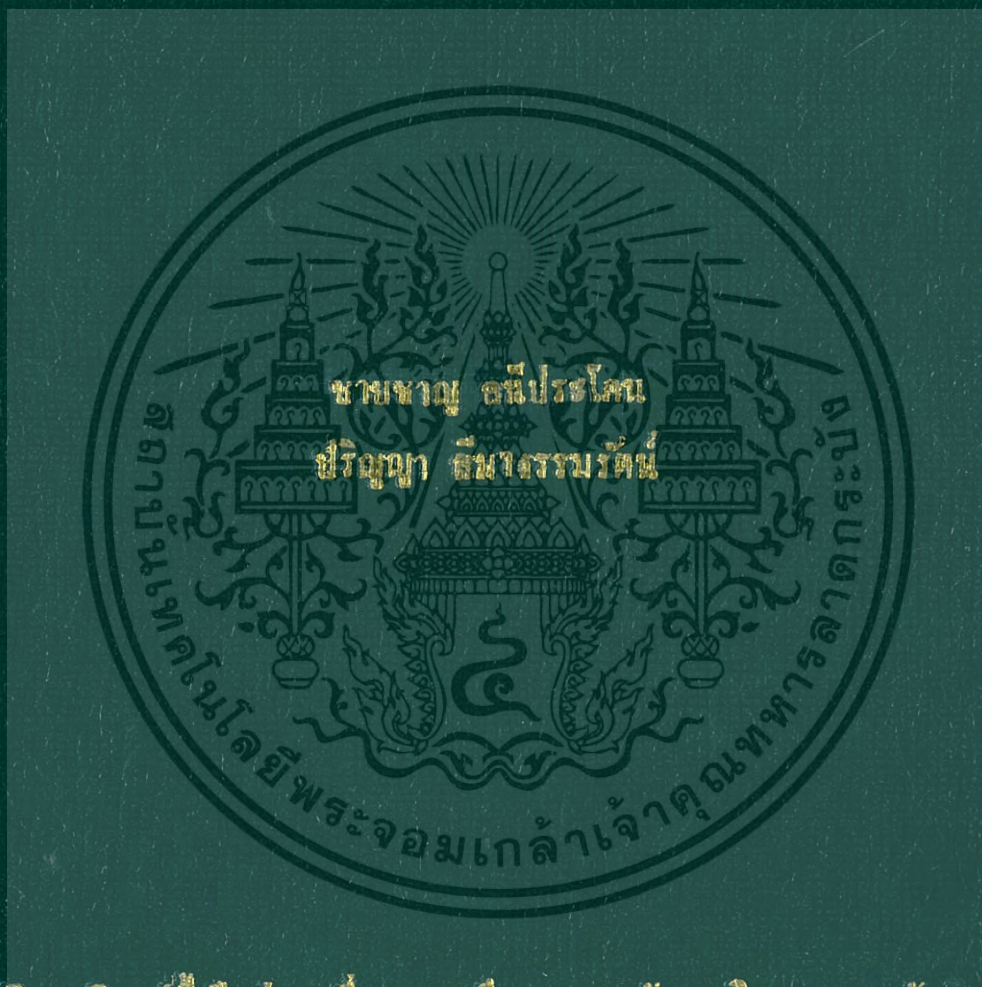


คลังข้อมูลสำหรับคณะเทคโนโลยีสารสนเทศบนระบบฮาปาเซฮาดูพ

FACULTY OF INFORMATION TECHNOLOGY DATA  
WAREHOUSE ON APACHE HADOOP SYSTEM



ปริญญาโท เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต

สาขาวิชาเทคโนโลยีสารสนเทศ

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ภาคเรียนที่ 2 ปีการศึกษา 2558

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

คลังข้อมูลสำหรับคณะเทคโนโลยีสารสนเทศระบบบออาปาเซฮาดูป

FACULTY OF INFORMATION TECHNOLOGY DATA  
WAREHOUSE ON APACHE HADOOP SYSTEM



T146231



เลขหมู่ 146231  
ลักษณะเป็น  
รวมเดือนปี 25 ๒๕๖๑-๖๕๖๑

b. 12840361  
l.

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต

สาขาวิชาเทคโนโลยีสารสนเทศ

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ภาคเรียนที่ 2 ปีการศึกษา 2558

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น มิได้อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**FACULTY OF INFORMATION TECHNOLOGY DATA  
WAREHOUSE ON APACHE HADOOP SYSTEM**



**A PROJECT SUBMITTED IN PARTIAL FULFILLMENT  
OF THE REQUIREMENT FOR THE DEGREE OF  
BACHELOR OF SCIENCE PROGRAM IN INFORMATION TECHNOLOGY  
FACULTY OF INFORMATION TECNOLOGY  
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

**2/2015**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



**COPYRIGHT 2016**

**FACULTY OF INFORMATION TECHNOLOGY**

**KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ใบรับรองปริญญาโท ประจำปีการศึกษา 2558

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง คัดตั้งข้อมูลสำหรับคณะเทคโนโลยีสารสนเทศบนระบบอาปาเซฮาดูป

FACULTY OF INFORMATION TECHNOLOGY DATA

WAREHOUSE ON APACHE HADOOP SYSTEM

ผู้จัดทำ

1. นายชายชาญ อณีประโคน รหัสนักศึกษา 55070029
2. นายปริญญา ธิมาธรรมรัตน์ รหัสนักศึกษา 55070070

  
..... อาจารย์ที่ปรึกษา  
(รองศาสตราจารย์ ดร.วรพจน์ กริสุระเดช)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อโครงการ	คลังข้อมูลสำหรับคณะเทคโนโลยีสารสนเทศระบบอาปาเซฮาดูป
นักศึกษา	นายชายชาญ อนิประโคน รหัสนักศึกษา 55070029 นายปริญญา สีมารธรรมรัตน์ รหัสนักศึกษา 55070070
ปริญญา	วิทยาศาสตร์บัณฑิต
สาขาวิชา	เทคโนโลยีสารสนเทศ
ปีการศึกษา	2558
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร. วรพจน์ กิริสุระเดช

## บทคัดย่อ

ในปัจจุบันการจัดการข้อมูลและวิเคราะห์ข้อมูลนักศึกษาและอาจารย์ของ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ผู้ใช้งานจะต้องกรอกข้อมูลลงในโปรแกรมไมโครซอฟต์เอกเซล (Microsoft Excel) ตามรูปแบบที่ได้สร้างไว้ โดยจะกรอกข้อมูลในทุกๆ เทอมของแต่ละชั้นปี เพื่อใช้ในการสรุปหรือประเมินผล

จากการศึกษากระบวนการทำงานในปัจจุบันพบว่าการจัดการและวิเคราะห์ข้อมูลนักศึกษาและอาจารย์มีความยุ่งยาก เนื่องจากผู้ใช้งานต้องวิเคราะห์ข้อมูลด้วยตัวเอง ซึ่งทำให้เกิดความผิดพลาดได้ง่าย และไม่สามารถตีความหมายหรือวิเคราะห์ข้อมูลออกในรูปแบบแผนภูมิภาพเพื่อใช้ประกอบการวิเคราะห์แนวโน้มที่จะเกิดขึ้นในอนาคตได้ รวมทั้งอาจมีความเสี่ยงต่อข้อมูลสูญหายและด้านความปลอดภัยของข้อมูล โดยเฉพาะข้อมูลส่วนตัวแต่ละบุคคล

ดังนั้นจึงได้พัฒนาคลังข้อมูลสำหรับคณะเทคโนโลยีสารสนเทศระบบอาปาเซฮาดูปขึ้นเพื่อช่วยในการเก็บข้อมูลหรือเรียกใช้ข้อมูลสารสนเทศของคณะได้อย่างรวดเร็วและแม่นยำมากยิ่งขึ้น ซึ่งขั้นตอนการทำงานในระบบจะเริ่มต้นจากการนำข้อมูลที่ผู้ใช้ป้อนลงบนแบบฟอร์มอัปโหลดเพิ่มข้อมูลขึ้นสู่ระบบ และจะสามารถนำเสนอข้อมูลในรูปแบบของแดชบอร์ด(Dashboard) ได้ง่ายขึ้น นอกจากนี้ยังมีฟังก์ชันเสริมอื่น ๆ อีก เช่น การจัดทำรายงานแผนภูมิภาพ ค้นหาข้อมูลที่เป็นเฉพาะส่วนของข้อมูล และสามารถนำข้อมูลออกจากระบบได้โดยง่าย โครงการนี้จะช่วยให้อำนวยการข้อมูลสารสนเทศของคณะเทคโนโลยีสารสนเทศ เพื่อใช้ประเมินผลหรือวิเคราะห์แนวโน้มในอนาคตได้อย่างมีประสิทธิภาพมากขึ้น อีกทั้งยังทำให้มีความยืดหยุ่นในการใช้งานเพิ่มมากยิ่งขึ้น

**Project Title** FACULTY OF INFORMATION TECHNOLOGY DATA  
WAREHOUSE ON APACHE HADOOP SYSTEM

**Student** Mr. Chaichan Aneprakhone Student ID 55070029  
Mr. Parinya Srimathammarat Student ID 55070070

**Degree** Bachelor of Science

**Program** Information Technology

**Academic Year** 2015

**Advisor** Associate Professor Dr. Worapoj Kreesuradej

## ABSTRACT

Currently, faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang uses Microsoft Excel for making reports and analyzing student and lecturer data. These have some difficulties in managing and analyzing the information because student and lecturer data have to be entered manually. In addition, the current system is at risk of data loss and security.

Therefore, we propose to develop Faculty of Information Technology Data Warehouse using Apache Hadoop technology. The proposed system has better efficient in data storage than the current system. The new system also has dashboard reports and ad hoc reports. This project will improve the efficiency of making reports and analyzing student and lecturer data.

## กิตติกรรมประกาศ

ปริญญานิพนธ์เรื่องคลังข้อมูลสำหรับคณะเทคโนโลยีสารสนเทศบนระบบอาปาเซฮาดูปฉบับนี้อาจจะไม่ประสบผลสำเร็จลุล่วงได้หากไม่ได้รับความกรุณาจากอาจารย์ที่ปรึกษาโครงการรองศาสตราจารย์ ดร.วรพจน์ กรีสระเดช, อาจารย์ผู้มีส่วนเกี่ยวข้องกับระบบคลังข้อมูลคณะเทคโนโลยีสารสนเทศ ดร. กนกวรรณ อัจฉริยะชาญวณิช และผู้เชี่ยวชาญทางด้าน Big Data คุณกิตติรักษ์ ม่วงมิ่งสุข เป็นอย่างสูง ซึ่งท่านได้ให้ความรู้ แนวคิด คำแนะนำ ที่คอยส่งเสริมช่วยเหลือในการดำเนินงาน อีกทั้งสนับสนุนโครงการ ตอบข้อสงสัยต่างๆ ตลอดระยะเวลาการจัดทำโครงการ ทั้งนี้อาจารย์คอยให้โอกาสในการเข้าร่วมในการอบรมเกี่ยวกับเทคโนโลยีที่เกี่ยวข้อง กำลังใจอยู่เสมอ เพื่อให้ต่อยอดในการทำโครงการได้ง่ายขึ้น

สุดท้ายนี้ขอขอบคุณ บิดา มารดา เพื่อนนักศึกษา ตลอดจนผู้ที่เกี่ยวข้องทุกท่านที่ไม่ได้กล่าวนามไว้ ณ ที่นี้ ที่ได้ให้กำลังใจและมีส่วนช่วยเหลือให้ความร่วมมือในการจัดทำโครงการครั้งนี้เป็นอย่างดี

ชายชาญ อนิประโคน

ปริญญา สีมาธรรมรัตน์

# สารบัญ

หน้า

บทคัดย่อ .....	I
ABSTRACT .....	II
กิตติกรรมประกาศ .....	III
สารบัญ .....	IV
สารบัญตาราง .....	VII
สารบัญรูป .....	VIII

บทที่ 1 บทนำ .....	1
1.1 ความเป็นมาและความสำคัญ .....	1
1.2 ความมุ่งหมายและวัตถุประสงค์ .....	2
1.3 ขอบเขต .....	2
1.4 ขั้นตอนและวิธีการดำเนินงาน .....	2
1.5 ประโยชน์ที่คาดว่าจะได้รับ .....	3
1.6 ส่วนประกอบของปริญญานิพนธ์ .....	3

บทที่ 2 การทบทวนทฤษฎีพื้นฐาน งานวิจัย และเทคโนโลยีที่เกี่ยวข้อง .....	5
2.1 ทฤษฎีและงานวิจัยที่เกี่ยวข้องกับระบบคลังข้อมูล .....	5
2.1.1 คำจำกัดความของคลังข้อมูล .....	5
2.1.2 คุณลักษณะเฉพาะของคลังข้อมูล .....	6
2.1.3 สถาปัตยกรรมของคลังข้อมูล .....	6
2.1.4 ขั้นตอนการสร้างคลังข้อมูล .....	8
2.2 เทคโนโลยีที่เกี่ยวข้องกับโครงการ .....	11
2.2.1 เทคโนโลยีอาปาเช่าฮาดูป .....	11

# สารบัญ (ต่อ)

หน้า

2.2.2 เทคโนโลยีที่ใช้ในการสร้างเว็บแอปพลิเคชัน .....	27
<b>บทที่ 3</b> วิธีการดำเนินการ .....	<b>33</b>
3.1 ระบบการทำงานในปัจจุบัน.....	33
3.2 ขั้นตอนการดำเนินงาน .....	34
3.3 การออกแบบคลังข้อมูล .....	35
3.3.1 Fact Table .....	36
3.3.2 Dimension Table .....	37
3.3.3 Other Table .....	39
3.4 การออกแบบยูสเคส ไดอะแกรม (Use Case Diagram).....	40
3.5 แผนภาพกิจกรรม (Activity Diagram).....	44
3.6 แผนภาพแสดงคลาส (Class Diagram) .....	47
3.7 แผนภาพแสดงลำดับการทำงาน (Sequence Diagram).....	49
3.8 ออกแบบส่วนติดต่อกับผู้ใช้.....	51
3.8.1 ไวร์เฟรม (Wireframe) .....	51
<b>บทที่ 4</b> ผลการดำเนินงานของโครงการ .....	<b>54</b>
4.1 เว็บแอปพลิเคชัน (Web Application) .....	54
4.2 ส่วนควบคุมการทำงานของระบบ .....	55
4.3 ส่วนการประมวลผลข้อมูล .....	59
<b>บทที่ 5</b> บทสรุปของโครงการ .....	<b>65</b>
5.1 สรุปผลโครงการ .....	65
5.2 ปัญหาและอุปสรรค.....	65

## สารบัญ (ต่อ)

หน้า

5.3 แนวทางพัฒนาต่อ.....	65
บรรณานุกรม .....	66
ภาคผนวก ก .....	69



# สารบัญตาราง

ตารางที่	หน้า
2.1 แสดงการเปรียบเทียบระหว่าง Pig, Hive และ Impala .....	26
2.2 แสดงการเปรียบเทียบระหว่าง RDBMS, Hive และ Impala .....	27
3.1 แสดงชนิดของข้อมูลใน Fact Table ของนักศึกษา ในตาราง Personal_Status_Fact.....	36
3.2 แสดงชนิดของข้อมูลใน Fact Table ของนักศึกษา ในตาราง Teacher_Workload_Fact .....	37
3.3 แสดงชนิดของข้อมูลใน Dimension Table ของจุดของเวลา ในตาราง Date.....	37
3.4 แสดงชนิดของข้อมูลใน Dimension Table ของวิชาที่เปิดสอน ในตาราง Course .....	37
3.5 แสดงชนิดของข้อมูลใน Dimension Table ของสถานะการศึกษา ในตาราง Status.....	38
3.6 แสดงชนิดของข้อมูลใน Dimension Table ของวุฒิการศึกษาที่เข้าศึกษา ในตาราง Degree.....	38
3.7 แสดงชนิดของข้อมูลใน Dimension Table ของนักศึกษา ในตาราง Student .....	38
3.8 แสดงชนิดของข้อมูลใน Dimension Table ของอาจารย์ผู้สอน ในตาราง Category.....	38
3.9 แสดงชนิดของข้อมูลใน Dimension Table ของเพศ ในตาราง Sex.....	39
3.10 แสดงชนิดของข้อมูลใน Dimension Table ของจังหวัด ในตาราง Province .....	39
3.11 แสดงชนิดของข้อมูลใน Dimension Table ของอาจารย์ผู้สอน ในตาราง Teacher.....	39
3.12 แสดงชนิดของข้อมูลในตาราง Staff_Account .....	40
3.13 แสดงรายละเอียดของ Use Case เพิ่มข้อมูลลงคลังข้อมูลในแต่ละตาราง.....	41
3.14 แสดงรายละเอียดของ Use Case ดูข้อมูลในคลังข้อมูลแต่ละตาราง .....	41
3.15 แสดงรายละเอียดของ Use Case ดูแดชบอร์ดโดยรวมของระบบ(Dashboard) .....	42
3.16 แสดงรายละเอียดของ ข้อมูลแผนภูมิภาพ ตามที่ผู้ใช้ต้องการ .....	42
3.17 แสดงรายละเอียดของ แผนภูมิภาพเชิงเปรียบเทียบ 2 ค่า ตามที่ผู้ใช้ต้องการ .....	43

# สารบัญรูป

รูปที่	หน้า
2.1 แสดงหลักการการทำงานของคลังข้อมูล .....	5
2.2 แสดงโครงสร้างของสถาปัตยกรรมคลังข้อมูลแบบพื้นฐาน .....	7
2.3 แสดงโครงสร้างของสถาปัตยกรรมคลังข้อมูลแบบมีจุดตรวจสอบข้อมูล .....	7
2.4 แสดงโครงสร้างของสถาปัตยกรรมคลังข้อมูลแบบมีจุดตรวจสอบข้อมูล และตลาดข้อมูล.....	8
2.5 แสดงแบบจำลองข้อมูลเชิงมิติแบบดาวที่แยกประเภทของตารางด้วยสี .....	9
2.6 แสดงแบบจำลองข้อมูลเชิงมิติแบบเกล็ดหิมะ .....	10
2.7 แสดงกระบวนการ ETL.....	11
2.8 แสดงโลโก้ของอาปาเซ่ฮาดูป.....	12
2.9 แสดงหลักการการทำงานของ HDFS.....	13
2.10 แสดงหลักการการทำงานของ Map Reduce .....	14
2.11 แสดงแอปพลิเคชันในระบบนิเวศเชื่อมโยงของอาปาเซ่ฮาดูป .....	15
2.12 แสดงโลโก้ของอาปาเซ่เอ็มบาร์รี่ .....	15
2.13 แสดงโลโก้ของอาปาเซ่อาไวร์ .....	16
2.14 แสดงโลโก้ของอาปาเซ่แคสแซนดร้า .....	16
2.15 แสดงโลโก้ของอาปาเซ่ซุกวา .....	17
2.16 แสดงโลโก้ของอาปาเซ่ฟลูม .....	17
2.17 แสดงโลโก้ของอาปาเซ่เอชเบส .....	18
2.18 แสดงโลโก้ของอาปาเซ่มาเฮาท์.....	18
2.19 แสดงโลโก้ของอาปาเซ่อูซี.....	19
2.20 แสดงโลโก้ของอาปาเซ่ฟิค .....	19
2.21 แสดงโลโก้ของอาปาเซ่สคูป .....	20
2.22 แสดงโลโก้ของอาปาเซ่สปาร์ค .....	20
2.23 แสดงโลโก้ของอาปาเซ่ซูคิปเปอร์.....	21

## สารบัญรูป (ต่อ)

รูปที่	หน้า
2.24 แสดงโลโก้ของอาปาเซ่โฮพี.....	21
2.25 แสดงตัวอย่างการเขียนด้วย HiveQL .....	22
2.26 แสดงโลโก้ของอิมพาลา.....	22
2.27 แสดงการเปรียบเทียบการทำงานของ Pig และ Impala .....	25
2.28 แสดงตัวอย่างการเขียนด้วย JSP.....	28
2.29 แสดงโลโก้ของ Java.....	28
2.30 แสดงตัวอย่างการเขียนด้วยภาษา Java.....	29
2.31 แสดงโลโก้ของ JavaScript.....	29
2.32 แสดงตัวอย่างการเขียนด้วย JavaScript.....	30
2.33 แสดงโลโก้ของ CSS.....	30
2.34 แสดงตัวอย่างการเขียนด้วย CSS .....	31
2.35 แสดงโลโก้ของ jQuery.....	31
2.36 แสดงตัวอย่างการเขียนด้วย jQuery .....	32
2.37 แสดงโลโก้ของ jQuery.....	32
2.38 แสดงตัวอย่างการเขียน Highcharts.....	32
3.1 แผนผังการคุมกำหนดงาน (Gantt chart).....	33
3.2 แผนภาพการจัดเก็บข้อมูลของคณะเทคโนโลยีสารสนเทศในปัจจุบัน .....	34
3.3 ขั้นตอนการดำเนินงาน.....	35
3.4 แสดง Star Schema สำหรับเก็บข้อมูลของคณะเทคโนโลยีสารสนเทศ .....	36
3.5 แสดงการออกแบบยูสเคสไดอะแกรม .....	40
3.6 แสดงแผนภาพกิจกรรม Use Case ของการเพิ่มข้อมูลเข้าระบบคลังข้อมูล .....	44
3.7 แสดงแผนภาพกิจกรรม Use Case ของการดู/ค้นหาข้อมูลในระบบคลังข้อมูล.....	45
3.8 แสดงแผนภาพกิจกรรม Use Case ของการดูแดชบอร์ดแสดงแผนภูมิภาพโดย สรุปในระบบคลังข้อมูล .....	45

## สารบัญรูป (ต่อ)

รูปที่	หน้า
3.9 แสดงแผนภาพกิจกรรม Use Case ของการดูรีพอร์ตแสดงแผนภูมิภาพ โดยสรุปตามที่ใช้กำหนดเงื่อนไขในระบบคลังข้อมูล .....	46
3.10 แสดงแผนภาพกิจกรรม Use Case ของการดูรีพอร์ตแสดงแผนภูมิภาพ โดยสรุปตามที่ใช้กำหนดเงื่อนไขในระบบคลังข้อมูล .....	46
3.11 แสดง Class Diagram ของระบบคลังข้อมูล (ส่วนที่1 นักศึกษา).....	47
3.12 แสดง Class Diagram ของระบบคลังข้อมูล (ส่วนที่2 ครู/อาจารย์) .....	48
3.13 แสดงแผนแสดงลำดับการทำงานของระบบเพิ่มข้อมูลลงระบบแต่ละตาราง .....	49
3.14 แสดงแผนแสดงลำดับการทำงานของกรเรียกดูหรือค้นหาข้อมูลบนระบบ .....	49
3.15 แสดงแผนแสดงลำดับการทำงานของกรแดชบอร์ด โดยแสดงแผนภูมิภาพแบบสรุป .....	50
3.16 แสดงแผนแสดงลำดับการทำงานของกรรีพอร์ตแสดงจะแผนภูมิภาพตามที่ใช้กำหนด .....	50
3.17 แสดงแผนแสดงลำดับการทำงานของกรเปรียบเทียบเพื่อแสดงแผนภูมิภาพ ตามที่ใช้กำหนด .....	51
3.18 แสดงไวร์เฟรมหน้าแรก (Login) ของระบบคลังข้อมูล .....	51
3.19 แสดงไวร์เฟรมหน้าแดชบอร์ด (Dashboard) และรีพอร์ต (Report) ของ ระบบคลังข้อมูล.....	52
3.20 แสดงไวร์เฟรมหน้าเพิ่ม/คิวรี่ข้อมูล (Put/Query Data) ของระบบคลังข้อมูล .....	52
4.1 แสดงส่วนของหน้าล็อกอิน .....	54
4.2 แสดงส่วนของหน้าแดชบอร์ด .....	54
4.3 แสดงส่วนเกี่ยวข้องกับผู้ใช้งานและแสดงชื่อหน้า ณ ปัจจุบัน .....	55
4.4 แสดงส่วนเกี่ยวข้องกับผู้ใช้งาน .....	55
4.5 แสดงชื่อหน้า ณ ปัจจุบัน .....	55
4.6 แสดงส่วนของเมนูการควบคุม .....	56
4.7 แสดงส่วนของเมนูแดชบอร์ดภาพรวมของคณะ .....	56
4.8 แสดงส่วนประกอบย่อยของแต่ละเมนูการควบคุม.....	57

## สารบัญรูป (ต่อ)

รูปที่	หน้า
4.9 แสดงส่วนควบคุมของการเลือกข้อมูล.....	57
4.10 แสดงส่วนของเมนูการควบคุม ในส่วน Put/Query Data .....	58
4.11 แสดงหน้าของสิทธิ์การเข้าถึงข้อมูล .....	59
4.12 แสดงส่วนของการเพิ่มข้อมูลลงระบบคลังข้อมูล .....	59
4.13 แสดงหน้าตารางบนคลังข้อมูล (โดยลักษณะข้อมูลจะต่างกัน ตามที่เลือกแสดงข้อมูล) .....	60
4.14 แสดงหน้าตารางบนคลังข้อมูล (เมื่อมีการเพิ่มข้อมูลเข้าชื้อนระบบจะมีการแจ้งเตือน).....	60
4.15 แสดงส่วนหน้าแดชบอร์ดนักศึกษา (Dashboard) สรุปรวมทั้งระบบ.....	61
4.16 แสดงส่วนหน้าแดชบอร์ดครู/อาจารย์ (Dashboard) สรุปรวมทั้งระบบ .....	62
4.17 แสดงส่วนประกอบหน้าแดชบอร์ดครู/อาจารย์ (Dashboard) ในการทำแผนภูมิภาพออกจากระบบ.....	63
4.18 แสดงส่วนหน้าเรียกดูข้อมูล ตามที่ผู้ใช้กำหนดการคัดกรอง.....	63
4.19 แสดงส่วนประกอบหน้าเรียกดูข้อมูล ตามที่ผู้ใช้กำหนดการคัดกรอง ในการนำข้อมูลดิบออกจากระบบ .....	63

# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาและความสำคัญ

มหาวิทยาลัยและสถาบันการศึกษาชั้นนำของประเทศไทย ต่างที่จะต้องการผลิตภัณฑ์ที่มีคุณภาพและเป็นที่ต้องการในตลาดแรงงาน ดังนั้นการพัฒนาล้างข้อมูลของมหาวิทยาลัยและสถาบันการศึกษาจึงเป็นสิ่งจำเป็น โดยคลังข้อมูลจะช่วยให้สามารถตัดสินใจได้อย่างแม่นยำ รวดเร็ว อีกทั้งลดภาระการทำงานของบุคลากรได้ และยังสามารถเปรียบเทียบแนวโน้มหรือพยากรณ์สิ่งที่จะเกิดขึ้นในอนาคตได้โดยง่าย ซึ่งทั้งนี้ทำให้สามารถตัดสินใจและประเมินผล อาจารย์และนักศึกษาได้ง่ายขึ้น สร้างข้อได้เปรียบทางการแข่งขันกับมหาวิทยาลัยและสถาบันการศึกษาชั้นนำอื่นๆ ได้อย่างทันท่วงที

การนำระบบเทคโนโลยีคลังข้อมูลเข้ามาช่วยภายในคณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังนั้น เป็นปัจจัยสำคัญที่ช่วยในการวิเคราะห์หรือจัดเก็บข้อมูลที่ซับซ้อนของคณะให้เป็นระบบและง่ายต่อการเข้าถึงข้อมูลมากยิ่งขึ้น โดยสามารถดึงสารสนเทศจากข้อมูลออกมาใช้ประโยชน์ได้อย่างเต็มประสิทธิภาพ สามารถเข้าถึงข้อมูลได้อย่างรวดเร็วและแม่นยำ เนื่องจากปัจจุบันบุคลากรมีการจัดเก็บข้อมูลต่าง ๆ ลงบน Microsoft Excel แบบแยกส่วนกัน ข้อมูลค่อนข้างกระจัดกระจายกันทำให้ค้นหาข้อมูลได้ค่อนข้างยาก รวมทั้งยังต้องสรุปผลข้อมูลที่เกี่ยวข้องกับนักศึกษาและอาจารย์ของคณะในแต่ละเทอมเองทั้งหมด โดยในการจัดทำระบบในแต่ละครั้งนั้น จะมีค่าใช้จ่ายทางด้านฮาร์ดแวร์และซอฟต์แวร์ที่มีราคาค่อนข้างสูง เนื่องจากการจัดทำคลังข้อมูลนั้นจะเน้นการดึงหรือเก็บข้อมูลที่ละมากๆ จึงต้องใช้ฮาร์ดแวร์และซอฟต์แวร์ที่มีประสิทธิภาพสูงและเหมาะสมกับการใช้งาน ดังนั้นทางผู้จัดทำจึงนำเทคโนโลยีอาปาเช่ฮาดูป (Apache Hadoop) เข้ามาช่วยทดแทน ให้คณะสามารถเก็บและดึงข้อมูลได้อย่างรวดเร็วและมีประสิทธิภาพการทำงานสูง โดยมีต้นทุนที่ค่อนข้างต่ำ โดยการนำคอมพิวเตอร์หลาย ๆ เครื่องเข้ามาช่วยกันในการทำงาน และประมวผลผลแทนการใช้เซิร์ฟเวอร์ ที่สำคัญคือ ไม่ต้องเสียค่าลิขสิทธิ์ซอฟต์แวร์เพราะเป็นระบบโอเพนซอร์ส (Open Source) ซึ่งสามารถนำมาจัดทำเป็นแดชบอร์ด โดยแดชบอร์ดจะมีการแสดงข้อมูลและแผนภาพกราฟิกต่าง ๆ ซึ่งจัดทำขึ้นสำหรับสนับสนุนการทำงานของบุคลากรในคณะ ทั้งนี้เพื่อนำสารสนเทศของคณะไปใช้งานและสรุปผลเพื่อพัฒนาปรับปรุงคณะต่อไป

## 1.2 ความมุ่งหมายและวัตถุประสงค์

- 1) ต้องศึกษาหลักการการออกแบบและการพัฒนาระบบคลังข้อมูล (Data Warehouse)
- 2) ต้องศึกษาขอบเขต ทฤษฎี และหลักการการทำงานของระบบอาปาเช่ฮาดูป (Apache Hadoop)
- 3) สามารถพัฒนาระบบคลังข้อมูล (Data Warehouse) สำหรับคณะเทคโนโลยีสารสนเทศบนระบบอาปาเช่ฮาดูป (Apache Hadoop) ได้

## 1.3 ขอบเขต

โครงการพัฒนาระบบคลังข้อมูลสำหรับคณะเทคโนโลยีสารสนเทศบนอาปาเช่ฮาดูป ดำเนินการศึกษา ลงมือปฏิบัติและทดลองภายใต้ขอบเขตของโครงการ ดังนี้

- 1) พัฒนาระบบคลังข้อมูล (Data Warehouse) สำหรับข้อมูลนักศึกษาและบุคลากรในคณะเทคโนโลยีสารสนเทศเท่านั้น
- 2) พัฒนาค้างข้อมูลของคณะเทคโนโลยีสารสนเทศบนระบบอาปาเช่ฮาดูป เพื่อสร้างแดชบอร์ดและรองรับข้อมูลปริมาณมากในอนาคต
- 3) ความเร็วของการดึงหรือเก็บข้อมูลขึ้นอยู่กับปริมาณข้อมูลที่ต้องการให้แสดงและขีดจำกัดของฮาร์ดแวร์ของคอมพิวเตอร์และเซิร์ฟเวอร์ที่ติดตั้งระบบอยู่ในปัจจุบัน
- 4) พัฒนาเพื่อใช้บนคอมพิวเตอร์/โน้ตบุ๊ก โดยเข้าผ่าน Browser: Chrome 46 ขึ้นไป
- 5) ไม่สามารถลบข้อมูลที่เพิ่มขึ้นระบบไปแล้วได้ ฉะนั้นข้อมูลที่เพิ่ม จำเป็นต้องข้อมูลที่ถูกต้อง ไม่มีการเปลี่ยนแปลง

## 1.4 ขั้นตอนและวิธีการดำเนินงาน

ในขั้นตอนการศึกษาของโครงการนี้ จะมีขั้นตอนและวิธีการดำเนินงานซึ่งได้แสดงลำดับการดำเนินงานตั้งแต่เริ่มต้นโครงการจนถึงสิ้นสุดโครงการ ดังรายละเอียดต่อไปนี้

- 1) คิดหัวข้อโครงการและศึกษาหลักการออกแบบ และการพัฒนาระบบคลังข้อมูลเบื้องต้น
- 2) ศึกษาการขอบเขต ทฤษฎี และหลักการการทำงานของระบบอาปาเช่ฮาดูปเบื้องต้น
- 3) เก็บความต้องการและความคาดหวังที่จะได้รับจากโครงการนี้ของผู้ใช้งานจริง
- 4) ออกแบบระบบคลังข้อมูลตามความต้องการและความคาดหวังของผู้ใช้งานจริง
- 5) ออกแบบระบบต้นแบบ (Prototype) ของระบบลงบนกระดาษ ก่อนนำไปพัฒนาระบบจริง
- 6) ติดตั้งระบบอาปาเช่ฮาดูปลงบนเซิร์ฟเวอร์และคอมพิวเตอร์ที่ใช้งานจริง
- 7) พัฒนาระบบคลังข้อมูลคณะเทคโนโลยีสารสนเทศบนระบบอาปาเช่ฮาดูป
- 8) จัดทำรายงานและแดชบอร์ด เพื่อแสดงผลการวิเคราะห์ข้อมูลในระบบคลังข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 9) ทดลองระบบ เพื่อหาข้อบกพร่องของระบบ แล้วนำไปปรับปรุงแก้ไขระบบให้มีความเสถียรมากขึ้น
- 10) นำระบบที่สมบูรณ์แล้วไปใช้งานจริงกับคณะเทคโนโลยีสารสนเทศของสถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหาร ลาดกระบัง

## 1.5 ประโยชน์ที่คาดว่าจะได้รับ

- 1) คณะเทคโนโลยีสารสนเทศ จะมีระบบสารสนเทศคลังข้อมูล สำหรับการวิเคราะห์ข้อมูล นักศึกษาและข้อมูลบุคลากรภายในคณะ บนระบบอาปาเซฮาดูปที่มีประสิทธิภาพสูง แต่มีค่าใช้จ่ายที่ค่อนข้างต่ำ
- 2) คณะผู้จัดทำโครงการได้ศึกษาทำความเข้าใจ เกี่ยวกับหลักการทำงาน และทฤษฎีเบื้องต้นของการทำคลังข้อมูล (Data Warehouse)
- 3) คณะผู้จัดทำโครงการได้ศึกษาทำความเข้าใจ เกี่ยวกับหลักการทำงาน และทฤษฎีเบื้องต้นของระบบ อาปาเซฮาดูป (Apache Hadoop)
- 4) คณะผู้จัดทำโครงการได้ศึกษาทำความเข้าใจเกี่ยวกับหลักการทำงานและคำสั่งของภาษาอิมพาลา (Impala)

## 1.6 ส่วนประกอบของปริญญานิพนธ์

ปริญญานิพนธ์ฉบับนี้จะประกอบไปด้วยองค์ประกอบทั้งหมด 5 บทด้วยกัน ดังนี้

บทที่ 1 บทนำ ในส่วนของบทนำจะกล่าวถึงความเป็นมาและความสำคัญของโครงการ ความมุ่งหมายและวัตถุประสงค์ของโครงการ ขอบเขตของโครงการ ขั้นตอนและวิธีการดำเนินงานของโครงการ ประโยชน์ที่คาดว่าจะได้รับของโครงการ และส่วนประกอบของปริญญานิพนธ์นี้

บทที่ 2 การทบทวนทฤษฎีพื้นฐาน งานวิจัยและเทคโนโลยีที่เกี่ยวข้อง ในบทนี้จะกล่าวถึงทฤษฎีพื้นฐาน งานวิจัยต่างๆ และเทคโนโลยีที่ใช้ในการทำโครงการที่เกี่ยวข้องกับทฤษฎีและงานวิจัยที่เกี่ยวข้องกับระบบคลังข้อมูล ไม่ว่าจะเป็นคำจำกัดความของคลังข้อมูล คุณลักษณะเฉพาะของคลังข้อมูล สถาปัตยกรรมของคลังข้อมูล และขั้นตอนการสร้างคลังข้อมูล รวมไปถึงเทคโนโลยีที่เกี่ยวข้องกับโครงการ ไม่ว่าจะเป็นเทคโนโลยีอาปาเซฮาดูป ที่จะกล่าวถึงคำจำกัดความของอาปาเซฮาดูป หลักการการทำงานของอาปาเซฮาดูป และระบบนิเวศเชื่อมโยงของอาปาเซฮาดูป รวมไปถึงเทคโนโลยีที่ใช้ในการสร้างเว็บแอปพลิเคชัน ไม่ว่าจะเป็น JSP, Java และ JavaScript เป็นต้น

บทที่ 3 เนื้อหาของโครงการ ในส่วนของเนื้อหาจะกล่าวถึงวิธีการดำเนินงานของโครงการ ซึ่งจะเกี่ยวกับการออกแบบและการพัฒนาระบบคลังข้อมูลของคณะเทคโนโลยีสารสนเทศบนอาปาเซฮาดูป และการจัดทำแดชบอร์ด (Dashboard) และรายงาน (Report) เพื่อแสดงผลการวิเคราะห์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข้อมูลในระบบคลังข้อมูล ไม่ว่าจะเป็นระบบการทำงานในปัจจุบัน ขั้นตอนการดำเนินงาน การออกแบบคลังข้อมูล การออกแบบยูสเคส ไดอะแกรม แผนภาพกิจกรรม แผนภาพแสดงคลาส แผนภาพแสดงลำดับการทำงาน และการออกแบบส่วนติดต่อกับผู้ใช้ เป็นต้น

บทที่ 4 ผลการดำเนินงานของ โครงการงาน ในส่วนของผลการดำเนินงานของโครงการงานจะกล่าวถึงตัวโครงการงานที่ถูกพัฒนาขึ้นจริง ไม่ว่าจะเป็นเว็บแอปพลิเคชัน ส่วนควบคุมการทำงานของระบบ และส่วนการประมวลผลข้อมูล เป็นต้น

บทที่ 5 บทสรุปของโครงการงาน ในส่วนของบทสรุปของโครงการงานจะกล่าวถึงการสรุปผลของโครงการงาน ปัญหาและอุปสรรคที่เกิดขึ้นของโครงการงาน และแนวทางการพัฒนาต่อของโครงการงานในอนาคต



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## การทบทวนทฤษฎีพื้นฐาน งานวิจัย และเทคโนโลยีที่เกี่ยวข้อง

เนื่องมาจากในปัจจุบันข้อมูลของคณะเทคโนโลยีสารสนเทศแห่งนี้มีขนาดใหญ่เพิ่มขึ้นเรื่อย ๆ ทุกปี ซึ่งทางงานทะเบียนนักศึกษาของคณะได้ประสบกับปัญหาในการจัดการกับข้อมูลเหล่านี้ เวลาจะค้นหาข้อมูลก็ยากต่อการค้นหาและใช้เวลานานในการประมวลผล จึงทำให้เกิดโครงการนี้ขึ้น ซึ่งในบทนี้จะกล่าวถึงทฤษฎี งานวิจัย และเทคโนโลยีที่ใช้ในการอ้างอิงเพื่อใช้สำหรับการทำโครงการฉบับนี้ ทั้งในส่วนของทฤษฎีและหลักการของการออกแบบและการพัฒนาระบบคลังข้อมูล รวมไปถึงทฤษฎีและหลักการการทำงานของระบบอาปาเซฮาคุป และเทคโนโลยีที่ใช้ในการสร้างเว็บแอปพลิเคชัน โครงการฉบับนี้จึงน่าจะเป็นประโยชน์ต่อการนำเสนอข้อมูล

### 2.1 ทฤษฎีและงานวิจัยที่เกี่ยวข้องกับระบบคลังข้อมูล

โครงการนี้จะมีทฤษฎีพื้นฐาน วรรณกรรม งานวิจัย และเทคโนโลยีที่เกี่ยวข้องที่ใช้ในการดำเนินโครงการดังนี้

