

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG
KHOA CÔNG NGHỆ THÔNG TIN 1



BÁO CÁO
MÔN HỌC: THỰC TẬP TỐT NGHIỆP
Doanh nghiệp: Công ty cổ phần VCCORP
Giảng viên hướng dẫn: Đỗ Thị Liên
Nhóm: 25

Họ tên: Hoàng Văn An

Mã sinh viên: B20DCCN045

Số điện thoại: 0867865001

Email: vananhoang10052002@gmail.com

Hà Nội – 2024

LỜI CẢM ƠN

Em xin gửi lời cảm ơn chân thành nhất đến Công ty Cổ phần VCCORP đã tạo điều kiện cho em có cơ hội được thực tập tại một môi trường chuyên nghiệp và năng động. Trong suốt quá trình 7 tuần thực tập, em đã học hỏi được rất nhiều từ các anh chị và đồng nghiệp tại công ty, không chỉ về kiến thức chuyên môn mà còn về cách làm việc hiệu quả, sáng tạo và trách nhiệm.

Đặc biệt, em muốn gửi lời cảm ơn sâu sắc đến anh Ngô Văn Vĩ và các anh chị trong team của anh đã luôn sẵn lòng hỗ trợ, hướng dẫn tận tình, giúp em vượt qua những khó khăn và hoàn thành tốt nhiệm vụ được giao. Sự chỉ dẫn và góp ý của các anh chị không chỉ giúp em nâng cao kỹ năng mà còn mở ra cho em những góc nhìn mới mẻ trong công việc và cuộc sống.

Cuối cùng, em cũng xin cảm ơn nhà trường và cô Đỗ Thị Liên đã luôn ủng hộ và động viên em trong suốt quá trình học tập và thực tập. Đây là một trải nghiệm quý báu, giúp em tích lũy được nhiều kiến thức và kinh nghiệm, chuẩn bị cho hành trình sắp tới trong sự nghiệp của mình.

Hà Nội, ngày 10 tháng 08 năm 2024

Sinh viên

**An
Hoàng Văn An**

This image shows a full page of a document template designed for handwriting practice. It features approximately 30 evenly spaced, horizontal grey lines across the entire width of the page. The lines are thin and light, providing a guide for letter height without being distracting. There are no margins, text, or other markings present on the page.

[illegible]

MỤC LỤC

LỜI CẢM ƠN	1
DANH MỤC THUẬT NGỮ VÀ TỪ VIẾT TẮT	5
I. GIỚI THIỆU CHUNG ĐƠN VỊ THỰC TẬP	6
1.1. Thông tin về đơn vị thực tập:	6
1.2. Cơ sở thực tập:	6
II. NỘI DUNG CÔNG VIỆC TRONG KÌ THỰC TẬP:	8
Nội dung thực tập trong 7 tuần (trao đổi với leader):	8
2.1. NỘI DUNG CÔNG VIỆC TUẦN 1 (24/6 – 1/7):	8
2.2. NỘI DUNG CÔNG VIỆC TUẦN 2 (1/7 – 7/7)	10
2.3. NỘI DUNG CÔNG VIỆC TUẦN 3 - 4 (7/7 – 21/7):	12
2.4. NỘI DUNG CÔNG VIỆC TUẦN 5 (21/7 – 28/7):	13
2.5. NỘI DUNG CÔNG VIỆC TUẦN 6 – 7 (28/7 – 11/8):	15
III. KẾT QUẢ THỰC TẬP:	16
3.1. Những điều đạt được trong kì thực tập:	16
3.2. Những điều chưa đạt được:	17
3.3. Bài học kinh nghiệm:	17
3.4. Đánh giá của người hướng dẫn(Mentor):	18
3.5. Tự đánh giá:	18
IV. TỔNG KẾT:	18
TÀI LIỆU THAM KHẢO.....	20

