

**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP.HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO THỰC TẬP DOANH NGHIỆP**  
**CÔNG TY CỔ PHẦN CÔNG NGHỆ TIA CHỚP XANH**

Người hướng dẫn tại công ty : Nguyễn Đình Thanh

Giảng viên hướng dẫn : ThS. Trương Vĩnh Linh

Sinh viên : Ngô Hồng Thông

MSSV : 22649011

Thành phố Hồ Chí Minh, tháng 10 năm 2025

## NHẬN XÉT CỦA ĐƠN VỊ THỰC TẬP

Họ và tên sinh viên: .....MSSV:.....  
Công ty thực tập: .....  
Thời gian thực tập: .....  
Vị trí thực tập: .....

### Nhận xét quá trình thực tập của sinh viên:

.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

....., ngày.....tháng .....năm .....

**Người hướng dẫn**

*(Ký tên và đóng dấu)*

## **NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN**

Họ và tên sinh viên: .....MSSV:.....

Điện thoại:.....Mã lớp: .....

Công ty thực tập: .....

Thời gian thực tập: .....

Vị trí thực tập: .....

### **Đánh giá quá trình thực tập của sinh viên:**

Các kết quả sinh viên đã thực hiện được:.....

.....

.....

.....

Điểm: ..... Bằng chữ:.....

....., ngày.....tháng .....năm .....

**Giáo viên hướng dẫn**

## LỜI CẢM ƠN

Trong thời đại công nghệ số, dữ liệu không chỉ là tài sản quý giá mà còn là nguồn lực chiến lược, đóng vai trò then chốt giúp doanh nghiệp đưa ra các quyết định quan trọng. Việc thu thập, lưu trữ và xử lý dữ liệu hiệu quả trở thành nền tảng vững chắc cho các hệ thống phân tích và ứng dụng trí tuệ nhân tạo (AI) hiện đại.

Được sự phân công của Khoa Công nghệ Thông tin – Trường Đại học Công nghiệp Thành phố Hồ Chí Minh và sự đồng ý tiếp nhận của Công Ty Cổ Phần Công Nghệ Tia Chớp Xanh, em đã có cơ hội thực tập trong thời gian 12 tuần tại công ty. Trong suốt quá trình thực tập, em được tham gia trực tiếp vào các dự án thực tế liên quan đến xây dựng hệ thống thu thập và xử lý dữ liệu lớn, triển khai các kho dữ liệu nhằm tối ưu hóa quy trình phân tích và hỗ trợ ra quyết định.

Bên cạnh việc trau dồi chuyên môn, em còn học hỏi và rèn luyện được tác phong làm việc chuyên nghiệp, tinh thần kỷ luật và đạo đức nghề nghiệp – những yếu tố quan trọng để trở thành một kỹ sư công nghệ thông tin trong tương lai.

Em xin chân thành cảm ơn Khoa Công nghệ Thông tin – Trường Đại học Công nghiệp TP.HCM cùng quý thầy cô đã tạo điều kiện thuận lợi để em thực hiện kỳ thực tập này. Đồng thời, em đặc biệt biết ơn Ban lãnh đạo Công Ty Cổ Phần Công Nghệ Tia Chớp Xanh, các anh/chị đồng nghiệp và đặc biệt là anh Nguyễn Đình Thanh người đã tận tình hướng dẫn, hỗ trợ em trong suốt quá trình thực tập.

Mặc dù đã nỗ lực hoàn thành tốt nhiệm vụ được giao, nhưng do kinh nghiệm thực tế còn hạn chế và kiến thức còn đang trong quá trình rèn luyện, báo cáo này chắc chắn không tránh khỏi những thiếu sót. Em rất mong nhận được sự góp ý của quý thầy cô để báo cáo được hoàn thiện hơn và em có thể nâng cao kiến thức, kỹ năng của bản thân.

Em xin chân thành cảm ơn.

