với các trạng thái dữ liệu, trong thời gian tiến hành mô phỏng dữ liệu thì giá trị vận tốc gió thực tế chưa vượt ngưỡng cảnh báo, tuy nhiên để minh họa cho thực nghiệm chúng ta sẽ thay đổi mức cảnh báo tốc độ gió cảnh báo

nguy hiểm từ 11m/s xuống thành 5m/s. Thì ta thấy có thời điểm đồ thị đã hiển thị màn hình với màu cảnh báo để cho biết tốc độ gió đã vượt qua mức cảnh báo nguy hiểm như Hình 4.11 và Hình 4.12.



Hình 4.11: Đồ thị hiển thị tốc độ gió vượt ngưỡng cảnh báo được cài đặt



Hình 4.12: Cảnh báo gửi về email khi giá trị vượt ngưỡng cảnh báo đã cài đặt

4.3.3. Báo cáo

Báo cáo sẽ được tổng hợp và hiển thị theo bảng với mốc thời gian lựa chọn theo chu kỳ tính tới thời điểm thực tế.

Báo cáo theo Tuần ta lựa chọn chu kỳ thời gian là 7 ngày trước hoặc là thời gian cố định theo tuần như hình minh họa bên dưới.



Báo cáo định kỳ theo tuần							
time	Nhiệt độ không khí (r)	Nhiệt độ không khí (r)	Nhiệt độ không khí (ma)	Lực căng cáp (min)	Lực căng cáp (mean)	Lực căng cáp (max)	Độ n
2022-05-11 03:02:04	28.3	30.3	34.9	6281	6313	6338	
2022-05-12 07:02:04	28	30.4	33.9	6254	6303	6335	
2022-05-13 11:02:04	28.6	30.6	34.2	6283	6332	6365	
2022-05-14 15:02:04	28.9	31.3	38.4	6294	6327	6357	
2022-05-15 19:02:04	28.9	31.1	34.2	6282	6326	6362	
2022-05-16 23:02:04	28.9	31	33.9	6290	6332	6388	

Hình 4.13: Báo cáo kết quả thực tế theo Tuần



Báo cáo định kỳ theo tuần							
time	Nhiệt độ không khí (r)	Nhiệt độ không khí (r)	Nhiệt độ không khí (ma)	Lực căng cáp (min)	Lực căng cáp (mean)	Lực căng cáp (max)	Độ n
2022-05-07 04:47:59	28.9	31.4	33.9	6290	6332	6388	
2022-05-08 05:35:59	29	31	34.5	6299	6330	6361	
2022-05-09 06:23:59	28.1	30.6	35	6287	6323	6349	
2022-05-10 07:11:59	27.7	31	34.3	6281	6308	6332	
2022-05-11 07:59:59	28.6	31	34.1	6288	6327	6386	
2022-05-12 08:47:59	29.1	34.4	38.7	6294	6324	6361	

Hình 4.14: Báo cáo kết quả thực tế theo Tháng

Báo cáo sẽ được gửi về kênh nhận báo cáo (email, slack...) kết quả báo cáo định kỳ theo ngày, tuần, tháng... như ảnh dưới.