DANH MỤC THUẬT NGỮ VÀ TỪ VIẾT TẮT

Thuật ngữ	Ý nghĩa
API	Application Programming Interface : Giao diện lập trình ứng dụng
Aerospike	Aerospike là một hệ thống cơ sở dữ liệu NoSQL hiệu năng cao được thiết kế để xử lý khối lượng dữ liệu lớn và cung cấp tốc độ truy xuất nhanh, hỗ trợ lưu trữ vĩnh viễn và khả năng mở rộng cao
Hbase	HBase là một cơ sở dữ liệu phân tán, mã nguồn mở, được thiết kế để cung cấp khả năng lưu trữ và truy xuất dữ liệu có cấu trúc lớn với hiệu suất cao, chạy trên hệ thống file HDFS của Hadoop.
Kafka	Kafka là một nền tảng streaming phân tán mạnh mẽ, được thiết kế để xử lý và truyền tải lượng lớn dữ liệu theo thời gian thực với độ tin cậy cao và khả năng mở rộng linh hoạt.
Spark	Spark là một framework xử lý dữ liệu phân tán mạnh mẽ, hỗ trợ xử lý dữ liệu lớn với hiệu suất cao thông qua các API để sử dụng cho SQL, streaming, machine learning, và graph processing.
Docker	Docker là một nền tảng mã nguồn mở giúp tự động hóa việc triển khai ứng dụng bên trong các container phần mềm, đảm bảo tính nhất quán, di động và hiệu quả trong việc phát triển và vận hành phần mềm.
MongoDB	MongoDB là một cơ sở dữ liệu NoSQL dạng tài liệu, cung cấp khả năng lưu trữ và truy vấn dữ liệu linh hoạt và mở rộng dễ dàng.
Design pattern	Design pattern là các giải pháp thiết kế tái sử dụng để giải quyết các vấn đề phổ biến trong phát triển phần mềm.
Hadoop	Hadoop là một khung phần mềm mã nguồn mở cho phép xử lý và lưu trữ khối lượng lớn dữ liệu phân tán trên các cụm máy tính sử dụng mô hình lập trình đơn giản.
Zookeeper	Zookeeper là một dịch vụ đồng bộ hóa phân tán cung cấp các công cụ để quản lý cấu hình, phối hợp dịch vụ và giữ cho các ứng dụng phân tán đồng bộ.
Kraft	Kraft là một thư viện mã nguồn mở giúp triển khai và quản lý các dịch vụ kết nối với Kafka trong môi trường Kubernetes.
RabbitMQ	RabbitMQ là một hệ thống quản lý hàng đợi tin nhắn mã nguồn mở, hỗ trợ giao tiếp giữa các ứng dụng và dịch vụ thông qua các hàng đợi tin nhắn.

I. GIỚI THIỆU CHUNG ĐƠN VỊ THỰC TẬP

Phần này cung cấp một cái nhìn tổng quan về đơn vị thực tập, bao gồm lịch sử hình thành và phát triển, tầm nhìn và sứ mệnh của công ty. Chương này cũng sẽ giới thiệu các sản phẩm và dịch vụ chính mà công ty cung cấp, nhấn mạnh vào những giá trị cốt lõi và cam kết của công ty trong việc mang lại các giải pháp công nghệ tiên tiến và hiệu quả cho khách hàng.

1.1. Thông tin về đơn vị thực tập:

1.1.1. Giới thiệu về VCCorp:

- Được thành lập vào năm 2006, sau 18 năm phát triển, Công ty CP VCCorp (VCCorp) là một trong những công ty tiên phong trong lĩnh vực công nghệ và công nghệ cao ở Việt Nam. VCCorp đã xây dựng và ứng dụng thành công hạ tầng cloud computing có khả năng lưu trữ và tính toán lượng dữ liệu lớn (Big Data, Data mining) hỗ trợ cho nền tảng công nghệ quảng cáo trực tuyến Admicro, phục vụ trên 50 triệu người dùng (chiếm trên 90% tổng số người dùng Internet tại Việt Nam) với hơn 200 website uy tín, gần 30 website trong số đó thuộc sở hữu độc quyền, phục vụ hàng tỷ lượt xem mỗi tháng. Ngoài ra, VCCorp nằm trong top 3 công ty phát hành game trên nền tảng di động với thương hiệu Sohagame, và mới đây hệ sinh thái chuyển đổi số Bizfly đang được VCCorp phát triển và triển khai mạnh mẽ. Đồng thời VCCorp cũng là đơn vị xây dựng Lotus - mạng xã hội do người Việt Nam làm chủ.



Hình 1.1. Tổng quan về các sản phẩm của VCCorp

1.1.2. Lịch sử phát triển:

-

1.2. Cơ sở thực tập:

1.2.1. Thông tin thực tập:

- Tên doanh nghiệp: Công ty cổ phần VCCORP
- Vị trí thực tập: Java Backend Developer
- Khối được phân: Khối Adtech của VCCORP
- Leader, mentor hướng dẫn: Leader Ngô Văn Vĩ
- Hình thức thực tập: Tại văn phòng công ty
- Mô tả công việc của team:
 - Xử lý các phần backend Server của các hệ thống phân tán, các hệ thống dữ liệu, data minning.
 - Tối ưu, xử lý dữ liệu.

1.2.2. Đội ngũ hướng dẫn thực tập sinh:

- Leader Ngô Văn Vĩ:
 - Chuyên gia với kiến thức chuyên sâu về các công nghệ mới nhất và xu hướng phát triển trong ngành.

1.2.3. Cơ sở vật chất:

- VCCorp có văn phòng được thiết kế theo tiêu chuẩn công nghệ cao, với không gian mở và các khu vực làm việc riêng biệt. Thiết kế văn phòng tối ưu hóa sự linh hoạt và sự giao tiếp giữa các bộ phận..
- Công nghệ, thiết bị, cập nhật liên tục, đảm bảo nhân viên luôn tiếp cận với những công nghệ mới nhất.
- Cơ sở vật chất của VCCorp được đảm bảo về mặt an ninh và an toàn, với hệ thống bảo vệ chuyên nghiệp và các biện pháp phòng cháy chữa cháy đầy đủ.

