TRƯỜNG ĐẠI HỌC KINH TẾ

**KHOA THỐNG KÊ – TIN HỌC**

–––––––––––––––––––––––––––––––



**BÁO CÁO THỰC TẬP NGHỀ NGHIỆP**

**NGÀNH HỆ THỐNG THÔNG TIN QUẢN LÝ**

**CHUYÊN NGÀNH QUẢN TRỊ HỆ THỐNG THÔNG TIN**

**DATA ENGINEER**

**THU THẬP DỮ LIỆU, XỬ LÝ VÀ ĐÁNH GIÁ NGÀNH CÔNG NGHIỆP ĐIỆN ẢNH**

Sinh viên thực hiện : Hồ Thị Hồng Thắm

Nguyễn Lê Hoài Long

Phạm Tấn Thiên

Lớp : 45K21.1

Đơn vị thực tập : TMA Solutions Bình Định

Cán bộ hướng dẫn : Nguyễn Bá Văn

Nguyễn Khương Duy

Nguyễn Tấn Vỹ

Giảng viên hướng dẫn : ThS. Cao Thị Nhâm

**Đà Nẵng, 8/2022**

**NHẬN XÉT CỦA ĐƠN VỊ THỰC TẬP**

Họ và tên sinh viên: Hồ Thị Hồng Thắm

Lớp: 45K21.1

Khoa: Thống kê – Tin học

Trường: Đại học Kinh tế - Đại học Đà Nẵng

Thực tập từ ngày: 06/06/2022 đến ngày: 12/06/2022

Tại: TMA Solutions Bình Định

Địa chỉ: 12 Đại lộ Khoa học, Thung lũng Sáng tạo, Ghềnh Ráng, Tp. Qui Nhơn, Bình Định

Sau quá trình thực tập tại đơn vị của sinh viên, chúng tôi có một số nhận xét, đánh giá như sau:

**1. Về thái độ, ý thức, đạo đức, kỷ luật**

**2. Kiến thức chuyên môn**

**3. Khả năng hòa nhập và thích nghi với công việc**

**4. Trách nhiệm, sáng tạo trong công việc**

**5. Các nhận xét khác**

**Đánh giá chung:**

**Điểm:**

Bình Định, ngày .......tháng ......năm 2022

**Xác nhận của đơn vị thực tập**

**NHẬN XÉT CỦA ĐƠN VỊ THỰC TẬP**

Họ và tên sinh viên: Nguyễn Lê Hoài Long

Lớp: 45K21.1

Khoa: Thống kê – Tin học

Trường: Đại học Kinh tế - Đại học Đà Nẵng

Thực tập từ ngày: 06/06/2022 đến ngày: 12/06/2022

Tại: TMA Solutions Bình Định

Địa chỉ: 12 Đại lộ Khoa học, Thung lũng Sáng tạo, Ghềnh Ráng, Tp. Qui Nhơn, Bình Định

Sau quá trình thực tập tại đơn vị của sinh viên, chúng tôi có một số nhận xét, đánh giá như sau:

**1. Về thái độ, ý thức, đạo đức, kỷ luật**

**2. Kiến thức chuyên môn**

**3. Khả năng hòa nhập và thích nghi với công việc**

**4. Trách nhiệm, sáng tạo trong công việc**

**5. Các nhận xét khác**

**Đánh giá chung:**

**Điểm:**

Bình Định, ngày .......tháng ......năm 2022

**Xác nhận của đơn vị thực tập**

**NHẬN XÉT CỦA ĐƠN VỊ THỰC TẬP**

Họ và tên sinh viên: Phạm Tấn Thiên

Lớp: 45K21.1

Khoa: Thống kê – Tin học

Trường: Đại học Kinh tế - Đại học Đà Nẵng

Thực tập từ ngày: 06/06/2022 đến ngày: 12/06/2022

Tại: TMA Solutions Bình Định

Địa chỉ: 12 Đại lộ Khoa học, Thung lũng Sáng tạo, Ghềnh Ráng, Tp. Qui Nhơn, Bình Định

Sau quá trình thực tập tại đơn vị của sinh viên, chúng tôi có một số nhận xét, đánh giá như sau:

**1. Về thái độ, ý thức, đạo đức, kỷ luật**

**2. Kiến thức chuyên môn**

**3. Khả năng hòa nhập và thích nghi với công việc**

**4. Trách nhiệm, sáng tạo trong công việc**

**5. Các nhận xét khác**

**Đánh giá chung:**

**Điểm:**

Bình Định, ngày .......tháng ......năm 2022

**Xác nhận của đơn vị thực tập**

# LỜI CẢM ƠN

Lời cảm ơn đầu tiên nhóm em xin gửi đến quý thầy, cô giáo trong khoa Thống kê – Tin học trường Đại học Kinh tế - Đại học Đà Nẵng lời cảm ơn chân thành vì đã tận tâm giảng dạy và truyền đạt những kiến thức, kinh nghiệm quý báu cho nhóm em. Đặc biệt nhóm em xin gửi đến cô Cao Thị Nhâm – người đã tận tình hướng dẫn, giúp đỡ nhóm hoàn thành chuyên đề báo cáo thực tập này lời cảm ơn sâu sắc nhất.

Nhóm xin chân thành cảm ơn ban lãnh đạo cùng với các phòng ban, các cô chú, anh chị trong Công ty TMA Solutions Bình Định – đơn vị đã tiếp nhận và đã tạo điều kiện thuận lợi cho nhóm tiếp cận thực tế, nắm bắt được các quy trình làm việc.

Cuối cùng nhóm xin cảm ơn các anh Nguyễn Khương Duy, Nguyễn Bá Văn, Nguyễn Tấn Vỹ - là những Mentors hướng dẫn tụi em trực tiếp tại công ty TMA. Các anh đã tận tình giúp đỡ, cung cấp những tài liệu cho tụi em để hoàn thành bài báo cáo trong suốt quá trình thực tập về mảng Data Engineer tại công ty.

Vì thời gian và kiến thức còn hạn hẹp nên bài báo cáo không thể tránh khỏi những thiếu sót, rất mong sự góp ý của công ty, quý thầy cô để tụi em rút kinh nghiệm và hoàn thành tốt hơn.

Nhóm em xin chân thành cảm ơn!

# LỜI CAM ĐOAN

Nhóm em xin cam đoan dự án “Thu thập dữ liệu, xử lý và đánh giá ngành công nghiệp điện ảnh” là kết quả nghiên cứu của nhóm dưới sự hướng dẫn của giảng viên Cao Thị Nhâm và các anh mentors Nguyễn Khương Duy, Nguyễn Bá Văn, Nguyễn Tấn Vỹ. Ngoài ra không có bất cứ sự sao chép nào của người khác.

Đề tài, nội dung báo cáo thực tập là sản phẩm mà nhóm em đã nỗ lực nghiên cứu trong suốt quá trình thực tập tại công ty TMA. Các dữ liệu trong bài báo cáo là hoàn toàn trùng thực, chúng em xin chịu hoàn toàn trách nhiệm, kỷ luật của nhà trường đề ra nếu có vấn đề nào xảy ra.

Quy định:

Soạn thảo trên trang A4 (trang dọc, lề trái: 3.5cm; trên, phải, dưới: 2.5 cm), Font Times New Roman, canh đều 2 bên, size 13, cách dòng 1.5, cách đoạn trên 6pt, cách đoạn dưới 3pt, hàng đâu tiên lùi vào 1.27 cm; hình và bảng soạn theo caption, chèn trích dẫn chéo (Cross-reference) cho bảng và hình; các danh mục hình, bảng, mục lục làm tự động; đánh số trang như file mẫu (bìa không có số trang, danh mục + mục lục số trang theo i, ii, iii…, nội dung chính theo 1,2,3…)

# MỤC LỤC

[NHẬN XÉT CỦA DOANH NGHIỆP THỰC TẬP ii](#_Toc98339732)

[LỜI CẢM ƠN iii](#_Toc98339733)

[LỜI CAM ĐOAN iv](#_Toc98339734)

[MỤC LỤC vi](#_Toc98339735)

[DANH MỤC HÌNH ẢNH viii](#_Toc98339736)

[DANH MỤC BẢNG BIỂU ix](#_Toc98339737)

[DANH MỤC CÁC TỪ VIẾT TẮT x](#_Toc98339738)

[LỜI MỞ ĐẦU 1](#_Toc98339739)

[CHƯƠNG 1. TỔNG QUAN … 2](#_Toc98339740)

[1.1. Mục 1.1 2](#_Toc98339741)

[1.1.1. Mục 1.1.1 2](#_Toc98339742)

[1.1.2. Mục 1.1.2 2](#_Toc98339743)

[1.2. Mục 1.2 2](#_Toc98339744)

[CHƯƠNG 2. LÝ THUYẾT 3](#_Toc98339745)

[2.1. Mục 2.1 3](#_Toc98339746)