Ngày	Nhiệt độ không khí			Lực căng cáp			Độ nghiêng đỉnh tháp		Dao động dầm
				Mar-22					
	(oC)	(oC)	(oC)	(KN)	(KN)	(KN)	(o)	(o)	(m/s ²)
Min	TB	Max	Min	TB	Max	Max	TB	Max – Min	
01	27.5	30.3	35.5	6252	6291	6320	-0.38	-0.57	0.002
02	28.3	30.3	34.9	6281	6313	6338	-0.38	-0.57	0.001
03	28	30.4	33.9	6254	6303	6335	-0.37	-0.57	0.001
04	28.6	30.6	34.2	6283	6332	6365	-0.3	-0.56	0.001
05	28.9	31.3	38.4	6294	6327	6357	-0.28	-0.56	0.001
06	28.9	31.1	34.2	6282	6326	6362	-0.32	-0.56	0.001
07	28.9	31	33.9	6290	6332	6388	-0.33	-0.56	0.001
08	29	31.4	34.5	6299	6330	6361	-0.34	-0.56	0.001
09	28.1	31	35	6287	6323	6349	-0.35	-0.56	0.001
10	27.7	30.6	34.3	6281	6308	6332	-0.32	-0.57	0.001
11	28.6	31	34.1	6288	6327	6386	-0.29	-0.57	0.001
12	29.1	31.4	38.7	6294	6324	6361	-0.31	-0.56	0.001
13	0	19.4	34.4	6284	6319	6342	-0.28	-0.57	0.729
14	0	0	0	6302	6332	6368	-0.33	-0.56	0.002
15	0	0	0	6290	6328	6374	-0.34	-0.57	0.001
16	0	0	0	6286	6326	6368	-0.32	-0.56	0.001

Hình 4.15: Báo cáo kết quả tháng 3 năm 2022

4.4. Nhận xét kết quả thử nghiệm ứng dụng

Kết quả thử nghiệm ứng dụng đã xây dựng được bảng điều khiển hiển thị kết quả dựa trên dữ liệu cảm biến thu được. Màn hình điều khiển minh họa các số liệu trực quan, dễ nhìn và theo dõi. Phần cài đặt các ngưỡng cảnh báo và tin nhắn cảnh báo cũng tối ưu để người vận hành dễ dàng điều chỉnh theo các số liệu được cập nhật mới nhất.

Các ưu điểm của hệ thống

- Hệ thống sử dụng phần mềm mã nguồn mở và các thư viện hỗ trợ mới nhất, từ đó góp phần giảm thời gian cài đặt cũng như tích hợp.
- Hệ thống sử dụng công cụ truyền phát trực tiếp Spark Structured Streaming - phiên bản API mới nâng cấp của Apache Spark 2.0. Spark Structured Streaming có nhiều tính năng cải thiện hơn so với các công cụ truyền phát khác giúp hệ thống xử lý một cách nhất quán và tăng cường khả năng chịu lỗi.
- Hệ thống có tính bảo mật cao vì các phần mềm sử dụng trong mô hình của bài luận văn đều hỗ trợ phân quyền cấu hình cho các user truy cập vào hệ thống và cấp quyền điều chỉnh các thông số theo quyền được cấp.

- Hệ thống xử lý dữ liệu Big data thường rất phức tạp và gồm nhiều thành phần và nhiều khâu xử lý, luận văn cơ bản đã thiết kế một mô hình tinh gọn nhất các thành phần thực sự cần thiết cho hệ thống phù hợp với đặc thù dữ liệu là các thông số quan trọng của Cầu dây văng bằng cách kết hợp một cách hợp lý nhất các nền tảng mã nguồn mở.

Bên cạnh ưu điểm thì hệ thống cũng còn điểm cần cân nhắc

- Hệ thống cơ sở dữ liệu là InfluxDB là dạng cơ sở dữ liệu chuỗi thời gian nên chỉ sử dụng được cho các loại dữ liệu chuỗi thời gian, đối với các dữ liệu không mang dấu thời gian sẽ không hỗ trợ.
- Bên cạnh ưu điểm là mã nguồn mở mới nhất giúp giảm thời gian tích hợp thì đó cũng coi là một nhược điểm nhỏ. Vì hiện tại các hệ thống có sự hỗ trợ kết nối và tương thích với nhau nhưng nếu tương lai một trong các phần mềm cập nhật theo hướng không tương thích với một trong các phần mềm khác thì chúng ta phải kiểm tra lại trước hết là nội dung cập nhật có cần phải update hay không, nếu nó liên quan tới bảo mật thì phải update, khi đó chúng ta sẽ phải kiểm tra đối ứng với các phần mềm khác trong hệ thống.