1.2.4. Dịch vụ hỗ trợ thực tập sinh:

- Đào tạo và hướng dẫn:
 - Cung cấp chương trình đào tạo ban đầu để giúp thực tập sinh làm quen với công việc và các công cụ cần thiết.
 - Được hướng dẫn bởi các chuyên gia có kinh nghiệm trong ngành, giúp thực tập sinh nâng cao kỹ năng chuyên môn.
- Môi trường làm việc chuyên nghiệp:
 - Cung cấp môi trường làm việc hiện đại, tiện nghi với đầy đủ trang thiết bị cần thiết.
 - Không gian làm việc thân thiện, khuyến khích sự sáng tạo và hợp tác giữa các nhân viên.
- Phản hồi và đánh giá:
 - Thường xuyên cung cấp phản hồi và đánh giá hiệu quả làm việc, giúp thực tập sinh nhận biết được điểm mạnh và điểm cần cải thiện.
 - Hỗ trợ phát triển cá nhân thông qua các buổi mentoring và coaching.
- Cơ hội phát triển nghề nghiệp:
 - Tạo điều kiện để thực tập sinh tham gia vào các dự án thực tế, tích lũy kinh nghiệm quý báu.
 - Cơ hội trở thành nhân viên chính thức của VCCorp sau khi hoàn thành chương trình thực tập xuất sắc.

1.2.5. Thời gian thực tập:

- Thực tập từ 26/6/2024 đến 11/8/2024
- Lên công ty theo lịch đã đăng kí, buổi sáng từ 9h đến 12h, buổi chiều từ 13h30 đến 18h

II. NỘI DUNG CÔNG VIỆC TRONG KÌ THỰC TẬP:

Nội dung thực tập trong 7 tuần (trao đổi với leader):

- Tuần 1 (24/6 – 30/6): Tìm hiểu, làm bài tập và giải đáp thắc mắc về java basic, java core.
- Tuần 2 (30/6 – 6/7): Tìm hiểu, làm bài tập và giải đáp thắc mắc về các công nghệ được áp dụng trong dự án thực tế của công ty.
- Tuần 3 - 4(6/7 – 12/7): Tìm hiểu, làm bài tập và giải đáp thắc mắc về các công nghệ được áp dụng trong dự án thực tế của công ty.
- Tuần 5 – Tuần 7(12/7 – 11/8): Leader sẽ giao một dự án nhỏ cho mỗi thành viên trong nhóm thực tập, với yêu cầu chi tiết và tư vấn hướng dẫn. Nếu đạt yêu cầu và đánh giá của leader thì sẽ được giao vào một dự án thực tế.

Tuần	STT	Họ Và Tên	Team	Thứ 2	Thứ 3	Thứ 4	Thứ 5	Thứ 6	Sáng Thứ 7	Note	Mãng dạn
tuần 1	1	Nguyễn Quang Hạnh					Full	Full	Sáng		
	2	Hoàng Văn An			Chiều	Full	Full	Full			
	3	Vũ Đức Chính				Full	Full	Sáng			
	4	Nguyễn Đức Trường Giang			Sáng	Full	Full	Full			
	5	Đào Văn Lâm				Full	Full	Full			
	6	Vũ Thị Thanh Ngân				Full	Full	Full			
	7	Khúc Văn Vinh				Full	Full	Full	Sáng		
	8	Lê Minh Sơn				Sáng	Sáng	Sáng			
	9										
	10										
	11										
	12										

2.1. NỘI DUNG CÔNG VIỆC TUẦN 1 (24/6 – 1/7):

2.1.1. Tìm hiểu, làm bài tập và giải đáp các thắc mắc về java basic:

- Task 1: Tìm hiểu và viết báo cáo về các thuật toán tìm kiếm và sắp xếp, độ phức tạp thuật toán, nghiên cứu làm bài tập với độ phức tạp thuật toán nhỏ nhất.
- Task 2: Tìm hiểu và viết báo cáo về http và url trong phương thức HTTP, Tìm hiểu về xử lý chuỗi trong java, cách sử dụng regex để cho phép bạn tìm kiếm, so khớp, và thao tác với các mẫu chuỗi văn bản cụ thể, nghiên cứu làm bài tập về validate url hợp lệ trong HTTP.
- Task 3: Tìm hiểu và viết báo cáo về Prefix Sum và ứng dụng trong bài toán counting sort

- Task 4: Tìm hiểu và viết báo cáo về bài toán xếp các từ thành câu có nghĩa và phương pháp đệ quy kết hợp với lưu trữ kết quả trung gian (memoization) để tránh tính toán lại các phần đã được xử lý, áp dụng vào bài toán viết một phương thức add các khoảng trắng vào chuỗi s sao cho thành các câu có thể, với một từ điền các từ đã được cho sẵn
- Task 5: Tìm hiểu và viết báo cáo về các design pattern trong java. Mô tả và implement lại chúng (phân tích so sánh dựa trên 1 case bài toán thực tế)
- Task 6: Tìm hiểu về các quy tắc clean code và viết báo cáo