[2.1.1. Mục 2.1.1 3](#_Toc98339747)

[2.1.2. Mục 2.1.2 3](#_Toc98339748)

[2.2. Mục 2.2 3](#_Toc98339749)

[CHƯƠNG 3. TRIỂN KHAI … 4](#_Toc98339750)

[3.1. Mục 3.1 4](#_Toc98339751)

[3.1.1. Mục 3.1.1 4](#_Toc98339752)

[3.1.2. Mục 3.1.2 4](#_Toc98339753)

[3.2. Mục 3.2 4](#_Toc98339754)

[CHƯƠNG 4. KẾT QUẢ 5](#_Toc98339755)

[4.1. Mục 4.1… 5](#_Toc98339756)

[4.2. Mục 4.2… 5](#_Toc98339757)

[KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN 6](#_Toc98339758)

[TÀI LIỆU THAM KHẢO 7](#_Toc98339759)

[PHỤ LỤC 8](#_Toc98339760)

*(Mục lục này chỉ là ví dụ)*

# DANH MỤC HÌNH ẢNH

[Hình 1.1 Ngôn ngữ lập trình Python 2](#_Toc98336120)

[Hình 4.1 Kiến trúc của mô hình 5](#_Toc98336121)

# DANH MỤC BẢNG BIỂU

[Bảng 4.1 Kiến trúc 5](#_Toc74235471)

# DANH MỤC CÁC TỪ VIẾT TẮT

**AI** : Artificial Intelligence

**…**

# LỜI MỞ ĐẦU

1. **Mục tiêu nghiên cứu của đề tài**

* Đề tài này nghiên cứu …

1. **Nhiệm vụ của đề tài**

* Nghiên cứu ...
* ...

1. **Phương pháp nghiên cứu**

* ...

1. **Đối tượng và phạm vi nghiên cứu**

* ...

1. **Kết cấu của đề tài**

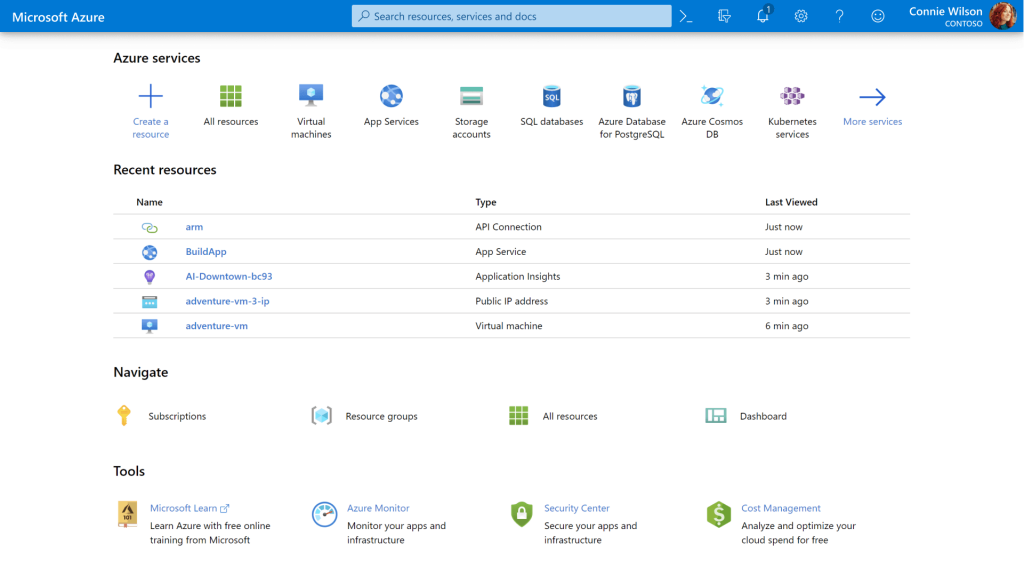
Đề tài được tổ chức gồm phần mở đầu, x chương nội dung và phần kết luận...

* Mở đầu
* **Chương 1**:
* **Chương 2**:
* **Chương 3**:
* **Chương 4**:
* Kết luận và hướng phát triển

# CƠ SỞ LÝ THUYẾT

## Microsoft Azure

### Microsoft Azure là gì?



Azure là một nền tảng điện toán đám mây (cloud computing platform) và một cổng trực tuyến (online portal) cho phép truy cập và quản lý các dịch vụ (service) và tài nguyên (resource) đám mây do Microsoft cung cấp. Các dịch vụ và tài nguyên này bao gồm lưu trữ và chuyển đổi dữ liệu, tùy thuộc vào yêu cầu là gì. Để có quyền truy cập vào các tài nguyên và dịch vụ này, tất cả những là kết nối internet và khả năng kết nối với Azure portal.