Hướng nghiên cứu tiếp theo

Từ số liệu của cảm cảm biến, đề tài luận văn hiện tại đã tính toán và cho ra được các chỉ số và cảnh báo dựa vào từng chỉ số. Hướng nghiên cứu tiếp theo của đề tài sẽ phân tích sự ảnh hưởng qua lại giữa các chỉ số kết hợp với phân tích chuyên sâu về số liệu của tình trạng kết cấu cầu để đánh giá được tác động lên toàn bộ cầu, từ đó lên kế hoạch vận hành, sửa chữa và bảo trì hiệu quả hơn.

KẾT LUẬN

Hiện nay có rất nhiều hướng nghiên cứu để triển khai hệ thống cơ sở dữ liệu quan trắc phục vụ tốt cho công tác duy tu bảo dưỡng cầu dây văng trong giai đoạn khai thác. Luận văn đã trình bày một hướng tiếp cận có hiệu quả đó là ứng dụng công nghệ Big Data vào việc phát triển hệ thống cảnh báo bất thường cho cầu dây văng. Đồng thời phương pháp đề xuất sẽ được phân tích và đánh giá bằng mô hình thực nghiệm dựa trên dữ liệu thực tế của công trình cầu Phú Mỹ.

Những đóng góp của luận văn

- Tìm hiểu tổng quan về kiến trúc và công nghệ Big Data.
- Tìm hiểu về cầu dây văng và hệ quan trắc cầu dây văng.
- Tìm hiểu về các công nghệ Big Data là Kafka, Apache Spark Structured Streaming, InfluxDB, Grafana.
- Thiết kế mô hình cảnh báo bất thường cho cầu dây văng và triển khai ứng dụng thực nghiệm.

Hướng phát triển tiếp theo của đề tài

- Từ những đóng góp của luận văn, học viên cũng đặt ra một số vấn đề nghiên cứu như sau:
 - Kết hợp các thông số của các cảm biến và tìm hiểu sự liên quan giữa các thông số với nhau để đánh giá sự ảnh hưởng qua lại giữa các chỉ số và đưa ra ứng xử cho cầu dây văng.
 - Phân tích dữ liệu để dự báo sự thay đổi của cầu dựa trên các chỉ số theo các mùa và sự thay đổi của kết cấu theo thời gian.

DANH MỤC TÀI LIỆU THAM KHẢO

- [1]. Nguyễn Danh Thắng, H.T.H., Trần Nguyễn Nhật Nam, *Hệ thống quan trắc công trình cầu nhịp lớn*. Tạp chí Khoa học Đại học Thủ Dầu Một, 2017.
- [2]. Nam, B.G.T.V.T.v., *Thông tư 52/2013/TT - BGTVT quy định về quản lý, khai thác và bảo trì công trình đường bộ*, B.G.T.V.T.v. Nam, Editor. 2013.
- [3]. Chính, L.M., *Các phương pháp quan trắc và xác định chuyển vị trụ tháp cầu dây văng của hệ thống quan trắc công trình cầu (SHMS)*. Khoa học kỹ thuật Thủy lợi và Môi trường, 2015.
- [4]. Nguyễn Viết Trung, B.X.H., *Hệ thống quan trắc cầu dây văng và cầu dây võng*, NXB Xây Dựng, Hà Nội. 2017.
- [5]. Chinh, L.M., *Analysis and Assessment of Existing Structural Health Monitoring Systems (Shms) of Cable-Stayed Bridge in Vietnam*. Structure and Environment, 2019. **11**(3).
- [6]. Jeong, S., et al., *A Big Data Management and Analytics Framework for Bridge Monitoring*, in *Structural Health Monitoring 2017*. 2017.
- [7]. Jeong, S., et al., *A data management infrastructure for bridge monitoring*. Vol. 9435. 2015.
- [8]. Gunner, S. & Vardanega, Paul & Tryfonas, T & Macdonald, J.H.G. & Wilson, R.E.. (2017). *Rapid Deployment of a WSN on the Clifton Suspension Bridge, UK*. 170. 59-71. 10.1680/jsmic.17.00014.
- [9]. Scullion, A., *Harnessing open source big data technology to transform structural health monitoring from data logging into decision support information*. Structural Faults and Repair. 2018.
- [10]. David Reinsel, J.G., John Rydning, *The Digitization of the World From Edge to Core*. IDC White Paper - #US44413318, 2018.