Sinh viên

Ngô Hồng Thông

## MỤC LỤC

LỜI CẢM ƠN.....	3
CHƯƠNG 1. GIỚI THIỆU .....	8
1.1. Mục tiêu thực tập.....	8
1.2. Tổng thời gian thực tập và các mốc thời gian .....	8
1.3. Người hướng dẫn tại doanh nghiệp.....	8
CHƯƠNG 2. DOANH NGHIỆP THỰC TẬP .....	9
2.1. Giới thiệu.....	9
2.2. Cơ cấu tổ chức.....	9
2.3. Văn hóa làm việc ở công ty.....	10
2.3.1. Môi trường làm việc.....	10
2.3.2. Mức độ hỗ trợ/hướng dẫn từ đồng nghiệp và quản lý .....	10
2.3.3. Tần suất họp/review công việc.....	10
CHƯƠNG 3. VAI TRÒ KHI THỰC TẬP .....	12
3.1. Vai trò, chức danh khi thực tập của bạn Chức danh.....	12
3.2. Mô tả về phòng ban/ nhóm mà bạn được phân công.....	12
3.3. Trách nhiệm được công ty giao.....	12
3.4. Trách nhiệm được công ty giao.....	12
CHƯƠNG 4. PROJECTS & TASKS .....	13
4.1 Mô tả project .....	13
4.2. Những nhiệm vụ cụ thể được giao .....	13
4.3. Công cụ, phương pháp hoặc công nghệ được sử dụng.....	14
4.4. Nhận xét cá nhân của bạn đối với mỗi project mà bạn tham gia.....	14
CHƯƠNG 5. NHẬT KÝ CÔNG VIỆC .....	15

5.1. Nhật ký công việc theo tuần.....	15
5.2. Các mốc thời gian quan trọng đạt được .....	16
5.3. Những thách thức gặp phải, và cách đương đầu .....	16
CHƯƠNG 6. KẾT LUẬN .....	17
6.1. Tóm tắt quá trình thực tập .....	17
6.2. Tự đánh giá.....	17
6.2.1. Ưu điểm, khuyết điểm.....	17
6.2.2. Bài học (Học được gì: hard skill, soft skill) .....	17
6.3. Sự ảnh hưởng .....	18
CHƯƠNG 7. THAM KHẢO .....	21

## DANH MỤC HÌNH

<i>Hình 1 Logo Công ty</i> .....	9
----------------------------------	---

## **DANH MỤC BẢNG**

<i>Bảng 1. Trình bày các nhiệm vụ theo giai đoạn.....</i>	13
<i>Bảng 2. Trình bày công việc theo tuần.....</i>	15
<i>Bảng 3. Trình bày thách thức gặp phải.....</i>	16



# CHƯƠNG 1. GIỚI THIỆU

## 1.1. Mục tiêu thực tập

Mục tiêu chính của kỳ thực tập này là làm quen và đi sâu vào lĩnh vực Data Engineering. Cụ thể, tôi mong muốn đạt được các mục tiêu sau:

- Nắm vững kiến thức chuyên môn: Áp dụng các kiến thức lý thuyết đã học tại trường về nguyên tắc hoạt động, kỹ thuật xử lý, lưu trữ dữ liệu lớn vào môi trường thực tế.
- Tham gia vào dự án thực tế: Trực tiếp tham gia vào quá trình xây dựng hệ thống thu thập và xử lý dữ liệu lớn (Big Data), từ đó tích lũy kinh nghiệm thực tiễn quý báu.
- Rèn luyện kỹ năng mềm: Hoàn thiện tác phong làm việc chuyên nghiệp, nâng cao tinh thần kỷ luật và ý thức trách nhiệm để chuẩn bị cho vai trò kỹ sư công nghệ thông tin trong tương lai.

## 1.2. Tổng thời gian thực tập và các mốc thời gian

Tổng thời gian thực tập là 12 tuần, bắt đầu từ ngày 01/06/2025 và kết thúc vào ngày 23/08/2025.

## 1.3. Người hướng dẫn tại doanh nghiệp

Người hướng dẫn trực tiếp của tôi trong suốt quá trình thực tập là anh Nguyễn Đình Thanh, Trưởng nhóm AI tại công ty.