#### 2.1.1 คำจำกัดความของคลังข้อมูล



รูปที่ 2.1 แสดงหลักการการทำงานของคลังข้อมูล

คลังข้อมูล (Data Warehouse) [1] คือ ฐานข้อมูลขนาดใหญ่ที่ถูกออกแบบขึ้นมาเพื่อใช้เก็บรวบรวมข้อมูลรายการการเปลี่ยนแปลงประจำวัน (Transaction) ขององค์กรหรือหน่วยงานต่าง ๆ ตั้งแต่อดีตจนถึงปัจจุบัน โดยไม่มีการลบข้อมูลหรืออัปเดตข้อมูลถึงแม้ว่าข้อมูลนั้นจะมีการเปลี่ยนแปลงหรือไม่ตรงกับข้อมูลในปัจจุบันก็ตาม เพื่อที่สามารถเรียกดูข้อมูลย้อนหลังได้ เช่น นายชายชาญได้ทำการเปลี่ยนชื่อเป็นนายชาญชาญ โดยระบบคลังข้อมูลจะต้องทราบได้ว่า นายชายชาญกับนายชาญชาญ คือบุคคลคนเดียวกัน เป็นต้น โดยคลังข้อมูลมีวัตถุประสงค์เพื่อใช้ในการเรียกดูข้อมูล (Query) และสามารถนำมาใช้ในการวิเคราะห์ข้อมูลเพื่อใช้ประโยชน์ได้ทั้งในรูปแบบรายงาน (Report) หรือรูปแบบข้อมูลภาพรวม (Dashboard) โดยแสดงผลลัพธ์เป็นกราฟหรือแผนภูมิ เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปภาพ เพื่อใช้ช่วยสนับสนุนการตัดสินใจภายในองค์กรของผู้บริหารได้ หรือทำเหมืองข้อมูล (Data Mining) เพื่อหาความสัมพันธ์ของข้อมูล และยังเป็นศูนย์รวมของข้อมูลของทั้งองค์กร ที่พร้อมต่อการใช้งาน เข้าถึงข้อมูลได้ง่าย มีความปลอดภัย ข้อมูลถูกต้อง ครบถ้วน สอดคล้องไม่ขัดแย้งกัน ซึ่งจะเข้าใจถึงหลักการการทำงานของคลังข้อมูลเบื้องต้นได้จากรูปที่ 2.1

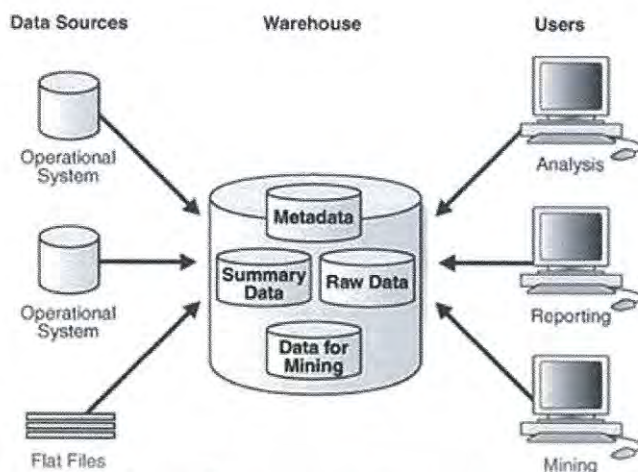
### 2.1.2 คุณลักษณะเฉพาะของคลังข้อมูล

- 1) การแบ่งโครงสร้างตามเนื้อหาที่สนใจ (Subject Oriented) [1] คือ คลังข้อมูลจะถูกออกแบบมาเพื่อช่วยในการวิเคราะห์ข้อมูลที่ทางองค์กรกำลังสนใจ ตัวอย่างเช่น ข้อมูลนักศึกษา ข้อมูลอาจารย์ และข้อมูลรายวิชาเรียน เป็นต้น โดยสามารถใช้คลังข้อมูลในการตอบคำถามเหล่านี้ เช่น จำนวนนักศึกษาที่ไม่จบการศึกษาในปีที่ผ่านมา หรือ อาจจะนำคลังข้อมูลใช้ประกอบการวิเคราะห์หาแนวโน้มจำนวนนักศึกษาที่จะไม่จบการศึกษาในปีถัดไป เป็นต้น
- 2) การจัดเก็บรวบรวมข้อมูล (Integrated) [1] คือ การจัดเก็บรวบรวมข้อมูลเข้าด้วยกันเป็นกลุ่มของข้อมูลที่อยู่ในรูปแบบที่ตรงกันหรือมีความสอดคล้องกันจากหลายๆ แหล่งข้อมูล โดยจะต้องทำการแก้ไขปัญหา เช่น การตั้งชื่อที่ขัดแย้งกัน และความไม่สอดคล้องของหน่วยของการวัด เป็นต้น เมื่อแก้ปัญหาเหล่านี้สำเร็จ ก็จะสามารถทำการจัดเก็บรวบรวมข้อมูลเข้าด้วยกันเป็นกลุ่มของข้อมูลได้
- 3) ความสัมพันธ์กับเวลา (Time Variant) [2] คือ การจัดเก็บรวบรวมข้อมูลในอดีตไว้ ณ คลังข้อมูล ตัวอย่างเช่น เราสามารถทำการเรียกดูข้อมูล 3 เดือน 6 เดือน 12 เดือน หรือข้อมูลเก่าๆ จากคลังข้อมูลได้ ซึ่งจะแตกต่างจากระบบการทำธุรกรรมที่จะทำการจัดเก็บรวบรวมข้อมูลเฉพาะข้อมูลล่าสุดเท่านั้น คลังข้อมูลจะทำการจัดเก็บรวบรวมข้อมูลจากอดีตจนถึงปัจจุบัน เพื่อสามารถนำข้อมูลในอดีตมาใช้ในการวิเคราะห์หาแนวโน้ม เปรียบเทียบ หรือพยากรณ์ข้อมูลขององค์กรได้
- 4) ข้อมูลไม่ควรได้รับการแก้ไข (Nonvolatile) [1] คือ ข้อมูลที่อยู่ในคลังข้อมูลจะไม่สามารถแก้ไขหรือเปลี่ยนแปลงค่าของข้อมูลได้ ผู้ใช้จะสามารถทำได้เพียงเรียกดูข้อมูลเพื่อมาวิเคราะห์ได้เท่านั้น

### 2.1.3 สถาปัตยกรรมของคลังข้อมูล

- 1) สถาปัตยกรรมของคลังข้อมูล โดยปัจจุบันจะสามารถแบ่งสถาปัตยกรรมของคลังข้อมูลออกเป็น 3 แบบ ดังนี้
  - a) สถาปัตยกรรมคลังข้อมูลแบบพื้นฐาน (Data Warehouse Architecture Basic)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

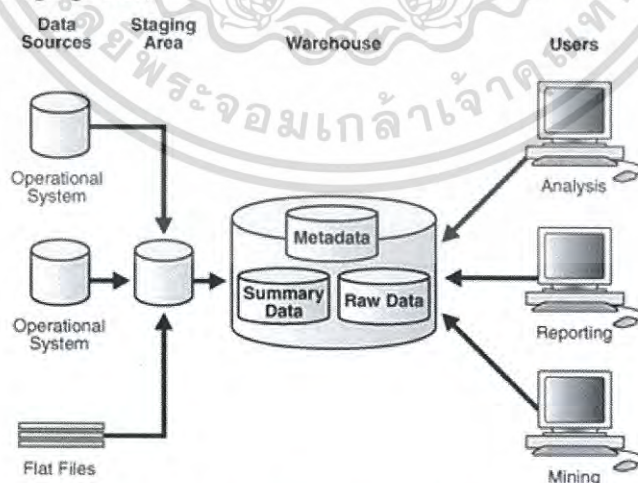


รูปที่ 2.2 แสดงโครงสร้างของสถาปัตยกรรมคลังข้อมูลแบบพื้นฐาน

ที่มา: [http://docs.oracle.com/cd/B28359\\_01/server.111/b28313/dwhsg013.gif](http://docs.oracle.com/cd/B28359_01/server.111/b28313/dwhsg013.gif)

จากรูปที่ 2.2 [1] จะแสดงให้เห็นถึงสถาปัตยกรรมอย่างง่ายของคลังข้อมูล โดยที่ผู้ใช้จริง (End Users) สามารถเข้าถึงข้อมูลจากแหล่งต่าง ๆ โดยตรงผ่านคลังข้อมูล ซึ่งจะเก็บเมตาดาต้า (Metadata) และข้อมูลดิบ (Raw data) ของระบบการประมวลผลธุรกรรมออนไลน์เดิม (Online Transaction Processing : OLTP) โดยจะทำการเพิ่มชนิดของข้อมูล (Type of data) และจะสรุปผลของข้อมูล (Summary data) ผลสรุปของข้อมูลนี้จะมีค่ามากและถูกจัดเก็บไว้ในคลังข้อมูล เนื่องจากระบบคลังข้อมูลเป็นระบบที่สามารถดำเนินการการคำนวณผลข้อมูลแบบล่วงหน้าเป็นระยะเวลานานได้

b) สถาปัตยกรรมคลังข้อมูลแบบมีจุดพักชั่วคราวสำหรับตรวจสอบข้อมูล (Data Warehouse Architecture with a Staging Area)



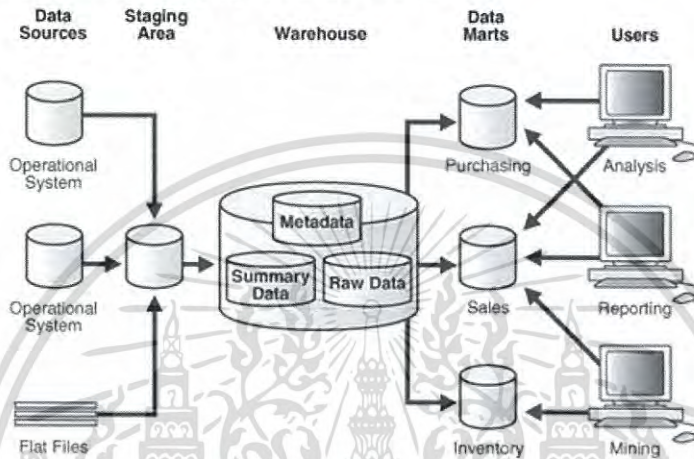
รูปที่ 2.3 แสดงโครงสร้างของสถาปัตยกรรมคลังข้อมูลแบบมีจุดตรวจสอบข้อมูล

ที่มา: [https://docs.oracle.com/cd/E11882\\_01/server.112/e25554/dwhsg015.gif](https://docs.oracle.com/cd/E11882_01/server.112/e25554/dwhsg015.gif)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 2.3 [1] จะแสดงให้เห็นถึงสถาปัตยกรรมของคลังข้อมูลแบบมีจุดพักชั่วคราว สำหรับตรวจสอบข้อมูล (Staging Area) โดยก่อนที่จะใส่ข้อมูลลงไปในคลังข้อมูลจะต้องทำการจัดกรองและประมวลผลข้อมูลที่จุดพักชั่วคราวสำหรับตรวจสอบข้อมูลก่อน โดยจุดพักชั่วคราว สำหรับตรวจสอบข้อมูลนี้จะช่วยลดความยุ่งยากในการสรุปผลและจัดการคลังข้อมูลแบบทั่วไป

c) สถาปัตยกรรมคลังข้อมูลแบบมีจุดพักชั่วคราวสำหรับตรวจสอบข้อมูล และตลาดข้อมูล (Data Warehouse Architecture with a Staging Area and Data Marts)



รูปที่ 2.4 แสดงโครงสร้างของสถาปัตยกรรมคลังข้อมูลแบบมีจุดตรวจสอบข้อมูลและตลาดข้อมูล  
ที่มา: [https://docs.oracle.com/cd/E11882\\_01/server.112/e25554/dwhsg064.gif](https://docs.oracle.com/cd/E11882_01/server.112/e25554/dwhsg064.gif)

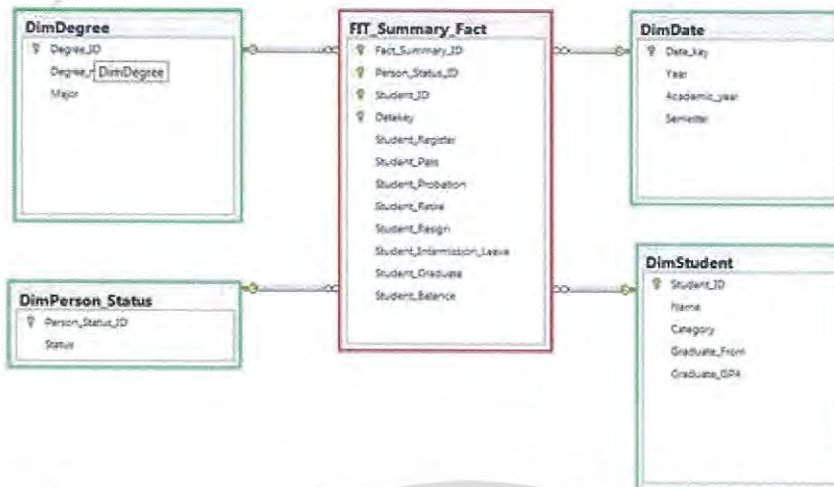
จากรูปที่ 2.4 [1] จะแสดงให้เห็นถึงสถาปัตยกรรมของคลังข้อมูลแบบมีจุดพักชั่วคราว สำหรับตรวจสอบข้อมูล โดยจะทำการเพิ่มตลาดข้อมูล (Data Mart) เข้าไป โดยสถาปัตยกรรมนี้ถูก ออกแบบมาเพื่อใช้ในสายธุรกิจโดยเฉพาะ แสดงให้เห็นถึงตัวอย่างของข้อมูลการจัดซื้อ การขาย และสินค้าคงเหลือที่จะถูกแยกออก ตัวอย่างเช่น นักวิเคราะห์ทางการเงินต้องการที่วิเคราะห์ข้อมูล การซื้อขายในอดีตเพื่อใช้พยากรณ์แนวโน้มเกี่ยวกับพฤติกรรมของลูกค้า

#### 2.1.4 ขั้นตอนการสร้างคลังข้อมูล

1) การออกแบบคลังข้อมูล คือ การออกแบบการเชื่อมโยงตารางภายในคลังข้อมูล ซึ่งจะมีลักษณะ ดังต่อไปนี้

a) การออกแบบคลังข้อมูลจะออกแบบโดยใช้แบบจำลองหลายมิติ (Multidimensional Modeling) ซึ่งจะประกอบไปด้วยตาราง 2 ประเภท คือ ตารางข้อเท็จจริง (Fact Table) และตารางมิติ (Dimension Table)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.5 แสดงแบบจำลองข้อมูลเชิงมิติแบบดาวที่แยกประเภทของตารางด้วยสี่ โดยมีลักษณะดังนี้

- ตารางข้อเท็จจริง (Fact Table) คือ ตารางหลักหรือตารางศูนย์กลางของแบบจำลองหลายมิติที่จะเก็บข้อมูลข้อเท็จจริง (Fact) ของคลังข้อมูลไว้ และยังเก็บคีย์ (Key) ทั้งคีย์หลัก (Primary Keys) และคีย์นอก (Foreign Keys) ที่เป็นค่าคีย์หลักของตารางมิติที่เชื่อมโยงกับตารางข้อเท็จจริง โดยจะจับคู่กันเป็นคีย์คู่แข่ง (Candidate Keys) เพื่อมาใช้เป็นคีย์หลักของตารางข้อเท็จจริง ซึ่งข้อมูลข้อเท็จจริงจะเป็นชุดของค่าที่จับคู่กันระหว่างจำนวนคอลัมน์ที่เป็นข้อมูลชนิดตัวเลข (Measure) ในตารางข้อเท็จจริงกับข้อมูลที่เป็นมุมมองในการวิเคราะห์ข้อมูล (Dimension) ในตารางมิติ เพื่อให้เกิดข้อเท็จจริงที่สามารถเข้าใจและวิเคราะห์ข้อมูลได้ เวลาจะใช้ตารางข้อเท็จจริงจะทำการรวมตารางข้อมูล (Join) ระหว่างจำนวนคอลัมน์เป็นข้อมูลชนิดตัวเลขของตารางข้อเท็จจริงเข้ากับข้อมูลที่เป็นมุมมองในการวิเคราะห์ข้อมูลของตารางมิติ โดยจำนวนคอลัมน์ที่เป็นข้อมูลชนิดตัวเลขจะใช้เพื่อวัดทั้งในเชิงปริมาณ (Quantitative) และเชิงคุณภาพ (Qualitative) ของข้อมูล ทำหน้าที่เก็บข้อมูลที่เกิดขึ้นในแต่ละรายการการเปลี่ยนแปลง รวมทั้งยังแสดงให้เห็นถึงค่าของข้อมูลที่ผ่านมาการวิเคราะห์แล้ว เช่น เกรดเฉลี่ย และจำนวนหน่วยกิต เป็นต้น ส่วนข้อมูลที่เป็นมุมมองในการวิเคราะห์ข้อมูล จะเป็นข้อมูลที่ใช้ในการอธิบายเพื่อให้เข้าใจในการวิเคราะห์ข้อมูล เช่น ชื่อนักศึกษาและนามสกุลนักศึกษา เป็นต้น ซึ่งจากรูปที่ 2.5 ตารางข้อเท็จจริงจะแสดงอยู่ในกรอบสีแดง
- ตารางมิติ (Dimension Table) คือ ตารางหลาย ๆ ตารางที่เชื่อมโยงกับตารางข้อเท็จจริง โดยจะมีข้อมูลทั้งที่เป็นข้อมูลที่เป็นมุมมองในการวิเคราะห์ข้อมูลและจำนวนคอลัมน์ที่เป็นข้อมูลชนิดตัวเลขเป็นส่วนประกอบ เพื่อใช้อธิบายรายละเอียดของข้อเท็จจริงที่อยู่ในตารางข้อเท็จจริง เพราะข้อมูลส่วนใหญ่ในตารางข้อเท็จจริงจะเป็นข้อมูลชนิด

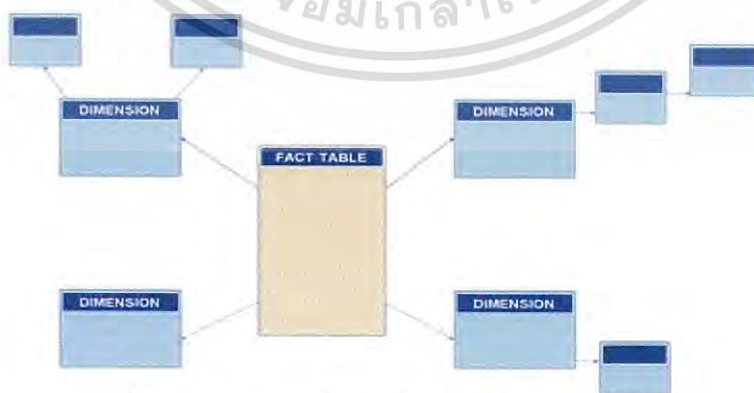
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวเลข ทำให้ไม่เข้าใจความหมายของข้อมูลเหล่านั้น จึงต้องมีข้อมูลที่เป็นมุมมองในการวิเคราะห์ข้อมูลเข้ามาช่วย เพื่อให้สามารถเข้าใจและวิเคราะห์ข้อมูลได้ โดยจะมีคีย์มิติ (Dimension Keys) ซึ่งจะเป็นคีย์หลักของตารางมิติไปเป็นคีย์นอกในตารางข้อเท็จจริงหรือตารางมิติที่เชื่อมโยงกับตารางมิตินั้นๆ ตารางมิติที่สำคัญที่สุดคือ ตารางเวลา (Time Dimension Table) ซึ่งจากรูปที่ 2.5 ตารางมิติจะอยู่ในกรอบสี่เหลี่ยม

b) การออกแบบคลังข้อมูลยังสามารถออกแบบเป็นแบบจำลองข้อมูลหลายมิติตามคุณลักษณะได้ 2 แบบ คือ แบบจำลองข้อมูลเชิงมิติแบบดาว (Star Schema Model) และแบบจำลองข้อมูลเชิงมิติแบบเกล็ดหิมะ (Snowflake Schema Model) โดยมีลักษณะดังนี้

i) แบบจำลองข้อมูลเชิงมิติแบบดาว (Star Schema Model) คือ เป็นแบบจำลองข้อมูลเชิงมิติที่ประกอบไปด้วยตารางมิติหลาย ๆ ตาราง ล้อมรอบตารางข้อเท็จจริงที่อยู่ตรงกลาง มีลักษณะและความสัมพันธ์ที่คล้ายกับดวงดาว มีข้อดีคือ สามารถวิเคราะห์และเรียกดูข้อมูลได้อย่างมีประสิทธิภาพสูง ได้จำนวนมาก ได้ง่าย และรวดเร็ว เนื่องจากมีจำนวนตารางมิติไม่มากนัก โดยแบบจำลองนี้จะเหมาะกับงานที่ค่อนข้างไม่มีความซับซ้อน ซึ่งจะแสดงในรูปที่ 2.5

ii) แบบจำลองข้อมูลเชิงมิติแบบเกล็ดหิมะ (Snowflake Schema Model) จะมีลักษณะคล้ายกับแบบจำลองข้อมูลเชิงมิติแบบดาว คือ มีตารางข้อเท็จจริงเป็นศูนย์กลางและเชื่อมโยงกับตารางมิติหลายๆ ตาราง โดยตารางมิติที่เชื่อมกับตารางข้อเท็จจริงอยู่นี้จะสามารถแตกเป็นตารางย่อยๆ ลงไปได้อีกเป็นผลมาจากการทำออร์มัลไลเซชัน (Normalization) จะมีลักษณะคล้ายกับเกล็ดหิมะ ซึ่งตารางแบบนี้จะเหมาะกับคลังข้อมูลขนาดใหญ่ที่ค่อนข้างมีความซับซ้อน มีข้อดีคือ เก็บข้อมูลได้น้อยลง ไม่ซ้ำซ้อน แต่มีข้อเสียคือ ค่อนข้างซับซ้อน ทำความเข้าใจได้ยาก เวลาเรียกดูข้อมูลจะใช้เวลาค่อนข้างนาน เพราะจะต้องทำการรวมตารางหลายตารางที่จะเรียกดูเข้าด้วยกัน ซึ่งแบบจำลองข้อมูลเชิงมิติแบบเกล็ดหิมะจะแสดงในรูปที่ 2.6



รูปที่ 2.6 แสดงแบบจำลองข้อมูลเชิงมิติแบบเกล็ดหิมะ

ที่มา: <https://upload.wikimedia.org/wikipedia/commons/thumb/b/b2/Snowflake-schema.png/350px-Snowflake-schema.png>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2) กระบวนการสกัดข้อมูล แปลงข้อมูล และ โหลดข้อมูลลงคลังข้อมูล (ETL)

กระบวนการอีทีแอล (ETL) [3] ย่อมาจากจาก สกัดข้อมูล (Extract) แปลงข้อมูล (Transform) และ โหลดข้อมูล (Load) กระบวนการทั้งสามนี้เป็นขั้นตอนในการใช้งานฐานข้อมูลและคลังข้อมูล ซึ่งจะช่วยในการรวบรวมและวิเคราะห์ข้อมูลที่ถูกจัดเก็บไว้ในฐานข้อมูลที่แตกต่างกันและรูปแบบที่แตกต่างกัน โดยข้อมูลเหล่านี้ได้มาจากการจัดเก็บรวบรวมข้อมูลจากหลายๆแหล่งข้อมูล เราจะเรียกกระบวนการนี้ว่า การสกัดข้อมูล (Extraction) เมื่อจัดเก็บรวบรวมข้อมูลแล้ว ข้อมูลจะถูกคัดกรอง ตรวจสอบแก้ไขให้ถูกต้อง และจัดรูปแบบใหม่ให้เป็นไปตามความต้องการของการดำเนินงาน เราจะเรียกกระบวนการนี้ว่า การแปลงข้อมูล (Transformation) และขั้นตอนสุดท้ายคือ ข้อมูลที่ได้จะถูกโหลดเข้าสู่ฐานข้อมูล คลังข้อมูลหรือตลาดข้อมูล (Data Mart) เป้าหมายเพื่อนำข้อมูลที่ได้ไปใช้ในการวิเคราะห์ข้อมูล เราจะเรียกกระบวนการนี้ว่า การโหลดข้อมูล (Load) โดยส่วนใหญ่เมื่อเราสกัดข้อมูล แปลงข้อมูล และ โหลดข้อมูลลงเป้าหมายก็จะถือเป็นการสิ้นสุดการทำงาน แต่ยกเว้นคลังข้อมูลและระบบธุรกิจอัจฉริยะ (Business Intelligence) ที่ ETL ยังสามารถใช้ในการย้ายข้อมูลจากระบบการดำเนินงานหนึ่งไปยังอีกระบบหนึ่งได้ ซึ่งกระบวนการอีทีแอลจะแสดงในรูปที่ 2.7



รูปที่ 2.7 แสดงกระบวนการ ETL

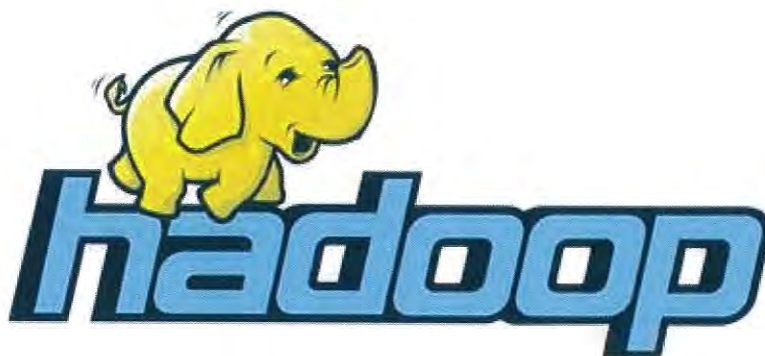
ที่มา: <http://blog.appliedinformaticsinc.com/wp-content/uploads/2015/05/ETL-Process.png>

## 2.2 เทคโนโลยีที่เกี่ยวข้องกับโครงการ

### 2.2.1 เทคโนโลยีอาปาเซฮาดูป

#### 2.2.1.1 คำจำกัดความของอาปาเซฮาดูป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.8 แสดงโลโก้ของอาปาเซฮาดูป

ที่มา: [http://www.lvsvn.com/sites/default/files/styles/post\\_image/public/Untitled-1.png?itok=Tw\\_0WAwE](http://www.lvsvn.com/sites/default/files/styles/post_image/public/Untitled-1.png?itok=Tw_0WAwE)

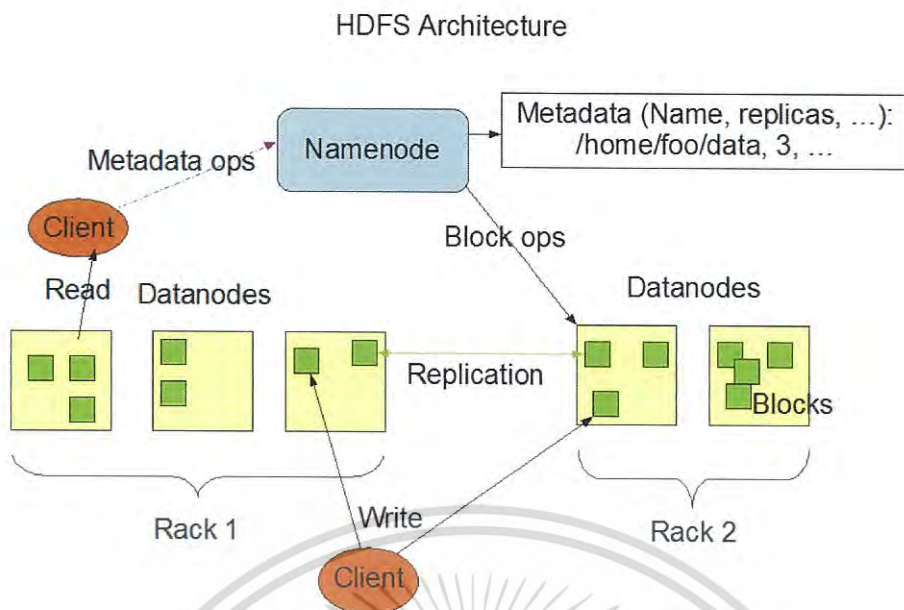
อาปาเซฮาดูป (Apache hadoop) เป็นซอฟต์แวร์แบบโอเพนซอร์ส (Open source software) ของอาปาเซ มีความน่าเชื่อถือและมีความยืดหยุ่นสามารถปรับขนาดได้ตามข้อมูลที่มีการเปลี่ยนแปลงอยู่ตลอดเวลา เหมาะสำหรับการประมวลผลข้อมูลแบบกระจาย (Distributed Computing) รองรับการจัดเก็บและประมวลผลข้อมูลขนาดใหญ่ด้วยการเชื่อมต่อกับเครื่องเซิร์ฟเวอร์จำนวนมากเข้าด้วยกัน ซึ่งอาปาเซฮาดูปนั้นจะถูกออกแบบมาให้สามารถเชื่อมต่อเซิร์ฟเวอร์ได้ตั้งแต่หนึ่งเครื่องจนถึงพันเครื่อง โดยการส่งข้อมูลจะสามารถทำงานได้ถึงแม้ข้อมูลในระบบจะมีการทำงานล้มเหลว เนื่องจากข้อมูลจะถูกทำการสำรองไว้ในเซิร์ฟเวอร์อย่างน้อย 3 เซิร์ฟเวอร์ เมื่อมีข้อมูลใดข้อมูลหนึ่งเกิดล้มเหลวขึ้นก็จะทำการไปเรียกข้อมูลสำรองจากเซิร์ฟเวอร์อื่นมาใช้งานได้อย่างทันที เลยทำให้ระบบทำงานได้โดยข้อมูลไม่มีการสูญหายเกิดขึ้น ซึ่งอาปาเซฮาดูปจะมีส่วนในการทำงานอยู่ 2 ส่วนหลัก ๆ คือ ส่วนของระบบจัดการไฟล์แบบกระจายของอาปาเซฮาดูป (Hadoop Distributed File System : HDFS) และส่วนของการประมวลผลข้อมูลด้วยแมปรีดิวซ์ (Map Reduce) ซึ่งจะกล่าวถึงในหัวข้อ 2.2.3 ต่อไป

### 2.2.1.2 หลักการทำงานของอาปาเซฮาดูป

อาปาเซฮาดูปจะมีส่วนในการทำงานอยู่ 2 ส่วนหลัก ๆ คือ ส่วนของระบบจัดการไฟล์แบบกระจายของอาปาเซฮาดูป (Hadoop Distributed File System : HDFS) และส่วนของการประมวลผลข้อมูลด้วยแมปรีดิวซ์ (Map Reduce) ซึ่งจะมีลักษณะดังนี้

1. ระบบจัดการไฟล์แบบกระจายของอาปาเซฮาดูป (Hadoop Distributed File System: HDFS) คือ ระบบที่ถูกออกแบบมาสำหรับใช้จัดเก็บข้อมูลขนาดใหญ่ โดยจะทำงานบนระบบประมวลผลคลัสเตอร์ (Cluster Computing System) ทั่วไป โดยมีโหนดการทำงานอยู่ 2 แบบด้วยกัน คือ Name node กับ Data node ดังภาพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.9 แสดงหลักการการทำงานของ HDFS

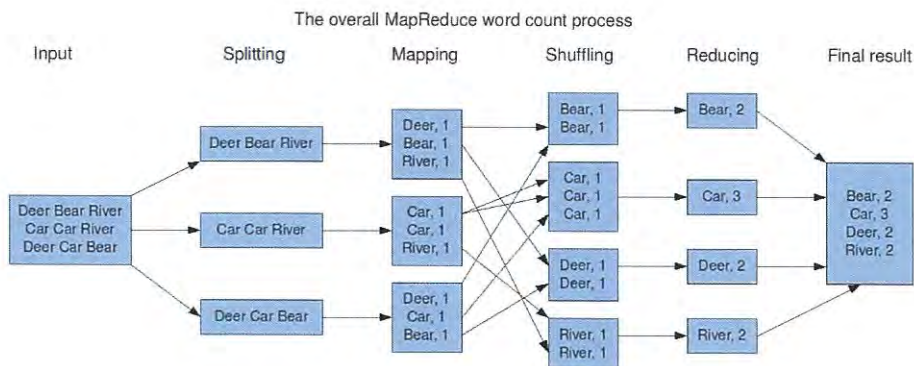
ที่มา: <https://hadoop.apache.org/docs/r1.2.1/images/hdfsarchitecture.gif>

โดยมีลักษณะดังนี้

a. Name node คือ โหนดศูนย์กลางการควบคุมของ HDFS จะคอยเก็บที่อยู่ของแต่ละไฟล์ที่ทำการแบ่ง ไฟล์ออกเป็นส่วนๆ และทำสำเนากระจายไปยัง Data node ต่างๆ เมื่อเครื่องไคลเอนต์ต้องการไฟล์ข้อมูล Name node ก็จะไปเรียกไฟล์ข้อมูลจากที่อยู่ของ Data node ที่เก็บกลุ่มไฟล์นั้นไว้

b. Data node คือ โหนดที่เก็บไฟล์แต่ละไฟล์ทั้งที่แบ่งและที่ทำสำเนาไว้ คอยบอกสถานะของไฟล์ว่ายังทำงานปกติหรือไม่แก่ Name node ถ้าหาก Name node ไม่ได้รับสัญญาณตอบกลับจาก Data node ของไฟล์นั้นก็จะตัดสินใจว่า Data node นั้นไม่สามารถทำงานได้อีก จึงทำการไปสำเนาไฟล์จากไฟล์กลุ่มเดียวกันใน Data node อื่น ไปไว้ใน Data node อีกอันที่ไม่มีกลุ่มไฟล์นั้นแทน

2. การประมวลผลข้อมูลด้วยแมปรีดิวซ์ (Map Reduce) คือ แบบจำลองการเขียนโปรแกรมสำหรับการประมวลผลข้อมูล ซึ่งเป็นแบบจำลองอย่างง่าย มักพบในภาษา Ruby, Python, Java และ C++ จะประกอบไปด้วย 2 ส่วน คือ ส่วนของแมป (Map) และส่วนของรีดิวซ์ (Reduce)



รูปที่ 2.10 แสดงหลักการการทำงานของ Map Reduce

ที่มา: [http://xiaochongzhang.me/blog/wp-](http://xiaochongzhang.me/blog/wp-content/uploads/2013/05/MapReduce_Work_Structure.png)

[content/uploads/2013/05/MapReduce\\_Work\\_Structure.png](http://xiaochongzhang.me/blog/wp-content/uploads/2013/05/MapReduce_Work_Structure.png)

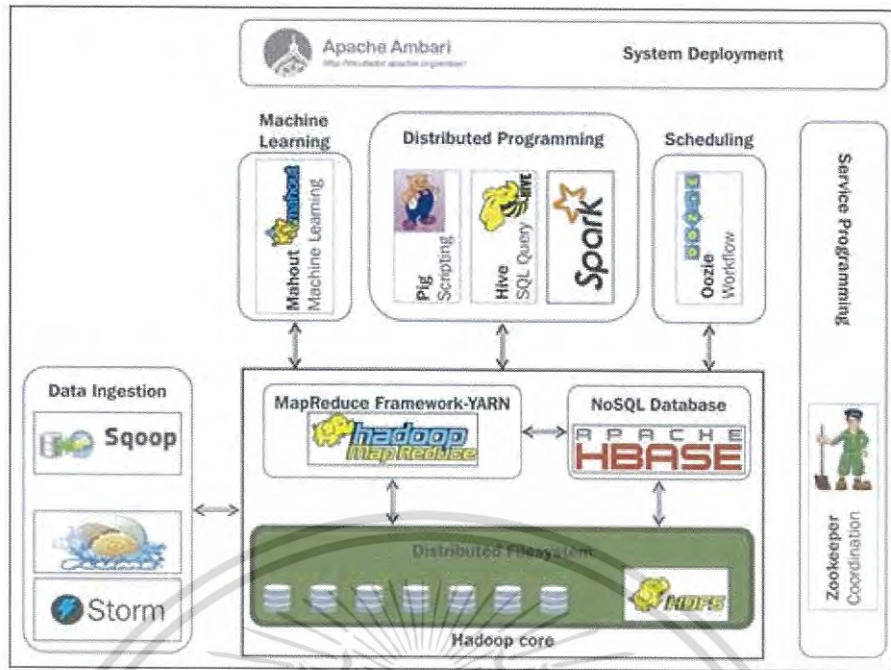
โดยมีลักษณะดังนี้

แมป (Map) คือ การที่ข้อมูลเข้าสู่เครื่องจะทำการแบ่งข้อมูลแล้วทำการกระจายไปยังโหนดต่าง ๆ โดยแต่ละโหนดจะทำการแมปข้อมูลที่เหมือนกันเข้าด้วยกันเป็น Key/Value เมื่อข้อมูลผ่านการแมปแล้วก็จะทำการ Shuffling ข้อมูล โดยนำข้อมูลที่เหมือนกันในแต่ละโหนดมาไว้ด้วยกัน แล้วทำการลดจำนวน Key/Value ลง เรียกว่า การรีดิวซ์ (Reduce)

### 2.2.1.3 ระบบนิเวศเชื่อมโยงของอาปาเซฮาดูป (Apache Hadoop Ecosystem)

อาปาเซฮาดูปกำลังเป็นสิ่งที่องค์กรทั่วโลกกำลังสนใจและจะนำเข้ามาใช้ในองค์กรของตน เนื่องมาจากอาปาเซฮาดูปมีค่าใช้จ่ายที่ต่ำ แต่กลับมีประสิทธิภาพที่ค่อนข้างสูงในการประมวลผลข้อมูลขนาดใหญ่ และก็มีการพัฒนาต่อยอดอย่างต่อเนื่องให้เข้าถึงผู้ใช้งานได้มากขึ้น ซึ่งอาปาเซฮาดูปจะมีแอปพลิเคชันที่เข้าร่วมกันอยู่หลายตัว ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.11 แสดงแอปพลิเคชันในระบบนิเวศเชื่อมโยงของอาปาเซฮาดูป

ที่มา: [https://www.safaribooksonline.com/library/view/hadoop-essentials/9781784396688/graphics/3765\\_02\\_06.jpg](https://www.safaribooksonline.com/library/view/hadoop-essentials/9781784396688/graphics/3765_02_06.jpg)

### 1.อาปาเซอัมบาร์

รูปที่ 2.12 แสดงโลโก้ของอาปาเซอัมบาร์

ที่มา: <http://bigdateref.github.io/icons/ambari.png>

อาปาเซอัมบาร์ (Apache Ambari) [4] เป็นซอฟต์แวร์โอเพนซอร์สที่ถูกพัฒนาขึ้นเพื่อช่วยให้ผู้ดูแลระบบจัดหา จัดการ และตรวจสอบกลุ่มฮาดูป (Hadoop Cluster) โดยในการจัดหาจะไม่ให้ความสำคัญในเรื่องขนาดของกลุ่มฮาดูป แต่จะเน้นที่การติดตั้ง การใช้งาน และการดูแลรักษาที่ง่าย ในเรื่องการจัดการจะมีเครื่องมือที่ทำให้การจัดการและการตั้งค่ากับกลุ่มฮาดูปง่ายขึ้น ส่วนการตรวจสอบกลุ่มฮาดูปจะมีหน้าต่างที่แสดงสถานะแจ้งเตือนเมื่อเกิดปัญหาขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2. อาปาเช่อะโวว์



รูปที่ 2.13 แสดงโลโก้ของอาปาเช่อะโวว์

ที่มา: <https://avro.apache.org/images/avro-logo.png>

อาปาเช่อะโวว์ (Apache Avro) [5] เป็นเฟรมเวิร์คที่ช่วยจัดการกับข้อมูลต่อเนื่องในรูปแบบที่มีสคีมาในตัว ข้อมูลต่อเนื่องอยู่ในรูปแบบไบนารีขนาดเล็กที่ไม่ต้องใช้พรีออกซ์หรือการสร้างรหัส แทนที่อาโวว์จะใช้การสร้างไลบรารีและความชำนาญในการพิมพ์ อาโวว์จะอาศัยสคีมาที่ถูกส่งไปพร้อมกับข้อมูลต่อเนื่องแทน ซึ่งสคีมาจะช่วยในการแปลงกลับของข้อมูล