2.1.2. Kết quả đạt được:

- Hiểu và cài đặt các thuật toán
- Thực hiện các bài tập với độ phức tạp thuật toán nhỏ nhất.
- Hiểu rõ cơ bản về HTTP, các phương thức HTTP (GET, POST, PUT, DELETE), tìm hiểu về cấu trúc URL và cách sử dụng.
- Thực hiện bài tập validate URL hợp lệ trong HTTP.
- Hiểu rõ thuật toán Prefix Sum và cách áp dụng nó, cài đặt và thử nghiệm thuật toán Counting Sort với ứng dụng của Prefix Sum.
- Hiểu về phương pháp đệ quy kết hợp với lưu trữ kết quả trung gian (memoization).
- Tìm hiểu và mô tả các design pattern phổ biến như Singleton, Factory, Adapter, Chain of Responsibility, cài đặt và thử nghiệm các design pattern này dựa trên các case thực tế.
- Hiểu rõ các nguyên tắc cơ bản của Clean Code, thực hiện các bài tập để viết mã sạch và dễ bảo trì.

2.1.3. Bài học kinh nghiệm:

- Hiểu rõ hơn về cách đánh giá độ phức tạp của các thuật toán và áp dụng vào các bài toán thực tế.
- Hiểu rõ hơn về đệ quy và memoization, và cách áp dụng vào các bài toán xử lý chuỗi.
- Nắm bắt được các design pattern và cách áp dụng chúng vào các dự án thực tế.
- Nhận ra tầm quan trọng của Clean Code và cách áp dụng nó để tạo ra mã nguồn dễ hiểu và dễ bảo trì.

2.1.4. Review, minh chứng:



2.2. NỘI DUNG CÔNG VIỆC TUẦN 2 (1/7 – 7/7)

2.2.1. Tìm hiểu về các công nghệ thực tế được sử dụng trong công ty

- **Track 1: Tìm hiểu và viết báo cáo về Aerospike, sau đó thực hiện các task sử dụng công nghệ đã tìm hiểu đó, mentor giải đáp các thắc mắc về công nghệ**
 - Cài đặt Aerospike
 - Ghi 1 bản ghi vào Aerospike (Ghi vĩnh viễn/có expire time)
 - Đọc 1 bản ghi từ Aerospike
 - Đọc nhiều bản ghi từ Aerospike
 - Sử dụng AQL
 - Tìm hiểu về các operation khác của Aerospike
 - Tìm hiểu về EventLoop và Callback trong Aerospike
 - Scan và Query
 - Tìm hiểu index trong Aerospike
 - So sánh giữa Aerospike với Redis
 - Tìm hiểu các chiến lược caching cho cả đọc và ghi
- **Track 2: Tìm hiểu và viết báo cáo về HBase, sau đó thực hiện các task sử dụng công nghệ đã tìm hiểu đó, mentor giải đáp các thắc mắc về công nghệ**
 - Cài đặt HBase (có thể dùng docker)
 - Thao tác bằng HBase shell

- Thao tác bằng Java API
- Get
- Exist bin
- Exist row
- Put
- Delete bin
- Delete Row
- Bulk
- Time To Live
- Scan + Filter
- Pagination using HBase Scan
- Tìm hiểu vai trò của ZooKeeper trong HBase
- Tìm hiểu về Compact trong Hbase
- Tìm hiểu các thành phần cấu hình khi cài đặt Hbase

2.2.2. Những kết quả đạt được:

- Hoàn thành cài đặt Aerospike và HBase.
- Thực hiện thành công các thao tác cơ bản với Aerospike và HBase.
- Viết báo cáo chi tiết về các công nghệ, cách cài đặt, và các thao tác cơ bản.
- So sánh Aerospike và Redis, nắm rõ các ưu nhược điểm của từng công nghệ.
- Áp dụng các chiến lược caching hiệu quả.

2.2.3. Những điều chưa đạt được:

- Chưa tối ưu hóa toàn diện cho các thao tác với dữ liệu lớn trong Aerospike và HBase.
- Chưa triển khai các kịch bản phức tạp sử dụng EventLoop và Callback trong Aerospike.
- Cần tìm hiểu thêm về các mô hình dữ liệu phức tạp và các phương pháp tối ưu hóa hiệu suất trong HBase.

2.2.4. Bài học kinh nghiệm:

- Cần tìm hiểu kỹ lưỡng về cấu trúc và mô hình dữ liệu của từng công nghệ trước khi triển khai.
- Việc sử dụng Docker giúp giảm thiểu thời gian cài đặt và cấu hình môi trường phát triển.
- Cần thường xuyên tham khảo tài liệu và hướng dẫn chính thức của từng công nghệ để nắm bắt các tính năng mới và các phương pháp tối ưu hóa.
- Việc so sánh và lựa chọn công nghệ cần dựa trên yêu cầu cụ thể của dự án và khả năng mở rộng trong tương lai.