## CHƯƠNG 2. DOANH NGHIỆP THỰC TẬP

### 2.1. Giới thiệu

- Tên công ty: Bluebolt Software (Công ty Cổ Phần Công Nghệ Tia Chớp Xanh).
- Trụ sở chính: 250 Nguyễn Đình Chiểu, P. Võ Thị Sáu, Quận 3, TP. Hồ Chí Minh.
- Website: <https://blueboltsoftware.com>

Bluebolt Software là một công ty công nghệ tiên phong, chuyên cung cấp các giải pháp công nghệ toàn diện và tư vấn chiến lược chuyển đổi số cho doanh nghiệp. Công ty tập trung vào việc tùy chỉnh và cung cấp các giải pháp công nghệ, cũng như nhân sự IT phù hợp với nhu cầu riêng biệt của từng khách hàng.

Các lĩnh vực hoạt động chính bao gồm:

- Phần mềm Công nghệ Thông tin.
- Cho thuê nhân sự IT.
- Tư vấn, triển khai chuyển đổi số.
- Tư vấn, xây dựng kế hoạch kinh doanh công nghệ.



*Hình 1 Logo Công ty*

### 2.2. Cơ cấu tổ chức

Cơ cấu tổ chức của Bluebolt Software được xây dựng theo mô hình tinh gọn, với sự quản lý của CEO anh Đạt, điều hành trực tiếp các phòng ban. Tôi là thành viên của Phòng AI, nơi chịu trách nhiệm phát triển và triển khai các dự án liên quan đến Trí tuệ nhân tạo (AI) và Khoa học Dữ liệu (Data Science)

### **2.3. Văn hóa làm việc ở công ty**

Trong quá trình thực tập, em đã học hỏi và rèn luyện được tác phong làm việc chuyên nghiệp, tinh thần kỷ luật và đạo đức nghề nghiệp. Công ty tạo điều kiện để sinh viên có cơ hội học hỏi từ những chuyên gia giàu kinh nghiệm trong lĩnh vực công nghệ thông tin, đồng thời tham gia các hoạt động ngoại khóa góp phần xây dựng môi trường làm việc năng động và gắn kết.

#### **2.3.1. Môi trường làm việc**

Môi trường làm việc tại Công ty Công nghệ Tia Chớp Xanh là sự kết hợp giữa tính linh hoạt cao của mô hình startup và tinh thần kỷ luật trong nghiệp vụ kỹ thuật.

- Tính kỷ luật cao trong chuyên môn: Team AI Engineer đòi hỏi sự chính xác và tuân thủ chặt chẽ các quy trình kỹ thuật, đặc biệt là trong việc xử lý dữ liệu lớn (Kafka pipeline, thiết kế Star Schema). Việc tuân thủ đạo đức nghề nghiệp và tác phong làm việc chuyên nghiệp được đề cao để đảm bảo chất lượng sản phẩm.
- Không gian mở và linh hoạt: Công ty khuyến khích sự chủ động và sáng tạo trong việc đề xuất giải pháp (ví dụ: việc lựa chọn FastAPI, Docker/Nginx do team kỹ thuật tự quyết định). Không gian làm việc thoải mái, tạo điều kiện cho việc trao đổi kiến thức giữa các thành viên.

#### **2.3.2. Mức độ hỗ trợ/hướng dẫn từ đồng nghiệp và quản lý**

Em nhận được sự hỗ trợ rất lớn từ anh Nguyễn Đình Thanh – người hướng dẫn trực tiếp, và các chuyên gia giàu kinh nghiệm khác trong công ty.

- Mentoring Style: Hướng dẫn tại công ty mang tính đồng hành và thực chiến (hands-on). Anh Thanh không chỉ giao việc mà còn phân tích các thách thức kỹ thuật (ví dụ: cách tối ưu truy vấn bằng Partitioning, hay chiến lược sử dụng Kafka) và hướng dẫn em cách tìm kiếm giải pháp tối ưu nhất, thay vì chỉ cung cấp đáp án.
- Hỗ trợ đa chiều: Em có cơ hội học hỏi từ các chuyên gia giàu kinh nghiệm thuộc nhiều mảng (DevOps, Data), giúp em hiểu rõ hơn về cách triển khai hệ thống lên Production (Tuần 11)

#### **2.3.3. Tần suất họp/review công việc**

Công ty áp dụng quy trình làm việc nhanh chóng, tập trung vào kết quả và tiến độ rõ ràng:

- Daily Stand-up Meeting: Team AI Engineer thường tổ chức họp ngắn (10-15 phút) vào đầu mỗi ngày để cập nhật tiến độ công việc trong 24 giờ qua (đã làm gì, dự định làm gì, gặp khó khăn gì), đảm bảo mọi người đều nắm được tình hình chung.
- Weekly Review/Planning: Cuối tuần, team tổ chức họp dài hơn để review kết quả đã đạt được trong tuần (dựa trên nhật ký công việc từ Tuần 1 đến Tuần 12) và lập kế hoạch chi tiết cho tuần tiếp theo, bao gồm việc phân tích yêu cầu từ team sản phẩm (Tuần 5, 8)

## **CHƯƠNG 3. VAI TRÒ KHI THỰC TẬP**

### **3.1. Vai trò, chức danh khi thực tập của bạn Chức danh**

Thực tập sinh team AI Engineer

### **3.2. Mô tả về phòng ban/ nhóm mà bạn được phân công**

Em được phân công vào team AI Engineer. Team này tập trung vào việc xây dựng các trình thu thập dữ liệu (crawler) từ các nền tảng mạng xã hội. Công việc của team bao gồm thiết kế, phát triển, tối ưu và triển khai hệ thống backend phục vụ cho việc phân tích cảm xúc bài viết, bình luận và tạo báo cáo dữ liệu

### **3.3. Trách nhiệm được công ty giao**

Các công việc chính được công ty giao bao gồm:

- Xây dựng và vận hành các trình thu thập dữ liệu (crawler) từ nhiều nguồn khác nhau như Facebook, TikTok, Youtube, Website, v.v..
- Tham gia phát triển, tối ưu và mở rộng hệ thống thu thập – phân tích dữ liệu thời gian thực.
- Đảm nhận vai trò chính trong việc thiết kế, phát triển, tối ưu và triển khai hệ thống backend phục vụ phân tích cảm xúc và báo cáo dữ liệu.

### **3.4. Trách nhiệm được công ty giao**

Trách nhiệm của em được nhìn nhận là cầu nối quan trọng giữa luồng dữ liệu thô (từ crawler) và giao diện người dùng cuối thông qua việc xây dựng hệ thống Backend API. Vai trò này yêu cầu sinh viên phải có kiến thức về Data Architecture, Data Driven, và hiểu biết về Computer Network, DevOps để đảm bảo hệ thống hoạt động ổn định và có khả năng mở rộng

## CHƯƠNG 4. PROJECTS & TASKS

**Project: Hệ thống Thu thập và Xử lý Dữ liệu Lớn cho nền tảng Mention.vn**

### 4.1 Mô tả project

Project này nhằm xây dựng một hệ thống hoàn chỉnh có khả năng thu thập, xử lý và phân tích dữ liệu theo thời gian thực từ các nền tảng mạng xã hội (ban đầu là Facebook, sau đó mở rộng sang Twitter/X) dựa trên các từ khóa do người dùng nhập vào. Hệ thống này hỗ trợ việc phân tích cảm xúc, xu hướng và cung cấp báo cáo dữ liệu cho người dùng cuối.

### 4.2. Những nhiệm vụ cụ thể được giao

Nhiệm vụ được chia thành 4 giai đoạn chính:

Giai đoạn	Nhiệm vụ cụ thể
Giai đoạn 1: Khởi động và Thu thập Dữ liệu Ban đầu	Tìm hiểu nghiệp vụ, nghiên cứu phương pháp thu thập dữ liệu Facebook (Selenium, Scrapy, BeautifulSoup). Triển khai crawler thu thập dữ liệu Facebook theo từ khóa, xử lý kỹ thuật phân trang (infinite scroll) và kiểm soát trùng lặp dữ liệu (Post ID).
Giai đoạn 2: Tích hợp Xử lý Dữ liệu Realtime	Tích hợp service Kafka vào hệ thống crawler để nhận keyword realtime từ người dùng. Thiết kế luồng xử lý (Producer/Consumer) với 2 topic (keywords-topic và facebook-posts-topic). Xử lý dữ liệu thô và lưu trữ vào MongoDB.
Giai đoạn 3: Phát triển và Tối ưu Hệ thống Backend	Phân tích, thiết kế Star Schema cho cơ sở dữ liệu phân tích (PostgreSQL). Xây dựng và tái cấu trúc hệ thống Backend API bằng FastAPI. Tối ưu hiệu suất bằng cách áp dụng Indexing, Partitioning, sử dụng Redis cho caching layer và background tasks.
Giai đoạn 4: Hoàn thiện và Triển khai Production	Hoàn thiện các API báo cáo nâng cao (Sentiment, Top Authors, Source Distribution). Đóng gói toàn bộ hệ thống Backend trên môi trường Docker. Triển khai hệ thống lên môi trường Production thông qua Nginx làm reverse proxy và SSL termination trên server Linux.