## 3. อาปาเซ่แคสแซนดร้า



รูปที่ 2.14 แสดงโลโก้ของอาปาเซ่แคสแซนดร้า

ที่มา: [https://upload.wikimedia.org/wikipedia/commons/thumb/5/5e/Cassandra\\_logo.svg/2000px-](https://upload.wikimedia.org/wikipedia/commons/thumb/5/5e/Cassandra_logo.svg/2000px-Cassandra_logo.svg.png)

[Cassandra\\_logo.svg.png](https://upload.wikimedia.org/wikipedia/commons/thumb/5/5e/Cassandra_logo.svg/2000px-Cassandra_logo.svg.png)

อาปาเซ่แคสแซนดร้า (Apache Cassandra) [6] เป็นระบบที่มีประสิทธิภาพสูง สามารถปรับขนาดได้ สามารถแก้ปัญหาฐานข้อมูลเชิงไม่สัมพันธ์แบบกระจาย แคสแซนดร้าจะรวมข้อดีทั้งหมดของ Google Bigtable และ Amazon Dynamo ในการจัดการประเภทของความต้องการการจัดการฐานข้อมูลที่ถูกจัดจำหน่าย RDBMS แบบดั้งเดิมไม่สามารถสนับสนุนได้ โดยถูกออกแบบมาสำหรับจัดการกับข้อมูลจำนวนมากในเซิร์ฟเวอร์ ซึ่งจะสนับสนุนระบบศูนย์ข้อมูลหลาย ๆ ทอดที่อยู่บนคลัสเตอร์ ด้วยวิธีการไม่ทำสำเนาพร้อม

กัน ทำให้กลุ่มลูกค้ามีการแอบแฝงค่อนข้างต่ำ เขาเท่านั้น ไม่อนุญาตให้เข้าไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 4. อาปาเช่ชุกวา



รูปที่ 2.15 แสดงโลโก้ของอาปาเช่ชุกวา

ที่มา: [https://chukwa.apache.org/images/chukwa\\_logo\\_small.jpg](https://chukwa.apache.org/images/chukwa_logo_small.jpg)

อาปาเช่ชุกวา (Apache Chukwa) [7] เป็นซอฟต์แวร์โอเพนซอร์สที่ถูกออกแบบมาเพื่อใช้แก้ปัญหาการเก็บรวบรวมข้อมูลซึ่งสามารถนำมาใช้สำหรับการตรวจสอบระบบการจัดการฐานข้อมูลแบบกระจายขนาดใหญ่ เหมาะสำหรับการวิเคราะห์และจัดการไฟล์ล็อก

## 5. อาปาเช่ฟลูม



รูปที่ 2.16 แสดงโลโก้ของอาปาเช่ฟลูม

ที่มา: [https://flume.apache.org/\\_static/flume-logo.png](https://flume.apache.org/_static/flume-logo.png)

อาปาเช่ฟลูม (Apache Flume) [8] เป็นซอฟต์แวร์โอเพนซอร์สที่เป็นแบบกระจายนำเชื่อถือ และสามารถใช้งานจริงในการจัดเก็บรวบรวมและเคลื่อนย้ายข้อมูลขนาดใหญ่ได้อย่างมีประสิทธิภาพ อาปาเช่ฟลูมมีสถาปัตยกรรมที่ง่ายและยืดหยุ่นขึ้นอยู่กับกระแสข้อมูลรองรับการเกิดความผิดพลาดของข้อมูลด้วยกลไก Tunable Reliability ซึ่งสามารถนำไปพัฒนาต่อเป็นแอปพลิเคชันการวิเคราะห์ข้อมูลแบบออนไลน์ได้

## 6. อาปาเช่เอชเบส



### รูปที่ 2.17 แสดงโลโก้ของอาปาเช่เอชเบส

ที่มา: [https://hbase.apache.org/0.94/images/hbase\\_logo.png](https://hbase.apache.org/0.94/images/hbase_logo.png)

อาปาเช่เอชเบส (Apache HBase) [9] เป็นฐานข้อมูลฮาดูปที่มีการกระจายและขยายขีดความสามารถการจัดเก็บข้อมูลขนาดใหญ่ จะใช้เอชเบสเมื่อต้องการสุ่มและเข้าถึงแบบอ่านหรือเขียนแบบเรียลไทม์ไปยังข้อมูลขนาดใหญ่ เป้าหมายของเอชเบส คือ การจัดการกับตารางขนาดใหญ่หลายๆ ที่มีพันล้านแถวคูณด้วยล้านคอลัมน์ เอชเบสมีความสามารถคล้ายกับ Google's Bigtable บนฮาดูปและ HDFS

## 7. อาปาเช่มาเฮาท์



### รูปที่ 2.18 แสดงโลโก้ของอาปาเช่มาเฮาท์

ที่มา: <https://mahout.apache.org/images/mahout-logo-poweredby-100.png>

เป็นไลบรารีของอัลกอริธึมเพื่อขยายขีดความสามารถของการเรียนรู้เครื่อง (Machine learning) ในอาปาเช่ฮาดูป การพัฒนาปรับปรุงบนอาปาเช่ฮาดูป และการใช้กระบวนการทัศน์แมปรีดิวซ์ การเรียนรู้เครื่องเป็นส่วนหนึ่งของปัญญาประดิษฐ์ที่มุ่งเน้นไปที่การเปิดใช้งานเครื่องที่จะเรียนรู้ได้โดยไม่ถูกตั้งโปรแกรมไว้อย่างชัดเจนและเป็นที่ยอมรับในการปรับปรุงการดำเนินงานในอนาคตขึ้นอยู่กับผลลัพธ์ที่ก่อนหน้านี้ เมื่อข้อมูลมีขนาดใหญ่จะถูกเก็บไว้ใน HDFS มาเฮาท์จะมีเครื่องมือวิทยาศาสตร์ข้อมูล เพื่อหารูปแบบที่มีความหมายในบรรดาชุดข้อมูลขนาดใหญ่โดยอัตโนมัติ จุดมุ่งหมายของมาเฮาท์ คือ การทำให้เร็วขึ้นและง่ายต่อการทำให้ข้อมูลขนาดใหญ่กลายเป็นสารสนเทศขนาดใหญ่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 8. อาปาเช่อูซี่



รูปที่ 2.19 แสดง โลโก้ของอาปาเช่อูซี่

ที่มา: <https://www.mapr.com/sites/default/files/oozie-image.png>

อาปาเช่อูซี่ (Apache Oozie) [11] เป็นซอฟต์แวร์โอเพนซอร์สที่เป็นเครื่องมือช่วยในการเขียนเวิร์กโฟลว (Work Flow) จะช่วยลดความยุ่งยากในการเขียน โดยการนำคำสั่งต่าง ๆ ของอาปาเช่ฮาดูปมาเชื่อมต่อกันในรูปของเวิร์กโฟลว อาปาเช่อูซี่แบ่งงานออกเป็น 2 ชนิด คือ กระแสงานของอูซี่ (Oozie Workflow) และการประสานงานของอูซี่ (Oozie Coordinator)

## 9. อาปาเซ่พิก



รูปที่ 2.20 แสดง โลโก้ของอาปาเซ่พิก

ที่มา: <https://www.mapr.com/sites/default/files/pig-image.png>

อาปาเซ่พิก (Apache Pig) [12] เป็นแพลตฟอร์มการเขียนสคริปต์สำหรับการประมวลผลและการวิเคราะห์ชุดข้อมูลขนาดใหญ่ ด้วย YARN เป็นศูนย์กลางสถาปัตยกรรมของอาปาเช่ฮาดูป เครื่องมือในการเข้าถึงข้อมูลหลายอย่างเช่น การโต้ตอบกับอาปาเซ่พิกด้วยข้อมูลที่เก็บไว้ในคลัสเตอร์ อาปาเซ่พิกช่วยให้ผู้ใช้อาปาเช่ฮาดูปเขียนการแปลงแมปรีดิวซ์ที่ซับซ้อนโดยใช้ภาษาสคริปต์อย่างง่ายที่เรียกว่า “พิกละติน” พิกจะแปลงสคริปต์พิกละตินเข้าสู่แมปรีดิวซ์ เพื่อที่จะสามารถดำเนินการได้ด้วย YARN สำหรับการเข้าถึงชุดข้อมูลเดียวที่เก็บไว้ใน HDFS

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 10. อาปาเช่สคูป



รูปที่ 2.21 แสดงโลโก้ของอาปาเช่สคูป

ที่มา: <http://sqoop.apache.org/images/sqoop-logo.png>

อาปาเช่สคูป (Apache Sqoop) [13] เป็นระบบที่ถ่ายโอนข้อมูลระหว่างกลุ่มอาปาเช่ฮาดูปและดาต้าสโตร์ โครงสร้างอย่างมีประสิทธิภาพ เช่น ฐานข้อมูลเชิงสัมพันธ์ สคูปจะช่วยกำจัดงานบางอย่าง เช่น กระบวนการอีทีแอล ออกจากคลังข้อมูลฮาดูป เพื่อให้ได้การทำงานที่มีประสิทธิภาพในราคาที่ต่ำกว่ามาก นอกจากนี้สคูปยังสามารถใช้ในการดึงข้อมูลจากฮาดูปและส่งไปยังดาต้าสโตร์ โครงสร้างภายนอก สคูปจะทำงานร่วมกับฐานข้อมูลเชิงสัมพันธ์ เช่น Teradata, Netezza, Oracle, MySQL and HSQLDB เป็นต้น

## 11. อาปาเช่สปราร์ค



รูปที่ 2.22 แสดงโลโก้ของอาปาเช่สปราร์ค

ที่มา: <https://upload.wikimedia.org/wikipedia/commons/e/ea/Spark-logo-192x100px.png>

อาปาเช่สปราร์ค (Apache Spark) [14] เป็นซอฟต์แวร์โอเพนซอร์สบนระบบประมวลผลคลัสเตอร์ อาปาเช่สปราร์คจะต้องมีผู้จัดการกลุ่มแบบคลัสเตอร์และระบบจัดเก็บข้อมูลแบบกระจาย อาปาเช่สปราร์คจะประกอบด้วยแกนหลักของสปราร์คและชุดของไลบรารี แกนหลักจะเป็นเครื่องมือในการดำเนินการแบบกระจายบนแพลตฟอร์มของภาษาจาวา (Java) สกาลา (Scala) และไพทอน (Python) สำหรับการพัฒนาอีทีแอลแบบกระจาย

## 12. อาปาเช่ซูคิปเปอร์



รูปที่ 2.23 แสดงโลโก้ของอาปาเช่ซูคิปเปอร์

ที่มา: [http://tomaszdziurko.pl/images/blog/2014/07/zookeeper\\_logo.png](http://tomaszdziurko.pl/images/blog/2014/07/zookeeper_logo.png)

อาปาเช่ซูคิปเปอร์ (Apache ZooKeeper) [15] เป็นซอฟต์แวร์โอเพนซอร์สที่ใช้เป็นศูนย์กลางในการเชื่อมประสานการทำงานของแอปพลิเคชันแบบกระจาย ใช้เป็นศูนย์กลางในการให้บริการการตั้งชื่อ การดูแลและตั้งค่าต่าง ๆ ของข้อมูลสารสนเทศ การประสานการกระจาย และการบริการแบบกลุ่ม อินเทอร์เน็ตจะใช้ภาษาจาวา (Java) และซี (C) ส่วนเครื่องไคลเอนต์จะใช้ภาษาไพทอน (Python) รูบี้ (Ruby) และโก (Go) เมื่อเซิร์ฟเวอร์เครื่องใดเกิดข้อผิดพลาด (Bugs) ก็สามารถแก้ไขง่าย ๆ ด้วยการเปลี่ยนเซิร์ฟเวอร์ที่ต่อแทน

## 13. อาปาเช่ไฮฟ



รูปที่ 2.24 แสดงโลโก้ของอาปาเช่ไฮฟ

ที่มา: [http://hortonworks.com/wp-content/uploads/2013/05/hive\\_logo.png](http://hortonworks.com/wp-content/uploads/2013/05/hive_logo.png)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อาปาเซฮัยฟ์ (Apache Hive) [16] เป็นซอฟต์แวร์คลังข้อมูลอำนวยความสะดวกในการสอบถามและการจัดการชุดข้อมูลขนาดใหญ่ที่อยู่ในการจัดเก็บแบบกระจาย ฮัยฟ์มีกลไกในโครงสร้างโครงการบนข้อมูลนี้และค้นหาข้อมูลโดยใช้ภาษาที่คล้าย SQL เรียกว่า ฮัยฟ์คิวแอล (HiveQL) ภาษาฮัยฟ์คิวแอลยังช่วยในการแมปรีดิวซ์แบบดั้งเดิม ส่วนขยายโปรแกรมเมอร์อย่างง่ายสำหรับ plugging ในตัวแมปและตัวรีดิวซ์ที่กำหนดเองได้ เมื่อมันไม่สะดวกหรือไม่มีประสิทธิภาพในการแสดงตรรกะนี้ในฮัยฟ์คิวแอล

ตัวอย่างการใช้งานอาปาเซฮัยฟ์ด้วย HiveQL จะแสดงในรูปที่ 2.25

```
#hive
hive> CREATE TABLE foreign_aid
      (country STRING, sum BIGINT)
      ROW FORMAT DELIMITED
      FIELDS TERMINATED BY ','
      STORED AS TEXTFILE;
hive> SHOW TABLES;
hive> DESCRIBE foreign_aid;
hive> LOAD DATA INPATH 'econ_assist.csv'
      OVERWRITE INTO TABLE foreign_aid;
hive> SELECT * FROM foreign_aid LIMIT 10;
hive> SELECT country, SUM(sum) FROM foreign_aid
      GROUP BY country;
```

รูปที่ 2.25 แสดงตัวอย่างการเขียนด้วย HiveQL

ที่มา: [http://images.slideplayer.com/21/6241930/slides/slide\\_69.jpg](http://images.slideplayer.com/21/6241930/slides/slide_69.jpg)

#### 2.2.1.4 คลาวเดรา อิมพาลา



รูปที่ 2.26 แสดงโลโก้ของอิมพาลา

ที่มา: <http://impala.io/img/impala-logo.png>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คลาวเดรา อิมพาลา (Cloudera Impala) [17] มีประสิทธิภาพในการทำงานสูงและมีค่าความหน่วงเวลาในการคิวรี SQL ข้อมูลที่จัดเก็บไว้ในรูปแบบไฟล์อาปาเซฮาคุปที่ต่ำ สำหรับการคิวรีข้อมูลมาวิเคราะห์มีการตอบสนองในการค้นหา โต้ตอบ และปรับเข้าหากันได้อย่างรวดเร็ว

อิมพาลาจะทำงานร่วมกับฐานข้อมูลอาปาเซโฮฟี่เมต้าสโตร์ (Apache Hive metastore) เพื่อแบ่งปันตารางและฐานข้อมูลระหว่างองค์ประกอบของทั้งกลุ่ม สามารถทำงานระดับสูงร่วมกับอาปาเซโฮฟี่และเข้ากันได้กับไวยากรณ์ของโฮฟี่คิวแอล ช่วยให้คุณสามารถใช้ได้ทั้งอิมพาลาหรือโฮฟี่ในการสร้างตาราง คิวรีข้อมูล โหลดข้อมูล และอื่นๆ

ข้อได้เปรียบบางอย่างที่สำคัญของอิมพาลา มีดังต่อไปนี้

- อิมพาลาทำงานร่วมกับระบบนิเวศคลาวเดราแบบกระจายของอาปาเซฮาคุป(Cloudera distribution of Apache Hadoop : CDH) ที่มีอยู่ หมายถึง ข้อมูลที่สามารถจัดเก็บ แบ่งปัน และเข้าถึงได้ โดยใช้วิธีการแก้ไขปัญหาต่างๆที่มาพร้อมกับCDH นอกจากนี้ยังหลีกเลี่ยงการเก็บข้อมูลแบบไซโล (Data Silos) คือการเก็บข้อมูลไว้ แต่คนอื่นเอาไปใช้ไม่ได้ และช่วยลดการย้ายข้อมูลที่มีค่าใช้จ่ายสูง
- อิมพาลาสามารถเข้าถึงข้อมูลที่เก็บไว้ใน CDH โดยไม่จำเป็นต้องมีความรู้หรือทักษะภาษาจาวาในการเรียกใช้งานแมปรีคิวรี่ อิมพาลาสามารถเข้าถึงข้อมูลได้โดยตรงจากระบบไฟล์ HDFS อิมพาลายังมีส่วนแสดงผล SQL ในการเข้าถึงข้อมูลในระบบฐานข้อมูลเอชเบสหรือระบบการจัดเก็บข้อมูลอย่างง่ายอเมซอน (S3)
- อิมพาลาจะใช้เวลาในการคืนค่าผลลัพธ์เพียงไม่กี่วินาทีหรือไม่กี่นาที แต่การเรียกใช้โฮฟี่จนเสร็จสมบูรณ์จะใช้เวลามากกว่าอาจหลายนาทีหรือเป็นชั่วโมง
- อิมพาลาเป็นผู้ริเริ่มการใช้รูปแบบไฟล์พาร์เก้ต (Parquet file format) เป็นรูปแบบการจัดเก็บที่เหมาะสมกับการคิวรีข้อมูลขนาดใหญ่ใน scenarios ของคลังข้อมูล

#### 2.2.1.4.1 แนวคิดและสถาปัตยกรรมของอิมพาลา

ส่วนต่อไปนี้จะให้ข้อมูลพื้นฐานที่จะช่วยให้คุณสามารถใช้อิมพาลาและคุณลักษณะของมันอย่างมีประสิทธิภาพ รวมถึงคำอธิบายบริบทที่จะช่วยให้เข้าใจถึงลักษณะของอิมพาลาที่เกี่ยวข้องกับเทคโนโลยีอื่นๆ ที่คุณอาจจะคุ้นเคย เช่น ระบบการจัดการฐานข้อมูลเชิงสัมพันธ์และคลังข้อมูลหรือองค์ประกอบอื่นๆของฮาคุป เช่น Hive, HDFS และ HBase [18]

#### องค์ประกอบของเซิร์ฟเวอร์อิมพาลา

เซิร์ฟเวอร์อิมพาลา (Impala Server) [19] เป็นเครื่องมือฐานข้อมูลที่มีการประมวลผลแบบขนานขนาดใหญ่(Massively Parallel Processing : MPP) และแบบกระจาย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

มันจะประกอบไปด้วยกระบวนการดีมอน (Daemon Processes) ที่แตกต่างกันที่ทำงานบน โสสต์เฉพาะภายในกลุ่มคลัสเตอร์ CDH

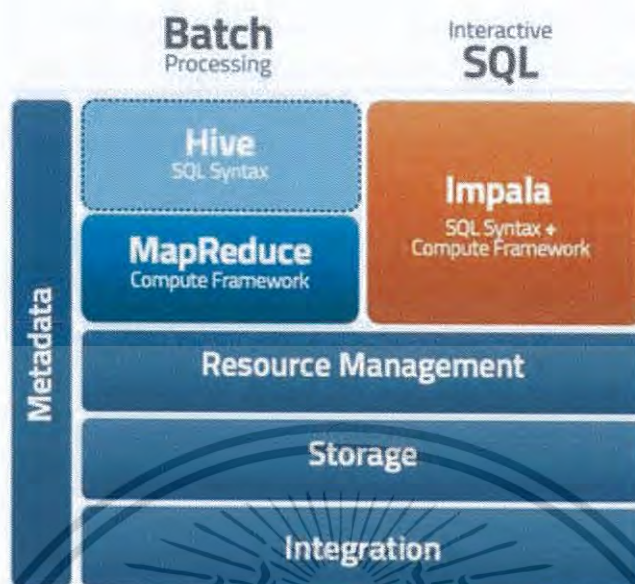
### อิมพาลาดีมอน (Impala Daemon)

องค์ประกอบหลักของอิมพาลา คือ กระบวนการดีมอนที่ทำงานบนแต่ละดาต้า โหนดของคลัสเตอร์ อิมพาลาจะอ่านและเขียนไปยังไฟล์ข้อมูล โดยจะทำตามคำสั่งที่ส่งมาจากคำสั่ง Impala shell, Hue, JDBC, หรือ ODBC มีการคิวรีแบบขนานและกระจายงานไปทั่วคลัสเตอร์ และส่งค่าผลลัพธ์ที่ได้กลับไปยังโหนดประสานงานส่วนกลาง

คุณสามารถที่จะคิวรีส่งไปยังอิมพาลาดีมอนที่ทำงานบน ใดๆ และตัวอย่างทำหน้าที่ของดีมอนคือ ทำหน้าที่เป็น โหนดประสาน (Coordinator node) สำหรับคิวรี โหนดอื่น โดยจะส่งผลลัพธ์บางส่วนกลับไปยังผู้ประสานงาน ซึ่งจะสร้างชุดผลลัพธ์สุดท้ายเพื่อใช้คิวรีเมื่อใช้ทดลองพร้อมกับการทำงานผ่านคำสั่ง Impala shell โดยจะต้องเชื่อมต่อกับอิมพาลาดีมอนเดียวกันเพื่อความสะดวก สำหรับการคลัสเตอร์ปริมาณการผลิตอาจจะต้องใช้การ Load Balance ด้วยการคิวรีแต่ละอันส่งไปยังอิมพาลาดีมอนที่แตกต่างกันในรูปแบบ Round Robin โดยใช้การแสดงผลแบบ JDBC หรือ ODBC

อิมพาลาดีมอนจะสื่อสารอย่างต่อเนื่องกับ Statestore เพื่อยืนยันว่า โหนดมีความพร้อมในการรับงานใหม่ เมื่อแต่ละ โหนดได้รับข้อความที่กระจายมาจาก Catalog Daemon บาง โหนดในอิมพาลาจะทำการคลัสเตอร์ creates, alters หรือ drops ออบเจ็คต์บางชนิด หรือ INSERT หรือ LOAD ข้อมูล จะประมวลผลผ่านอิมพาลา

### 2.2.1.5 อาปาเซโฮฟท์กับคราวเดราอิมพาลา



รูปที่ 2.27 แสดงการเปรียบเทียบการทำงานของ Hive และ Impala

อาปาเซโฮฟท์ [20] เป็นมาตรฐานที่มีประสิทธิภาพสำหรับ SQL ใน Hadoop โฮฟท์เป็นส่วนหนึ่งสำหรับการแยกคำสั่ง SQL, การสร้างแผนตรรกะ, การเพิ่มประสิทธิภาพแผนตรรกะ, การแปลงเป็นแผนทางกายภาพที่มีการดำเนินการโดยแมปรีดิวซ์ อาปาเซโฮฟท์ถูกออกแบบมาสำหรับระบบคลังข้อมูลเพื่อความสะดวกในการประมวลผลของคำสั่งการคิวรีข้อมูลแบบ Adhoc และชุดข้อมูลขนาดใหญ่ที่เก็บไว้ใน HDFS และความสะดวกในการรวมข้อมูล

อิมพาลาเป็นเครื่องมือในการคิวรี SQL แบบโอเพนซอร์ส พัฒนาขึ้นต่อจาก Google Dremel คราวเดรา อิมพาลาเป็นเครื่องมือ SQL สำหรับการประมวลผลข้อมูลที่เก็บไว้ในเอชเบสและ HDFS อิมพาลาใช้โฮฟท์เมก้าส โดร์และสามารถคิวรีตารางโฮฟท์ได้โดยตรง โดยแตกต่างจากโฮฟท์ตรงที่อิมพาลาไม่ได้แปลงคำสั่งเข้าไปในแมปรีดิวซ์ แต่สามารถดำเนินการได้ด้วยตัวเอง ซึ่งจะเห็นได้จากรูปที่ 2.27

อย่างไรก็ตามทั้งอาปาเซโฮฟท์และคราวเดราอิมพาลาจะสนับสนุนมาตรฐานทั่วไปของ HiveQL

- โฮฟท์เข้าแต่เหมาะที่จะเป็นตัวเลือกที่ดีสำหรับงาน ETL ที่หนักๆ มีบทบาทสำคัญที่น่าเชื่อถือ เช่น รวบรวมล็อก (Log) รายชั่วโมงสำหรับองค์กรโฆษณา อิมพาลาเป็นเครื่องมือในการคิวรี SQL แบบโอเพนซอร์ส ที่สามารถนำมาใช้อย่างมีประสิทธิภาพสำหรับการประมวลผลคำสั่งในข้อมูลที่มีปริมาณมาก อิมพาลาจะประมวลผลได้เร็วกว่าและจัดการปริมาณข้อมูลขนาดใหญ่ได้มากกว่าเครื่องมือคิวรีแบบโฮฟท์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- การคิวรี expressions ในไฮฟ์จะถูกสร้างขึ้นในช่วงของการคอมไพล์ ในขณะที่อิมพาลาจะสร้างขึ้นในช่วงของการเรียกใช้งาน โค้ดสำหรับรูปขนาดใหญ่ผ่าน LLVM ที่จะช่วยในการเพิ่มประสิทธิภาพของโค้ด
- การแปลงคำสั่งคิวรีของไฮฟ์จะดำเนินการเข้าสู่แมปรีดิวซ์ ตั้งแต่เริ่มต้น ในขณะที่อิมพาลาจะตอบสนองได้อย่างรวดเร็วผ่านการประมวลผลแบบขนาน

อิมพาลาจะเร็วกว่าอาปาเซไฮฟ์ แต่ไม่ได้หมายความว่ามันจะเป็นจะเป็นการแก้ปัญหา SQL สำหรับทุกปัญหาของข้อมูลขนาดใหญ่ได้ดีทั้งหมด เพราะอิมพาลาจะเน้นการใช้หน่วยความจำเป็นหลักและไม่ได้ทำงานอย่างมีประสิทธิภาพสำหรับการดำเนินงานข้อมูลอย่างหนักเช่นเดียวกับการ joins เพราะมันเป็นไปไม่ได้ที่จะทำทุกอย่างในหน่วยความจำ จึงนำไฮฟ์เข้ามาช่วย ถ้าแอปพลิเคชันที่มีชุดการประมวลผลประเภทความต้องการมากกว่าข้อมูลขนาดใหญ่เมื่อนั้นองค์กรจะต้องเลือกใช้ไฮฟ์ แต่ถ้าหากองค์กรต้องการการประมวลผลด้วยเวลาจริงของการคิวรีแบบ Adhoc ในสับเซตของข้อมูลแล้ว อิมพาลาจึงเป็นทางเลือกที่ดี ซึ่งเราสามารถดูการเปรียบเทียบการทำงานระหว่างไฮฟ์กับอิมพาลาได้ในตารางที่ 2.1 และ 2.2

ตารางที่ 2.1 แสดงการเปรียบเทียบระหว่าง Pig, Hive และ Impala

Feature	Pig	Hive	Impala
SQL-based query language	No	Yes	Yes
Optional schema and metastore	Yes	No	No
User-defined functions (UDFs)	Yes	Yes	Yes
Process data with external scripts	Yes	Yes	No
Extensible file format support	Yes	Yes	No
Complex data types	Yes	Yes	No
Query latency	High	High	Low
Built-in data partitioning	No	Yes	Yes
Accessible via ODBC / JDBC	No	Yes	Yes

ที่มา: <https://tulip.nectec.or.th/news/data/information/information1/23691/1434443921.pdf>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.2 แสดงการเปรียบเทียบระหว่าง RDBMS, Hive และ Impala

Feature	RDBMS	Hive	Impala
Insert individual records	Yes	No	Yes
Update and delete records	Yes	No	No
Transactions	Yes	No	No
Role-based authorization	Yes	Yes (Sentry)	Yes (Sentry)
Stored procedures	Yes	No	No
Index support	Extensive	Limited	None
Latency	Very low	High	Low
Data size	Terabytes	Petabytes	Petabytes
Complex data types	No	Yes	No
Storage cost	Very high	Very low	Very low

ที่มา: <https://tulip.nectec.or.th/news/data/information/information1/23691/1434443921.pdf>

## 2.2.2 เทคโนโลยีที่ใช้ในการสร้างเว็บแอปพลิเคชัน

### 1) เทคโนโลยี JSP

Java Server Page (JSP) [21] เป็นเทคโนโลยีสำหรับการควบคุมเนื้อหาหรือลักษณะของเว็บเพจโดยการใช้ Servlets ซึ่งเป็นโปรแกรมขนาดเล็กที่มีการระบุไว้ในเว็บเพจและทำงานบนเว็บเซิร์ฟเวอร์ในการปรับเปลี่ยนเว็บเพจก่อนที่จะถูกส่งไปยังผู้ใช้ที่ร้องขอ JSP ถูกพัฒนาขึ้นโดยนักพัฒนาชาวอเมริกันของบริษัท ซัน ไมโครซิสเต็มส์ (Sun Microsystems) นอกจากนี้เทคโนโลยี JSP ยังเป็น Servlet แสดงผลโปรแกรมแอปพลิเคชัน (Application Program Interface : API) อีกด้วย JSP ยังเทียบเท่ากับเทคโนโลยี Active Server Page (ASP) ของบริษัทไมโครซอฟต์ ในขณะที่ JSP จะเรียกใช้งานโปรแกรมจาวาที่ดำเนินการ (Executed) โดยเว็บเซิร์ฟเวอร์ แต่ ASP จะประกอบด้วยสคริปต์ที่จะถูกแปล (Interpreted) โดย Script Interpreter (เช่น VBScript หรือ JScript) ก่อนที่หน้าเพจจะถูกส่งให้กับผู้ใช้ ซึ่งตัวอย่างการเขียนด้วย JSP จะแสดงในรูปที่ 2.28

```

<%
Class.forName("org.apache.hive.jdbc.HiveDriver");
int count = 0;
Connection con = DriverManager.getConnection("jdbc:hive2://161.2
    "itkmitldw504");
Statement stmt = con.createStatement();

String Hivequery = "select * from dw.staff_account";
System.out.println("Running: " + Hivequery );
ResultSet res = stmt.executeQuery(Hivequery );
System.out.println("execute :"+res);
while (res.next()) {
    count += 1;
    String name = res.getString(1);
    String username = res.getString(2);
    String password = res.getString(3);
    String role = res.getString(4);
    request.setAttribute("count", count);
    request.setAttribute("name", name);
    request.setAttribute("username", username);
    request.setAttribute("password", password);
    request.setAttribute("role", role);
}
%>

```

รูปที่ 2.28 แสดงตัวอย่างการเขียนด้วย JSP

## 2) เทคโนโลยี Java



รูปที่ 2.29 แสดงโลโก้ของ Java

ที่มา: [http://docs.jpsh.io/image/sdk\\_java.png](http://docs.jpsh.io/image/sdk_java.png)

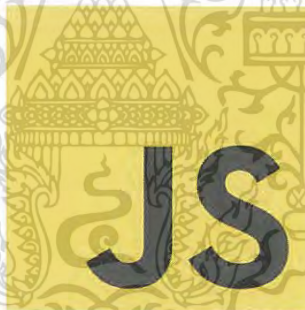
จาวา (Java) [22] เป็นภาษาโปรแกรมที่ออกแบบมาสำหรับการใช้งานในสภาพแวดล้อมที่มีการกระจายของอินเทอร์เน็ต จาวาถูกออกแบบมาให้มี "รูปลักษณะและความรู้สึก" แบบภาษา C ++ แต่มันจะง่ายกว่าที่จะใช้ภาษา C ++ และบังคับใช้รูปแบบการเขียนโปรแกรมเชิงวัตถุ จาวาสามารถใช้ในการสร้างแอปพลิเคชันที่สมบูรณ์แบบที่อาจทำงานบนคอมพิวเตอร์เครื่องเดียวหรือกระจายในเซิร์ฟเวอร์และไคลเอนต์ในเครือข่าย นอกจากนี้ยังสามารถใช้ในการสร้างโมดูลโปรแกรมขนาดเล็กหรือแอปพลิเคชันสำหรับการเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ใช้งานส่วนหนึ่งของเว็บเพจ แอปพลิเคชันจะสร้างความเป็นไปได้ที่ผู้ใช้เว็บเพจจะสามารถโต้ตอบกับหน้าเพจได้ ซึ่งตัวอย่างการเขียนด้วยภาษาจาวาจะแสดงในรูปที่ 2.30

```
public String getTeacherName() {
    return TeacherName;
}
public void setTeacherName(String teacherName) {
    TeacherName = teacherName;
}
public int getDate_key() {
    return Date_key;
}
public void setDate_key(int date_key) {
    Date_key = date_key;
}
public int getTotalHourPerSemester() {
    return totalHourPerSemester;
}
public void setTotalHourPerSemester(int totalHourPerSemester) {
    this.totalHourPerSemester = totalHourPerSemester;
}
public int getOver_Workload() {
    return Over_Workload;
}
```

รูปที่ 2.30 แสดงตัวอย่างการเขียนด้วยภาษา Java

### 3) เทคโนโลยี JavaScript



รูปที่ 2.31 แสดงโลโก้ของ JavaScript

ที่มา: [http://codepb.com/wp-content/uploads/2013/04/JavaScript\\_Logo.png](http://codepb.com/wp-content/uploads/2013/04/JavaScript_Logo.png)

จาวาสคริปต์ (JavaScript) [23] เป็นการข้ามแพลตฟอร์มและภาษาสคริปต์เชิงวัตถุ จาวาสคริปต์เป็นภาษาที่มีขนาดเล็กและน้ำหนักเบา ภายในสภาพแวดล้อมโฮสต์ (เช่น เว็บเบราว์เซอร์), JavaScript สามารถเชื่อมต่อกับวัตถุของสภาพแวดล้อมของตนเพื่อให้การควบคุมการเขียนโปรแกรมมากกว่าพวกเขา

จาวาสคริปต์ประกอบด้วยไลบรารีมาตรฐานของออบเจกต์ เช่น อาร์เรย์, วันที่, และคณิตศาสตร์และชุดองค์ประกอบหลักของภาษา เช่น คำนิยาม การสร้างการควบคุม และคำสั่ง จาวาสคริปต์โดยหลักแล้วจะสามารถขยายความหลากหลายของวัตถุประสงค์ได้ โดยการเสริมด้วยออบเจกต์เพิ่มเติม ซึ่งตัวอย่างการเขียนด้วยจาวาสคริปต์จะแสดงในรูปที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

§(function () {
    var date = new Date();
    var d = date.getDate();
    var m = date.getMonth();
    var y = date.getFullYear();

```

รูปที่ 2.32 แสดงตัวอย่างการเขียนด้วย JavaScript

#### 4) เทคโนโลยี CSS



รูปที่ 2.33 แสดงโลโก้ของ CSS

ที่มา: <http://php.quicoto.com/wp-content/uploads/2013/06/css3.jpg>

Cascading Style Sheet (CSS) [24] เป็นเว็บเพจที่ได้มาจากหลายแหล่งที่มาที่มีกำหนดลำดับความสำคัญไว้และมีการนิยามความขัดแย้งขององค์ประกอบรูปแบบใด ๆ Cascading Style Sheet ระดับที่ 1 (CSS1) ได้รับการรับรองจาก World Wide Web Consortium (W3C) ซึ่งจะนำมาใช้ในรุ่นล่าสุดของ Netscape และ Microsoft เว็บเบราว์เซอร์ กำหนด style sheets หรือ statements ที่เป็นไปได้ โดยอาจจะกำหนดองค์ประกอบที่ได้เป็นการนำเสนอในเว็บเพจ CSS ช่วยให้การควบคุมที่มากกว่าลักษณะของเว็บเพจ เพื่อสร้างเว็บเพจมากกว่าที่จะออกแบบเบราว์เซอร์หรือส่วนของผู้ชม

โดยทั่วไป style sheets ของการสร้างเว็บเพจจะมีความสำคัญ แต่ก็ขอแนะนำวิธีการให้เบราว์เซอร์สำหรับส่วนของผู้ชมที่สามารถแทนที่ด้วยรูปแบบคุณลักษณะในบางประการ เพราะมันเป็นไปได้ว่าเบราว์เซอร์ที่แตกต่างกันจะเลือกที่จะใช้ CSS ก่อนข้างแตกต่าง ผู้สร้างเว็บเพจจึงต้องทดสอบหน้าเพจด้วยเบราว์เซอร์ที่แตกต่างกัน ซึ่งตัวอย่างการเขียนด้วย CSS จะแสดงในรูปที่ 2.34

```

.errorContainer {
  background: url(../images/patterns/6.png) repeat;
  border:1px solid #c4c4c4;
  border-radius: 2px;
  -moz-border-radius:2px;
  -webkit-border-radius:2px;
  padding:15px;
  -webkit-box-shadow: 0px 0px 1px 1px rgba(0, 0, 0, 0.1);
  -moz-box-shadow: 0px 0px 1px 1px rgba(0, 0, 0, 0.1);
  box-shadow: 0px 0px 1px 1px rgba(0, 0, 0, 0.1);
  position: absolute;
  /* top:50%;*/
  left:50%;
  margin-left:-286px;
  margin-top: -155px;
  width:540px;
}

```

รูปที่ 2.34 แสดงตัวอย่างการเขียนด้วย CSS

## 5) เทคโนโลยี jQuery



รูปที่ 2.35 แสดงโลโก้ของ jQuery

ที่มา: <http://i2.wp.com/evanjmg.com/wp-content/uploads/2015/07/jQuery.gif>

jQuery [25] เร็ว มีขนาดเล็ก และมีไลบรารีจาวาสคริปต์ที่มีคุณลักษณะหลากหลาย jQuery จะสร้างสิ่งที่คล้ายกับการเข้าไปและจัดการเอกสาร HTML การจัดการเหตุการณ์ ภาพเคลื่อนไหว และ AJAX ที่ง่ายมาก ด้วย API ที่ง่ายต่อการใช้งานที่ทำงานในเบราว์เซอร์ที่หลากหลาย ด้วยการรวมกันของความคล่องตัวและการขยาย jQuery จะมีการเปลี่ยนแปลงวิธีการใช้งานของคนนับล้านที่เขียน JavaScript ซึ่งตัวอย่างการเขียนด้วย jQuery จะแสดงในรูปที่ 2.36

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

$(document).ready(function() {

    //----- Full calendar ----
    $(function () {
        var date = new Date();
        var d = date.getDate();
        var m = date.getMonth();
        var y = date.getFullYear();

        //front page calendar|
        $('#calendar').fullCalendar({

```

รูปที่ 2.36 แสดงตัวอย่างการเขียนด้วย jQuery

#### 6) เทคโนโลยี Highcharts

รูปที่ 2.37 แสดงโลโก้ของ jQuery

ที่มา: <http://api.highcharts.com/resources/images/Highcharts.svg>

Highcharts [26] เป็นไลบรารีที่ใช้ในการเขียนแผนภูมิด้วยจาวาสคริปต์ ที่มีวิธีนำเสนออย่างง่าย ด้วยการเพิ่มแผนภูมิการโต้ตอบไปยังเว็บไซต์หรือเว็บแอปพลิเคชัน Highcharts ที่สนับสนุน ได้แก่ เส้น, โคน, พื้นที่, พื้นที่โค้ง, คอลัมน์, แท่ง, พาย, กระจาย, มาตรฐานวง, ช่วงพื้นที่, ช่วงพื้นที่โค้ง, ช่วงคอลัมน์, แผนภาพฟอง, แผนภาพกล่อง, แผนภูมิแท่งบอกความคลาดเคลื่อน, แผนภูมิน้ำตกและแผนภูมิขั้วโลก ซึ่งตัวอย่างการเขียนด้วย Highcharts จะแสดงในรูปที่ 2.38

```

$('#RankTeacherDoWorkload').highcharts({
  chart: {
    type: 'column'
  },
  title: {
    text: 'จัดลำดับการใช้งานประมาณของ ปีการศึกษา ${chartInDa
  },
  subtitle: {
    text: 'Source: IT KMITL Data Warehouse'
  },