2.2.5. Review, minh chứng:

-

2.2.6. Review:



Vũ Anh Đức - DMining AT <ducvuanh@tech.admicro.vn>

đến tôi, Ngo ▾

10:30 Th 7, 3 thg 8 (1 ngày trước)



Dear An,

Anh có bổ sung thêm một số ý trong báo cáo của em, em tìm hiểu thêm bổ sung cho báo cáo nhé:

Chủ đề Aerospike

- Policy trong aerospike: Write policy, bin policy, ignore bin

- Kiểu dữ liệu trong aerospike

- Một số giới hạn của aerospike (giới hạn kích thước dữ liệu, số lượng, độ dài):

+ Namespace

+ Set

+ Bin

+ Cluster name

+ Index

+ Record

- Các khái niệm used ram, high water mark, stop write có ý nghĩa gì. Trường hợp tài nguyên đạt trạng thái high water mark thì việc đọc ghi dữ liệu sẽ bị ảnh hưởng như thế nào?

Thank An!

2.3. NỘI DUNG CÔNG VIỆC TUẦN 3 - 4 (7/7 – 21/7):

2.3.1. Tìm hiểu về các công nghệ thực tế được sử dụng trong công ty

- **Track 3: Tìm hiểu và viết báo cáo về Kafka, sau đó thực hiện các task sử dụng công nghệ đã tìm hiểu đó, mentor giải đáp các thắc mắc về công nghệ.**
 - Cài đặt Kafka
 - Produce message
 - Consume message
 - Consumer group
 - Tìm hiểu về partition và replicate trong Kafka
 - Tìm hiểu vai trò của ZooKeeper trong Kafka (tìm hiểu cả các loại khác thay thế zookeeper)
 - So sánh với rabbitmq

- **Track 4: Tìm hiểu và viết báo cáo về Spark, sau đó thực hiện các task sử dụng công nghệ đã tìm hiểu đó, mentor giải đáp các thắc mắc về công nghệ**
 - Tìm hiểu cách thức hoạt động của Spark
 - Tìm hiểu RDD
 - Tìm hiểu mô hình MapReduce
 - Tìm hiểu Spark Streaming
 - Tìm hiểu HDFS, Parquet file
 - Spark SQL
 - Spark SQL Optimization

2.3.2. Kết quả đạt được:

- Hoàn thành cài đặt và cấu hình Kafka và Spark.
- Thực hiện thành công các thao tác cơ bản với Kafka và Spark.
- Viết báo cáo chi tiết về các công nghệ, cách cài đặt, và các thao tác cơ bản.

- So sánh Kafka với RabbitMQ, nắm rõ các ưu nhược điểm của từng công nghệ.
- Hiểu cách tối ưu hóa truy vấn và xử lý dữ liệu với Spark.

2.3.3. Những điều chưa đạt được

- Chưa tối ưu hóa toàn diện cho các thao tác với dữ liệu lớn trong Kafka và Spark.
- Chưa triển khai các kịch bản phức tạp sử dụng Spark Streaming và các mô hình phức tạp trong Kafka.
- Cần tìm hiểu thêm về các phương pháp tối ưu hóa hiệu suất trong Kafka và Spark.

2.2.3. Bài học kinh nghiệm

- Cần tìm hiểu kỹ lưỡng về cấu trúc và mô hình dữ liệu của từng công nghệ trước khi triển khai.
- Việc sử dụng Docker giúp giảm thiểu thời gian cài đặt và cấu hình môi trường phát triển.
- Cần thường xuyên tham khảo tài liệu và hướng dẫn chính thức của từng công nghệ để nắm bắt các tính năng mới và các phương pháp tối ưu hóa.
- Việc so sánh và lựa chọn công nghệ cần dựa trên yêu cầu cụ thể của dự án và khả năng mở rộng trong tương lai.

2.2.4. Review, minh chứng:



2.4. NỘI DUNG CÔNG VIỆC TUẦN 5 (21/7 – 28/7):

2.4.1. Tìm hiểu về công nghệ và project:

- **Track 5: Tìm hiểu về MongoDB và viết báo cáo làm các task liên quan:**
 - Replica
 - Sharding
 - WiredTiger (storage)
 - xem phân pluggable storage
 - Deployment and administration
 - Design patterns
 - Queue, tree, transactions
 - Optimization query
 - Tìm hiểu về hệ thống Elastic Search

- Tìm hiểu Debezium
- Công nghệ trên thường được sử dụng trong các hệ thống quản lý dữ liệu, hệ thống phân phối quảng cáo, hệ thống phân tích và báo cáo quảng cáo của công ty. Những công nghệ này giúp tăng cường hiệu suất, độ tin cậy, và khả năng mở rộng của các hệ thống AdTech, giúp công ty quản lý và tối ưu hóa các chiến dịch quảng cáo hiệu quả hơn.