- [11]. boyd, d. and K. Crawford, *CRITICAL QUESTIONS FOR BIG DATA*. Information, Communication & Society, 2012. **15**(5): p. 662-679.
- [12]. De Paolis, V. De Luca and R. Paiano, "*Sensor data collection and analytics with thingsboard and spark streaming*," 2018 IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems (EESMS), 2018, pp. 1-6, doi: 10.1109/EESMS.2018.8405822.
- [13]. *Apache Kafka*. Available from: <https://kafka.apache.org/>
- [14]. *Apache Spark*. 2020; Available from: <https://spark.apache.org/>.
- [15]. *Influxdb*. 2020; Available from: <https://www.influxdata.com/>.
- [16]. Nasar, M. and M. Abu Kausar, *Suitability Of Influxdb Database For IoT Applications*. 2019.
- [17]. *DB-Engines. Db-engines ranking*. Available from: <https://db-engines.com/en/ranking/time+series+dbms/>.
- [18]. *Grafana*. 2020; Available from: <https://grafana.com/>.
- [19]. *Câu Phú Mỹ*. 2020; https://vi.wikipedia.org/wiki/Câu_Phú_Mỹ/.
- [20]. Kafka Connect connector <https://github.com/jcustenborder/kafka-connect-spooldir>

BẢN CAM ĐOAN

Tôi cam đoan đã thực hiện việc kiểm tra mức độ tương đồng nội dung luận văn qua phần mềm DoIT một cách trung thực và đạt kết quả mức độ tương đồng 7% toàn bộ nội dung luận văn. Bản luận văn kiểm tra qua phần mềm là bản cứng luận văn đã nộp để bảo vệ trước hội đồng. Nếu sai tôi xin chịu hình thức kỷ luật theo quy định hiện hành của Học viện.

TP.HCM, ngày 16 tháng 07 năm 2022

Học viên thực hiện luận văn

Phạm Thị Xuyên



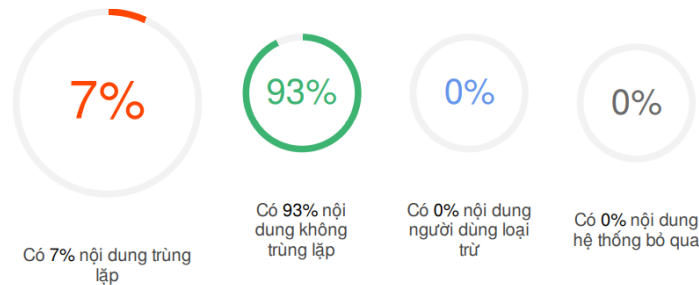
BÁO CÁO KIỂM TRA TRÙNG LẶP

Thông tin tài liệu

Tên tài liệu:	PhamThiXuyenLuanVanThacSi_16072022_final
Tác giả:	Phạm Thị Xuyên
Điểm trùng lặp:	7
Thời gian tải lên:	16:06 16/07/2022
Thời gian sinh báo cáo:	16:09 16/07/2022
Các trang kiểm tra:	67/67 trang



Kết quả kiểm tra trùng lặp



Nguồn trùng lặp tiêu biểu

123docz.net tailieu.vn vjol.info.vn

HỌC VIÊN CAO HỌC

NGƯỜI HƯỚNG DẪN KHOA HỌC

Phạm Thị Xuyên

TS. Huỳnh Trọng Thưa