```

รูปที่ 2.38 แสดงตัวอย่างการเขียน Highcharts

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 3

### วิธีการดำเนินการ

เนื่องจากปัจจุบันเป็นยุคของ “โลกาภิวัตน์” เทคโนโลยีสารสนเทศเข้ามามีบทบาทและมี ส่วนช่วยในการทำงานภายในองค์กรอย่างมาก โดยมีเครื่องมือที่ช่วยรองรับการเปลี่ยนแปลงของ ข้อมูลได้อย่างรวดเร็ว แม่นยำ และสามารถช่วยในการตัดสินใจในการลงทุน มองเห็นปัญหาภายใน องค์กร ได้อย่างชัดเจนและตรงจุดมากยิ่งขึ้น เพื่อเพิ่มศักยภาพในการแข่งขันของคณะเทคโนโลยี สารสนเทศให้มีความเหนือกว่ามหาวิทยาลัยและสถาบันการศึกษาอื่น โดยมองถึงแนวโน้มปริมาณ ของการเก็บข้อมูลเพิ่มสูงขึ้นอย่างไม่หยุดยั้ง ดังนั้นจึงนำเสนอการเก็บข้อมูลบนระบบอาปาเช่ฮาดูป และการแสดงผลข้อมูลในรูปแบบแดชบอร์ด (Dashboard) เพื่อให้สรุปผลการดำเนินงานได้อย่าง รวดเร็ว รวมทั้งสามารถมองเห็นผลรวมการดำเนินงานของนักศึกษาและบุคลากรของ คณะ เทคโนโลยีสารสนเทศได้อย่างชัดเจนมากยิ่งขึ้น

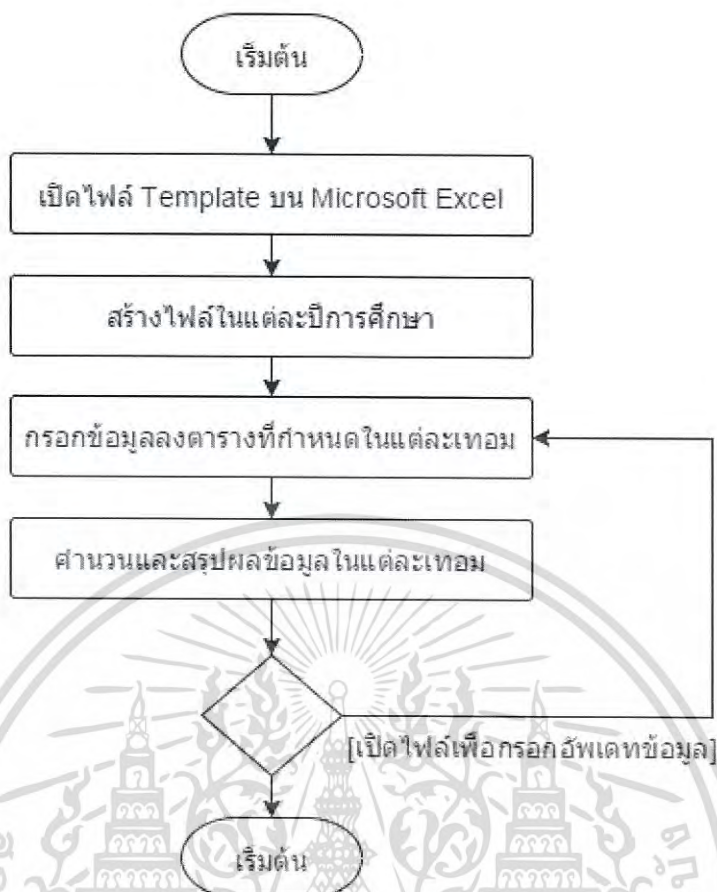
ลำดับ	หัวข้องาน	มิถุนายน	กรกฎาคม	สิงหาคม	ตุลาคม	พฤศจิกายน	มกราคม	กุมภาพันธ์	มีนาคม	เมษายน	พฤษภาคม
1	เก็บข้อมูลจากผู้ใช้งาน/บุคคลที่เกี่ยวข้อง										
2	ศึกษาและเพิ่มระบบ Hadoop เบื้องต้น										
3	ทดลองมุ่งศึกษา Solution ที่เหมาะสมกับการทำคลังข้อมูล										
4	ออกแบบคลังข้อมูล										
5	ออกแบบส่วนติดต่อกับผู้ใช้ระบบ										
6	เขียนโปรแกรม ส่วน Insert Data (Import CSV)										
7	เขียนโปรแกรม ส่วน Query Data (View/Query data)										
8	เขียนโปรแกรม ส่วน Dashbord (Show Dashbord chart)										
9	เขียนโปรแกรม ส่วน Report (Show chart)										
10	ทดสอบระบบแอปพลิเคชัน										
11	ทดสอบให้ผู้ใช้ระบบจริง(เพื่อประเมินประสิทธิภาพ)										
12	จัดทำรายงาน										

รูปที่ 3.1 แผนผังการคุมกำหนดงาน (Gantt chart)

### 3.1 ระบบการทำงานในปัจจุบัน

○ แผนภาพกิจกรรมการจัดเก็บข้อมูลของคณะเทคโนโลยีสารสนเทศในปัจจุบัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

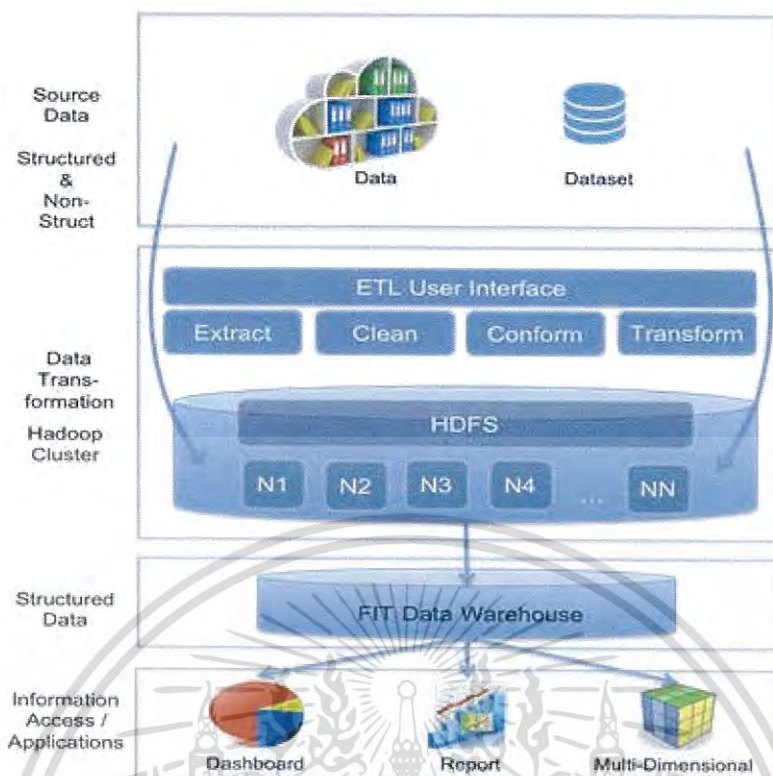


รูปที่ 3.2 แผนภาพการจัดเก็บข้อมูลของคณะเทคโนโลยีสารสนเทศในปัจจุบัน

### 3.2 ขั้นตอนการดำเนินงาน

เนื่องจากชุดข้อมูลเป็นของ คณะเทคโนโลยีสารสนเทศ ซึ่งจัดเก็บอยู่ในรูปแบบไฟล์ Microsoft Excel หลายๆ ไฟล์ โดยเก็บข้อมูลนักศึกษา กับบุคลากรคณาจารย์แยกย่อยเป็นส่วนๆ ทำให้เกิดปัญหาในการเรียกใช้ข้อมูลและค้นหาสารสนเทศที่มีอยู่ในไฟล์ข้อมูลแต่ละไฟล์ทำได้ช้าและยากลำบาก และมีความเสี่ยงต่อข้อมูลสูญหายและไม่ปลอดภัย โดยการนำข้อมูลนักศึกษา กับบุคลากรคณาจารย์ทั้งหมดเก็บในรูปแบบไฟล์ ในคลังข้อมูล (Data Warehouse) โดยจะนำข้อมูลมาวิเคราะห์และสกัดข้อมูลที่เข้าซึ่จนแล้วนำมาเฉพาะข้อมูลที่เป็นประโยชน์ เพื่อนำมาใช้แสดงผลวิเคราะห์ข้อมูลหลายมิติและการค้นหาความรู้ในคลังข้อมูล ในรูปแบบแผนภูมิหรือ Dashboard ในรูปแบบง่ายต่อความเข้าใจ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



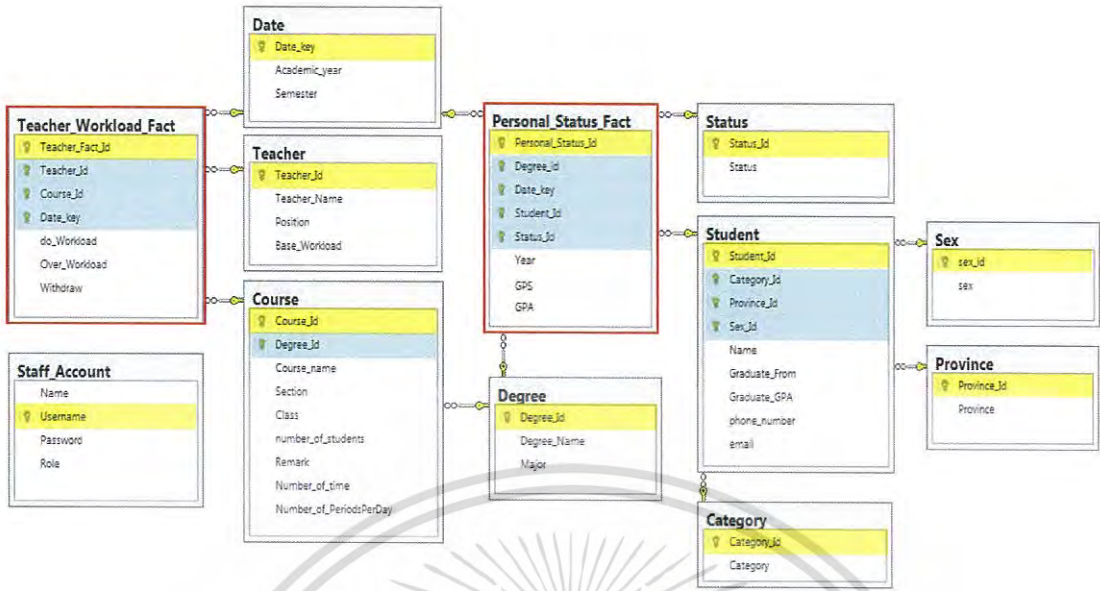
รูปที่ 3.3 ขั้นตอนการดำเนินงาน

ภาพข้อมูลข้างต้น แสดงถึงการแปลงและสกัดข้อมูล (Data Transformation) ข้อมูลไฟล์ Excel และเอกสาร เพื่อนำมาเฉพาะข้อมูลที่เป็นประโยชน์ เพื่อนำมาใช้แสดงผลวิเคราะห์ข้อมูลหลายมิติโดยผ่านระบบ Hadoop Cluster ระบบจะจัดการไฟล์แบบกระจายของอาปาเซฮาคุป (Hadoop Distributed File System : HDFS) แล้วจะทำสำเนาไฟล์ไว้ทั้งหมด 3 DataNode เพื่อให้ข้อมูลไม่เสียหาย (เพื่อป้องกันปัญหาทางด้านฮาร์ดแวร์หาก DataNode ใดไม่สามารถใช้งานได้ก็ยังสามารถใช้ระบบดึง/เก็บได้ปกติ) จากนั้นจึงจัดทำส่วนของการแสดงผลข้อมูลให้อยู่ในรูปแบบอย่างง่ายในรูปแบบต่างๆ เช่น แดชบอร์ด รีพอร์ต เป็นต้น

### 3.3 การออกแบบคลังข้อมูล

สร้างแผนภาพ ER (ER Diagram) เพื่อแสดงความสัมพันธ์ระหว่าง มิติของตาราง กับตารางข้อเท็จจริง โดยการออกแบบ Logical Data Model Star schema เป็นเทคนิคที่ใช้ ฐานข้อมูลหลายมิติ (Multidimensional Model) เน้นที่ประสิทธิภาพความเร็วในการดึง/เขียนข้อมูลในระบบเป็นหลัก และการออกแบบนี้ยังสามารถดึงข้อมูลหลายมิติได้ โดยการออกแบบ Star Schema (snowflake) ที่ใช้ในโครงการนี้มีทั้งหมด 2 Star Schema ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.4 แสดง Star Schema สำหรับเก็บข้อมูลของคณะเทคโนโลยีสารสนเทศ (โดยกรอบสีแดงคือ Fact Table และที่ไม่มีกรอบคือ Dimension)

จากรูป Star Schema ข้างต้น เป็นแสดงภาพรวมของข้อมูลทั้งระบบ ซึ่งรายละเอียดชื่อตัวแปร ชนิดตัวแปร รวมถึงรายละเอียดใน Fact Table และ Dimension Table ในตารางที่ 3.1 - 3.3 ดังนี้

### 3.3.1 Fact Table

ตารางที่ 3.1 แสดงชนิดของข้อมูลใน Fact Table ของนักศึกษา ในตาราง Personal Status Fact เพื่อแสดงผลสรุปข้อมูล เกี่ยวกับข้อมูลรายบุคคลของนักศึกษา

ชื่อตัวแปร	ชนิดข้อมูล	รายละเอียด
Personal_status_Id	String	Index ของตาราง Fact_Personal_Fact
Degree_Id	int	Index ของตาราง Dimension Student
Date_key	String	Index ของตาราง Dimension Date
Student_Id	String	Index ของตาราง Dimension Course
Year	String	แสดงถึงชั้นปี(1-4) ณ ปัจจุบัน
GPS	float	แสดงถึงเกรดเฉลี่ยในแต่เทอมของนักศึกษาแต่ละบุคคล
GPA	float	แสดงถึงเกรดเฉลี่ยสะสมของนักศึกษาแต่ละบุคคล
Status_Id	int	แสดงถึงสถานภาพทางการศึกษาของนักศึกษาแต่ละบุคคล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.2 แสดงชนิดของข้อมูลใน Fact Table ของอาจารย์ ในตาราง Teacher\_Workload\_Fact เพื่อแสดงผลสรุปข้อมูล เกี่ยวกับข้อมูลรายบุคคลของอาจารย์

ชื่อตัวแปร	ชนิดข้อมูล	รายละเอียด
Teacher_Workload_Fact	String	Index ของตาราง Fact_Teacher_Fact
Teacher_Id	int	Index ของตาราง Dimension Teacher
Course_Id	int	Index ของตาราง Dimension Course
Date_key	int	Index ของตาราง Dimension Date
Do_Workload	int	แสดงถึงภาระงานสะสมในแต่ละ Course
Withdraw	int	แสดงถึงจำนวนเงินที่ของเบิก

### 3.3.2 Dimension Table

ตารางที่ 3.3 แสดงชนิดของข้อมูลใน Dimension Table ของจุดของเวลา ในตาราง Date เพื่อแสดง ช่วงเวลาของข้อมูลนั้นๆ

ชื่อตัวแปร	ชนิดข้อมูล	รายละเอียด
Date_key	int	Index ของตาราง Date
Academic_year	int	แสดงถึงปีการศึกษา
Semester	int	แสดงถึงภาคเรียนการศึกษา

ตารางที่ 3.4 แสดงชนิดของข้อมูลใน Dimension Table ของวิชาที่เปิดสอน ในตาราง Course เพื่อ แสดงรายละเอียดของนักศึกษาในแต่ละบุคคล

ชื่อตัวแปร	ชนิดข้อมูล	รายละเอียด
Course_Id	int	Index ของตาราง Course
Degree_Id	int	Index ของตาราง Dimension Degree
Course_Name	String	แสดงถึงชื่อวิชาที่เปิดสอน
Class	String	แสดงถึงห้องเรียน
Number_Of_Students	int	แสดงถึงจำนวนนักศึกษา
Remark	String	แสดงถึงหมายเหตุเพิ่มเติม
Number_PeriodsPerWeek	int	แสดงถึงจำนวนคาบเรียน ต่อ สัปดาห์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ตารางที่ 3.5** แสดงชนิดของข้อมูลใน Dimension Table ของสถานะการศึกษา ในตาราง Status เพื่อระบุสถานะการศึกษาในแต่ละภาคเรียน

ชื่อตัวแปร	ชนิดข้อมูล	รายละเอียด
Status_Id	int	Index ของตาราง Status
Status	String	แสดงถึงชื่อสถานะ

**ตารางที่ 3.6** แสดงชนิดของข้อมูลใน Dimension Table ของวุฒิการศึกษาที่เข้าศึกษา ในตาราง Degree เพื่อแสดงรายละเอียดของนักศึกษาในแต่ละบุคคล

ชื่อตัวแปร	ชนิดข้อมูล	รายละเอียด
Degree_Id	int	Index ของตาราง Degree
Degree_Name	String	แสดงถึงชื่อวุฒิการศึกษา
Major	String	แสดงถึงสาขาวิชาเอก

**ตารางที่ 3.7** แสดงชนิดของข้อมูลใน Dimension Table ของ นักศึกษา ในตาราง Student เพื่อแสดงรายละเอียดของนักศึกษาในแต่ละบุคคล

ชื่อตัวแปร	ชนิดข้อมูล	รายละเอียด
Student_Id	String	Index ของตาราง Student
Category_Id	int	Index ของตาราง Category แสดงถึงประเภทการเข้าศึกษาต่อของนักศึกษา
Sex_Id	int	Index ของตาราง Sex แสดงถึงเพศของนักศึกษา
Name	String	แสดงถึงชื่อนักศึกษา
Graduate_From	String	แสดงถึงสถานศึกษาที่จบการศึกษาเดิม
Graduate_GPA	float	แสดงถึงเกรดเฉลี่ยที่จบ จบการศึกษาจากสถานศึกษาเดิม
Phone_Number	String	แสดงถึงเบอร์โทรศัพท์ของนักศึกษา
Email	String	แสดงถึงอีเมลของนักศึกษา

**ตารางที่ 3.8** แสดงชนิดของข้อมูลใน Dimension Table ของอาจารย์ผู้สอน ในตาราง Category เพื่อแสดงรายละเอียดของประเภทการรับเข้าศึกษาต่อของนักศึกษา

ชื่อตัวแปร	ชนิดข้อมูล	รายละเอียด
Category_Id	int	Index ของตาราง Category
Category	String	แสดงถึงประเภทการรับเข้าของนักศึกษา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ตารางที่ 3.9** แสดงชนิดของข้อมูลใน Dimension Table ของเพศ ในตาราง Sex เพื่อแสดงรายละเอียดของเพศของนักศึกษา

ชื่อตัวแปร	ชนิดข้อมูล	รายละเอียด
Sex_Id	int	Index ของตาราง Sex
Sex	String	แสดงถึงชื่อนักศึกษา

**ตารางที่ 3.10** แสดงชนิดของข้อมูลใน Dimension Table ของจังหวัด ในตาราง Province เพื่อแสดงรายละเอียดของจังหวัดตามภูมิศาสตร์

ชื่อตัวแปร	ชนิดข้อมูล	รายละเอียด
Province_Id	int	Index ของตาราง Province
Province	String	แสดงถึงจังหวัดของนักศึกษา

**ตารางที่ 3.11** แสดงชนิดของข้อมูลใน Dimension Table ของอาจารย์ผู้สอน ในตาราง Teacher เพื่อแสดงรายละเอียดของอาจารย์ในแต่ละบุคคล

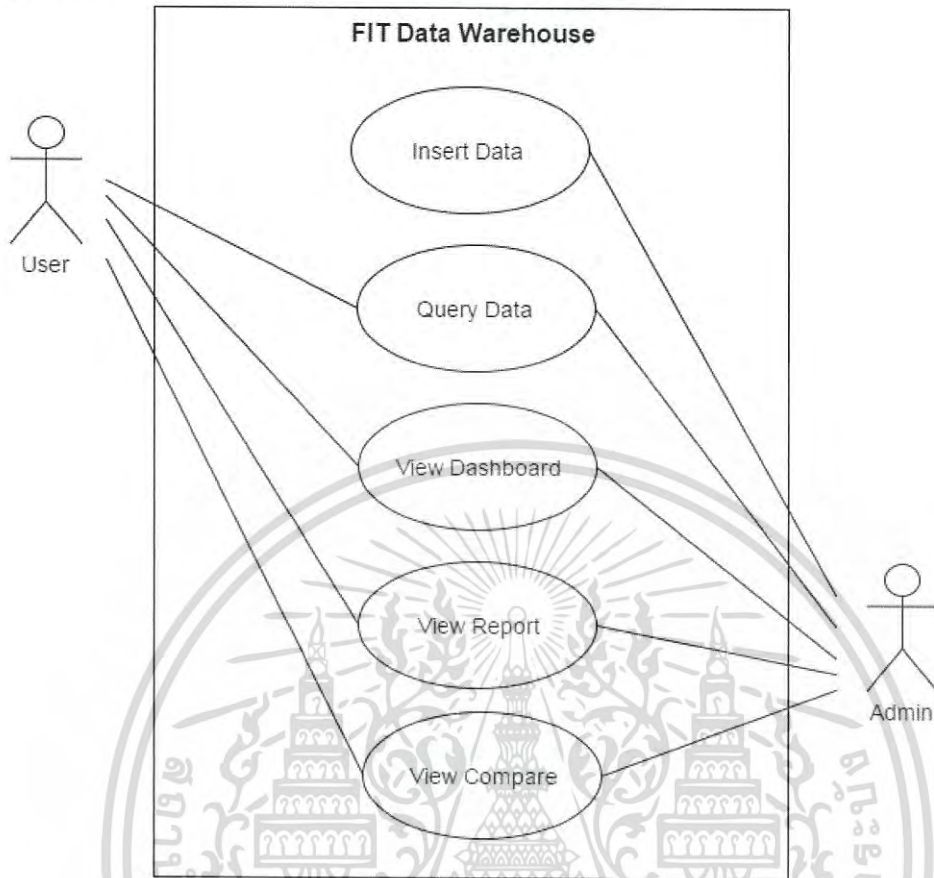
ชื่อตัวแปร	ชนิดข้อมูล	รายละเอียด
Teacher_Id	int	Index ของตาราง Teacher
Teacher_Name	String	แสดงถึงชื่ออาจารย์
Position	String	แสดงถึงตำแหน่งของอาจารย์
Base_Workload	int	แสดงถึงภาระงาน(จำนวนครั้งการสอน)

### 3.3.3 Other Table

**ตารางที่ 3.12** แสดงชนิดของข้อมูลในตาราง Staff\_Account เพื่อใช้ในการยืนยันตัวตนก่อนเข้าระบบ

ชื่อตัวแปร	ชนิดข้อมูล	รายละเอียด
Name	String	แสดงถึงชื่อผู้ใช้งาน
Username	String	แสดงถึงชื่อผู้ใช้งานในการเข้าสู่ระบบ
Password	String	แสดงถึงรหัสในการเข้าสู่ระบบ [MD5]
Role	String	แสดงถึงบทบาทในการเข้าถึงข้อมูล

### 3.4 การออกแบบยูสเคสไดอะแกรม (Use Case Diagram)



รูปที่ 3.5 แสดงการออกแบบยูสเคส ไดอะแกรม

#### คำอธิบาย Use Case ระบบคลังข้อมูลของคณะเทคโนโลยีสารสนเทศ

จากรายละเอียดของ “ระบบคลังข้อมูลของคณะเทคโนโลยีสารสนเทศ” และ Use Case Diagram ที่แสดงในรูปที่ 3.5 คำอธิบายของแต่ละ Use Case มีดังนี้

ตารางที่ 3.13 แสดงรายละเอียดของ Use Case เพิ่มข้อมูลลงคลังข้อมูลในแต่ละตาราง

Use Case name : Insert Data		Use Case ID : 1
Stakeholder Actor : ผู้ใช้ดูข้อมูลบนระบบคลังข้อมูล		
Stakeholder Actor : ผู้ใช้ที่มีสิทธิ์ในการเข้าถึงแบบ "Admin" เท่านั้น		
Trigger Event : ผู้ใช้เลือกเมนู "เพิ่ม/ดูข้อมูลทั้งหมด"		
Pre Conditions : ต้องเชื่อมต่อกับอินเทอร์เน็ตโดยเข้าผ่าน Browser		
Post conditions : -		
Main Flow : ผู้ใช้งานสามารถเพิ่มข้อมูลในแต่ละตารางในคลังข้อมูลได้ โดยการ โหลด Template เพื่อใช้ในการ Import Data ลงไปในระบบ		
Flow of Events :	<ol style="list-style-type: none"> <li>1. ผู้ใช้เข้าสู่ระบบ</li> <li>2. เข้าสู่ระบบหน้า "เพิ่ม/ดูข้อมูลทั้งหมด"</li> <li>3. ผู้ใช้ดาวน์โหลด Template และเลือกไฟล์.csv ที่ต้องการ</li> <li>4. กดปุ่ม "Update Data &gt;"</li> <li>5. จะแสดงมาหน้าเดิม พร้อมบอกสถานะการเพิ่มข้อมูลลงระบบ</li> </ol>	
Alternative Flow of Events :	<ol style="list-style-type: none"> <li>1A. หากผู้ใช้เลือกไฟล์.csv ที่กรอกข้อมูลไม่ถูกต้อง               <ol style="list-style-type: none"> <li>1A.1) มีการแจ้งเตือนผู้ใช้ว่าเพิ่มข้อมูลสำเร็จ</li> <li>1A.2) หากเพิ่มข้อมูลผิดพลาด จะมีข้อความแจ้งเตือนว่าเพิ่มข้อมูลผิดพลาด ให้ลองเพิ่มข้อมูลใหม่อีกครั้ง</li> </ol> </li> </ol>	

ตารางที่ 3.14 แสดงรายละเอียดของ เรียกดูข้อมูล ตามที่ผู้ใช้งานต้องการ

Use Case name : Query Data		Use Case ID : 2
Stakeholder Actor : ผู้ใช้ดูข้อมูลบนระบบคลังข้อมูล		
Stakeholder Actor : -		
Trigger Event : ผู้ใช้เลือกเมนู "เรียกดูข้อมูล"		
Pre Conditions : ต้องเชื่อมต่อกับอินเทอร์เน็ตโดยเข้าผ่าน Browser		
Post conditions : -		
Main Flow : ผู้ใช้งานสามารถค้นหาข้อมูลในแต่ละตารางได้ โดยการพิมพ์ในช่องค้นหา		
Flow of Events :	<ol style="list-style-type: none"> <li>1. ผู้ใช้เข้าสู่ระบบ</li> <li>2. เข้าสู่ระบบหน้า "เรียกดูข้อมูล"</li> <li>3. ผู้ใช้เลือกข้อมูลที่ต้องการจะดู</li> <li>4. กรอกข้อมูลตรงช่องค้นหา</li> <li>5. จะแสดงข้อมูลที่ผู้ใช้ค้นหา</li> </ol>	
Alternative Flow of Events :		

ตารางที่ 3.15 แสดงรายละเอียดของ Use Case ดูแดชบอร์ดโดยรวมของระบบ (Dashboard)

Use Case name : View Dashboard	Use Case ID : 3
Stakeholder Actor : ผู้ใช้ดูข้อมูลบนระบบคลังข้อมูล	
Stakeholder Actor : -	
Trigger Event : ผู้ใช้เลือกเมนู “แดชบอร์ด”	
Pre Conditions : ต้องเชื่อมต่อกับอินเทอร์เน็ตโดยเข้าผ่าน Browser	
Post conditions : -	
Main Flow : ผู้ใช้งานเข้าสู่หน้า “แดชบอร์ด” เพื่อให้ระบบคำนวณ แล้วแสดงข้อมูลเป็นแผนภูมิภาพในรูปแบบ ง่ายสรุปมาแล้ว โดยจะดึงข้อมูลของปีการศึกษาที่แล้วมาแสดง	
Flow of Events :	<ol style="list-style-type: none"> <li>1. ผู้ใช้เข้าสู่ระบบ</li> <li>2. เข้าสู่ระบบหน้า “แดชบอร์ด”</li> <li>3. เลือกระดับนักศึกษา/อาจารย์ ที่ต้องการดู</li> <li>3. แสดงข้อมูลในรูปแบบแผนภูมิสรุปข้อมูล</li> </ol>
Alternative Flow of Events :	-

ตารางที่ 3.16 แสดงรายละเอียดของ ข้อมูลแผนภูมิภาพ ตามที่ผู้ใช้งานต้องการ

Use Case name : View Report	Use Case ID : 4
Stakeholder Actor : ผู้ใช้ดูข้อมูลบนระบบคลังข้อมูล	
Stakeholder Actor : -	
Trigger Event : ผู้ใช้เลือกเมนู “รายงาน”	
Pre Conditions : ต้องเชื่อมต่อกับอินเทอร์เน็ตโดยเข้าผ่าน Browser	
Post conditions : -	
Main Flow : ผู้ใช้งานเข้าสู่หน้า “รายงาน” เพื่อดูข้อมูลแบบแผนภูมิภาพ ตามที่ผู้ใช้งานกำหนดค่า	
Flow of Events :	<ol style="list-style-type: none"> <li>1. ผู้ใช้เข้าสู่ระบบ</li> <li>2. เข้าสู่ระบบหน้า “รายงาน”</li> <li>3. เลือก ข้อมูลที่ต้องการสร้าง นักศึกษา หรือ อาจารย์</li> <li>4. เลือกข้อมูลที่ต้องการที่จะสร้าง แผนภูมิภาพ</li> <li>5. กดปุ่ม “สร้างรายงาน” เพื่อแสดงแผนภูมิภาพ</li> </ol>
Alternative Flow of Events :	-

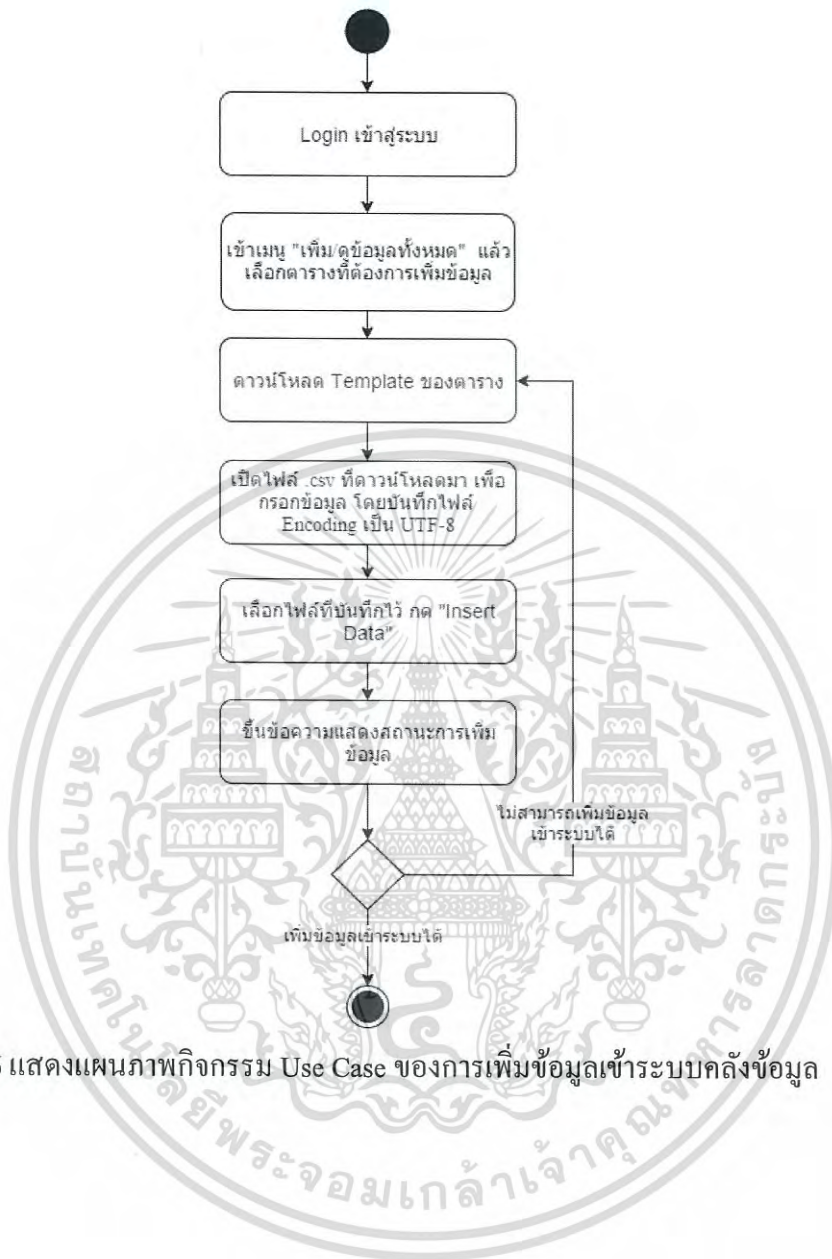
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.17 แสดงรายละเอียดของ แผนภูมิภาพเชิงเปรียบเทียบ 2 ค่า ตามที่ผู้ใช้งานต้องการ

Use Case name : View Compare		Use Case ID : 5
Stakeholder Actor : ผู้ใช้ข้อมูลบนระบบคลังข้อมูล		
Stakeholder Actor : -		
Trigger Event : ผู้ใช้เลือกเมนู “เปรียบเทียบ”		
Pre Conditions : ต้องเชื่อมต่อกับอินเทอร์เน็ตโดยเข้าผ่าน Browser		
Post conditions : -		
Main Flow : ผู้ใช้งานเข้าสู่หน้า “เปรียบเทียบ” เพื่อดูข้อมูลแบบแผนภูมิภาพเปรียบเทียบ ตามที่ผู้ใช้งานกำหนดค่า		
Flow of Events :	<ol style="list-style-type: none"> <li>1. ผู้ใช้เข้าสู่ระบบ</li> <li>2. เข้าสู่ระบบหน้า “เปรียบเทียบ”</li> <li>3. เลือก “กำหนดค่าเปรียบเทียบ1” และ ใน “กำหนดค่าเปรียบเทียบ1” เพื่อกำหนดค่าเปรียบเทียบข้อมูลนักศึกษา</li> <li>4. กดปุ่ม “สร้างรายงานการเปรียบเทียบ” เพื่อแสดงแผนภูมิภาพ</li> </ol>	
Alternative Flow of Events :	-	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

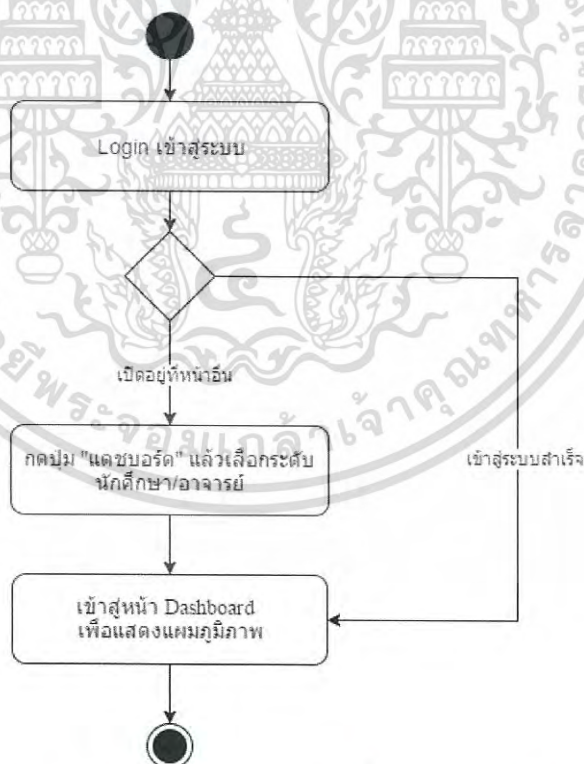
### 3.5 แผนภาพกิจกรรม (Activity Diagram)



รูปที่ 3.6 แสดงแผนภาพกิจกรรม Use Case ของการเพิ่มข้อมูลเข้าระบบคลังข้อมูล



รูปที่ 3.7 แสดงแผนภาพกิจกรรม Use Case ของการเรียกดู/ค้นหาข้อมูลในระบบคลังข้อมูล



รูปที่ 3.8 แสดงแผนภาพกิจกรรม Use Case ของการดูแคชบอร์ดแสดงแผนภูมิภาพ  
โดยสรุปในระบบคลังข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.9 แสดงแผนภาพกิจกรรม Use Case ของการดูรีพอร์ตแสดงแผนภูมิภาพ โดยสรุปตามที่ใช้กำหนดเงื่อนไขในระบบคลังข้อมูล