*Bảng 1. Trình bày các nhiệm vụ theo giai đoạn*

#### **4.3. Công cụ, phương pháp hoặc công nghệ được sử dụng**

- Ngôn ngữ/Framework: Python, FastAPI, SQLAlchemy.
- Thu thập dữ liệu: Selenium, BeautifulSoup, Scrapy.
- Kiến trúc Dữ liệu Lớn: Kafka (sử dụng 2 topic: keywords-topic và facebook-posts-topic), Star Schema (PostgreSQL).
- Hệ thống Dữ liệu: PostgreSQL (cho dữ liệu phân tích), MongoDB (cho dữ liệu thô), Structured Query Language (SQL).
- Tối ưu Hiệu suất DB: Indexing, Partitioning cho bảng fact, Materialized View cho thống kê nặng.
- Hệ thống/DevOps: Docker, Docker Compose, Nginx (Reverse Proxy, SSL Termination), Linux server (Ubuntu), Redis (caching layer, background tasks).

#### **4.4. Nhận xét cá nhân của bạn đối với mỗi project mà bạn tham gia**

Dự án Mention.vn là một trải nghiệm toàn diện, giúp em áp dụng kiến thức từ Data Architecture, Data Engineering đến Backend Development. Việc tham gia vào toàn bộ chu trình phát triển (từ thiết kế DB, tích hợp Kafka, xây dựng API hiệu suất cao, đến triển khai Production bằng Docker/Nginx) đã cung cấp kinh nghiệm thực tế về cách xây dựng một hệ thống dữ liệu có tính chịu tải, ổn định và khả năng mở rộng cao.

## CHƯƠNG 5. NHẬT KÝ CÔNG VIỆC

### 5.1. Nhật ký công việc theo tuần

Thực hiện 12 tuần: Từ ngày 01/06/2025

Tuần	Thời gian	Nội dung công việc thực hiện
Tuần 1	01/06/2025– 07/06/2025	- Tìm hiểu nghiệp vụ và các yêu cầu mà công ty đặt ra cho sinh viên thực tập. - Tham gia đào tạo, hội nhập công ty.
Tuần 2	08/06/2025 – 14/06/2025	- Nghiên cứu phương pháp thu thập dữ liệu cho nền tảng Facebook.
Tuần 3	15/06/2025 – 21/06/2025	- Triển khai crawler thu thập dữ liệu Facebook theo từ khóa do người dùng nhập.
Tuần 4	22/06/2025 – 28/06/2025	- Tích hợp service Kafka vào crawler Facebook để nhận keyword realtime từ người dùng.
Tuần 5	29/06/2025 – 05/07/2025	- Phân tích và thiết kế cơ sở dữ liệu cho hệ thống.
Tuần 6	06/07/2025 – 12/07/2025	- Xây dựng hệ thống Backend API cho dự án Mention.vn.
Tuần 7	13/07/2025 – 19/07/2025	- Tái cấu trúc backend để cải thiện hiệu suất và khả năng chịu tải.
Tuần 8	20/07/2025 – 26/07/2025	- Thiết kế lại cơ sở dữ liệu để hỗ trợ khả năng mở rộng dài hạn.
Tuần 9	27/07/2025 – 02/08/2025	- Cập nhật và bổ sung các tính năng qua API backend.
Tuần 10	03/08/2025 – 09/08/2025	- Đóng gói hệ thống backend trên môi trường Docker.
Tuần 11	10/08/2025 – 16/08/2025	- Triển khai hệ thống lên môi trường production thông qua Nginx trên server Linux.
Tuần 12	17/08/2025 – 23/08/2025	- Viết báo cáo thực tập và trình bày kết quả với giảng viên hướng dẫn.