Những điều cần biết về Azure:

* Azure được phát hành vào ngày 1 tháng 2 năm 2010, muộn hơn đáng kể so với đối thủ cạnh tranh chính AWS.
* Azure theo mô hình miễn phí để bắt đầu sử dụng và sau đó trả tiền cho mỗi lần sử dụng, có nghĩa là bạn chỉ trả tiền cho các dịch vụ mà bạn chọn chạy trên Azure.
* 80% trong số 500 công ty trong danh sách Fortune sử dụng dịch vụ Azure cho nhu cầu điện toán đám mây.
* Azure hỗ trợ nhiều ngôn ngữ lập trình, bao gồm Java, Node Js và C #.
* Một lợi ích khác của Azure là số lượng trung tâm dữ liệu hiện có trên khắp thế giới. Có 42 (con số này vẫn đang tăng) trung tâm dữ liệu Azure trải rộng trên toàn cầu, đây là số lượng trung tâm dữ liệu cao nhất cho bất kỳ nền tảng đám mây nào. Ngoài ra, Azure cũng đang có kế hoạch mở thêm 12 trung tâm dữ liệu, điều này sẽ sớm nâng số trung tâm dữ liệu lên 54 trung tâm.

### Cloud Computing (Điện toán đám mây)

a. Cloud Computing là gì?

Cloud Computing là việc cung cấp các dịch vụ điện toán hoàn toàn qua Internet. Hay nói đúng hơn là việc cung cấp tài nguyên phù hợp với nhu cầu người dùng hoàn toàn thông qua Internet. Các dịch vụ ở đây có thể bao gồm máy chủ, lưu trữ, phần mềm, …



b. Ưu điểm của Cloud Computing

- Có khả năng mở rộng linh hoạt về quy mô.

- Nhanh nhẹn: Là quá trình phân bổ hay phân bổ nguồn lực một cách nhanh chóng.

- Khả năng chịu lỗi: Khả năng duy trì hoạt động trong các lỗi của bộ phận và dịch vụ được gọi là khả năng chịu lỗi.

- Phục hồi sau thảm họa: Thảm họa xảy ra khi có thiên tai. Vì vậy, chúng sao chép các tài nguyên tại các địa điểm khác nhau để có thể hoạt động ngay cả trong thời gian thảm họa.

- Tính sẵn sàng cao: Đó là khả năng giữ cho các tài nguyên có thể truy cập được trong một thời gian dài với thời gian ngừng hoạt động tối thiểu. Tính khả dụng càng lớn thì càng tốt. Hầu hết các nguồn tài nguyên trong màu xanh đều có tính sẵn sàng cao.

c. Các dạng Cloud Computing

* Public Cloud

Public Cloud là loại hình được sở hữu và vận hành bởi nhà cung cấp dịch vụ trung gian. Họ cung cấp tài nguyên điện toán của h, như máy chủ và bộ nhớ hoàn toàn thông qua Internet.

Với Public Cloud, tất cả phần cứng, phần mềm và cơ sở hạ tầng hỗ trợ khác đều do nhà cung cấp sở hữu và quản lý.

* Private Cloud

Private Cloud là một dạng tài nguyên được sử dụng riêng cho một tổ chức hoặc doanh nghiệp nào đó. Lúc này các tài nguyên của hệ thống công nghệ thông tin trong doanh nghiệp đều được chia sẻ về hệ thống máy chủ tính toán, giúp doanh nghiệp tận dụng được các tài nguyên này một cách hiệu quả và bảo mật nhất, cũng như tối ưu được nguồn tài chính, giảm thiểu rủi ro đầu tư cho chi phí cố định.

Ngoài ra, đối với hình thức Private Cloud doanh nghiệp cũng có thể tận dụng các hạ tầng có sẵn của các đơn vị để triển khai các dịch vụ của riêng mình.

* Hybrid Cloud

Là sự kết hợp giữa cả hai hình thức trên. Chúng được ràng buộc với nhau bằng công nghệ để đảm bảo sự riêng tư giữa chúng. Lựa chọn này sẽ giúp cho doanh nghiệp có nhiều lựa chọn hơn, giúp tối ưu hóa cơ sở hạ tầng hiện tại.