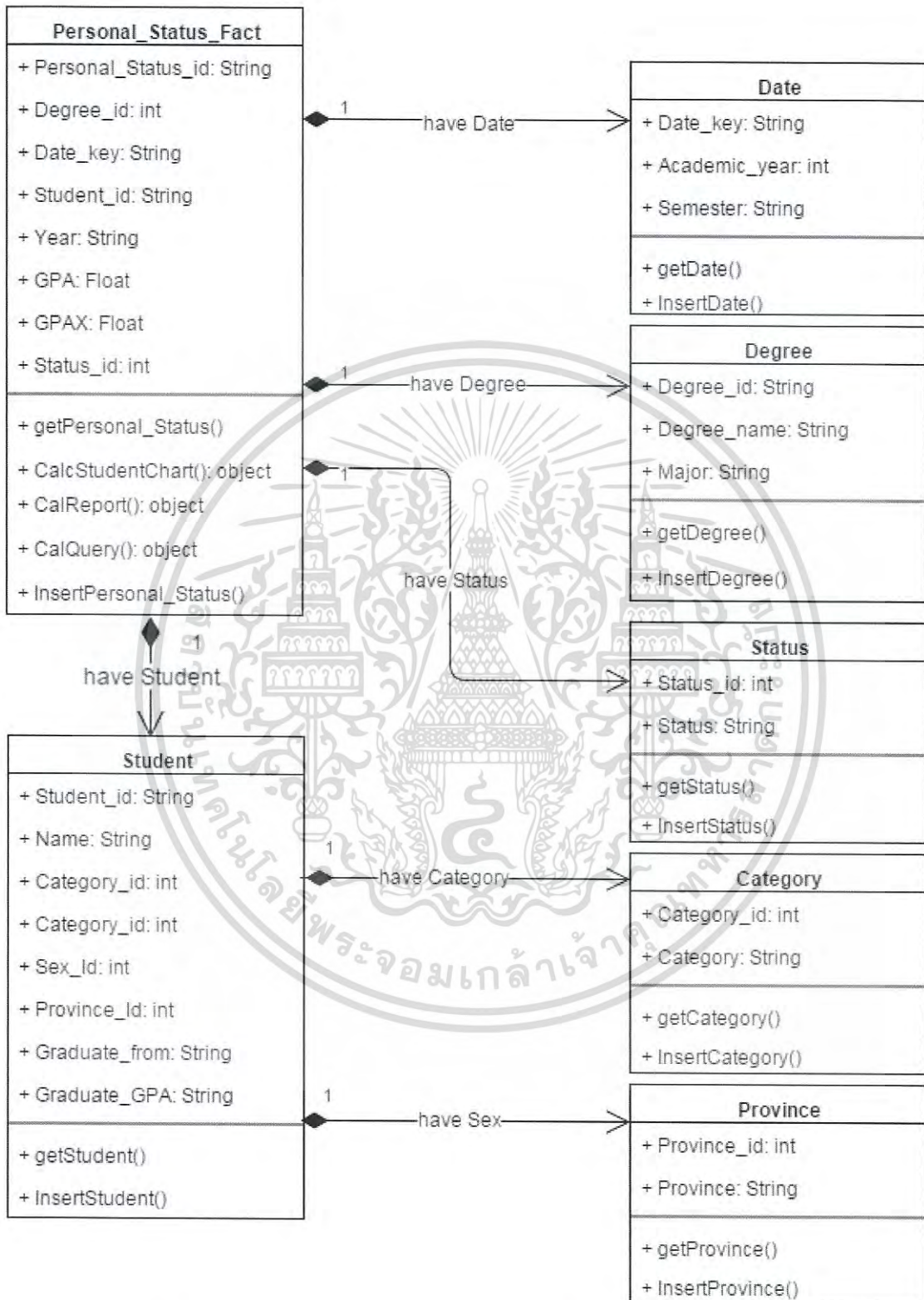


รูปที่ 3.10 แสดงแผนภาพกิจกรรม Use Case ของการดูรีพอร์ตแสดงแผนภูมิภาพ

โดยสรุปตามที่ใช้กำหนดเงื่อนไขในระบบคลังข้อมูล

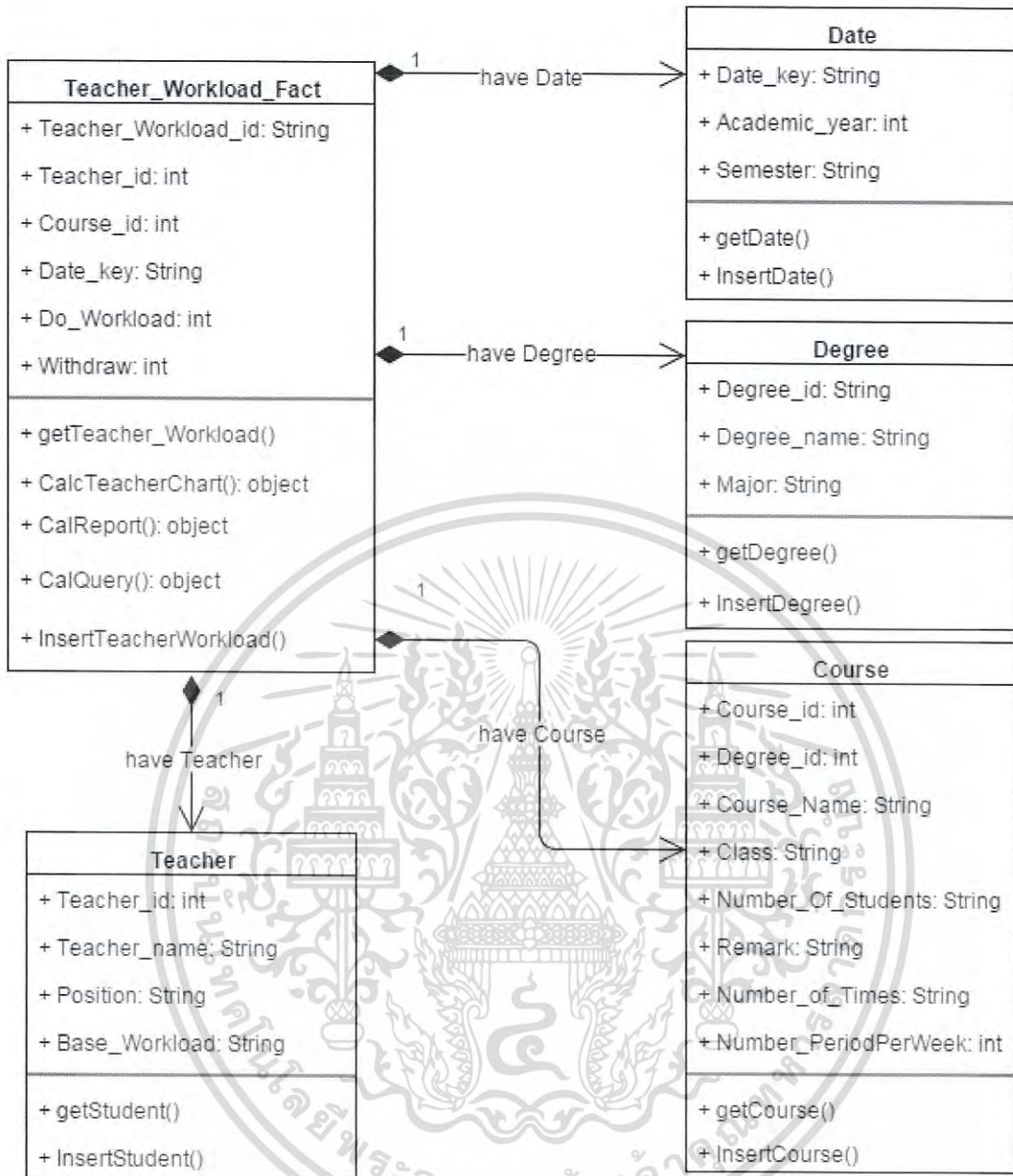
เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้กับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.6 แผนภาพแสดงคลาส (Class Diagram)



รูปที่ 3.11 แสดง Class Diagram ของระบบคลังข้อมูล(ส่วนที่ 1 นักศึกษา)

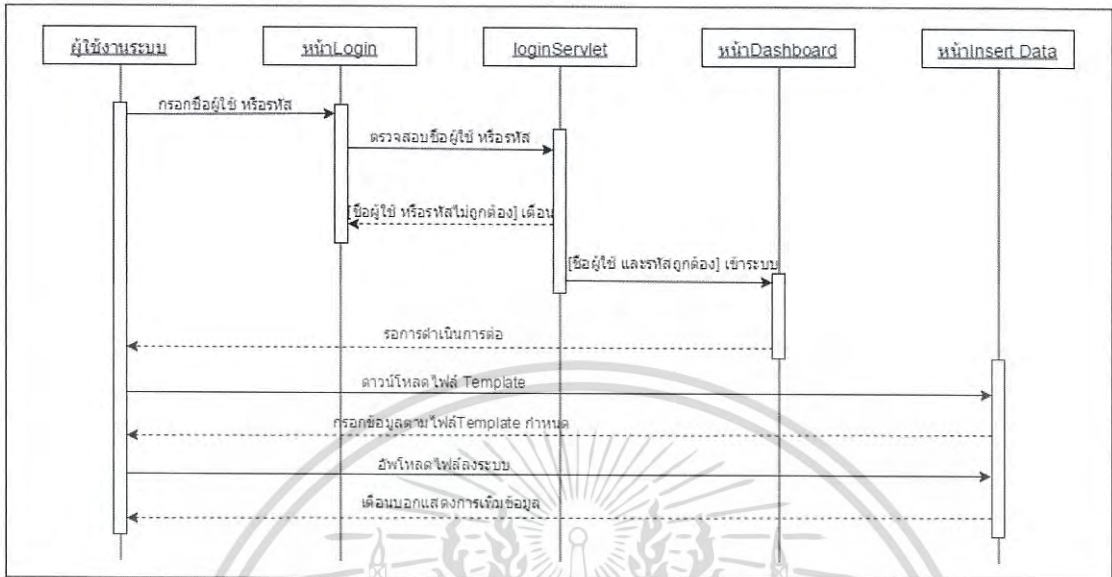
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



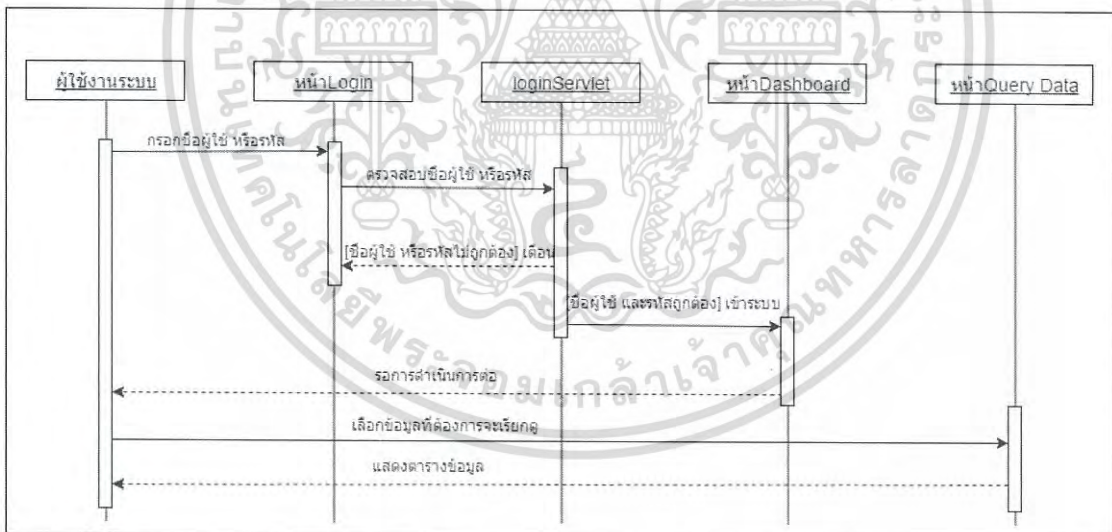
รูปที่ 3.12 แสดง Class Diagram ของระบบคลังข้อมูล(ส่วนที่2 ครู/อาจารย์)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.7 แผนภาพแสดงลำดับการทำงาน (Sequence Diagram)

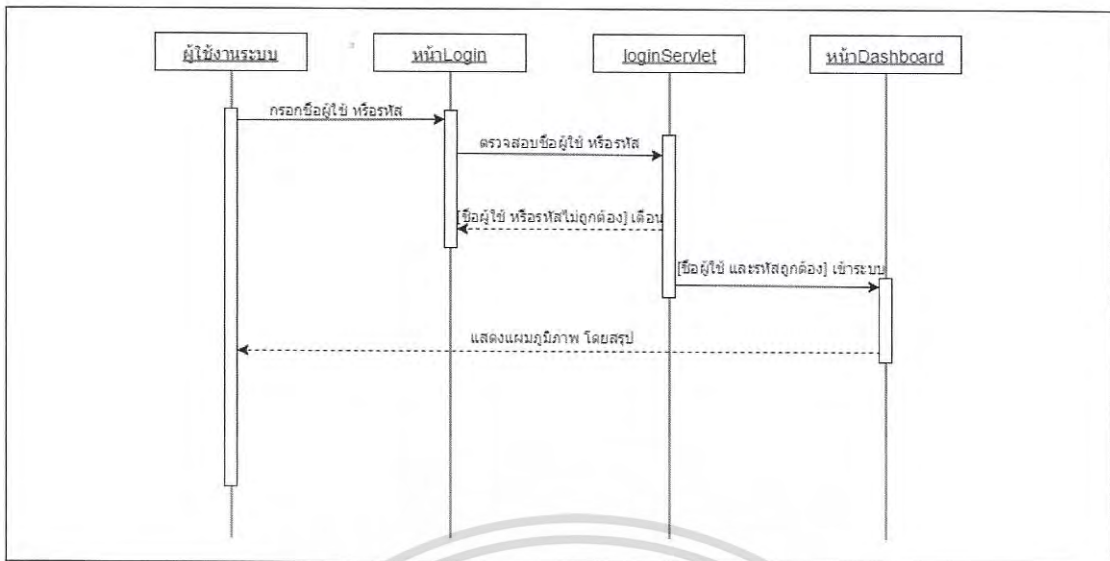


รูปที่ 3.13 แสดงแผนแสดงลำดับการทำงานของการเพิ่มข้อมูลลงระบบแต่ละตาราง

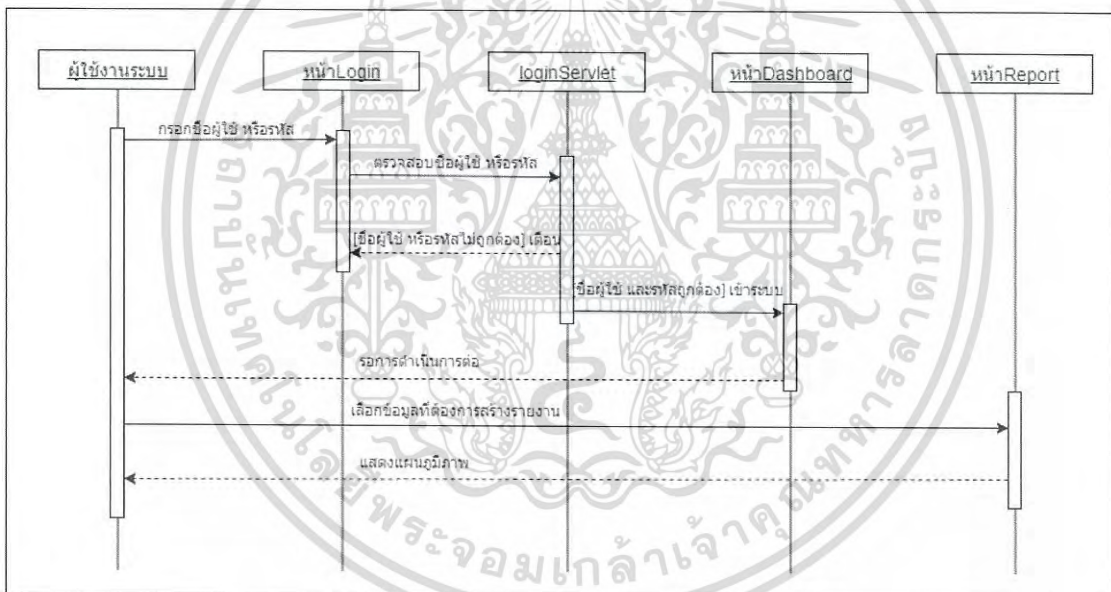


รูปที่ 3.14 แสดงแผนแสดงลำดับการทำงานของการเรียกดูหรือค้นหาข้อมูลบนระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

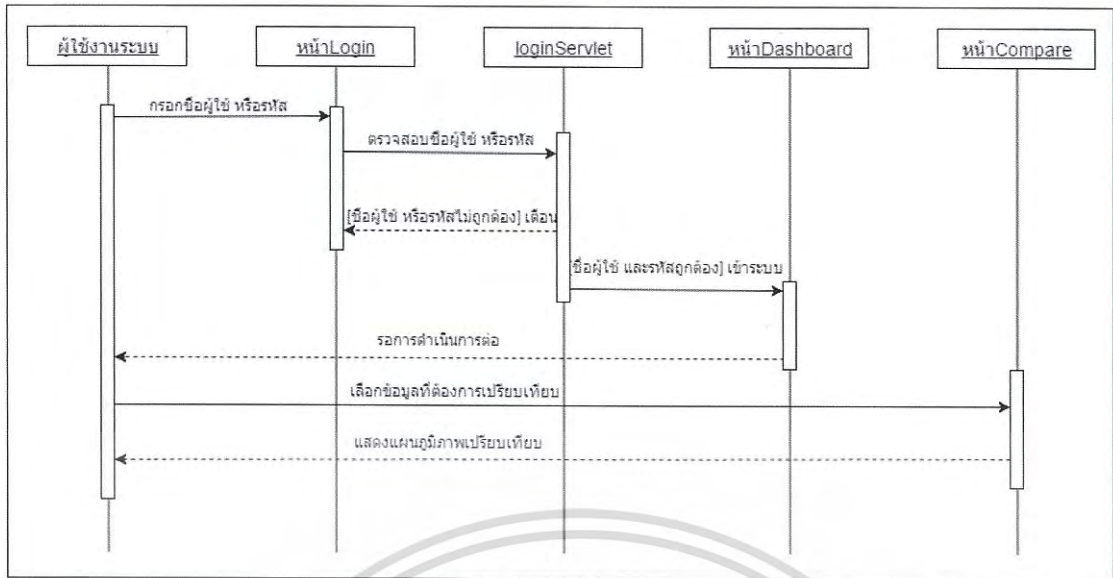


รูปที่ 3.15 แสดงแผนแสดงลำดับการทำงานของการทำงานของการแคชบอร์ด โดยจะแสดงแผนภูมิภาพแบบสรุป



รูปที่ 3.16 แสดงแผนแสดงลำดับการทำงานของการทำงานการรีพอร์ตเพื่อแสดงแผนภูมิภาพตามที่ผู้ใช้งานกำหนด

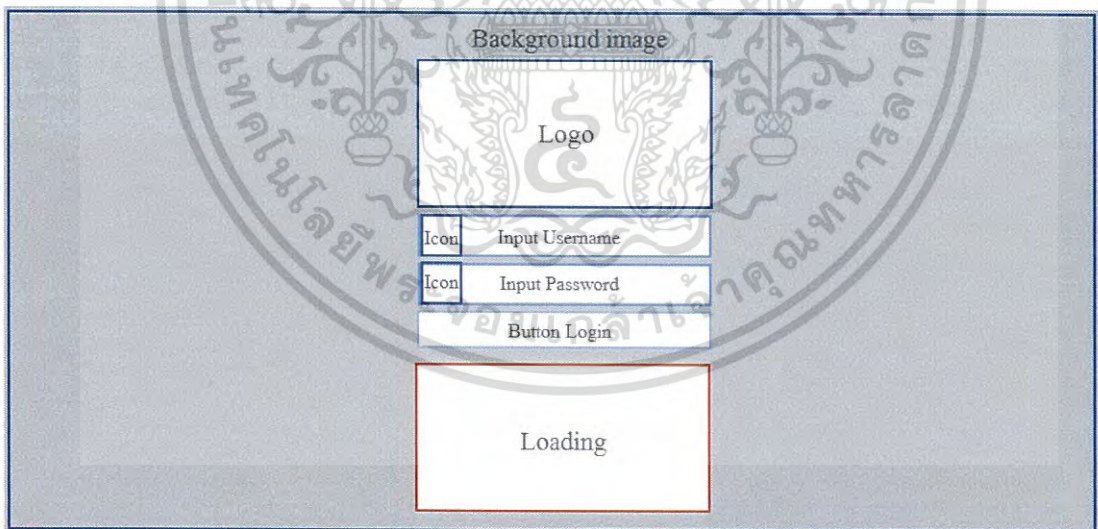
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.17 แสดงแผนแสดงลำดับการทำงานของการทำงานของการเปรียบเทียบเพื่อแสดงแผนภูมิภาพตามที่ผู้ใ้กำหนด

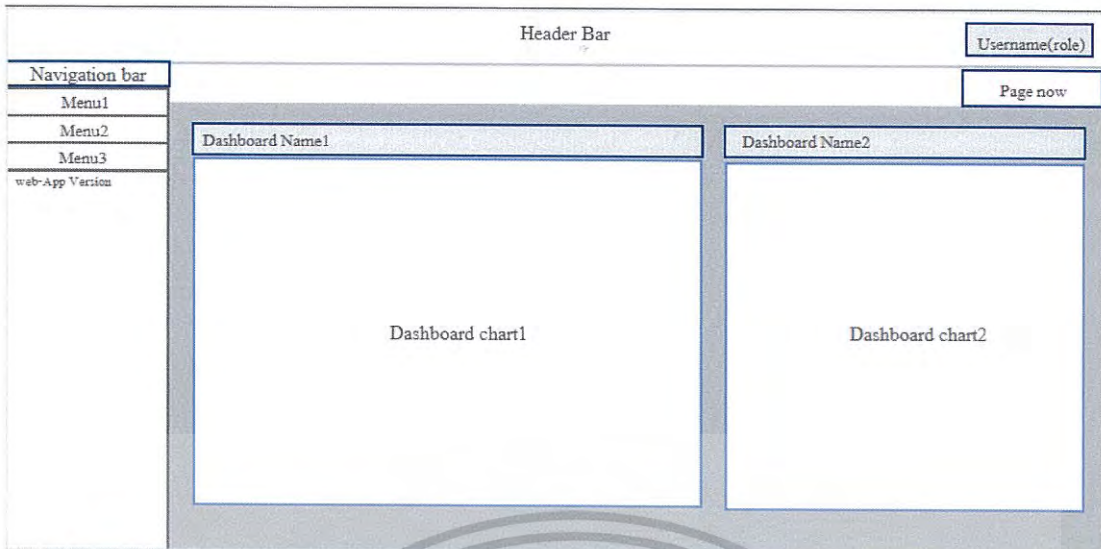
### 3.8 ออกแบบส่วนติดต่อกับผู้ใช้

#### 3.8.1 ไวร์เฟรม (Wireframe)

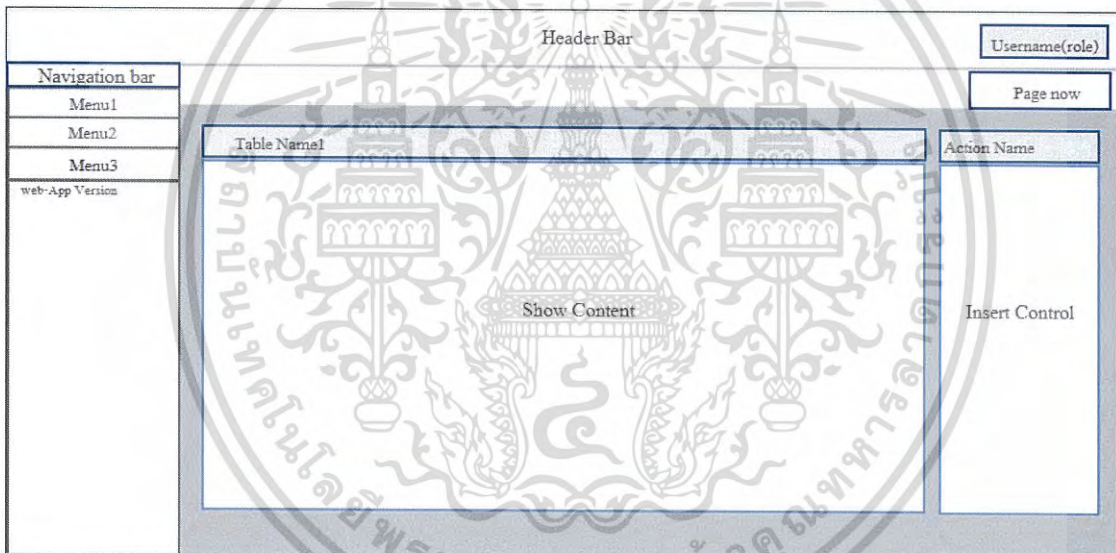


รูปที่ 3.18 แสดงไวร์เฟรมหน้าแรก(Login) ของระบบคลังข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.19 แสดงไวร์เฟรมหน้าแดชบอร์ด(Dashboard)และรีพอร์ต(Report) ของระบบคลังข้อมูล



รูปที่ 3.20 แสดงไวร์เฟรมหน้าเพิ่ม/คิวรีข้อมูล(Put/Query Data) ของระบบคลังข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 4

### ผลการดำเนินงานของโครงการ

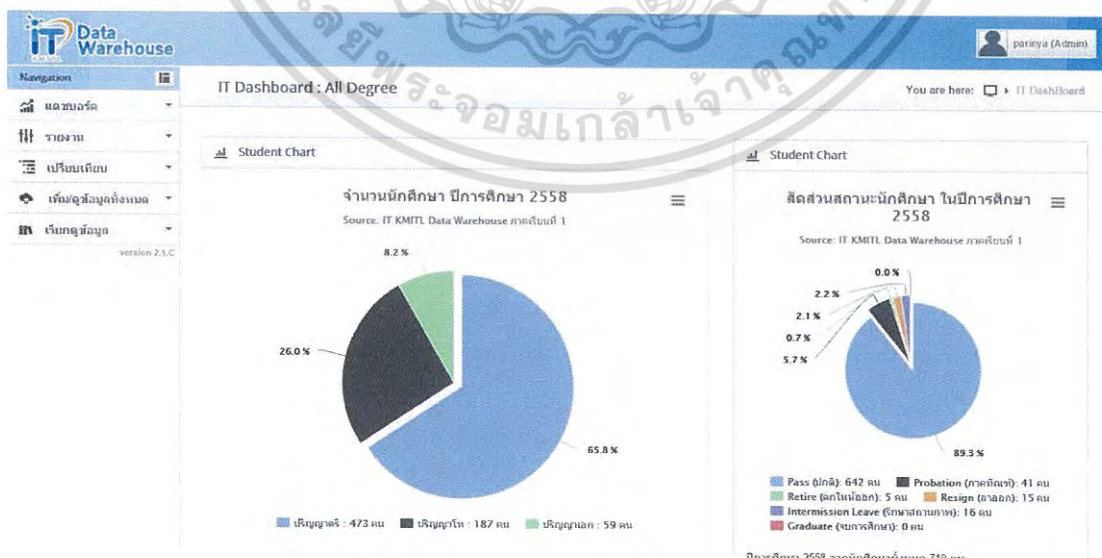
#### 4.1 เว็บแอปพลิเคชัน (Web Application)

4.1.1 ส่วนของหน้าล็อกอิน (Login) เพื่อตรวจสอบสิทธิ์การเข้าถึงระบบไว้ได้รับสิทธิ์ในการเข้าระบบหรือไม่



รูปที่ 4.1 แสดงส่วนของหน้าล็อกอิน

#### 4.1.2 ส่วนของหน้าแดชบอร์ด (Dashboard) ซึ่งเป็นหน้าหลักในการใช้งาน



รูปที่ 4.2 แสดงส่วนของหน้าแดชบอร์ด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 4.2 ส่วนควบคุมการทำงานของระบบ

### 4.2.1 ส่วนรายละเอียดในแต่ละหน้า (Header Bar)



รูปที่ 4.3 แสดงส่วนเกี่ยวข้องกับผู้ใช้งานและแสดงชื่อหน้า ณ ปัจจุบัน

#### 4.2.1.1 ส่วนเกี่ยวข้องกับผู้ใช้งาน



รูปที่ 4.4 แสดงส่วนเกี่ยวข้องกับผู้ใช้งาน

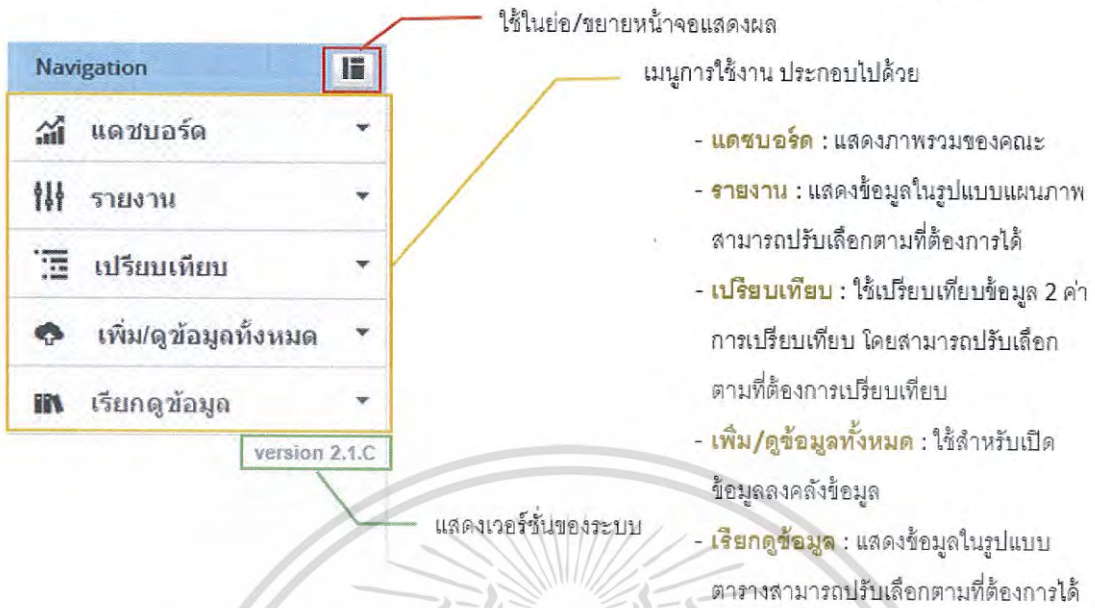
#### 4.2.1.2 ส่วนแสดงชื่อหน้า ณ ปัจจุบัน (Title Bar)



รูปที่ 4.5 แสดงชื่อหน้า ณ ปัจจุบัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

#### 4.2.2 ส่วนเมนูหน้าในแต่ละหน้า (Navigation Bar) และรายละเอียดแต่ละเมนู



รูปที่ 4.6 แสดงส่วนของเมนูการควบคุม



รูปที่ 4.7 แสดงส่วนของเมนูแดชบอร์ดภาพรวมของคณะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รายงาน

สถานะนักศึกษา

ปีการศึกษา :

ปีการศึกษาเริ่มต้น

ถึง :

ปีการศึกษาสุดท้าย

ภาคเรียน :

ทั้งหมด

รุ่นปี :

ทั้งหมด

ปี :

ทั้งหมด

ระดับการศึกษา :

ทั้งหมด

ประเภทการรับเข้า :

ทั้งหมด

สถานศึกษา :

ทั้งหมด

สร้างรายงานนักศึกษา >

เมื่อคลิกปุ่ม "รายงาน" แล้วกดเลือก นักศึกษา หรือ อาจารย์ ที่ต้องการ  
เรียกดูรายงานแผนภูมิภาพ

โดยเลือกข้อมูลที่ต้องการให้ระบบแสดงแผนภูมิภาพ เมื่อเลือกเสร็จให้  
กดปุ่ม "สร้างรายงาน >"

ปุ่มสำหรับกดเพื่อแสดงผลรายงานแผนภูมิ  
ภาพตามที่กำหนดด้านบน

รูปที่ 4.8 แสดงส่วนประกอบย่อยของแต่ละเมนูการควบคุม  
(เมนู: รายงาน, เปรียบเทียบ, เรียกดูข้อมูล จะมีลักษณะการทำงานที่เหมือนกัน)

อาจารย์ :

ทั้งหมด

ทั้งหมด

รศ.ดร. นพพร

ผศ.ดร. กิติสุชาติ

อาจารย์ศรีนวล

รศ.ดร. โชติพิชัย

ดร.มานพ

ดร.สุนทร

ผศ.ดร. สุขสันต์

ดร.สุภกิจ

ผศ.ดร. กิ่งตพงษ์

รศ.ดร. พรฤดี

ผศ.ดร. สมเกียรติ

ดร.ธีรพงษ์

ผศ.ดร. กิทธิชัย

รศ.ดร. จันทร์บุรณี

ผศ.บุญประเสริฐ

ดร.นล

ดร.กนกวรรณ

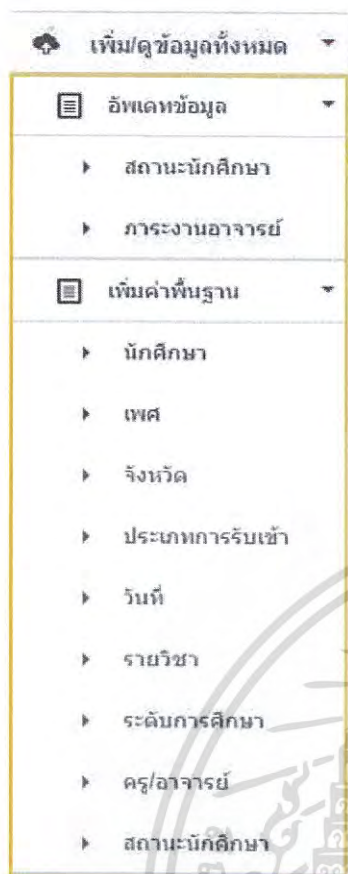
ผศ.ดร.ธนิศา

ผศ.ดร. ณเรศพล

เมื่อคลิกปุ่ม เพื่อเลือกข้อมูลที่ต้องการให้ระบบทำรายงาน  
[โดยข้อมูลที่แสดงระบบจะดึงข้อมูลมาจากคลังข้อมูล]

รูปที่ 4.9 แสดงส่วนควบคุมของการเลือกข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



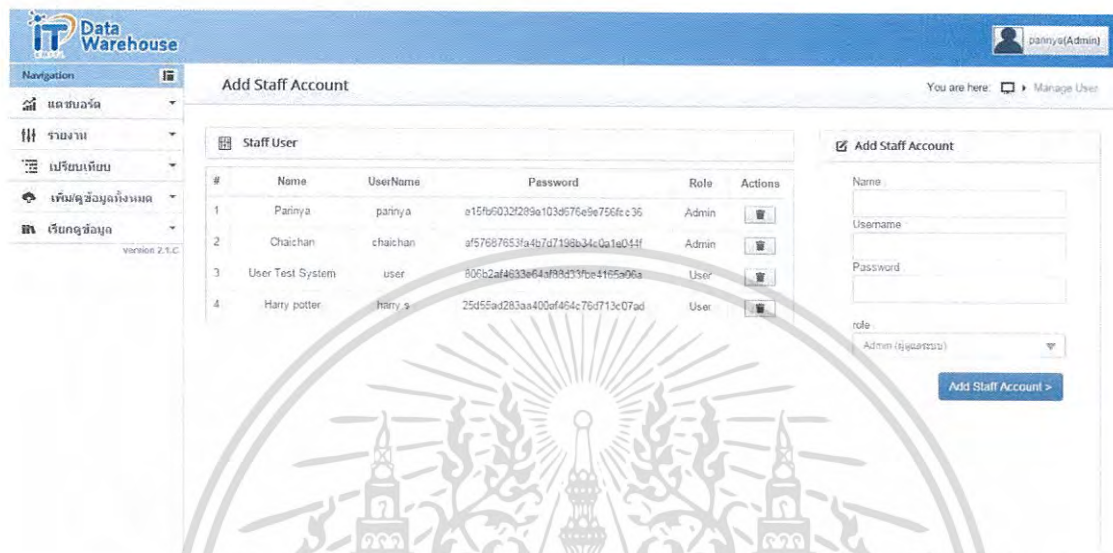
เมื่อกดปุ่ม "เพิ่ม/เรียกดูข้อมูล" แล้วกด อับเดทข้อมูล หรือ เพิ่มค่าพื้นฐาน เพื่อเลือกตารางที่ต้องการ เพิ่มข้อมูล ลงระบบ

รูปที่ 4.10 แสดงส่วนของเมนูการควบคุม ในส่วน Put/Query Data

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

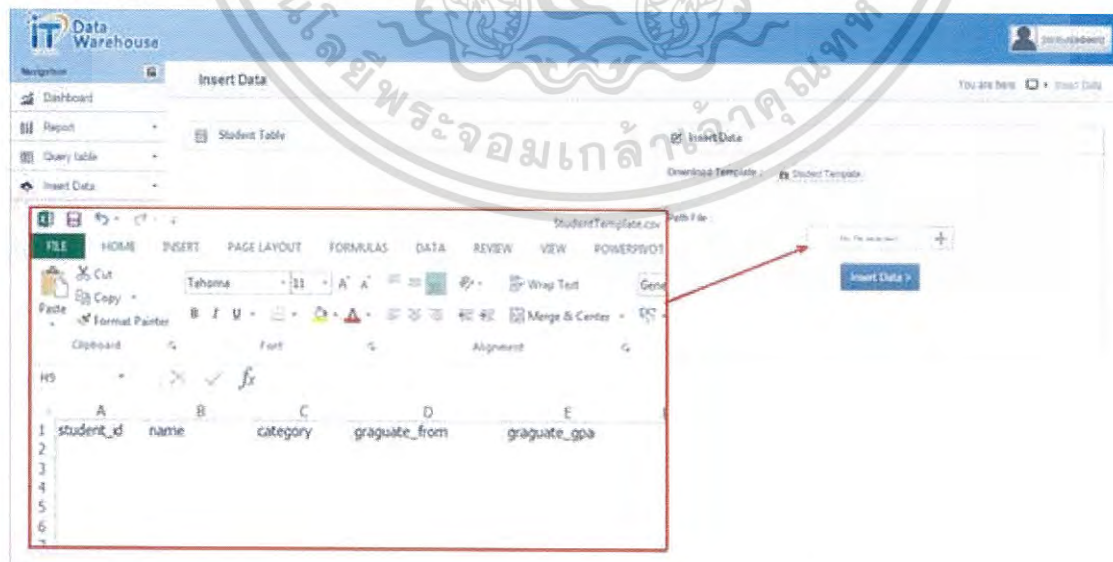
## 4.3 ส่วนการประมวลผลข้อมูล

4.3.1 ส่วนของหน้าล็อกอิน (Login) สามารถดึงข้อมูลจากคลังข้อมูล โดยใช้ ภาษาอิมพาลา (Impala) ได้ โดย รหัสผ่านจะถูกเข้ารหัสด้วย MD5 เพื่อความปลอดภัยของข้อมูลและมีตัวกรองไม่ให้ผู้ใช้ที่ไม่ได้ล็อกอินเข้าหน้าอื่นๆบนระบบได้



รูปที่ 4.11 แสดงหน้าของสิทธิ์การเข้าถึงข้อมูล

4.3.2 ส่วนการเพิ่ม/อัปเดตข้อมูลลงคลังข้อมูล โดยการดาวน์โหลด Template ที่มีให้และกรอกข้อมูลแล้วทำการอัปโหลดไฟล์ .csv เข้าไปในระบบ โดยการกดปุ่ม Update / Insert Data เข้าไปในระบบ



รูปที่ 4.12 แสดงส่วนของการเพิ่มข้อมูลลงระบบคลังข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Navigation: แสดงบอร์ด, รายงาน, เปรียบเทียบ, เพิ่ม/ดูข้อมูลทั้งหมด, เริ่มดูข้อมูล

version 2.1.C

You are here: Insert/Query Data

**Insert Data**

Download Template: Student

Path File: No file selected

Insert Data >

Student ID	Name	Category ID	Graduate From	Graduate GPA	Province	Sex	Phone Number	E-mail
55070001	นายสุริย นาคเทม	[1] รับตรง (ป.ล.)	เซนต์จอห์น	3.24	[21] นครราชสีมา	[1] ชาย	081-9237824	a001@gmail.com
55070002	นางฉัตรณี นนทสุข	[2] Admission (ป.ล.)	พนมพรเกษม พนมฉวีวิทยา	2.92	[1] กรุงเทพมหานคร	[1] ชาย	081-9237825	a002@gmail.com
55070003	นายสุทัศน์ สุขเกษม	[4] โฉดจา (ป.ล.)	เตรียมอุดมศึกษาพัฒนา	3.38	[18] นครนายก	[1] ชาย	081-9237826	a003@gmail.com
55070004	นายฉัตร อังมลระพี	[2] Admission (ป.ล.)	หัวหินวิทยาลัย	3.31	[60] สมุทรปราการ	[1] ชาย	081-9237827	harry.p@gmail.com
55070005	นายภูธร วัฒนวิเศษ	[1] รับตรง (ป.ล.)	วิเศษวิทย	3.4	[62] สมุทรสาคร	[1] ชาย	081-9237828	love.u@hotmail.com
55070006	นายสุวิทย์ นนทสุข	[1] Admission (ป.ล.)	เซนต์จอห์น	3.24	[21] นครราชสีมา	[1] ชาย	081-9237824	a001@gmail.com

รูปที่ 4.13 แสดงหน้าตารางบนคลังข้อมูล (โดยลักษณะข้อมูลจะต่างกัน ตามที่เลือกแสดงข้อมูล)

Navigation: แสดงบอร์ด, รายงาน, เปรียบเทียบ, เพิ่ม/ดูข้อมูลทั้งหมด, เริ่มดูข้อมูล

version 2.1.C

You are here: Insert/Query Data

**Update Data**

Download Template: อาจารย์, ศึกษานิเทศก์, ศึกษานิเทศก์, 2554, ภาคเรียน: T.Workload Download

Path File: No file selected

Update Data >

Add Teacher Workload Fact Data Fail, Please try again!