*Bảng 2. Trình bày công việc theo tuần*



## 5.2. Các mốc thời gian quan trọng đạt được

- Hoàn thành Crawler Facebook cơ bản và cơ chế chống trùng lặp dữ liệu (Tuần 3).
- Tích hợp thành công Kafka Realtime Pipeline cho luồng dữ liệu (Tuần 4).
- Hoàn thiện thiết kế mô hình Star Schema cho cơ sở dữ liệu phân tích (Tuần 5).
- Hoàn thành Backend API nền tảng bằng FastAPI và Docker hóa (Tuần 6).
- Tối ưu hiệu suất backend bằng Redis Caching và Partitioning DB (Tuần 7).
- Triển khai hệ thống thành công lên môi trường Production sử dụng Docker và Nginx (Tuần 11)

## 5.3. Những thách thức gặp phải, và cách đương đầu

Thách thức gặp phải	Cách đương đầu và giải pháp kỹ thuật
Xử lý trang web động và cơ chế chống crawl phức tạp của Facebook.	Sử dụng Selenium để mô phỏng hành vi người dùng (cuộn trang, đăng nhập) và đặt delay hợp lý (5-10s) để tránh bị block.
Đảm bảo hiệu suất API khi truy vấn dữ liệu lớn từ các bảng fact/dimension.	Áp dụng mô hình Star Schema, sử dụng Partitioning và Indexing trong PostgreSQL. Đồng thời, triển khai lớp Caching bằng Redis cho các truy vấn thống kê nặng.
Đảm bảo tính ổn định và khả năng mở rộng (scalability) của hệ thống.	Sử dụng Kafka để tách biệt luồng điều khiển và luồng dữ liệu. Đóng gói dịch vụ bằng Docker và sử dụng Nginx để phân phối tải (load balancing) và chạy multi-worker.

*Bảng 3. Trình bày thách thức gặp phải*

## CHƯƠNG 6. KẾT LUẬN

### 6.1. Tóm tắt quá trình thực tập

Quá trình thực tập kéo dài 12 tuần tại Công ty Công nghệ Tia chớp xanh đã cung cấp cho em trải nghiệm quý giá, giúp em tham gia phát triển một hệ thống dữ liệu hoàn chỉnh, từ việc xây dựng công cụ thu thập dữ liệu (crawler), tích hợp luồng xử lý thời gian thực (Kafka), thiết kế cơ sở dữ liệu phân tích (Star Schema/PostgreSQL), phát triển Backend API hiệu suất cao (FastAPI), cho đến triển khai và vận hành hệ thống trên môi trường Production sử dụng Docker và Nginx

### 6.2. Tự đánh giá

#### 6.2.1. Ưu điểm, khuyết điểm

##### - Ưu điểm:

1. Kiến thức chuyên môn vững chắc: Nắm vững kiến thức nền tảng về Data Engineer, bao gồm các nguyên tắc hoạt động, hệ thống, kỹ thuật xử lý và lưu trữ dữ liệu lớn.
2. Kỹ năng lập trình và hệ thống: Có kỹ năng lập trình tốt (Python, FastAPI) và thành thạo kỹ năng hệ thống thực tế như triển khai trên Linux server, cấu hình Nginx, và quản lý service bằng Docker Compose.
3. Kỹ năng mềm: Có khả năng giao tiếp và hợp tác hiệu quả với các thành viên trong nhóm.

##### - Khuyết điểm:

1. Kinh nghiệm thực tế còn hạn chế và kiến thức còn đang trong quá trình rèn luyện.
2. Cần nâng cao kinh nghiệm trong việc quản lý quy mô dữ liệu cực lớn và tối ưu hóa các mô hình phân tích sâu hơn.

#### 6.2.2. Bài học (Học được gì: hard skill, soft skill)

##### Hard Skill (Kỹ năng chuyên môn):

1. Kỹ năng Dữ liệu: Thiết kế và tối ưu hóa cơ sở dữ liệu (sử dụng chỉ mục, partition, materialized view) để tăng tốc truy vấn.
2. Kỹ năng Lập trình: Phát triển crawler và xây dựng API backend hiệu suất cao với FastAPI.
3. Kỹ năng Hệ thống/DevOps: Tạo và quản lý container bằng Docker, triển khai cụm dịch vụ (Kafka, MongoDB, PostgreSQL, Redis) trong môi trường Docker Compose.