2.4.2. Kết quả đã đạt được:

- Hiểu và triển khai mô hình Replica Set, kiểm tra và xác nhận việc failover giữa các node trong Replica Set, thực hiện các thao tác đọc/ghi trong môi trường Replica Set.
- Tìm hiểu cách phân mảnh dữ liệu (sharding) để quản lý khối lượng dữ liệu lớn.
- Hiểu về WiredTiger Storage Engine và các ưu điểm của nó, cấu hình và kiểm tra hiệu suất của WiredTiger so với các storage engine khác.
- Nghiên cứu về khả năng thay đổi storage engine trong MongoDB, thử nghiệm với các storage engine khác nhau và so sánh hiệu suất.
- Triển khai MongoDB trong các môi trường khác nhau (local, cloud).
- Áp dụng các mẫu thiết kế Design pattern phổ biến trong việc xây dựng ứng dụng với MongoDB.
- Cài đặt và kiểm tra các cấu trúc hàng đợi (queue) và cây (tree) trong MongoDB, sử dụng các giao dịch (transactions) để đảm bảo tính nhất quán dữ liệu.
- Tối ưu hóa các truy vấn MongoDB để cải thiện hiệu suất, tìm hiểu và tích hợp Elastic Search với MongoDB.
- Nghiên cứu Debezium và cách nó hoạt động với MongoDB, cấu hình và thử nghiệm việc stream dữ liệu từ MongoDB sang các hệ thống khác.

2.4.3. Kết quả chưa đạt được:

- Vẫn còn khó khăn trong việc tối ưu hóa failover để giảm thiểu downtime.
- Chưa thể kiểm tra hết các kịch bản sử dụng WiredTiger trong môi trường sản xuất.
- Chưa có đủ thời gian để thử nghiệm sâu hơn với nhiều loại storage engine khác nhau.
- Còn gặp khó khăn trong việc đồng bộ dữ liệu real-time.
- Tích hợp Elastic Search với MongoDB vẫn còn một số hạn chế về hiệu suất.

2.4.4. Bài học kinh nghiệm:

- Tìm hiểu kỹ về các công cụ và công nghệ mới để đảm bảo hiệu suất và tính ổn định.
- Luôn cập nhật và áp dụng các best practices trong quản trị và triển khai hệ thống MongoDB.
- Luôn kiểm tra và tối ưu hóa cấu trúc dữ liệu trước khi triển khai hệ thống.

2.4.5. Review, minh chứng:


Vi Ngo Van 23:33, Th 5

Debezium

Lúc tìm hiểu e tìm hiểu xem nó là gì? kiến trúc nó ra sao? có cách nào khác ko? nó work ntn?


Vi Ngo Van 23:30, Th 5

Elasticsearch

kỹ thuật ranking ntn? cơ chế index ra sao? tokenizer thì ntn?


Vi Ngo Van 23:30, Th 5

Caching: sử dụng cơ chế caching để lưu trữ kết...

cơ chế cache ntn?


Vi Ngo Van 23:30, Th 5

Optimization Query

từng query này có mô tả phù hợp cho use case ntn ko

2.5. NỘI DUNG CÔNG VIỆC TUẦN 6 – 7 (28/7 – 11/8):

2.5.1. Project:

- Xây dựng một hệ thống search engine:

- Có luồng stream CDC (Data Change Capture) từ một hệ thống database khác
- Có tính năng autocomplete (tìm hiểu thuật toán liên quan) 293989
- Có tính năng search, và có thể tùy chọn ranking
- Có giao diện thực hiện việc search

2.5.2. Kết quả đạt được

- Hệ thống đã được tích hợp thành công với một nguồn dữ liệu khác bằng cách sử dụng một giải pháp CDC như Debezium.
- Hệ thống đã triển khai tính năng autocomplete sử dụng thuật toán cây tiền tố (Trie) để gợi ý các kết quả tìm kiếm khi người dùng nhập liệu.
- Đã triển khai thành công tính năng search với khả năng tùy chọn ranking. Người dùng có thể tìm kiếm và nhận kết quả dựa trên nhiều tiêu chí khác nhau, và hệ thống hỗ trợ sắp xếp các kết quả tìm kiếm dựa trên các thuật toán ranking như TF-IDF hoặc PageRank.
- Đã phát triển giao diện người dùng đơn giản và thân thiện, cho phép thực hiện các truy vấn tìm kiếm và hiển thị kết quả một cách trực quan.

2.5.3. Kết quả chưa đạt được

- Khi dữ liệu tăng lên, hệ thống search engine bắt đầu gặp phải các vấn đề về hiệu suất. Quá trình tìm kiếm và gợi ý tự động trở nên chậm hơn, đặc biệt là khi xử lý một lượng lớn dữ liệu trong thời gian thực.
- Mặc dù hệ thống đã hỗ trợ ranking cơ bản, nhưng các thuật toán sắp xếp vẫn chưa thực sự tối ưu.
- Giao diện hiện tại mới chỉ đáp ứng được các nhu cầu cơ bản và chưa có tính tương tác cao.