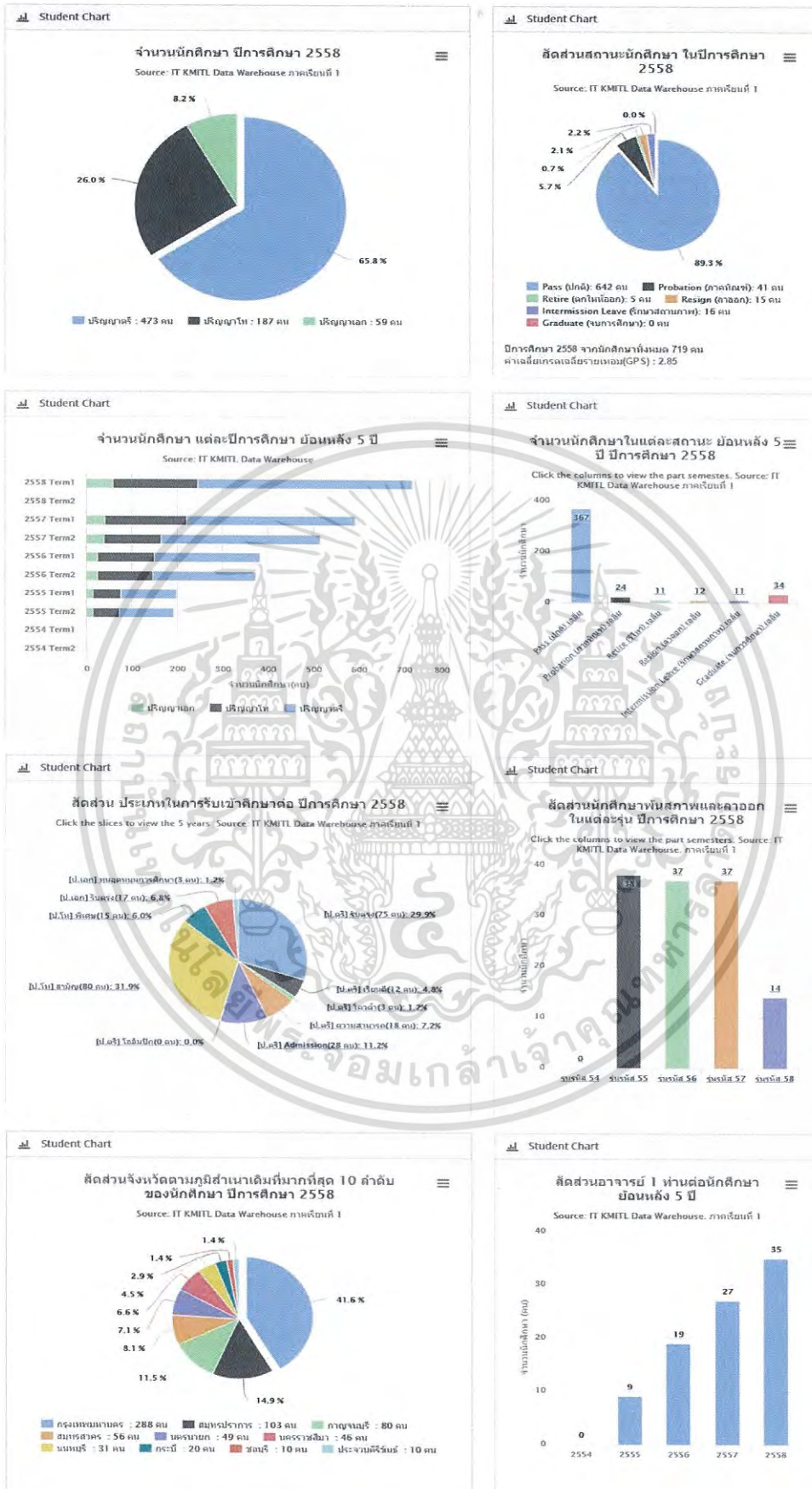
Can't Insert, Duplicated Teacher\_Fact ID : 110001.601622101.255401, 110001.601622102.255401

Teacher Fact ID	Teacher ID	Date Key	Course ID	DoWorkload(time)	Withdraw(Baht)
110001.601611601.255402	110001	255402	601611601	15	3000
110001.601611601.255502	110001	255502	601611601	15	3000
110001.601611601.255602	110001	255602	601611601	15	3000
110001.601611601.255702	110001	255702	601611601	15	3000
110001.601611601.255802	110001	255802	601611601	15	3000
110001.601611701.255501	110001	255501	601611701	15	3000
110001.601611701.255601	110001	255601	601611701	15	3000
110001.601611701.255701	110001	255701	601611701	15	3000
110001.601611701.255801	110001	255801	601611701	15	3000

รูปที่ 4.14 แสดงหน้าตารางบนคลังข้อมูล (เมื่อมีการเพิ่มข้อมูลซ้ำซ้อนระบบจะมีการแจ้งเตือน)

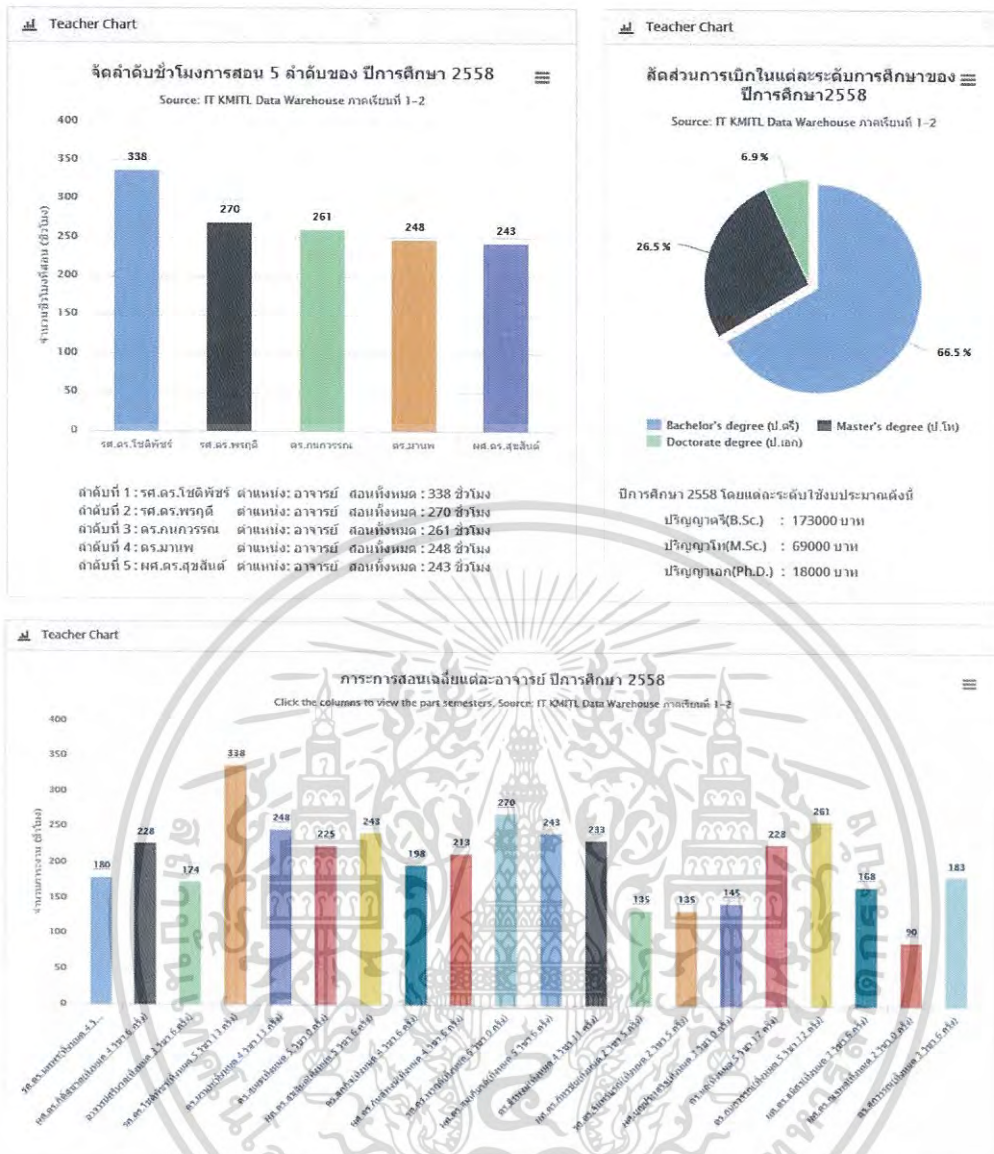
4.3.3 ส่วนของหน้าแดชบอร์ด (Dashboard) ซึ่งเป็นหน้าหลักในการใช้งาน ซึ่งการดึงข้อมูลจากคลังข้อมูล โดยการใช้ jQuery ในการดึงเพื่อมาแสดงผ่าน JavaScript (Highchart)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.15 แสดงส่วนหน้าแดชบอร์ดนักศึกษา (Dashboard) สรุปรวมทั้งระบบ

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้กับการแข่งขันเพื่อการพัฒนาเท่านั้น เมื่อผู้ผู้เห็นใช้ขอขอร้องขอขานการดำเนินการ  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

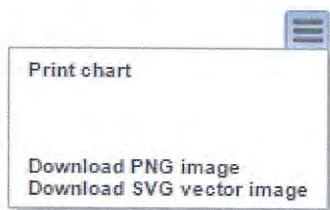


รูปที่ 4.16 แสดงส่วนหน้าแดชบอร์ดครู/อาจารย์ (Dashboard) สรุปรวมทั้งระบบ

โดยผู้ใช้งานจะสามารถนำแผนภูมิภาพออกจากระบบได้ 3 แบบ ดังนี้

- Print chart: เพื่อส่งคำสั่งปริ้นแผนภูมิภาพ
- Download PNG image: เพื่อส่งคำสั่งเซฟเป็นไฟล์รูปภาพสกุล PNG
- Download SVG vector image: เพื่อส่งคำสั่งเซฟเป็นไฟล์รูปภาพสกุล SVG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.17 แสดงส่วนประกอบหน้าแดชบอร์ดครู/อาจารย์ (Dashboard) ในการทำแผนภูมิภาพออกจากระบบ

4.3.4 ส่วนของหน้าเรียกดูข้อมูล ซึ่งการดึงข้อมูลจากคลังข้อมูล โดยการใช้ java ในการดึงข้อมูลตามที่ผู้ใช้งานกำหนดการคัดกรอง เพื่อมาแสดงหน้า JSP

ปีการศึกษา 2555 ถึง 2555, ภาคเรียน : 1, ระดับการศึกษา: ทั้งหมด, ชุม: ทั้งหมด, ประเภทวิชา: ทั้งหมด, ปี: ทั้งหมด, สถาบัน: ทั้งหมด

Person Status Table										
Personal Status ID	Category	Student ID	Name	Degree Name	Major	Date Key	Year	GP5	GPA	Status
55070001.1000.255501.1	รับสง (ป.ตรี)	55070001	นางศศิณีย์ นาคพิณ	B.Sc.		255501	1	2.31	2.31	Pass(ปกติ)
55070002.1000.255501.1	Admission (ป.ตรี)	55070002	นายฉัตรกรดี มะสิงคัม	B.Sc.		255501		3.13	3.13	Pass(ปกติ)
55070003.1000.255501.1	โหวง (ป.ตรี)	55070003	นายศศิณี สรภกน	B.Sc.		255501		3.07	3.07	Pass(ปกติ)
55070004.1000.255501.1	Admission (ป.ตรี)	55070004	นายฉัตรเดช ฉินวรสณ	B.Sc.		255501		2.47	2.47	Pass(ปกติ)
55070005.1000.255501.1	รับสง (ป.ตรี)	55070005	นายเสถียรพงษ์ ทนภสิตร	B.Sc.		255501	1	2.43	2.43	Pass(ปกติ)
55070006.1000.255501.1	รับสง (ป.ตรี)	55070006	นายจารุ ทนภามวง	B.Sc.		255501		2.52	2.52	Pass(ปกติ)
55070007.1000.255501.1	Admission (ป.ตรี)	55070007	นายชัชชนร์ สรภกรณ	B.Sc.		255501		2.88	2.88	Pass(ปกติ)
55070008.1000.255501.1	รับสง (ป.ตรี)	55070008	นายชัชชนร์ สรภกรณ	B.Sc.		255501		2.9	2.9	Pass(ปกติ)
55070009.1000.255501.1	Admission (ป.ตรี)	55070009	นายเอกฉน จิตร์พวง	B.Sc.		255501	1	3.5	3.5	Pass(ปกติ)
55070010.1000.255501.1	Admission (ป.ตรี)	55070010	นายวิวัฒน์ ชนุชเขต	B.Sc.		255501	1	3.98	3.98	Pass(ปกติ)

รูปที่ 4.18 แสดงส่วนหน้าเรียกดูข้อมูล ตามที่ผู้ใช้งานกำหนดการคัดกรอง

โดยผู้ใช้งานจะสามารถนำข้อมูลดิบออกตามที่ผู้ใช้งานได้กำหนดการคัดกรองจากระบบได้ 3 แบบ ดังนี้

- Copy: เพื่อส่งคำสั่งคัดลอกข้อมูลในตาราง (โดยแต่ละแถวจะคั่นด้วย Tab)
- CSV: เพื่อส่งคำสั่งเซฟเป็นไฟล์เอกสารสกุล CSV
- Excel: เพื่อส่งคำสั่งเซฟเป็นไฟล์เอกสารสกุล Excel
- Print: เพื่อส่งคำสั่งปริ้นตารางข้อมูล



รูปที่ 4.19 แสดงส่วนประกอบหน้าเรียกดูข้อมูล ตามที่ผู้ใช้งานกำหนดการคัดกรอง ในการนำข้อมูลดิบออกจากระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 5

# บทสรุปของโครงการ

### 5.1 สรุปผลโครงการ

ในบทนี้จะกล่าวถึงสรุปผลการดำเนินโครงการ สืบเนื่องมาจากปัญหาที่ทางคณะมีข้อมูลนักศึกษาและข้อมูลอาจารย์เพิ่มขึ้นเรื่อยๆ ทุกปี ทำให้ยากต่อการจัดการและการที่จะนำมาใช้งานทางเราจึงได้เริ่มโครงการนี้ โดยการทำคลังข้อมูลสำหรับคณะเทคโนโลยีสารสนเทศแห่งนี้ขึ้น เพื่อจัดการข้อมูลให้สามารถนำมาใช้งานได้สะดวกตอบโจทย์การวิเคราะห์และการประเมินของคณะ โดยใช้ผ่านเว็บแอปพลิเคชันสามารถสรุปผลเป็นแผนภูมิภาพแดชบอร์ดและรายงานที่ทางคณะต้องการเรียกดู โดยทำการเลือกข้อมูลที่ต้องการรวมทั้งนำข้อมูลดิบจะให้แสดงผลหรือนำออกจากระบบได้อย่างง่ายดาย

### 5.2 ปัญหาและอุปสรรค

- Server มีปัญหาค่อนข้างบ่อย ทำให้ไม่สามารถรีโมตเข้าใช้งานได้ในช่วงเวลา
- ไฮฟ์ (Hive), อิมพาลา (Impala) เป็นเทคโนโลยีใหม่อาจจะมีข้อจำกัดบางประการ ซึ่งอาจจะให้เป็นอุปสรรคต่อการจัดทำโครงการบ้างในบางกรณี
- ในตอนต้นเราใช้ภาษาไฮฟ์ (Hive) ในการพัฒนาโครงการของเรา แต่เมื่อพัฒนาไปแล้วเราพบภาษาไฮฟ์ (Hive) ไม่เหมาะกับโครงการของเรา เพราะใช้เวลาในการ Query คำสั่งในแต่ละครั้งใช้ระยะเวลานาน เราจึงเปลี่ยนมาใช้ อิมพาลา (Impala) ในการพัฒนาแทน

### 5.3 แนวทางพัฒนาต่อ

- หาวิธีการเพิ่มฟังก์ชันการลบข้อมูลที่ระายการ/ไฟล์ที่อัพโหลดแต่ละไฟล์ ผ่านอิมพาลา (Impala) ในกรณีที่ผู้ใช้งานกรอกข้อมูลผิด

## บรรณานุกรม

- [1] Oracle **“Data Warehousing Concepts”** [Online]  
 [URL]: [https://docs.oracle.com/cd/E11882\\_01/server.112/e25554/concept.htm.2559](https://docs.oracle.com/cd/E11882_01/server.112/e25554/concept.htm.2559).
- [2] 1keydata **“Data Warehouse Definition”** [Online]  
 [URL]: <http://www.1keydata.com/datawarehousing/data-warehouse-definition.html.2559>.
- [3] ETLtools **“Extract Transform Load”** [Online]  
 [URL]: <http://www.etltools.org.2559>.
- [4] Hortonworks **“Apache Ambari”** [Online]  
 [URL]: <http://hortonworks.com/hadoop/ambari.2559>.
- [5] IBM **“Apache Avro”** [Online]  
 [URL]: <http://www.ibm.com/developerworks/library/avro-hadoop.2556>.
- [6] Brian O’Neill **“Apache Cassandra”** [Online]  
 [URL]: <https://dzone.com/refcardz/apache-cassandra.2559>.
- [7] Apache **“Apache Chukwa”** [Online]  
 [URL]: <https://chukwa.apache.org/docs/r0.5.0/admin.html.2556>.
- [8] Apache **“Apache Flume”** [Online]  
 [URL]: <https://flume.apache.org.2558>.
- [9] Otis Gospodnetic **“Apache HBase”** [Online]  
 [URL]: <https://dzone.com/refcardz/hbase.2559>.
- [10] Hortonworks **“Apache Mahout”** [Online]  
 [URL]: <http://hortonworks.com/apache/mahout.2559>.
- [11] Hortonworks **“Apache Oozie”** [Online]  
 [URL]: <http://hortonworks.com/hadoop/oozie.2559>.
- [12] Hortonworks **“Apache Pig”** [Online]  
 [URL]: <http://hortonworks.com/apache/pig.2559>.
- [13] Hortonworks **“Apache Sqoop”** [Online]  
 [URL]: <http://hortonworks.com/apache/sqoop.2559>.
- [14] Hortonworks **“Apache Spark”** [Online]  
 [URL]: <http://hortonworks.com/hadoop/spark.2559>.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
 ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บรรณานุกรม (ต่อ)

- [15] Apache “**Apache ZooKeeper**” [Online]  
[URL]: <https://zookeeper.apache.org>. 2559.
- [16] Apache “**Apache Hive**” [Online]  
[URL]: <https://hive.apache.org>. 2557.
- [17] Cloudera “**Cloudera Impala Guide**” [Online]  
[URL]: <http://www.cloudera.com/content/www/en-us/documentation/enterprise/latest/topics/impala.html>. 2559.
- [18] Cloudera “**Impala Concepts and Architecture**” [Online]  
[URL]: [http://www.cloudera.com/content/www/en-us/documentation/enterprise/latest/topics/impala\\_concepts.html](http://www.cloudera.com/content/www/en-us/documentation/enterprise/latest/topics/impala_concepts.html). 2559.
- [19] Cloudera “**Components of the Impala Server**” [Online]  
[URL]: [http://www.cloudera.com/content/www/en-us/documentation/enterprise/latest/topics/impala\\_components.html](http://www.cloudera.com/content/www/en-us/documentation/enterprise/latest/topics/impala_components.html). 2559.
- [20] Dezyre “**Hive vs Impala**” [Online]  
[URL]: <https://www.dezyre.com/article/-hive-vs-impala-sql-war-in-the-hadoop-ecosystem/148>. 2558.
- [21] SearchSOA “**Java Server Page (JSP) definition**” [Online]  
[URL]: <http://searchsoa.techtarget.com/definition/Java-Server-Page>. 2559.
- [22] SearchSOA “**Java definition**” [Online]  
[URL]: <http://searchsoa.techtarget.com/definition/Java>. 2557.
- [23] MDN “**Introduction**” [Online]  
[URL]: <https://developer.mozilla.org/en-US/docs/Web/JavaScript/Guide/Introduction>. 2559.
- [24] SearchSOA “**cascading style sheet (CSS) definition**” [Online]  
[URL]: <http://searchsoa.techtarget.com/definition/cascading-style-sheet-CSS>. 2558.
- [25] jQuery “**What is jQuery?**” [Online]  
[URL]: <https://jquery.com>. 2559.
- [26] jQuery “**Highcharts Product**” [Online]  
[URL]: <http://www.highcharts.com/products/highcharts>. 2559.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## คู่มือการใช้งาน

### ระบบคลังข้อมูลคณะเทคโนโลยีสารสนเทศบนระบบอาปาเซฮาดูป

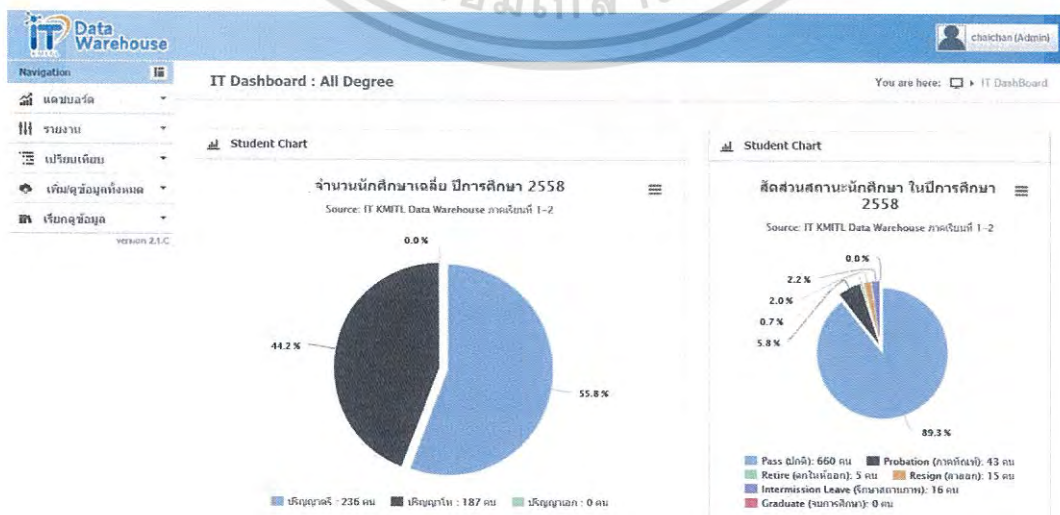
#### 1. เว็บแอปพลิเคชัน (Web Application)

##### 1) ส่วนของ Login



รูปที่ ก.1 แสดงส่วนของหน้า Login

##### 2) ส่วนของหน้า Dashboard



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับรูปที่ ก.2 แสดงส่วนของหน้า Dashboard ญาติให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2. ส่วนควบคุมการทำงานของระบบ

### 1) ส่วนรายละเอียดในแต่ละหน้า (Header Bar)



รูปที่ ก.3 แสดงส่วนของรายละเอียดในแต่ละหน้า (Header Bar)

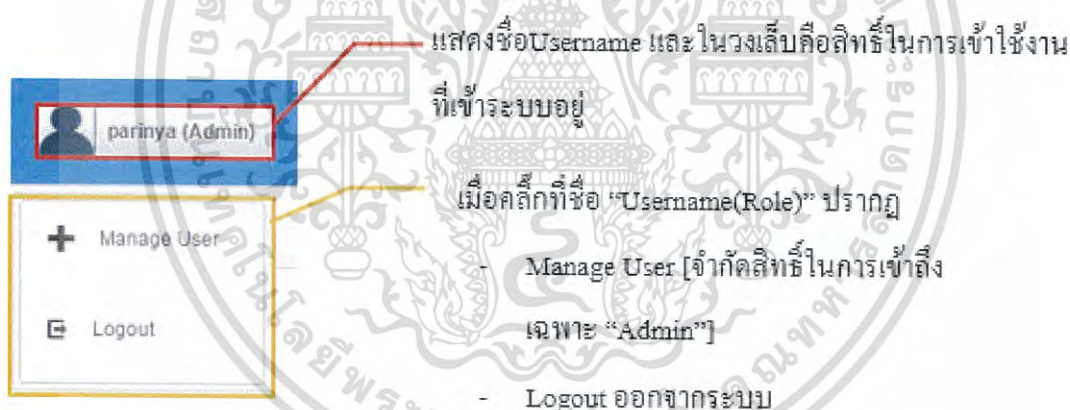
#### 1.1 ส่วนของการลิงค์ไปยังหน้าแรก



เมื่อกดที่ส่วนนี้จะลิงค์ไปยังหน้า IT Dashboard : All Degree

รูปที่ ก.4 แสดงส่วนของการลิงค์ไปยังหน้าแรก

#### 1.2 ส่วนเกี่ยวข้องกับผู้ใช้งาน



รูปที่ ก.5 แสดงส่วนเกี่ยวข้องกับผู้ใช้งาน

#### 1.3 ส่วนแสดงชื่อหน้า ณ ปัจจุบัน (Title Bar)

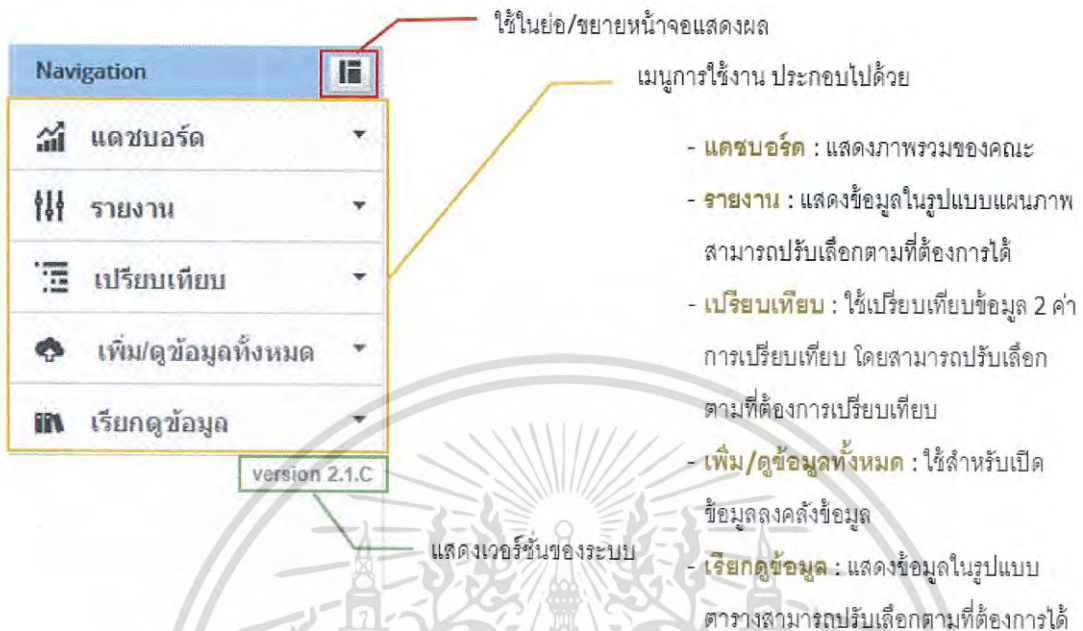


รูปที่ ก.6 ส่วนแสดงชื่อหน้า ณ ปัจจุบัน (Title Bar)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

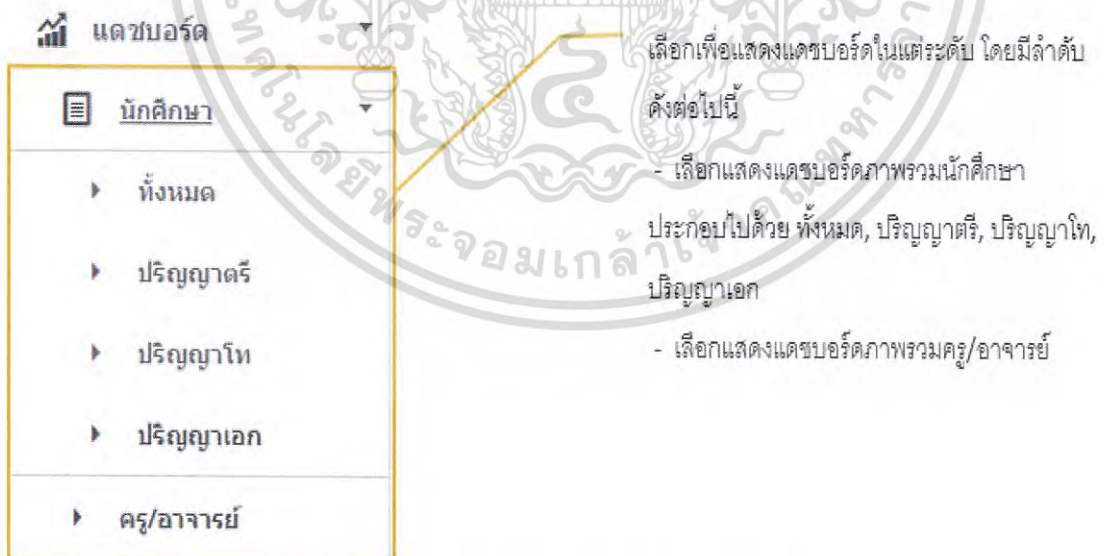
## 2) ส่วนเมนูหน้าในแต่ละหน้า (Navigation Bar) และรายละเอียดแต่ละเมนู

### 2.1 ส่วนของเมนูการควบคุม



รูปที่ ก.7 ส่วนแสดงชื่อหน้า ณ ปัจจุบัน (Title Bar)

### 2.2 ส่วนของเมนูแดชบอร์ดภาพรวมของคณะ



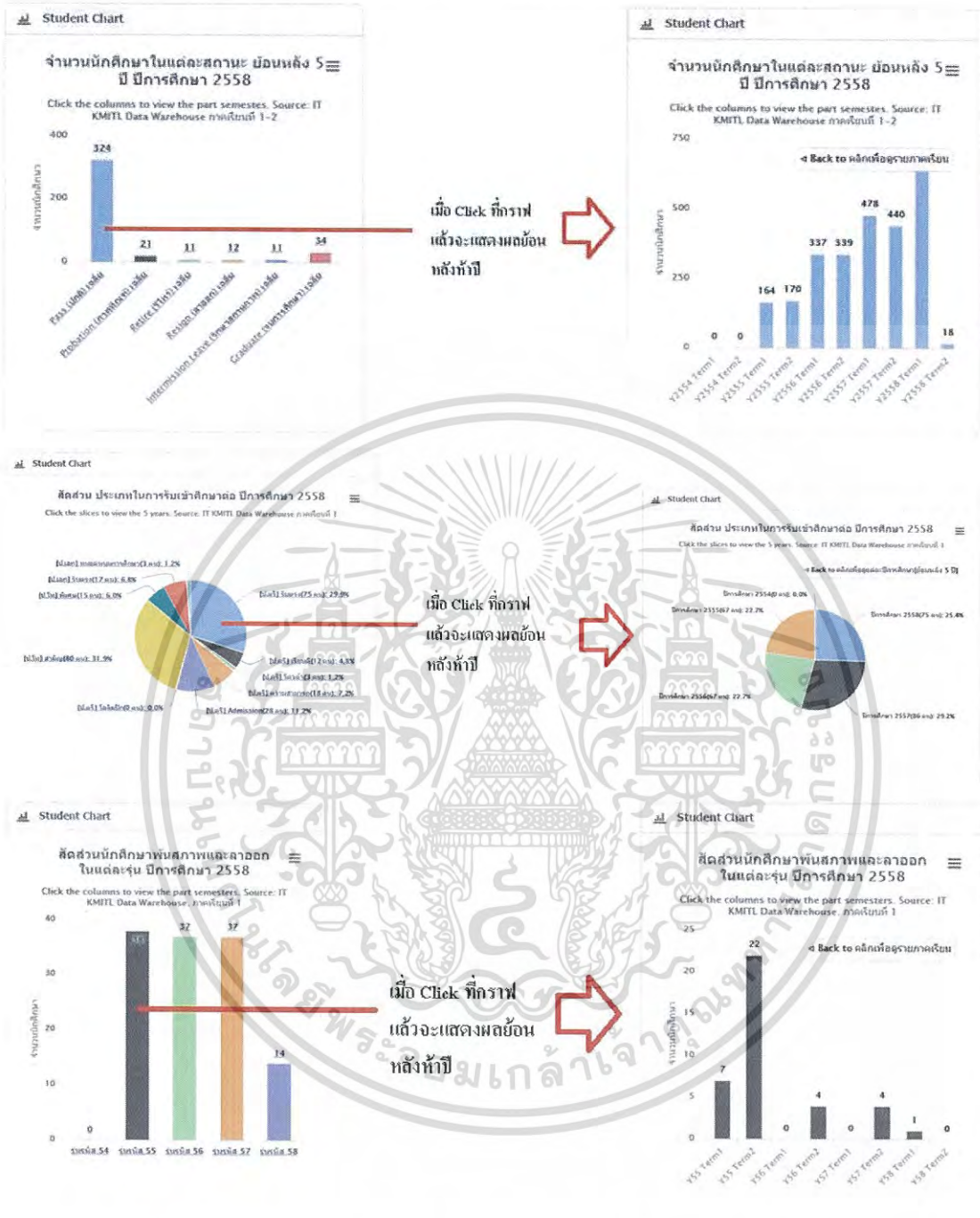
รูปที่ ก.8 ส่วนแสดงชื่อหน้า ณ ปัจจุบัน (Title Bar)

- แดชบอร์ดแบ่งออกเป็นสองส่วนคือ ส่วนของนักศึกษาและส่วนของครู/อาจารย์ ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



- เมื่อ Click บางกราฟจะสามารถแสดงผลย้อนหลังห้าปี ดังนี้

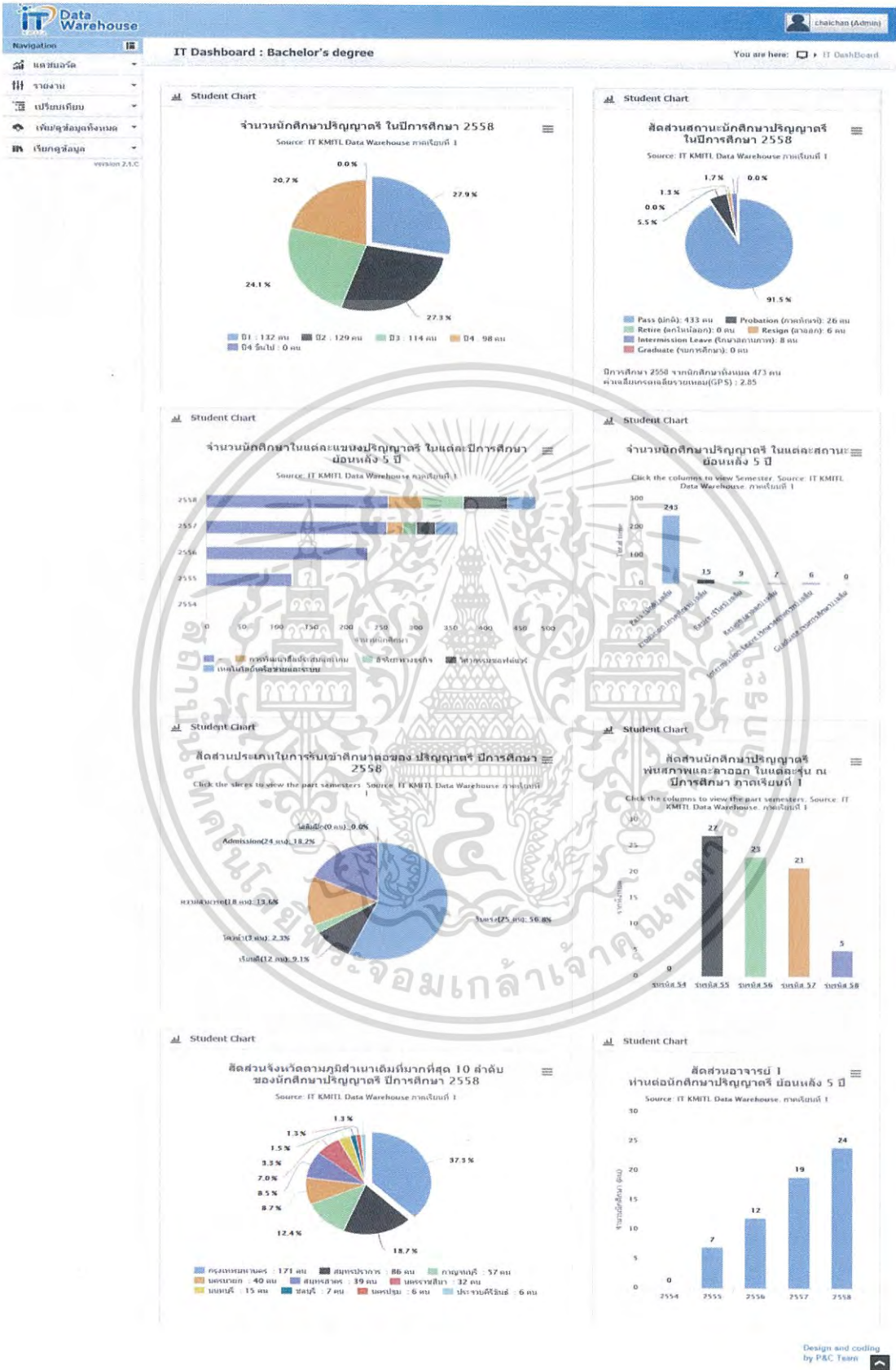


รูปที่ ก.10 ส่วนแสดงผลย้อนหลังห้าปีของแดชบอร์ดรวมของนักศึกษาทั้งหมด

- เมื่อต้องการกลับไปทีกราฟปีการศึกษา 2558 ให้กดปุ่ม < Back toคลิกเพื่อดูรายงานภาคเรียน

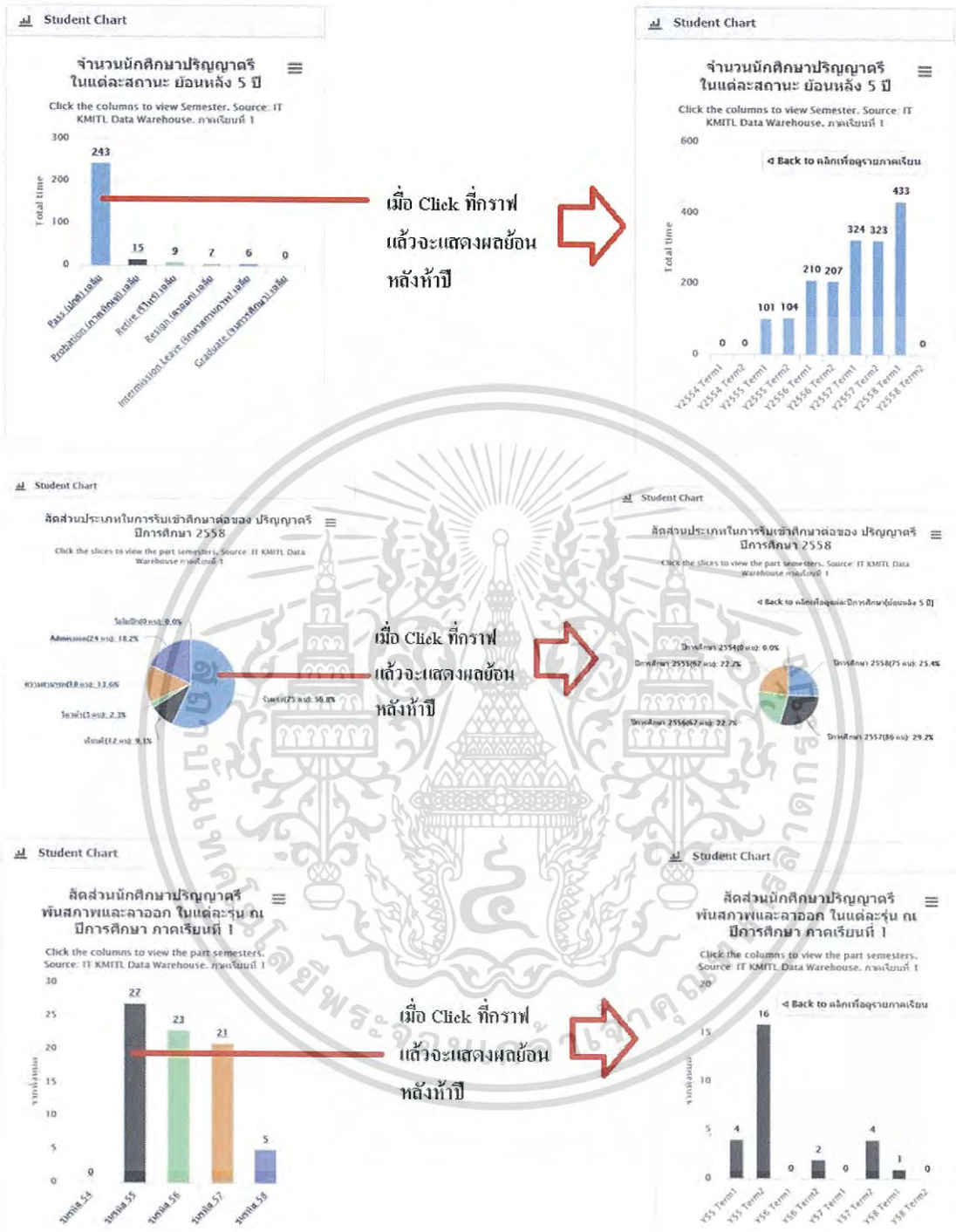
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 2.2.2 แสดงแดชบอร์ดเฉพาะนักศึกษาปริญญาตรี



เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้สำหรับใช้ภายในเท่านั้น ไม่ควรเผยแพร่โดยไม่ได้รับอนุญาตจากเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- เมื่อ Click บางกราฟจะสามารถแสดงผลย้อนหลังห้าปี ดังนี้