Có kiến thức thực tế về triển khai hệ thống trên Linux server và cấu hình Nginx reverse proxy.

#### **Soft Skill (Kỹ năng mềm):**

1. Kỹ năng làm việc nhóm: Có khả năng hợp tác hiệu quả để hoàn thành mục tiêu chung.
2. Kỹ năng giải quyết vấn đề: Có khả năng phân tích và giải quyết các vấn đề phát sinh trong quá trình thực tập.
3. Kỹ năng học tập và thích nghi: Có khả năng tự học hỏi và cập nhật kiến thức mới liên tục, thích nghi với môi trường làm việc mới

### **6.3. Sự ảnh hưởng**

Kỳ thực tập tại Bluebolt Software không chỉ là một trải nghiệm làm việc mà còn là một bước ngoặt quan trọng, củng cố mạnh mẽ định hướng nghề nghiệp của em trở thành một Kỹ sư Dữ liệu (Data Engineer). Việc trực tiếp tham gia vào quy trình phát triển một hệ thống dữ liệu hoàn chỉnh đã giúp em nhận ra rằng, vai trò của một Kỹ sư Dữ liệu không chỉ dừng lại ở việc xây dựng các đường ống dữ liệu (data pipeline) mà còn là người đặt nền móng kiến trúc cho toàn bộ hệ sinh thái dữ liệu của một tổ chức.

Từ đó, em đã chủ động tìm hiểu sâu hơn về bối cảnh của ngành, những yêu cầu của thị trường lao động và định hình một lộ trình phát triển rõ ràng cho bản thân.

#### **• Bối cảnh Thị trường và Sự Chuyển dịch của Ngành**

Qua quan sát, em nhận thấy nhu cầu tuyển dụng Kỹ sư Dữ liệu tại thị trường Việt Nam đang tăng trưởng rất nóng, đặc biệt là ở các công ty công nghệ, tài chính, ngân hàng và thương mại điện tử. Tuy nhiên, đây không phải là một xu hướng riêng lẻ mà là một phần trong làn sóng chuyển đổi số toàn cầu. Doanh nghiệp ngày nay không chỉ coi dữ liệu là tài sản mà còn xem khả năng khai thác dữ liệu là lợi thế cạnh tranh cốt lõi.

Điều này dẫn đến sự thay đổi lớn trong vai trò của Kỹ sư Dữ liệu:

- **Từ người xây dựng trở thành kiến trúc sư:** Vai trò này đang dịch chuyển từ việc thuần túy viết code cho các tác vụ ETL (Extract - Transform - Load) sang việc thiết kế các hệ thống có khả năng mở rộng, chịu lỗi cao và tối ưu về chi phí. Tư duy kiến

trúc hệ thống, khả năng đánh đổi giữa xử lý theo lô (batch) và thời gian thực (real-time) trở thành kỹ năng được đánh giá cao nhất.

- **Sự hội tụ của Dữ liệu và Trí tuệ Nhân tạo (AI/ML):** Ranh giới giữa Kỹ sư Dữ liệu và Kỹ sư Học máy (Machine Learning Engineer) đang mờ dần. Kỹ sư Dữ liệu hiện đại phải là người xây dựng nền tảng vững chắc cho MLOps, đảm bảo dữ liệu luôn sẵn sàng, sạch sẽ và có thể truy xuất hiệu quả để huấn luyện và triển khai các mô hình AI vào vận hành thực tế.
- **Lấy đám mây làm trung tâm (Cloud-centric):** Các hệ thống dữ liệu tại chỗ (on-premise) đang dần được thay thế bởi các nền tảng đám mây như AWS, Google Cloud (GCP), Azure. Việc thành thạo các dịch vụ đám mây chuyên dụng (như S3, Redshift, BigQuery, Kafka/Kinesis) đã trở thành yêu cầu bắt buộc.
- **Những Kỹ năng Thiết yếu cho Tương lai**

Dựa trên những phân tích về sự chuyển dịch của thị trường, em nhận định rằng để trở thành một Kỹ sư Dữ liệu có năng lực cạnh tranh trong 5-10 năm tới, việc nắm vững các kỹ năng nền tảng như Python, SQL là chưa đủ. Lộ trình phát triển của em cần tập trung vào các nhóm kỹ năng chiến lược sau:

1. **Chuyên sâu về Công nghệ Dữ liệu lớn & Xử lý Thời gian thực:** Nắm vững các framework như **Apache Spark** để xử lý dữ liệu phân tán quy mô lớn và các công cụ như **Apache Kafka**, **Apache Flink** để xây dựng các hệ thống có khả năng phản ứng ngay lập tức với luồng dữ liệu mới.
2. **Kiến trúc Dữ liệu trên Đám mây:** Thành thạo việc thiết kế và triển khai giải pháp trên ít nhất một nền tảng đám mây lớn (ví dụ: AWS). Mục tiêu không chỉ là sử dụng dịch vụ mà là kết hợp chúng để tạo ra các kiến trúc tối ưu (ví dụ: xây dựng Data Lake, Data Warehouse, Lakehouse).
3. **Tự động hóa và Vận hành Dữ liệu (DataOps):** Áp dụng các nguyên tắc của DevOps vào quy trình làm việc với dữ liệu. Sử dụng các công cụ điều phối tác vụ như **Apache Airflow**, các công cụ quản lý hạ tầng bằng code (Infrastructure as Code) như **Terraform**, **Docker** để tự động hóa hoàn toàn việc triển khai và giám sát hệ thống dữ liệu.

#### **4. Quản trị và Chất lượng Dữ liệu (Data Governance & Quality):**

Hiểu và triển khai các quy trình để đảm bảo tính toàn vẹn, bảo mật và chất lượng của dữ liệu, một yếu tố ngày càng trở nên quan trọng khi các quy định về dữ liệu ngày càng chặt chẽ.

Kỳ thực tập này đã tạo một bước đệm vững chắc, giúp em tự tin bước tiếp trên con đường đã chọn. Em hiểu rằng công nghệ luôn thay đổi, và tinh thần tự học hỏi không ngừng chính là chìa khóa quan trọng nhất. Mục tiêu của em là trở thành một Kỹ sư Dữ liệu có khả năng không chỉ giải quyết các bài toán kỹ thuật phức tạp mà còn có thể đưa ra những giải pháp kiến trúc mang lại giá trị chiến lược cho doanh nghiệp trong kỷ nguyên của dữ liệu và AI.

## CHƯƠNG 7. THAM KHẢO

1. Docker, Inc. (không có năm). *“Docker Documentation”*. docs.docker.com.
2. Richardson, Leonard (không có năm). *“Beautiful Soup Documentation”*. crummy.com.
3. SeleniumHQ (không có năm). *“Selenium Documentation”*. selenium.dev.
4. Ramírez, Sebastián (không có năm). *“FastAPI Documentation – Fast, High-Performance Web Framework for Building APIs”*. fastapi.tiangolo.com.
5. Patairya, Dilip Kumar & Whitfield, Brennan (2024). *“Star Schema vs. Snowflake Schema Explained”*. builtin.com. [builtin.com](https://builtin.com).
6. Tom (2024). *“Optimizing FastAPI for High Performance”*. medium.com.
7. tiangolo (không có năm). *“uvicorn-gunicorn-fastapi-docker”*. github.com.
8. Sematext (không có năm). *“PostgreSQL Performance Tuning: Partitioning and Indexing”*. sematext.com.
9. Bhuyan, Aditya Pratap (2024). *“Best Practices for PostgreSQL Database Design to Ensure Performance, Scalability, and Efficiency”*. dev.to [dev.to](https://dev.to).
10. Manoharan, Sivakumar (2024). *“Caching in FastAPI: Unlocking High-Performance Development”*. dev.to. [dev.to](https://dev.to).
11. F5 NGINX (không có năm). *“NGINX Reverse Proxy”*. docs.nginx.com. [docs.nginx.com](https://docs.nginx.com/docs.nginx.com).
12. Vivesh (2024). *“An In-Depth Look at Nginx: The Modern Web Server and Its Applications in Web and Application Backend Infrastructure”*. dev.to. [dev.to](https://dev.to).
13. Katyal, Anchal (2024). *“Optimizing Nginx for High-Traffic Systems: A Guide to Configuration & Monitoring”*. hackernoon.com. [hackernoon.com](https://hackernoon.com).