2.5.4. Bài học kinh nghiệm:

- Việc chọn đúng công nghệ như Debezium cho CDC hay Trie cho autocomplete đã giúp giải quyết các bài toán phức tạp một cách hiệu quả.
- Quá trình xây dựng hệ thống đã chỉ ra rằng khả năng mở rộng là một yếu tố cần được chú trọng ngay từ đầu.
- Hệ thống search engine liên quan đến nhiều yếu tố phức tạp và đòi hỏi phải thử nghiệm, điều chỉnh thường xuyên

III. KẾT QUẢ THỰC TẬP:

3.1. Những điều đạt được trong kì thực tập:

- Sau đợt thực tập tại Công ty Cổ phần VCCORP, em đã có cơ hội trải nghiệm và phát triển nhiều kỹ năng quan trọng.
 - Năm vững được các kỹ năng về java core và các công nghệ được áp dụng trong công ty cũng như cách hoạt động và kiến trúc của nó như Aerospike, Hbase, Kafka, Spark, MongoDB, ...
 - Em đã nắm vững hơn các kiến thức chuyên môn liên quan đến Back-End qua đó cải thiện đáng kể khả năng áp dụng lý thuyết vào thực tiễn.
 - Việc tham gia vào các công việc thực tế tại công ty đã giúp em hiểu rõ hơn về quy trình làm việc chuyên nghiệp, từ cách lập kế hoạch, quản lý thời gian, đến việc phối hợp và giao tiếp hiệu quả với đồng nghiệp.
 - Em đã học hỏi được rất nhiều kỹ năng mềm như kỹ năng làm việc nhóm, kỹ năng giải quyết vấn đề và khả năng tư duy sáng tạo.
 - Những thử thách trong công việc đã rèn luyện cho em tinh thần kiên nhẫn và sự linh hoạt trong cách tiếp cận và giải quyết vấn đề.
 - Em đã xây dựng được mối quan hệ tốt đẹp với các anh chị trong công ty, những người đã không ngừng hỗ trợ và chia sẻ kinh nghiệm quý báu với em.

- Những bài học từ đợt thực tập này không chỉ giúp em tự tin hơn khi bước vào thị trường lao động mà còn định hướng rõ ràng hơn cho con đường sự nghiệp sắp tới.

3.2. Những điều chưa đạt được:

- Mặc dù đợt thực tập tại Công ty Cổ phần VCCORP đã mang lại cho em nhiều kinh nghiệm quý báu, nhưng em nhận thấy vẫn còn một số điều mình chưa đạt được.
 - Do thời gian thực tập có giới hạn, em chưa có cơ hội để tham gia sâu vào toàn bộ các quy trình và dự án lớn của công ty, từ đó không thể nắm bắt đầy đủ các khía cạnh phức tạp và thực tiễn của công việc.
 - Em vẫn còn hạn chế trong việc quản lý thời gian và ưu tiên công việc, dẫn đến một số nhiệm vụ chưa được hoàn thành một cách tối ưu. Em nhận ra rằng cần phải cải thiện kỹ năng này để có thể làm việc hiệu quả hơn trong môi trường thực tế, nơi yêu cầu sự linh hoạt và chính xác trong quản lý thời gian.
 - Em cảm thấy mình vẫn cần phải nâng cao khả năng giao tiếp chuyên nghiệp, đặc biệt là trong việc trình bày ý tưởng và phản hồi nhanh chóng trong các tình huống cần ra quyết định. Em nhận thấy rằng kỹ năng này là rất quan trọng trong công việc nhóm và đóng góp vào sự thành công chung của dự án.
- Những điều chưa đạt được này là động lực để em tiếp tục học hỏi và hoàn thiện bản thân hơn nữa trong tương lai, chuẩn bị tốt hơn cho con đường sự nghiệp phía trước.

3.3. Bài học kinh nghiệm:

- Sau đợt thực tập tại Công ty Cổ phần VCCORP, em đã rút ra nhiều bài học kinh nghiệm quý báu mà em tin rằng sẽ có giá trị lâu dài trong sự nghiệp của mình.
 - Em nắm vững được các kỹ năng về java core và các công nghệ được áp dụng trong công ty cũng như cách hoạt động và kiến trúc của nó như Aerospike, Hbase, Kafka, Spark, MongoDB, ...
 - Em học được tầm quan trọng của việc chủ động trong công việc. Việc không ngại đặt câu hỏi và tìm kiếm sự hỗ trợ khi cần thiết đã giúp em giải quyết những vấn đề khó khăn một cách hiệu quả hơn, đồng thời mở rộng thêm kiến thức và kỹ năng của bản thân.
 - Em nhận thấy rằng kỹ năng làm việc nhóm đóng vai trò cực kỳ quan trọng trong môi trường công việc thực tế. Qua việc phối hợp với các đồng nghiệp trong dự án, em đã học được cách lắng nghe ý kiến của người khác, thấu hiểu và tôn trọng quan điểm đa dạng, từ đó đạt được sự đồng thuận và hợp tác hiệu quả. Điều này không chỉ giúp dự án tiến triển thuận lợi mà còn tạo ra một môi trường làm việc tích cực và sáng tạo.
 - Một bài học khác mà em rút ra được là khả năng quản lý thời gian và công việc. Em nhận ra rằng việc sắp xếp công việc theo mức độ ưu tiên và lên kế hoạch cụ thể cho từng nhiệm vụ là vô cùng cần thiết để đảm bảo tiến độ và

chất lượng công việc. Điều này cũng giúp em giảm thiểu áp lực và tránh những sai sót không đáng có.

- Cuối cùng, em hiểu rõ hơn về tầm quan trọng của sự linh hoạt và thích ứng trong công việc. Đôi khi, kế hoạch ban đầu có thể phải thay đổi để phù hợp với tình hình thực tế, và khả năng thích nghi nhanh chóng với những thay đổi đó là yếu tố quyết định thành công.