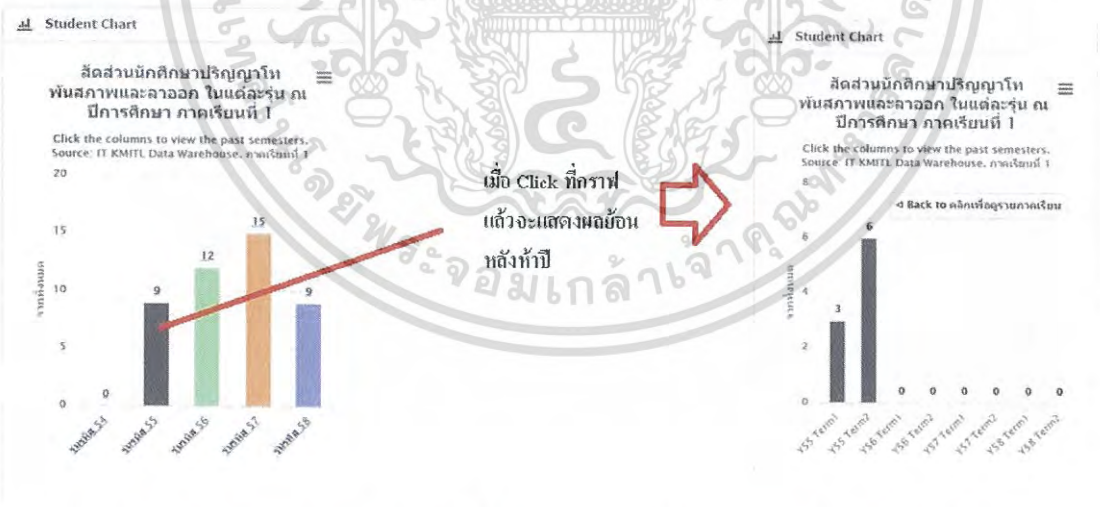
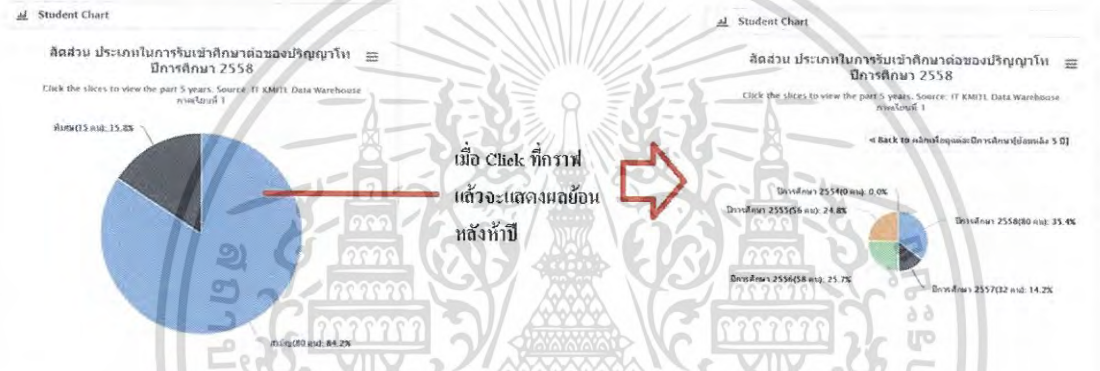
รูปที่ ก.12 ส่วนแสดงผลย้อนหลังห้าปีของแดชบอร์ดเฉพาะนักศึกษาปริญญาตรี

- เมื่อต้องการกลับไปทีกราฟปีการศึกษา 2558 ให้กดปุ่ม < Back to คลิกเพื่อดูรายภาคเรียน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



- เมื่อ Click บางกราฟจะสามารถแสดงผลย้อนหลังห้าปี ดังนี้

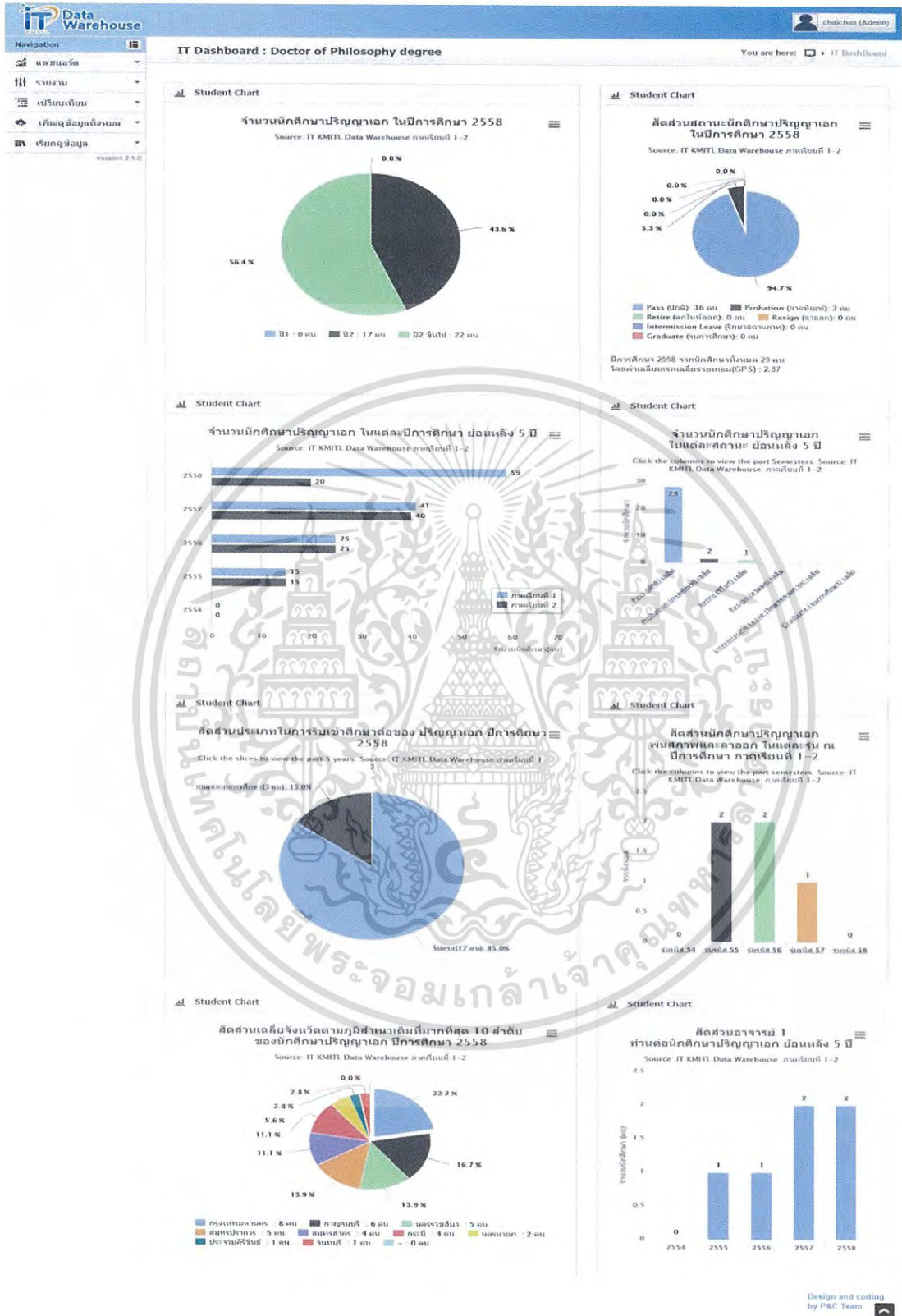


รูปที่ ก.14 ส่วนแสดงผลย้อนหลังห้าปีของแดชบอร์ดเฉพาะนักศึกษาปริญญาโท

- เมื่อต้องการกลับไปทีกราฟปีการศึกษา 2558 ให้กดปุ่ม < Back to คลิกเพื่อดูรายภาคเรียน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.4 แสดงแดชบอร์ดเฉพาะนักศึกษาปริญญาเอก



รูปที่ ก.15 ส่วนแสดงแดชบอร์ดเฉพาะนักศึกษาปริญญาเอก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- เมื่อ Click บางกราฟจะสามารถแสดงผลย้อนหลังห้าปี ดังนี้

**Student Chart**  
จำนวนนักศึกษาปริญญาเอก  
ในแต่ละสถานะ: เมื่อจนถึง 5 ปี

Click the columns to view the part Semesters. Source: IT KMUTL Data Warehouse. ภาพหน้าจอ 1-2

เมื่อ Click ที่กราฟ  
แล้วจะแสดงผลย้อน  
หลังห้าปี

**Student Chart**  
จำนวนนักศึกษาปริญญาเอก  
ในแต่ละสถานะ: เมื่อจนถึง 5 ปี

Click the columns to view the part Semesters. Source: IT KMUTL Data Warehouse. ภาพหน้าจอ 1-2

เมื่อ Click ที่กราฟ  
แล้วจะแสดงผลย้อน  
หลังห้าปี

**Student Chart**  
สัดส่วนประเภทในการรับเข้าศึกษาต่อของ ปริญญาเอก  
ปีการศึกษา 2558

Click the slices to view the part 5 years. Source: IT KMUTL Data Warehouse. ภาพหน้าจอ 1

เมื่อ Click ที่กราฟ  
แล้วจะแสดงผลย้อน  
หลังห้าปี

**Student Chart**  
สัดส่วนประเภทในการรับเข้าศึกษาต่อของ ปริญญาเอก  
ปีการศึกษา 2558

Click the slices to view the part 5 years. Source: IT KMUTL Data Warehouse. ภาพหน้าจอ 1

เมื่อ Click ที่กราฟ  
แล้วจะแสดงผลย้อน  
หลังห้าปี

**Student Chart**  
สัดส่วนนักศึกษาปริญญาเอก  
ที่นศึกษาและลาออก ในแต่ละรุ่น ณ  
ปีการศึกษา ภาคเรียนที่ 1

Click the columns to view the part semesters. Source: IT KMUTL Data Warehouse. ภาพหน้าจอ 1

เมื่อ Click ที่กราฟ  
แล้วจะแสดงผลย้อน  
หลังห้าปี

**Student Chart**  
สัดส่วนนักศึกษาปริญญาเอก  
ที่นศึกษาและลาออก ในแต่ละรุ่น ณ  
ปีการศึกษา ภาคเรียนที่ 1

Click the columns to view the part semesters. Source: IT KMUTL Data Warehouse. ภาพหน้าจอ 1

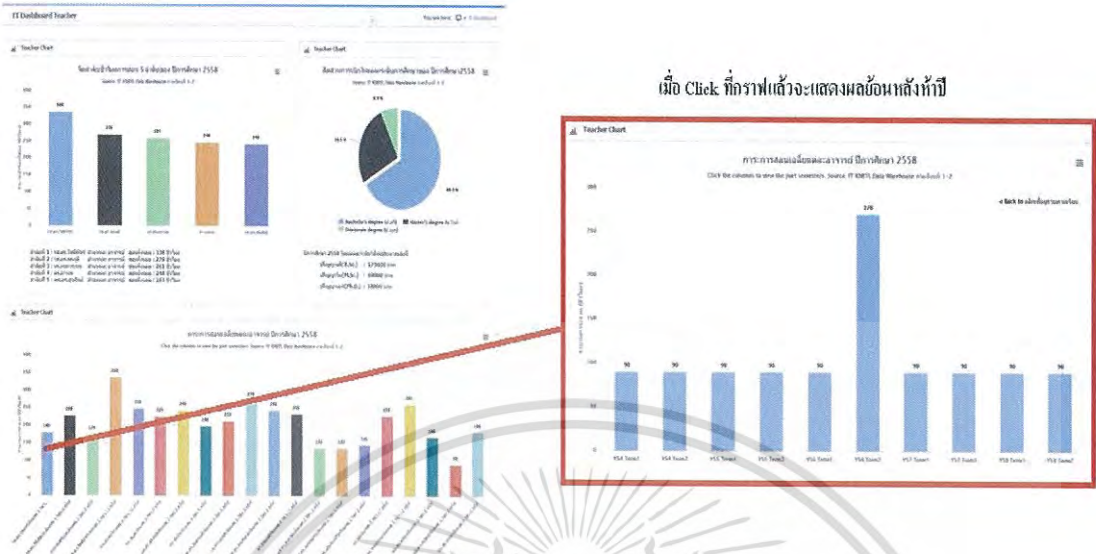
เมื่อ Click ที่กราฟ  
แล้วจะแสดงผลย้อน  
หลังห้าปี

รูปที่ ก.16 ส่วนแสดงผลย้อนหลังห้าปีของแดชบอร์ดเฉพาะนักศึกษาปริญญาเอก

- เมื่อต้องการกลับไปทีกราฟปีการศึกษา 2558 ให้กดปุ่ม < Back toคลิกเพื่อดูรายภาคเรียน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 2.2.5 แสดงแดชบอร์ดเฉพาะครูหรืออาจารย์



เมื่อ Click ที่กราฟแล้วจะแสดงผลย้อนหลังทันที

รูปที่ ก.17 ส่วนแสดงแดชบอร์ดเฉพาะครูหรืออาจารย์

### 2.3 ส่วนการสร้างรายงาน

#### 2.3.1 ส่วนประกอบย่อยของแต่เมนูการควบคุม

The screenshot shows a control menu for report generation with the following fields:

- รายงาน (Report)
- สถานะ: นักศึกษา (Status: Student)
- ปีการศึกษา: 2554 (Academic Year: 2554)
- ถึง: 2554 (To: 2554)
- ภาคเรียน: ทั้งหมด (Semester: All)
- รุ่นรหัส: ทั้งหมด (Batch: All)
- ปี: ทั้งหมด (Year: All)
- ระดับการศึกษา: ทั้งหมด (Education Level: All)
- ประเภทการรับเข้า: ทั้งหมด (Admission Type: All)
- สถานะการศึกษา: ทั้งหมด (Study Status: All)
- สร้างรายงานนักศึกษา > (Generate Student Report >)

เมื่อกดปุ่ม "รายงาน" แล้วกดเลือก นักศึกษา หรือ อาจารย์ ที่ต้องการ เรียกดูรายงานแผนภาพ โดยเลือกข้อมูลที่ต้องการให้ระบบ แสดงแผนภูมิภาพ เมื่อเลือกเสร็จ ให้กดปุ่ม "สร้างรายงาน >"

ปุ่มสำหรับกดเพื่อแสดงรายงาน แผนภูมิภาพตามที่กำหนดด้านบน

รูปที่ ก.18 แสดงส่วนประกอบย่อยของแต่เมนูการควบคุม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 2.3.2 ส่วนควบคุมของการเลือกข้อมูล

เมื่อคลิกปุ่ม  เพื่อเลือกข้อมูล  
ที่ต้องการให้ระบบทำรายงาน  
[ โดยข้อมูลที่แสดง ระบบจะดึงข้อมูล  
มาจากคลังข้อมูล ]

รูปที่ ก.19 แสดงส่วนควบคุมของการเลือกข้อมูล

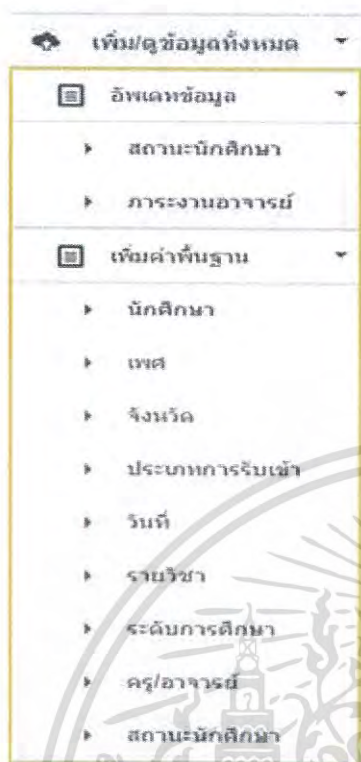
### 2.4 ส่วนของการเปรียบเทียบข้อมูล

เมื่อคลิกปุ่ม "เปรียบเทียบ" แล้วให้คลิกปุ่ม "กำหนดค่าเปรียบเทียบ1" ทำการเลือกข้อมูลที่เราต้องการใช้ในการเปรียบเทียบเป็น ข้อมูลชุดแรก เมื่อเลือกเสร็จ ให้ทำการคลิกปุ่ม "กำหนดค่าเปรียบเทียบ2" แล้วทำการเลือกข้อมูลที่เราต้องการใช้ในการเปรียบเทียบเป็น ข้อมูลชุดที่สอง  
เมื่อเลือกข้อมูลเสร็จทั้งสองชุดแล้วให้คลิกปุ่ม "สร้างรายงานเปรียบเทียบ"

ปุ่มสำหรับกดเพื่อแสดงผลรายงานแผนภูมิภาพ  
ตามที่กำหนดไว้ด้านบน

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ การแจ้งให้เพื่อเป็นการยกเว้นความผิดให้แก่นักเรียน นักศึกษา ผู้ปกครอง และบุคลากรที่เกี่ยวข้องให้ทราบว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

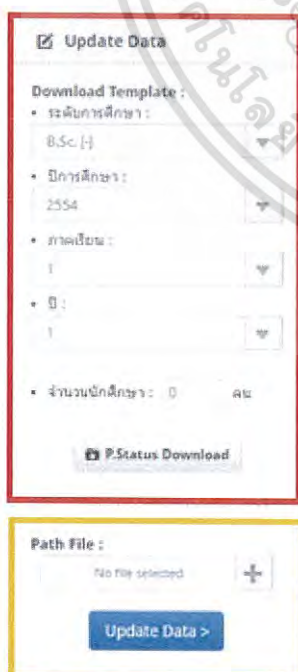
## 2.5 ส่วนของเมนูการควบคุม ในส่วน Put/Query Data



เมื่อกดปุ่ม "เพิ่ม/ดูข้อมูลทั้งหมด" แล้วกด อัปเดตข้อมูล หรือ เพิ่มค่าพื้นฐาน เพื่อเลือกตารางที่ต้องการเพิ่มข้อมูล ลงระบบ

รูปที่ ก.21 แสดงส่วนของเมนูการควบคุม ในส่วน Put/Query Data

### 2.5.1 ส่วนอัปเดตข้อมูลสถานะนักศึกษา



เมื่อกดปุ่ม "อัปเดตข้อมูล" แล้วเลือกกดปุ่ม "สถานะนักศึกษา" จะแสดงส่วนนี้ขึ้นมาให้ทางด้านขวา ให้ทำการเลือกข้อมูลที่ต้องการ เมื่อเลือกเสร็จให้ทำการกดปุ่ม "P.Status Download" เพื่อดาวน์โหลดเพิ่มเติม

เมื่อกรอกข้อมูลเพิ่มเติมเรียบร้อยแล้ว ให้ทำการอัปเดตไฟล์ csv ขึ้นระบบ โดยกดปุ่ม "+" จากนั้นเลือกไฟล์ที่ต้องการ แล้วกดปุ่ม "Update Data"

รูปที่ ก.22 แสดงส่วนอัปเดตข้อมูลสถานะนักศึกษา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

\*โดยลำดับการเพิ่มข้อมูลดังนี้

วันที่ > นักศึกษา > อาจารย์ > อัพเดทสถานะนักศึกษา/ภาระงานอาจารย์

### 2.5.2 ส่วนอัพเดทข้อมูลภาระงานอาจารย์

เมื่อกดปุ่ม "อัพเดทข้อมูล" แล้วเลือกกดปุ่ม "ภาระงานอาจารย์" จะแสดงส่วนนี้ขึ้นมาให้ทางด้านขวา ให้ทำการเลือกข้อมูลที่ต้องการ เมื่อเลือกเสร็จให้ทำการกดปุ่ม

"T.Workload Download"

เพื่อดาวน์โหลดเพิ่มเติม

เมื่อกรอกข้อมูลเพิ่มเติมเรียบร้อยแล้ว ให้ทำการอัพโหลดไฟล์ csv ขึ้นระบบ โดยกดปุ่ม "+" จากนั้นเลือกไฟล์ที่ต้องการ แล้วกดปุ่ม

"Update Data"

รูปที่ ก.23 แสดงส่วนอัพเดทข้อมูลภาระงานอาจารย์

### 2.5.3 ส่วนเพิ่มค่าพื้นฐาน

#### 1. ข้อมูลนักศึกษา

ทำการกดปุ่ม "Student" เพื่อทำการดาวน์โหลดเพิ่มเติม

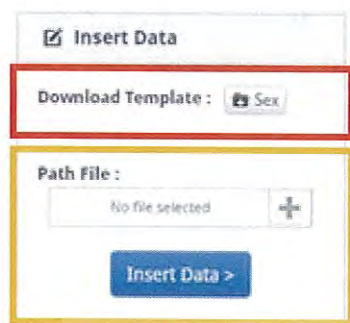
เมื่อกรอกข้อมูลเพิ่มเติมเรียบร้อยแล้ว ให้ทำการอัพโหลดไฟล์ csv ขึ้นระบบ โดยกดปุ่ม "+" จากนั้นเลือกไฟล์ที่ต้องการ แล้วกดปุ่ม

"Insert Data"

รูปที่ ก.24 แสดงส่วนเพิ่มค่าพื้นฐานของข้อมูลนักศึกษา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2. ข้อมูลเพศ



ทำการกดปุ่ม "Sex" เพื่อทำการดาวน์โหลด  
เทมเพลต ซึ่งประกอบไปด้วย

**EX**

รหัสเพศ	เพศ
1	ผู้ชาย
2	ผู้หญิง

เมื่อกรอกข้อมูลเทมเพลตเรียบร้อยแล้ว ให้ทำ  
การอัปโหลดไฟล์ csv ขึ้นระบบ โดยกดปุ่ม  
"+" จากนั้นเลือกไฟล์ที่ต้องการ แล้วกดปุ่ม  
"Insert Data"

รูปที่ ก.25 แสดงส่วนเพิ่มค่าพื้นฐานของข้อมูลเพศ

## 3. ข้อมูลจังหวัด



ทำการกดปุ่ม "Province" เพื่อทำการดาวน์โหลด  
เทมเพลต

เมื่อกรอกข้อมูลเทมเพลตเรียบร้อยแล้ว ให้ทำ  
การอัปโหลดไฟล์ csv ขึ้นระบบ โดยกดปุ่ม  
"+" จากนั้นเลือกไฟล์ที่ต้องการ แล้วกดปุ่ม  
"Insert Data"

รูปที่ ก.26 แสดงส่วนเพิ่มค่าพื้นฐานของข้อมูลจังหวัด

## โดยมีการกำหนดรหัสตัวแทนจังหวัดดังนี้

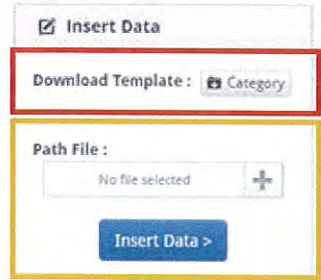
ตารางที่ ก.1 แสดงรหัสและชื่อของจังหวัดทั้งหมด

รหัส จังหวัด	ชื่อจังหวัด	รหัส จังหวัด	ชื่อจังหวัด	รหัส จังหวัด	ชื่อจังหวัด
1	กรุงเทพมหานคร	27	บึงกาฬ	53	ลำปาง
2	กระบี่	28	บุรีรัมย์	54	ลำพูน
3	กาญจนบุรี	29	ปทุมธานี	55	เลย
4	กาฬสินธุ์	30	ประจวบคีรีขันธ์	56	ศรีสะเกษ
5	กำแพงเพชร	31	ปราจีนบุรี	57	สกลนคร
6	ขอนแก่น	32	ปัตตานี	58	สงขลา
7	จันทบุรี	33	พระนครศรีอยุธยา	59	สตูล
8	ฉะเชิงเทรา	34	พังงา	60	สมุทรปราการ
9	ชลบุรี	35	พัทลุง	61	สมุทรสงคราม
10	ชัยนาท	36	พิจิตร	62	สมุทรสาคร
11	ชัยภูมิ	37	พิษณุโลก	63	สระแก้ว
12	ชุมพร	38	เพชรบุรี	64	สระบุรี
13	เชียงราย	39	เพชรบูรณ์	65	สิงห์บุรี
14	เชียงใหม่	40	แพร่	66	สุโขทัย
15	ศรีสะเกษ	41	พะเยา	67	สุพรรณบุรี
16	ตราด	42	ภูเก็ต	68	สุราษฎร์ธานี
17	ตาก	43	มหาสารคาม	69	สุรินทร์
18	นครนายก	44	มุกดาหาร	70	หนองคาย
19	นครปฐม	45	แม่ฮ่องสอน	71	หนองบัวลำภู
20	นครพนม	46	ยะลา	72	อ่างทอง
21	นครราชสีมา	47	ยโสธร	73	อุดรธานี
22	นครศรีธรรมราช	48	ร้อยเอ็ด	74	อุทัยธานี
23	นครสวรรค์	49	ระนอง	75	อุตรดิตถ์
24	นนทบุรี	50	ระยอง	76	อุบลราชธานี
25	นราธิวาส	51	ราชบุรี	77	อำนาจเจริญ
26	น่าน	52	ลพบุรี		

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

#### 4. ข้อมูลประเภทการรับเข้า

ทำการกดปุ่ม "Category" เพื่อทำการดาวน์โหลดเทมเพลต ซึ่งประกอบไปด้วย



EX

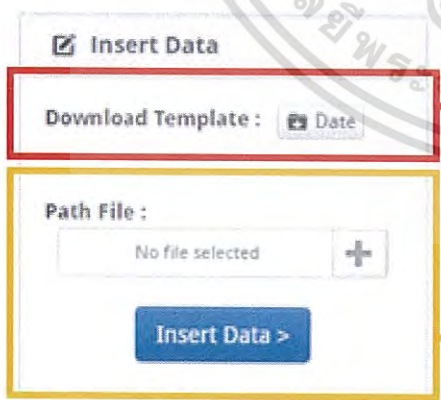
รหัสประเภทการรับเข้า	ประเภทการรับเข้า
1	รับตรง (ป.ตรี)
2	Admission (ป.ตรี)
3	ความสามารถ (ป.ตรี)
4	โควตา (ป.ตรี)
5	เรียนดี (ป.ตรี)
6	โอลิมปิก (ป.ตรี)
7	สามัญ (ป.โท)
8	พิเศษ (ป.โท)
9	รับตรง (ป.เอก)
10	ทุนอุดหนุนการศึกษา (ป.เอก)

เมื่อกรอกข้อมูลเทมเพลตเรียบร้อยแล้ว ให้ทำการอัปโหลดไฟล์ csv ขึ้นระบบ โดยกดปุ่ม "+" จากนั้นเลือกไฟล์ที่ต้องการ แล้วกดปุ่ม "Insert Data"

รูปที่ ก.27 แสดงส่วนเพิ่มค่าพื้นฐานของข้อมูลประเภทการรับเข้า

#### 5. ข้อมูลวันที่

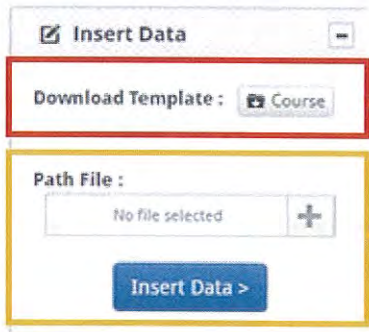
ทำการกดปุ่ม "Date" เพื่อทำการดาวน์โหลดเทมเพลต



เมื่อกรอกข้อมูลเทมเพลตเรียบร้อยแล้ว ให้ทำการอัปโหลดไฟล์ csv ขึ้นระบบ โดยกดปุ่ม "+" จากนั้นเลือกไฟล์ที่ต้องการ แล้วกดปุ่ม "Insert Data"

รูปที่ ก.28 แสดงส่วนเพิ่มค่าพื้นฐานของข้อมูลวันที่

## 6. ข้อมูลรายวิชา




ทำการกดปุ่ม "Course" เพื่อทำการดาวน์โหลดเทมเพลต

เมื่อกรอกข้อมูลเทมเพลตเรียบร้อยแล้ว ให้ทำการอัปโหลดไฟล์ csv ขึ้นระบบ โดยกดปุ่ม "+" จากนั้นเลือกไฟล์ที่ต้องการ แล้วกดปุ่ม "Insert Data"

รูปที่ ก.29 แสดงส่วนเพิ่มค่าพื้นฐานของข้อมูลรายวิชา

## 7. ข้อมูลระดับการศึกษา



ทำการกดปุ่ม "Degree" เพื่อทำการดาวน์โหลดเทมเพลต ซึ่งประกอบไปด้วย

รหัสระดับการศึกษา	ระดับการศึกษา	วิชาเอก (Major)
1000	B.Sc.	-
1001	B.Sc.	Network and System Technology
1002	B.Sc.	Software Engineering
1003	B.Sc.	Business Intelligence
1004	B.Sc.	Multimedia and Game Development
		Philosophy
2000	M.Sc.	-
2001	M.Sc.	Information Science
2002	M.Sc.	Information Science Technology
2003	M.Sc.	Information Technology
		Management
2004	M.Sc.	Network and System Technology
3000	Ph.D.	-
3001	Ph.D.	Information Science

เมื่อกรอกข้อมูลเทมเพลตเรียบร้อยแล้ว ให้ทำการอัปโหลดไฟล์ csv ขึ้นระบบ โดยกดปุ่ม "+" จากนั้นเลือกไฟล์ที่ต้องการ แล้วกดปุ่ม "Insert Data"

รูปที่ ก.30 แสดงส่วนเพิ่มค่าพื้นฐานของข้อมูลระดับการศึกษา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 8. ข้อมูลครู/อาจารย์

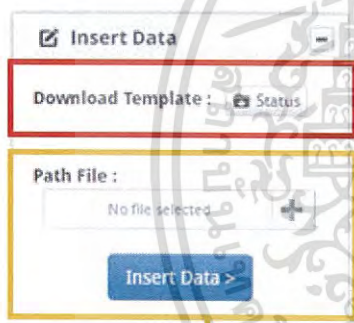


ทำการกดปุ่ม "Teacher" เพื่อทำการดาวน์โหลดเทมเพลต

เมื่อกรอกข้อมูลเทมเพลตเรียบร้อยแล้ว ให้ทำการอัปโหลดไฟล์ csv ขึ้นระบบ โดยกดปุ่ม "+" จากนั้นเลือกไฟล์ที่ต้องการ แล้วกดปุ่ม "Insert Data"

รูปที่ ก.31 แสดงส่วนเพิ่มค่าพื้นฐานของข้อมูลครู/อาจารย์

## 9. ข้อมูลสถานะนักศึกษา



ทำการกดปุ่ม "Status" เพื่อทำการดาวน์โหลดเทมเพลต ซึ่งประกอบไปด้วย

EX

รหัสสถานะนักศึกษา	สถานะนักศึกษา
0	ลงทะเบียน
1	ปกติ
2	ภาคทัณฑ์
3	รีไทร์
4	ลาออก
5	รักษาสถานภาพ
6	จบการศึกษา

เมื่อกรอกข้อมูลเทมเพลตเรียบร้อยแล้ว ให้ทำการอัปโหลดไฟล์ csv ขึ้นระบบ โดยกดปุ่ม "+" จากนั้นเลือกไฟล์ที่ต้องการ แล้วกดปุ่ม "Insert Data"

รูปที่ ก.32 แสดงส่วนเพิ่มค่าพื้นฐานของข้อมูลสถานะนักศึกษา

## อธิบายสถานะการแจ้งเตือนการเพิ่ม/อัปเดตข้อมูล

**กรณีที่ 1** ถ้าหากการเพิ่ม/อัปเดตข้อมูลเข้าระบบสำเร็จแล้ว ระบบจะแสดงข้อความ

สีเขียว ดังภาพ

รูปที่ ก.33 แสดงส่วนการเพิ่ม/อัปเดตข้อมูลเข้าระบบสำเร็จแล้ว

**กรณีที่ 2** ถ้าหากการเพิ่ม/อัปเดตข้อมูลเข้าระบบไม่สำเร็จแล้วระบบจะแสดง เนื่องจากไฟล์ข้อมูลที่เพิ่ม/อัปเดตไปมีปัญหาบางประการ โดยจะมีข้อความ

**สีแดง** โดยไม่มีข้อความสีเทาเชื่อมต่อด้านล่าง หรือ ขึ้นแสดงเป็น ข้อผิดพลาดที่เกิดขึ้น แสดงว่า ดังภาพ

- HTTP Status 500 คือ ผู้ใช้อาจจะอัปเดตไฟล์ข้อมูลไม่ตรง ตามสกุลที่ระบบกำหนด
- Add Person Status Fact Data Fail, Please try again! คือ ผู้ใช้อาจจะเลือกอัปเดตไฟล์ข้อมูลผิด

รูปแบบ

**กรณีที่ 3** ถ้าหากการเพิ่ม/อัปเดตข้อมูลเข้าระบบไม่สำเร็จแล้วระบบจะแสดง ข้อความ

**สีแดง** แสดงว่ามีข้อมูลทั้งหมดซ้ำโดยระบบไม่นำข้อมูลที่ซ้ำเหล่านั้นเข้าระบบ ดังภาพ

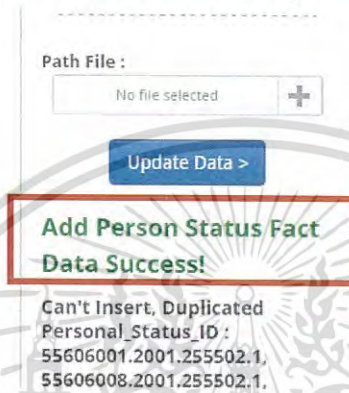
รูปที่ ก.34 แสดงส่วนการเพิ่ม/อัปเดตข้อมูลเข้าระบบไม่สำเร็จ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

\* โดยถ้าผู้ใช้งานระบบมีการเพิ่ม/อัปเดตข้อมูลซ้ำ

- ระบบจะแสดง ID ข้อมูลที่ผู้ใช้เพิ่มซ้ำ ข้อความสีเทาเข้ม เนื่องจากมีข้อมูลซ้ำทั้งหมดที่อัปเดตระบบไปแล้ว

**กรณีที่ 4** ถ้าหากการเพิ่ม/อัปเดตข้อมูล เข้าระบบสำเร็จแล้วระบบจะแสดง ข้อความสีเขียว แสดงว่า มีข้อมูลบางส่วนซ้ำโดยระบบ ไม่นำข้อมูลที่ซ้ำเหล่านั้นเข้าระบบ ดังภาพ



รูปที่ ก.35 แสดงส่วนการเพิ่ม/อัปเดตข้อมูลเข้าระบบสำเร็จ โดยระบบ ไม่นำข้อมูลที่ซ้ำเข้าระบบ

\* โดยถ้าผู้ใช้งานระบบมีการเพิ่ม/อัปเดตข้อมูลซ้ำ

- ระบบจะแสดง ID ข้อมูลที่ผู้ใช้เพิ่มซ้ำ ข้อความสีเทาเข้ม เนื่องจากมีข้อมูลซ้ำบางส่วนที่อัปเดตระบบไปแล้ว

## 2.6 ส่วนของการเรียกดูข้อมูล

เรียกดูข้อมูล

สถานะนักศึกษา

ปีการศึกษา : 2554 ถึง : 2554

ภาคเรียน : ทั้งหมด

รุ่นรหัส : ทั้งหมด

ปี : ทั้งหมด

ระดับการศึกษา : ทั้งหมด

ประเภทการรับเข้า : ทั้งหมด

สถานการศึกษา : ทั้งหมด

เรียกดูข้อมูลนักศึกษา >

เมื่อกดปุ่ม "เรียกดูข้อมูล" แล้วกดเลือก สถานะ นักศึกษา หรือ ภาระงานอาจารย์ ที่ต้องการดู ตารางข้อมูล โดยเลือกข้อมูลที่ต้องการให้ระบบแสดง เมื่อเลือกเสร็จให้กดปุ่ม "เรียกดูข้อมูล"

ปุ่มสำหรับกดเพื่อแสดงข้อมูลตามที่ กำหนดด้านบน

รูปที่ ก.36 แสดงส่วนของการเรียกดูข้อมูล

## 3) ส่วนกำหนดสิทธิ์ผู้ใช้

parinya (Admin)

+ Manage User

Logout

แสดงชื่อ Username และ ใ้มองถึงคือสิทธิ์ในการเข้าใช้งาน ที่เข้าระบบอยู่

เมื่อคลิกที่ชื่อ "Username(Role)" ปรากฏ

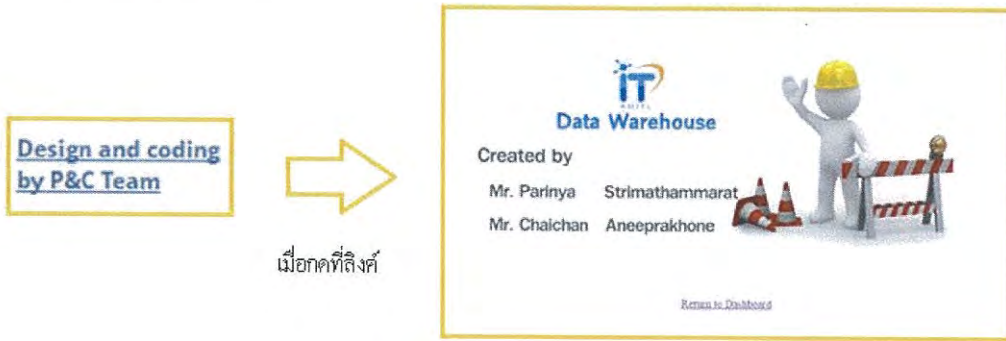
- Manage User [จำกัดสิทธิ์ ในการเข้าถึง เฉพาะ "Admin"]
- Logout ออกจากระบบ

รูปที่ ก.37 แสดงส่วนของการกำหนดสิทธิ์ผู้ใช้

- สิทธิ์ของ User จะไม่สามารถใช้งานในส่วนของการ Manage User และการเพิ่มหรือดู ข้อมูลทั้งหมดได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 4) ส่วนของผู้จัดทำ



รูปที่ ก.38 แสดงส่วนของผู้จัดทำ



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## ประวัติผู้เขียน

ชื่อ-นามสกุล	นายชายชาญ อนิประโคน
วัน เดือน ปี เกิด	18 สิงหาคม 2536
ที่อยู่	601/28 ถ.ริมคลองลาดพร้าว เขตห้วยขวาง แขวงบางกะปิ กรุงเทพฯ 10310
โทรศัพท์	090-393-5667
ประวัติการศึกษา	พ.ศ. 2558 วิทยาศาสตรบัณฑิต สาขาวิชาเทคโนโลยี สารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหาร ลาดกระบัง
ชื่อ-นามสกุล	นายปริญญา สีมารมรัตน์
วัน เดือน ปี เกิด	30 ตุลาคม 2535
ที่อยู่	48 ซ.เอกชัย 44 ถ.เอกชัย แขวงบางบอน เขตบางบอน กรุงเทพฯ 10150
โทรศัพท์	092-535-5656
ประวัติการศึกษา	พ.ศ. 2558 วิทยาศาสตรบัณฑิต สาขาวิชาเทคโนโลยี สารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหาร ลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# คลังข้อมูลสำหรับคณะเทคโนโลยีสารสนเทศ

## บนระบบอาปาเซฮาดูป

ชายชาญ อนิประโคน<sup>1</sup>, ปริญญา สีมารธรรมรัตน์<sup>2</sup>

<sup>1</sup>คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง กรุงเทพฯ

<sup>2</sup>คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง กรุงเทพฯ

Email: chaichananeepakhone@gmail.com<sup>1</sup>, parinya\_srimathammarat@hotmail.com<sup>2</sup>