### Understand Azure Core Services

##### Vùng Azure

* Azure có các khu vực trên toàn thế giới có thể được sử dụng để lưu trữ tài nguyên.
* Bạn có thể lưu trữ tài nguyên trong Azure ở bất kỳ khu vực nào có sẵn.
* Vùng khả dụng:

+ Đây là những địa điểm riêng biệt về mặt vật lý trong một vùng Azure.

+ Mỗi vùng khả dụng được tạo thành từ một hoặc nhiều trung tâm dữ liệu.

+ Mỗi máy đều được trang bị nguồn, làm mát và kết nối mạng độc lập.

+ Có thể sử dụng vùng sẵn sàng để chạy các ứng dụng quan trọng với tính khả dụng cao và sao chép độ trễ thấp.

##### Mạng ảo

Máy ảo Azure được lưu trữ trong Mạng ảo Azure. Trong Azure, bạn tạo mạng ảo của mình. Nếu cần, bạn có thể xây dựng một số mạng ảo trong Azure. Mỗi mạng ảo trong azure cần được gán cho một không gian địa chỉ.

Một số mạng con được gọi có không gian địa chỉ là một tập con của mạng ảo. Vì có thể có nhiều tập con cho một tập hợp số, nên có thể có nhiều mạng con cho một mạng ảo. Bạn có thể tạo một máy ảo trong một mạng con, mỗi máy được khởi chạy trong một mạng con sẽ có một địa chỉ IP riêng. Địa chỉ IP riêng tạo điều kiện giao tiếp giữa các mạng con.

### Understand Security, Privacy, Compliance, And Trust

### Understand Azure Pricing And Support

## Azure Databricks

* Đây là nền tảng phân tích dựa trên Apache Spark được tối ưu hóa cho đám mây Microsoft Azure
* Với Azure Databricks, bạn có thể tạo các cụm Apache Spark được quản lý hoàn toàn
* Bạn có thể động các cụm tỷ lệ tự động. Bạn cũng có thể tạo các cụm máy chủ và cũng có thể chia sẻ chúng giữa các nhóm
* Bạn dễ dàng khám phá dữ liệu với sự trợ giúp của sổ ghi chép bằng R, Python, Scala và SQL
* Bạn có thể sử dụng trang tổng quan tương tác để tạo báo cáo động
* Nó có các khía cạnh bảo mật khi nói đến kiểm soát truy cập dựa trên vai trò và tích hợp với Azure Active Directory
* Apache Spark là một công cụ đa ngôn ngữ để thực thi kỹ thuật dữ liệu, khoa học dữ liệu và học máy trên các máy hoặc nút đơn.
* Các tính năng chính:

+ Dữ liệu hàng loạt/ Truyền trực tuyến:

Hợp nhất việc xử lý dữ liệu của bạn theo lô và phát trực tuyến theo thời gian thực, sử dụng ngôn ngữ ưa thích của bjan: python, SQL, Scala, Java haowjc R.

+ Phân tích SQL:

Thực thi các truy vấn ANSI SQL phân tán, nhanh chóng để lập bảng điều khiển và báo cáo đặc biệt. Chạy nhanh hơn hầu hết các kho dữ liệu.

+ Khoa học dữ liệu ở quy mô:

Thưc thi phân tích dữ liệu khám phá (EDA) trên dữ liệu quy mô petabyte mà không cần phải lấy mẫu xuống.

+ Máy học:

Đào tạo các thuật toán học máy trên máy tính xách tay sử dụng cùng một mã để chia tỷ lệ thành các cụm có khả năng chịu looic của hàng nghìn máy.

# LÝ THUYẾT

## Mục 2.1

Nội dung văn bản….

### Mục 2.1.1

### Mục 2.1.2

## Mục 2.2

# TRIỂN KHAI …

## Mục 3.1

### Mục 3.1.1

### Mục 3.1.2

## Mục 3.2

# KẾT QUẢ

## Mục 4.1…

## Mục 4.2…

Kết quả được xây dựng dựa trên ngôn ngữ lập trình thể hiện như (Bảng 4.1).

Bảng . Kiến trúc

|  |  |  |
| --- | --- | --- |
| Lớp | Ý nghĩa | Tham số |
|  |  |  |
|  |  |  |
|  | | |

Hình . Kiến trúc của mô hình

# KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Đề tài đã thực hiện được …

# TÀI LIỆU THAM KHẢO

* + - 1. Tên tác giả, Tên tác giả… *Tên công trình*. Nơi xuất bản, năm, trang
      2. Tên tác giả, Tên tác giả… *Tên công trình*. Nơi xuất bản, năm, trang

# PHỤ LỤC