3.4. Đánh giá của người hướng dẫn(Mentor):

- Trong suốt quá trình thực tập tại Công ty Cổ phần VCCORP, em đã nhận được sự hướng dẫn tận tình và những đánh giá chân thành từ người hướng dẫn của mình, anh Ngô Văn Vĩ. Anh Vĩ đánh giá cao tinh thần học hỏi và sự chủ động của em trong công việc. Anh nhận thấy rằng em đã nỗ lực rất nhiều để làm quen với môi trường làm việc thực tế, cũng như không ngừng cải thiện kiến thức và kỹ năng chuyên môn. Tuy nhiên, anh Vĩ cũng chỉ ra rằng em cần phải rèn luyện thêm về kỹ năng quản lý thời gian và xử lý công việc dưới áp lực cao.
- Anh Vĩ còn nhấn mạnh rằng nếu tiếp tục duy trì thái độ tích cực và tinh thần cầu tiến thì một người sẽ còn phát triển vào mai sau. Sự đánh giá này không chỉ khích lệ em mà còn là động lực để em không ngừng phấn đấu, học hỏi thêm từ những trải nghiệm trong công việc sau này.

3.5. Tự đánh giá:

- Qua quá trình thực tập tại Công ty Cổ phần VCCORP, em tự nhận thấy mình đã có những tiến bộ nhất định trong việc áp dụng kiến thức lý thuyết vào thực tiễn công việc.
- Em đã tích cực tham gia vào các công việc và chủ động tìm hiểu, học hỏi từ những kinh nghiệm của các anh chị đồng nghiệp.
- Khả năng giao tiếp, làm việc nhóm của em cũng đã được cải thiện, giúp em hòa nhập tốt hơn vào môi trường công ty.
- Khả năng quản lý thời gian của em đôi khi chưa hiệu quả, dẫn đến việc một số công việc chưa được hoàn thành đúng tiến độ mong muốn.
- Em vẫn cần nâng cao thêm kỹ năng xử lý tình huống dưới áp lực cao, để có thể đưa ra những quyết định nhanh chóng và chính xác hơn trong công việc.
- Em đánh giá đợt thực tập này là một trải nghiệm vô cùng quý giá, giúp em không chỉ hiểu rõ hơn về lĩnh vực mà mình đang theo đuổi, mà còn nhận ra những điểm mạnh và yếu của bản thân.

IV. TỔNG KẾT:

Qua đợt thực tập tại Công ty Cổ phần VCCORP, em đã có cơ hội áp dụng những kiến thức đã học vào thực tế, từ đó tích lũy được nhiều kinh nghiệm quý báu. Đây là khoảng thời gian không chỉ giúp em trưởng thành hơn về mặt chuyên môn mà còn giúp em hiểu rõ hơn về môi trường làm việc chuyên nghiệp, cách thức vận hành của một công ty và những kỹ năng cần thiết để phát triển trong lĩnh vực mà mình đang theo đuổi.

Những kiến thức thực tế và bài học kinh nghiệm em có được từ đợt thực tập này sẽ là hành trang quý giá cho em trong sự nghiệp sau này. Dù còn một số hạn chế và thử thách chưa hoàn toàn khắc phục được, nhưng nhờ vào sự hướng dẫn tận tình của các anh chị đồng nghiệp, em đã rút ra được nhiều bài học quan trọng, đồng thời nhận thức rõ hơn về những điểm mạnh và yếu của bản thân.

Tóm lại, đợt thực tập này đã giúp em không chỉ củng cố nền tảng kiến thức mà còn trang bị cho em những kỹ năng và kinh nghiệm thực tế để tự tin hơn khi bước vào môi trường làm việc chính thức. Em xin gửi lời cảm ơn chân thành đến Công ty Cổ phần VCCORP, các anh chị hướng dẫn, cũng như thầy cô và gia đình đã luôn đồng hành, hỗ trợ em trong suốt quá trình thực tập. Em hy vọng rằng những kiến thức và kinh nghiệm thu được từ đợt thực tập này sẽ góp phần giúp em đạt được những thành công trong tương lai.

TÀI LIỆU THAM KHẢO

- [1] Phòng truyền thông Vccorp, Trang web chính thức của VCCORP, <https://vccorp.vn/>
- [2] Lê Văn Các, Báo cáo thực tập tốt nghiệp, <https://www.ebookbkmt.com/2020/09/bao-cau-thuc-tap-tot-nghiep-ung-dung.html>
- [3] Ngân hàng PTIT, Mẫu báo cáo thực tập tốt nghiệp ,
https://drive.google.com/drive/u/0/folders/1ip0i_G9NI_98WI8QgXY8lfEeyMKKf4N?fbclid=IwY2xjawEkiFllHRuA2FlbQIxMAABHWQV9yfY4azXxk5dKl0l6i2kW2Iajrl5otk0ZR7hDVu4zPnGVqdf6HG2Gg_aem_wgOWvW8uIUT0Qv4iPXJdmQ

-