### บทคัดย่อ

ในปัจจุบันการจัดการข้อมูลและวิเคราะห์ข้อมูลนักศึกษาและอาจารย์ของคณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ผู้ใช้งานจะต้องกรอกข้อมูลลงในโปรแกรมไมโครซอฟต์เอกเซล (Microsoft Excel) ตามรูปแบบที่ได้สร้างไว้ โดยจะกรอกข้อมูลในทุกๆ เทอมของแต่ละชั้นปี เพื่อใช้ในการสรุปหรือประเมินผล จากการศึกษากระบวนการทำงานในปัจจุบันพบว่าการจัดการและวิเคราะห์ข้อมูลนักศึกษาและอาจารย์มีความยุ่งยาก เนื่องจากผู้ใช้งานต้องวิเคราะห์ข้อมูลด้วยตัวเอง ซึ่งทำให้เกิดความผิดพลาดได้ง่าย และไม่สามารถตีความหมายหรือวิเคราะห์ข้อมูลออกในรูปแบบแผนภูมิภาพเพื่อใช้ประกอบการวิเคราะห์แนวโน้มที่จะเกิดขึ้นในอนาคตได้ รวมทั้งอาจมีความเสี่ยงต่อข้อมูลสูญหายและด้านความปลอดภัยของข้อมูลโดยเฉพาะข้อมูลส่วนตัวแต่ละบุคคล ดังนั้นจึงได้พัฒนาค้นคลังข้อมูลสำหรับคณะเทคโนโลยีสารสนเทศ บนระบบอาปาเซฮาดูป ขึ้นเพื่อช่วยในการเก็บข้อมูลหรือเรียกใช้ข้อมูลสารสนเทศของคณะได้อย่างรวดเร็วและแม่นยำมากยิ่งขึ้น โครงการนี้จะช่วยให้จัดการข้อมูลสารสนเทศของคณะเทคโนโลยีสารสนเทศ เพื่อใช้ประเมินผลหรือวิเคราะห์แนวโน้มในอนาคตได้อย่างมีประสิทธิภาพมากขึ้น อีกทั้งยังทำให้มีความยืดหยุ่นในการใช้งานเพิ่มมากยิ่งขึ้น

คำสำคัญ – คลังข้อมูล; อาปาเซฮาดูป; HDFS; Map Reduce; อาปาเซ ไฮฟ์; คราวเดรา อิมพาลา

### 1. บทนำ

มหาวิทยาลัยและสถาบันการศึกษาชั้นนำของประเทศไทยต่างต้องการที่จะผลิตบัณฑิตที่มีคุณภาพและเป็นที่ต้องการในตลาดแรงงาน ดังนั้น การพัฒนาค้นคลังข้อมูลของมหาวิทยาลัยและสถาบันการศึกษาจึงเป็นสิ่งจำเป็น โดยคลังข้อมูลจะช่วยให้สามารถตัดสินใจได้อย่างแม่นยำ รวดเร็ว อีกทั้งลดภาระการทำงานของบุคลากรได้ และสามารถเปรียบเทียบแนวโน้มหรือพยากรณ์สิ่งที่จะเกิดขึ้นในอนาคตได้โดยง่าย ซึ่งทั้งนี้ทำให้สามารถตัดสินใจและ

ทางการแข่งขันกับมหาวิทยาลัยและสถาบันการศึกษาชั้นนำอื่นๆได้อย่างทันท่วงที การนำระบบเทคโนโลยีคลังข้อมูลเข้ามาช่วยภายในคณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังนั้น เป็นปัจจัยสำคัญที่ช่วยในการวิเคราะห์หรือจัดเก็บข้อมูลที่ซับซ้อนของคณะให้เป็นระบบและง่ายต่อการเข้าถึงข้อมูลมากยิ่งขึ้น โดยสามารถดึงสารสนเทศจากข้อมูลออกมาใช้ประโยชน์ได้อย่างเต็มประสิทธิภาพ สามารถเข้าถึงข้อมูลได้อย่างรวดเร็วและแม่นยำ เนื่องจากปัจจุบันบุคลากรมีการจัดเก็บข้อมูลต่าง ๆ

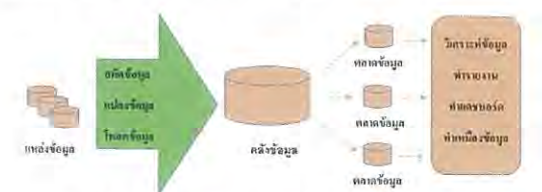
เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการใช้งานเพื่อการศึกษาระดับปริญญาโทและปริญญาเอกเท่านั้น ไม่สามารถนำข้อมูลไปใช้ในการค้า  
ฉบับ Microsoft Excel แบบแยกส่วนกัน ข้อมูลค่อนข้าง  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กระจัดกระจายกันทำให้ค้นหาข้อมูลได้ค่อนข้างยาก รวมทั้งยังต้องสรุปผลข้อมูลที่เกี่ยวข้องกับนักศึกษาและอาจารย์ของคณะในแต่ละเทอมเองทั้งหมดโดยในการจัดทำระบบในแต่ละครั้งนั้น จะมีค่าใช้จ่ายทางด้านฮาร์ดแวร์และซอฟต์แวร์ที่มีราคาค่อนข้างสูง เนื่องจากการจัดทำคลังข้อมูลนั้นจะเน้นการดึงหรือเก็บข้อมูลที่ละมากๆ จึงต้องใช้ฮาร์ดแวร์และซอฟต์แวร์ที่มีประสิทธิภาพสูงและเหมาะสมกับการใช้งาน ดังนั้นทางผู้จัดทำจึงนำเทคโนโลยีอาปาเซฮาดูป (Apache Hadoop) เข้ามาช่วยทดแทนให้คณะสามารถเก็บและดึงข้อมูลได้อย่างรวดเร็วและมีประสิทธิภาพการทำงานสูง โดยมีต้นทุนที่ค่อนข้างต่ำ โดยการนำคอมพิวเตอร์หลาย ๆ เครื่องเข้ามาช่วยกันในการทำงาน และประมวลผลแทนการใช้เซิร์ฟเวอร์ ที่สำคัญคือไม่ต้องเสียค่าลิขสิทธิ์ซอฟต์แวร์เพราะเป็นระบบโอเพนซอร์ส (Open Source) ซึ่งสามารถนำมาจัดทำเป็นแดชบอร์ด โดยแดชบอร์ดจะมีการแสดงข้อมูลและแผนภาพกราฟิกต่าง ๆ ซึ่งจัดทำขึ้นสำหรับสนับสนุนการทำงานของบุคลากรในคณะ ทั้งนี้เพื่อนำสารสนเทศของคณะไปใช้งานและสรุปผลเพื่อพัฒนา ปรับปรุงคณะต่อไป

## 2. การทบทวนทฤษฎีพื้นฐาน งานวิจัย และเทคโนโลยีที่เกี่ยวข้อง

### 2.1 ทฤษฎีและงานวิจัยที่เกี่ยวข้องกับระบบคลังข้อมูล

#### 2.1.1 คำจำกัดความของคลังข้อมูล



รูปที่ 1. แสดงหลักการการทำงานของคลังข้อมูล

คลังข้อมูล (Data Warehouse) [1] คือ ฐานข้อมูลขนาดใหญ่ที่ถูกออกแบบขึ้นมาเพื่อใช้เก็บรวบรวมข้อมูลรายการ การเปลี่ยนแปลงประจำวัน (Transaction) ขององค์กรหรือหน่วยงานต่าง ๆ ตั้งแต่อดีตจนถึงปัจจุบัน โดยไม่มีการลบ

ข้อมูลหรืออัปเดตข้อมูลถึงแม้ว่าข้อมูลนั้นจะมีการเปลี่ยนแปลงหรือไม่ตรงกับข้อมูลในปัจจุบันก็ตาม เพื่อที่สามารถเรียกดูข้อมูลย้อนหลังได้ เช่น นายชายชาญได้ทำการเปลี่ยนชื่อเป็นนายชาญชาญ โดยระบบคลังข้อมูลจะต้องทราบได้ว่า นายชายชาญกับนายชาญชาญ คือบุคคลคนเดียวกัน เป็นต้น โดยคลังข้อมูลมีวัตถุประสงค์เพื่อใช้ในการเรียกดูข้อมูล (Query) และสามารถนำมาใช้ในการวิเคราะห์ข้อมูลเพื่อใช้ประโยชน์ได้ทั้งในรูปแบบรายงาน (Report) หรือรูปแบบข้อมูลภาพรวม (Dashboard) โดยแสดงผลลัพธ์เป็นกราฟหรือแผนภูมิรูปภาพ เพื่อใช้ช่วยสนับสนุนการตัดสินใจภายในองค์กรของผู้บริหารได้ หรือทำเหมืองข้อมูล (Data Mining) เพื่อหาความสัมพันธ์ของข้อมูล และยังเป็นศูนย์รวมของข้อมูลของทั้งองค์กร ที่พร้อมต่อการใช้งาน เข้าถึงข้อมูลได้ง่าย มีความปลอดภัย ข้อมูลถูกต้อง ครบถ้วน สอดคล้องไม่ขัดแย้งกัน ซึ่งจะเข้าใจถึงหลักการการทำงานของคลังข้อมูลเบื้องต้นได้จากรูปที่ 1

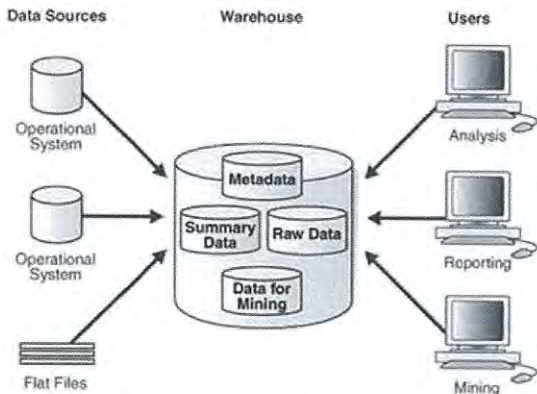
#### 2.1.2 คุณลักษณะเฉพาะของคลังข้อมูล

- 1) การแบ่งโครงสร้างตามเนื้อหาที่สนใจ (Subject Oriented) [1] คือ คลังข้อมูลถูกออกแบบมาเพื่อช่วยในการวิเคราะห์ข้อมูลที่ทางองค์กรกำลังสนใจ
- 2) การจัดเก็บรวบรวมข้อมูล (Integrated) [1] คือ การจัดเก็บข้อมูลเป็นกลุ่มของข้อมูลที่อยู่ในรูปแบบที่มีความสอดคล้องกันจากหลายๆ แหล่งข้อมูล โดยจะต้องทำการแก้ไขปัญหา เช่น การตั้งชื่อที่มีความขัดแย้งกัน เป็นต้น เมื่อแก้ไขปัญหาเหล่านี้สำเร็จ ก็จะสามารถทำการจัดเก็บรวบรวมข้อมูลเข้าด้วยกันเป็นกลุ่มของข้อมูลได้
- 3) ความสัมพันธ์กับเวลา (Time Variant) [2] คือ การจัดเก็บรวบรวมข้อมูลในอดีตไว้ในคลังข้อมูล เพื่อสามารถนำข้อมูลในอดีตมาใช้ในการวิเคราะห์หาแนวโน้ม เปรียบเทียบหรือพยากรณ์ข้อมูลขององค์กรได้
- 4) ข้อมูลไม่ควรได้รับการแก้ไข (Nonvolatile) [1] คือ ข้อมูลที่อยู่ในคลังข้อมูลจะไม่สามารถแก้ไขหรือเปลี่ยนแปลงค่าของข้อมูลได้ ผู้ใช้จะสามารถทำได้เพียงเรียกดูข้อมูลเพื่อนำมาวิเคราะห์

#### 2.1.3 สถาปัตยกรรมคลังข้อมูล

- 1) สถาปัตยกรรมคลังข้อมูลแบบพื้นฐาน (Data Warehouse Architecture Basic)

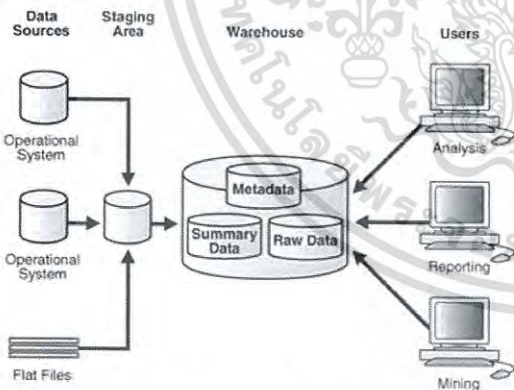
เอกสารนี้เป็นเอกสารลิขสิทธิ์ของมหาวิทยาลัยราชภัฏวไลยอลงกรณ์ การแก้ไขเพิ่มเติม เนื้อหาผู้จัดทำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2. แสดงโครงสร้างของสถาปัตยกรรมคลังข้อมูลแบบพื้นฐาน

จากรูปที่ 2 [1] จะแสดงให้เห็นถึงสถาปัตยกรรมอย่างง่ายของคลังข้อมูล โดยที่ผู้ใช้จริง (End Users) สามารถเข้าถึงข้อมูลจากแหล่งต่าง ๆ โดยตรงผ่านคลังข้อมูล ซึ่งจะเก็บเมตาดาต้า (Metadata) และข้อมูลดิบ (Raw data) ของระบบการประมวลผลธุรกรรมออนไลน์เต็ม (Online Transaction Processing : OLTP) โดยจะทำการเพิ่มชนิดของข้อมูล (Type of data) และจะสรุปผลของข้อมูล (Summary data)

2) สถาปัตยกรรมคลังข้อมูลแบบมีจุดพักชั่วคราวสำหรับตรวจสอบข้อมูล (Data Warehouse Architecture with a Staging Area)

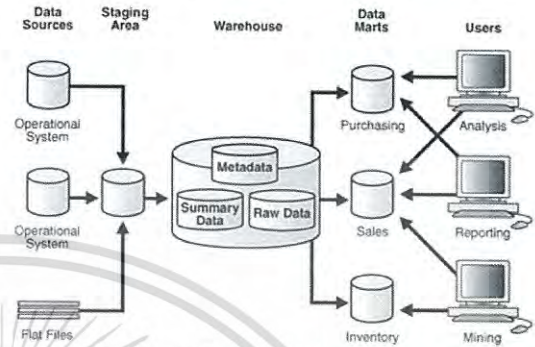


รูปที่ 3. แสดงโครงสร้างของสถาปัตยกรรมคลังข้อมูลแบบมีจุดตรวจสอบข้อมูล

จากรูปที่ 3 [1] จะแสดงให้เห็นถึงสถาปัตยกรรมของคลังข้อมูลแบบมีจุดพักชั่วคราวสำหรับตรวจสอบข้อมูล (Staging Area) โดยก่อนที่จะใส่ข้อมูลลงไปคลังข้อมูลจะต้องทำการคัดกรองและประมวลผลข้อมูลที่จุดพักชั่วคราวสำหรับตรวจสอบข้อมูลก่อน โดยจุดพักชั่วคราว

สำหรับตรวจสอบข้อมูลนี้จะช่วยลดความยุ่งยากในการสรุปผลและจัดการคลังข้อมูลแบบทั่วไป

3) สถาปัตยกรรมคลังข้อมูลแบบมีจุดพักชั่วคราวสำหรับตรวจสอบข้อมูล และตลาดข้อมูล (Data Warehouse Architecture with a Staging Area and Data Marts)

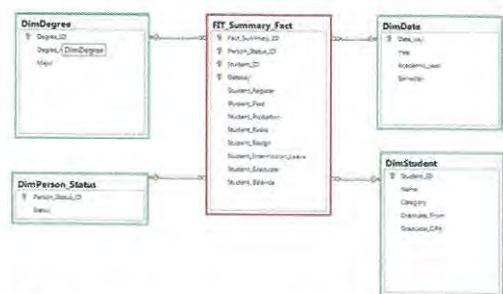


รูปที่ 4. แสดงโครงสร้างของสถาปัตยกรรมคลังข้อมูลแบบมีจุดตรวจสอบข้อมูลและตลาดข้อมูล

จากรูปที่ 4 [1] จะแสดงให้เห็นถึงสถาปัตยกรรมของคลังข้อมูลแบบมีจุดพักชั่วคราวสำหรับตรวจสอบข้อมูลโดยจะทำการเพิ่มตลาดข้อมูล (Data Mart) เข้าไป โดยสถาปัตยกรรมนี้ถูกออกแบบมาเพื่อใช้ในสายธุรกิจ โดยเฉพาะ แสดงให้เห็นถึงตัวอย่างของข้อมูลการจัดซื้อ การขาย และสินค้าคงเหลือที่จะถูกแยกออก ตัวอย่างเช่น นักวิเคราะห์ทางการเงินต้องการที่วิเคราะห์ข้อมูลการซื้อขายในอดีตเพื่อใช้พยากรณ์แนวโน้มเกี่ยวกับพฤติกรรมของลูกค้า

### 2.1.4 ขั้นตอนการสร้างคลังข้อมูล

1) การออกแบบคลังข้อมูลจะออกแบบโดยใช้แบบจำลองหลายมิติ (Multidimensional Modeling) แบ่งออกเป็น 2 แบบ ดังนี้



รูปที่ 5. แสดงแบบจำลองข้อมูลเชิงมิติแบบดาวที่แยกประเภทของตารางด้วยสี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางข้อเท็จจริง (Fact Table) คือ ตารางหลักที่จะเก็บข้อมูลข้อเท็จจริง (Fact) ของคลังข้อมูลไว้ ซึ่งข้อมูลข้อเท็จจริงจะเป็นชุดของค่าที่จับคู่กันระหว่างจำนวนคอลัมน์ที่เป็นข้อมูลชนิดตัวเลข (Measure) ในตารางข้อเท็จจริงกับข้อมูลที่เป็นมุมมองในการวิเคราะห์ข้อมูล (Dimension) ในตารางมิติ เพื่อให้เกิดข้อเท็จจริงที่สามารถเข้าใจและวิเคราะห์ข้อมูลได้ เวลาจะใช้ตารางข้อเท็จจริงจะทำการรวมตารางข้อมูล (Join) ซึ่งจากรูปที่ 5 ตารางข้อเท็จจริงจะแสดงอยู่ในกรอบสีแดง

ตารางมิติ (Dimension Table) คือ ตารางหลาย ๆ ตารางที่เชื่อมโยงกับตารางข้อเท็จจริง โดยจะมีข้อมูลทั้งที่เป็นข้อมูลที่เป็นมุมมองในการวิเคราะห์ข้อมูลและจำนวนคอลัมน์ที่เป็นข้อมูลชนิดตัวเลขเป็นส่วนประกอบ เพื่อใช้อธิบายรายละเอียดของข้อเท็จจริงที่อยู่ในตารางข้อเท็จจริง เพราะข้อมูลส่วนใหญ่ในตารางข้อเท็จจริงจะเป็นข้อมูลชนิดตัวเลข ทำให้ไม่เข้าใจความหมายของข้อมูลเหล่านั้น จึงต้องมีข้อมูลที่เป็นมุมมองในการวิเคราะห์ข้อมูลเข้ามาช่วย เพื่อให้สามารถเข้าใจและวิเคราะห์ข้อมูลได้ ซึ่งจากรูปที่ 5 ตารางมิติจะแสดงอยู่ในกรอบสีเขียว

2) การออกแบบคลังข้อมูลยังสามารถออกแบบเป็นแบบจำลองข้อมูลหลายมิติตามคุณลักษณะได้ 2 แบบ ดังนี้



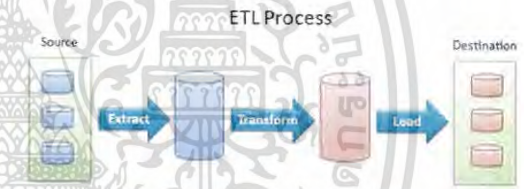
รูปที่ 6. แสดงแบบจำลองข้อมูลเชิงมิติแบบเกล็ดหิมะ

แบบจำลองข้อมูลเชิงมิติแบบดาว (Star Schema Model) คือ เป็นแบบจำลองข้อมูลเชิงมิติที่ประกอบไปด้วยตารางมิติหลาย ๆ ตาราง ล้อมรอบตารางข้อเท็จจริงที่อยู่ตรงกลาง มีลักษณะและความสัมพันธ์ที่คล้ายกับดวงดาว มีข้อดีคือสามารถวิเคราะห์และเรียกดูข้อมูลได้อย่างมีประสิทธิภาพสูงจำนวนมาก ได้ง่าย และรวดเร็ว โดยแบบจำลองนี้จะเหมาะกับงานที่ค่อนข้างไม่มีความซับซ้อน ซึ่งจะแสดงในรูปที่ 5

แบบจำลองข้อมูลเชิงมิติแบบเกล็ดหิมะ (Snowflake Schema Model) คือ มีตารางข้อเท็จจริงเป็นศูนย์กลางอยู่ตรงกลางและเชื่อมโยงกับตารางมิติหลายๆ ตาราง โดย

ตารางมิติที่เชื่อมกับตารางข้อเท็จจริงอยู่นี้จะสามารถแตกเป็นตารางย่อยๆ ลงไปได้ ซึ่งตารางแบบนี้จะเหมาะกับคลังข้อมูลขนาดใหญ่ที่ค่อนข้างมีความซับซ้อน มีข้อดีคือเก็บข้อมูลได้น้อยลง ไม่ซ้ำซ้อน แต่มีข้อเสียคือ ค่อนข้างซับซ้อน ทำความเข้าใจได้ยาก เวลาเรียกดูข้อมูลจะใช้เวลาค่อนข้างนาน เพราะจะต้องทำการรวมตารางหลายตารางที่จะเรียกดูเข้าด้วยกัน ซึ่งแบบจำลองข้อมูลเชิงมิติแบบเกล็ดหิมะจะแสดงในรูปที่ 6

3) กระบวนการอีทีแอล (ETL) [3] การสกัดข้อมูล (Extraction) คือการจัดเก็บรวบรวมข้อมูลจากแหล่งต่างๆ แล้วนำข้อมูลมาคัดกรอง การแปลงข้อมูล (Transformation) คือ การตรวจสอบแก้ไขให้ถูกต้อง และจัดรูปแบบใหม่ให้เป็นไปตามความต้องการของงาน ดำเนินงาน และการโหลดข้อมูล (Load) คือ ข้อมูลที่ได้จะถูกโหลดเข้าสู่คลังข้อมูลเป้าหมาย เพื่อนำข้อมูลที่ได้ไปใช้ในการวิเคราะห์ข้อมูล ซึ่งกระบวนการอีทีแอลแสดงในรูปที่ 7



รูปที่ 7. แสดงกระบวนการ ETL

## 2.2 เทคโนโลยีที่เกี่ยวข้องกับโครงการ

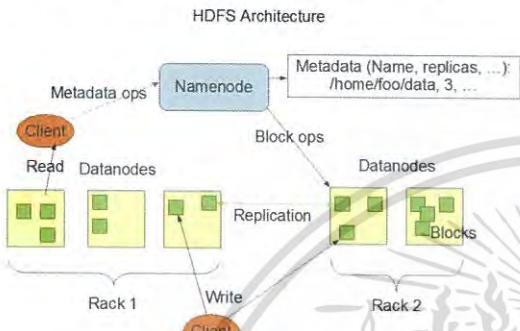
### 2.2.1 เทคโนโลยีอาปาเซฮาดูป

อาปาเซฮาดูป (Apache hadoop) เป็นซอฟต์แวร์แบบโอเพ่นซอร์ส (Open source software) ที่มีความน่าเชื่อถือและมีความยืดหยุ่นสามารถปรับขนาดได้ตามข้อมูลที่มีการเปลี่ยนแปลงอยู่ตลอดเวลา เหมาะสำหรับการประมวลผลข้อมูลแบบกระจาย (Distributed Computing) รองรับการจัดเก็บและประมวลผลข้อมูลขนาดใหญ่ด้วยการเชื่อมต่อกับเครื่องเซิร์ฟเวอร์จำนวนมากเข้าด้วยกัน โดยการส่งข้อมูลจะสามารถทำงานได้ถึงแม้ข้อมูลในระบบจะมีการทำงานล้มเหลว เนื่องจากข้อมูลจะถูกทำการสำรองไว้ในเซิร์ฟเวอร์อย่างน้อย 3 เซิร์ฟเวอร์ เมื่อมีข้อมูลใดข้อมูลหนึ่งเกิดล้มเหลวขึ้นก็จะทำการไปเรียกข้อมูลสำรองจากเซิร์ฟเวอร์อื่นมาใช้งานได้อย่างทันที ซึ่งอาปาเซฮาดูปจะมีส่วนในการทำงานอยู่ 2 ส่วนหลัก ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1) หลักการทำงานของฮาดูป

-ระบบจัดการไฟล์แบบกระจายของฮาดูป (Hadoop Distributed File System: HDFS) คือ ระบบที่ถูกออกแบบมาใช้จัดเก็บข้อมูลขนาดใหญ่ โดยจะทำงานบนระบบประมวลผลคลัสเตอร์ (Cluster Computing System) โดยมีโหนดการทำงานอยู่ 2 แบบด้วยกัน คือ Name node กับ Data node ดังภาพ

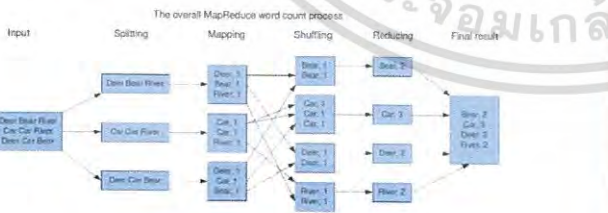


รูปที่ 8. แสดงหลักการการทำงานของ HDFS

Name node คือ โหนดศูนย์กลาง จะคอยเก็บที่อยู่ของแต่ละไฟล์ที่ทำการแบ่งไฟล์ เมื่อเครื่องไคลเอนต์ต้องการไฟล์ข้อมูล Name node ก็จะไปเรียกไฟล์ข้อมูลจากที่อยู่ของ Data node ที่เก็บกลุ่มไฟล์นั้นไว้

Data node คือ โหนดที่เก็บไฟล์แต่ละไฟล์ทั้งที่ถูกแบ่งและที่ทำสำเนาไว้ คอยบอกสถานะของไฟล์ว่ายังทำงานปกติดีหรือไม่ แก่ Name node

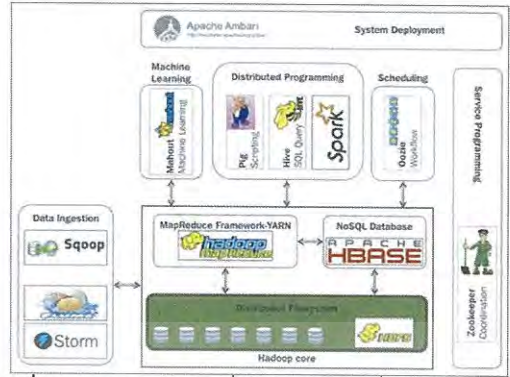
-การประมวลผลข้อมูลด้วยแมปรีดิวซ์ (Map Reduce) ซึ่งเป็นแบบจำลองอย่างง่ายจะประกอบไปด้วย 2 ส่วน ดังนี้



รูปที่ 9. แสดงหลักการการทำงานของ Map Reduce

แมป (Map) คือ การที่ข้อมูลเข้าสู่เครื่องจะแบ่งข้อมูลแล้วทำการกระจายไปยังโหนดต่าง ๆ โดยแต่ละโหนดจะทำการแมปข้อมูลที่เหมือนกันเข้าด้วยกันเป็น Key/Value เมื่อข้อมูลผ่านการแมปแล้วก็จะทำการ Shuffling ข้อมูล โดยนำข้อมูลที่เหมือนกันในแต่ละโหนดมาไว้ด้วยกัน แล้วทำการลดจำนวน Key/Value ลง เรียกว่า การรีดิวซ์ (Reduce)

2) ระบบนิเวศเชื่อมโยงของฮาดูป (Apache Hadoop Ecosystem)



รูปที่ 10. แสดงแอปพลิเคชันในระบบนิเวศเชื่อมโยงของฮาดูป

ฮาดูปมีค่าใช้จ่ายที่ต่ำ แต่กลับมีประสิทธิภาพที่ค่อนข้างสูงในการประมวลผลข้อมูลขนาดใหญ่ และก็มีการพัฒนาต่อยอดอย่างต่อเนื่องให้เข้าถึงผู้ใช้งานได้มากขึ้น ซึ่งฮาดูปจะมีแอปพลิเคชันที่ใช้ร่วมกันอยู่หลายตัว เช่น Apache Ambari, Apache Avro, Apache Cassandra, Apache Chukwa, Apache Flume, Apache HBase, Apache Mahout, Apache Oozie, Apache Pig, Apache Sqoop, Apache Spark, Apache ZooKeeper, Apache Hive และ Cloudera Impala เป็นต้น

3) ฮาดูปเชื่อมกับคราวเดราอิมพาลา



รูปที่ 11. แสดงการเปรียบเทียบการทำงานของ Hive และ Impala

ฮาดูป [4] ถูกออกแบบมาสำหรับระบบคลังข้อมูลเพื่อความสะดวกในการประมวลผลของคำสั่งการคิวรีข้อมูลแบบ Adhoc และชุดข้อมูลขนาดใหญ่ที่เก็บไว้ใน HDFS และความสะดวกในการรวมข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น มิใช่สัญญาที่นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คลาวเดรา อิมพาลาเป็นโอเพนซอร์สที่พัฒนาขึ้นต่อจาก Google Dremel อิมพาลาเป็นเครื่องมือสำหรับการประมวลผลข้อมูลที่เก็บไว้ใน HBase และ HDFS อิมพาลาใช้ Hive Megastore และสามารถคิวรีตารางโฮปได้โดยตรง โดยแตกต่างจากโฮปที่ตรงที่อิมพาลาไม่ได้แปลงคำสั่งเข้าไปใน MapReduce แต่สามารถดำเนินการได้ด้วยตัวเอง ซึ่งจะเห็นได้จากรูปที่ 11

อิมพาลาจะสามารถทำงานระดับสูงร่วมกับอปาเซโฮปและเข้ากันได้กับไวยากรณ์ของ HiveQL ช่วยให้สามารถใช้ได้ทั้งอิมพาลาหรือโฮปในการสร้างตาราง คิวรีข้อมูล โหลดข้อมูล และอื่นๆ

อย่างไรก็ตามทั้งอปาเซโฮปและคราวเดราอิมพาลาจะสนับสนุนมาตรฐานทั่วไปของ HiveQL

โฮปเข้าแต่เหมาะที่จะเป็นตัวเลือกที่ดีสำหรับงาน ETL ที่หนักๆ แต่น่าเชื่อถือ เช่น รวมล็อก (Log) รายชั่วโมงสำหรับองค์กรโฆษณา ที่สามารถนำมาใช้อย่างมีประสิทธิภาพสำหรับการประมวลผลคำสั่งในข้อมูลที่มีปริมาณมาก อิมพาลาจะประมวลผลได้เร็วกว่าและจัดการข้อมูลขนาดใหญ่ได้มากกว่าเครื่องมือคิวรีแบบโฮป

อิมพาลาจะเร็วกว่าอปาเซโฮป แต่ไม่ได้หมายความว่า จะเป็นการแก้ปัญหา SQL สำหรับทุกปัญหาของข้อมูลขนาดใหญ่ ได้ดีทั้งหมด เพราะอิมพาลาจะเน้นการใช้หน่วยความจำเป็นหลักและไม่ได้ทำงานอย่างมีประสิทธิภาพสำหรับการดำเนินงานข้อมูลอย่างหนักเช่นการ joins เพราะมันเป็นไปได้ที่จะทำทุกอย่างในหน่วยความจำ จึงนำโฮปเข้ามาช่วย ถ้าแอปพลิเคชันที่มีชุดการประมวลผลประเภท ความต้องการมากกว่าข้อมูลขนาดใหญ่เมื่อนั้นองค์กร จะต้องเลือกใช้โฮป แต่ถ้าหากองค์กรต้องการการประมวลผลด้วยเวลาจริงของการคิวรีแบบ Adhoc ในสับเซตของข้อมูลแล้ว อิมพาลาจึงเป็นทางเลือกที่ดี ซึ่งเราสามารถดูการเปรียบเทียบการทำงานระหว่างโฮปกับอิมพาลาได้ในตารางที่ 1 และ 2

ตารางที่ 1 แสดงการเปรียบเทียบระหว่าง Pig, Hive และ Impala

Feature	Pig	Hive	Impala
SQL-based query language	No	Yes	Yes
Optional schema and metastore	Yes	No	No
User-defined functions (UDFs)	Yes	Yes	Yes
Process data with external scripts	Yes	Yes	No
Extensible file format support	Yes	Yes	No
Complex data types	Yes	Yes	No
Query latency	High	High	Low
Built-in data partitioning	No	Yes	Yes
Accessible via ODBC / JDBC	No	Yes	Yes

ตารางที่ 2 แสดงการเปรียบเทียบระหว่าง RDBMS, Hive และ Impala

Feature	RDBMS	Hive	Impala
Insert individual records	Yes	No	Yes
Update and delete records	Yes	No	No
Transactions	Yes	No	No
Role-based authorization	Yes	Yes (Sentry)	Yes (Sentry)
Stored procedures	Yes	No	No
Index support	Extensive	Limited	None
Latency	Very low	High	Low
Data size	Terabytes	Petabytes	Petabytes
Complex data types	No	Yes	No
Storage cost	Very high	Very low	Very low

## 2.2.2 เทคโนโลยีที่ใช้ในการสร้างเว็บแอปพลิเคชัน

### 1) เทคโนโลยี JSP

Java Server Page (JSP) เป็นเทคโนโลยีสำหรับการควบคุมเนื้อหาหรือลักษณะของเว็บเพจโดยการใช้ Servlets ซึ่งเป็นโปรแกรมขนาดเล็กที่มีการระบุไว้ในเว็บเพจและทำงานบนเว็บเซิร์ฟเวอร์ในการปรับเปลี่ยนเว็บเพจก่อนที่จะถูกส่งไปยังผู้ใช้ที่ร้องขอ JSP นอกจากนี้เทคโนโลยี JSP ยังเป็น Servlet แสดงผลโปรแกรม แอปพลิเคชัน (Application Program Interface: API) อีกด้วย

### 2) เทคโนโลยี Java

จาวา (Java) เป็นภาษาโปรแกรมที่ออกแบบมาให้มี "รูปลักษณ์และความรู้สึก" แบบภาษา C++ แต่มันจะง่ายกว่า และบังคับใช้รูปแบบการเขียนโปรแกรมเชิงวัตถุ จาวาสามารถใช้ในการสร้างแอปพลิเคชันที่สมบูรณ์แบบที่อาจทำงานบนคอมพิวเตอร์เครื่องเดียวหรือกระจายในเซิร์ฟเวอร์และไคลเอนต์ในเครือข่าย

### 3) เทคโนโลยี JavaScript

จาวาสคริปต์ (JavaScript) ประกอบด้วยไลบรารีมาตรฐานของออบเจกต์ เช่น อาร์เรย์, วันที่, และคณิตศาสตร์และชุดองค์ประกอบหลักของภาษา เช่น คำนิยาม การสร้างการควบคุมและคำสั่ง จาวาสคริปต์โดยหลักแล้วจะสามารถขยายความหลากหลายของวัตถุประสงค์ได้โดยการเสริมด้วยออบเจกต์เพิ่มเติม

### 4) เทคโนโลยี CSS

Cascading Style Sheet (CSS) เป็นเว็บเพจที่ได้มาจากหลายแหล่งที่มาที่มีกำหนดลำดับความสำคัญไว้และมีการนิยามความขัดแย้งขององค์ประกอบ กำหนด style sheets หรือ statements ที่เป็นไปได้ โดยอาจจะกำหนดองค์ประกอบที่ได้เป็นการนำเสนอในเว็บเพจ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 5) เทคโนโลยี jQuery

jQuery เร็ว มีขนาดเล็ก และมีไลบรารีจาวาสคริปต์ที่มีคุณลักษณะหลากหลาย jQuery จัดการเอกสาร HTML การจัดการเหตุการณ์ ภาพเคลื่อนไหว และ AJAX ที่ง่ายมาก ด้วย API ที่ง่ายต่อการใช้งานที่ทำงานในเบราว์เซอร์ที่หลากหลายหลาย ด้วยการรวมกันของความคล่องตัวและการขยาย jQuery จะมีการเปลี่ยนแปลงวิธีการใช้งานของคนนับล้านที่เขียนด้วย JavaScript

### 6) เทคโนโลยี Highcharts

Highcharts เป็นไลบรารีที่ใช้ในการเขียนแผนภูมิด้วยจาวาสคริปต์ ที่มีวิธีนำเสนออย่างง่าย ด้วยการเพิ่มแผนภูมิการโต้ตอบไปยังเว็บไซต์หรือเว็บแอปพลิเคชัน Highcharts ที่สนับสนุน ได้แก่ เส้น, โคน, พื้นที่, พื้นที่โค้ง, คอลัมน์, แท่ง, พาย, กระจาย, มาตราวัดเกจ, ช่วงพื้นที่, ช่วงพื้นที่โค้ง, ช่วงคอลัมน์, แผนภาพฟอง, แผนภาพกล่อง, แผนภูมิแท่งบอกความคลาดเคลื่อน, แผนภูมิน้ำตกและแผนภูมิขั้วโลก เป็นต้น

## 3. วิธีการดำเนินงาน

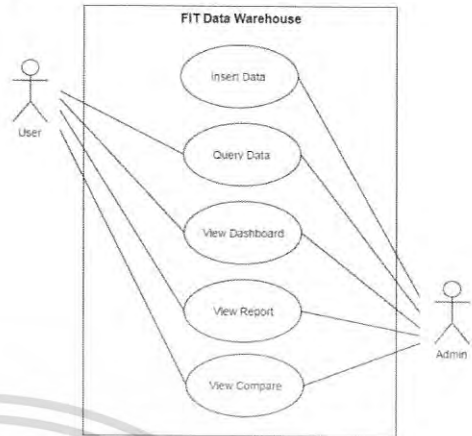
### 3.1 การออกแบบคลังข้อมูล

สร้างแผนภาพ ER (ER Diagram) เพื่อแสดงความสัมพันธ์ระหว่าง มิติของตาราง กับตารางข้อเท็จจริง โดยการออกแบบ Logical Data Model Star schema เป็นเทคนิคที่ใช้ ฐานข้อมูลหลายมิติ (Multidimensional Model) เน้นที่ประสิทธิภาพความเร็วในการดึง/เขียนข้อมูลในระบบเป็นหลัก และการออกแบบนี้ยังสามารถดึงข้อมูลหลายมิติได้ โดยการออกแบบ Star Schema (snowflake) ที่ใช้ในโครงการนี้มีทั้งหมด 2 Star Schema ดังนี้



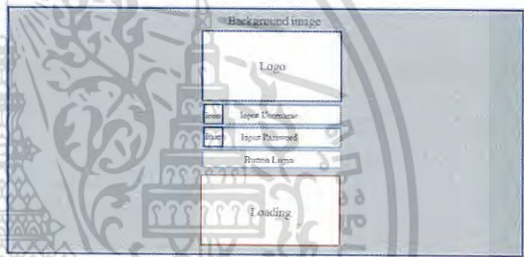
รูปที่ 12. แสดง Star Schema สำหรับเก็บข้อมูลของคณะเทคโนโลยีสารสนเทศ

### 3.2 การออกแบบยูสเคสไดอะแกรม (Use Case Diagram)

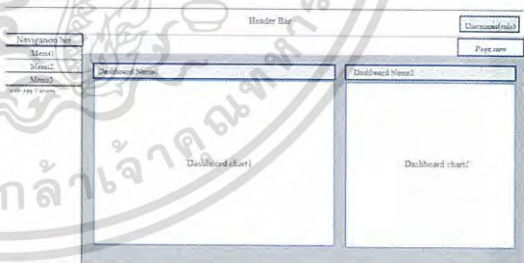


รูปที่ 13. แสดงการออกแบบยูสเคสไดอะแกรม

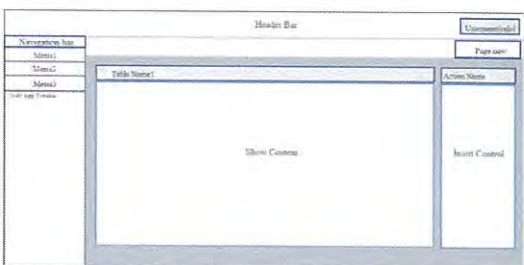
### 3.3 ออกแบบส่วนติดต่อกับผู้ใช้



รูปที่ 14. แสดงไวยากรณ์หน้าแรก (Login) ของระบบคลังข้อมูล



รูปที่ 15. แสดงไวยากรณ์หน้าแดชบอร์ด (Dashboard) และรีพอร์ต(Report) ของระบบคลังข้อมูล



รูปที่ 16. แสดงไวยากรณ์หน้าเพิ่ม/คิวรี่ข้อมูล (Put/Query Data) ของระบบคลังข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

