



Unit 9 Data Accumulation

Data Accumulation



Data แปลว่า ข้อมูล



Accumulation แปลว่า การรวบรวม, ลังสม, สะสม



หมายถึง การรวบรวม เปลี่ยนแปลง เก็บข้อมูล

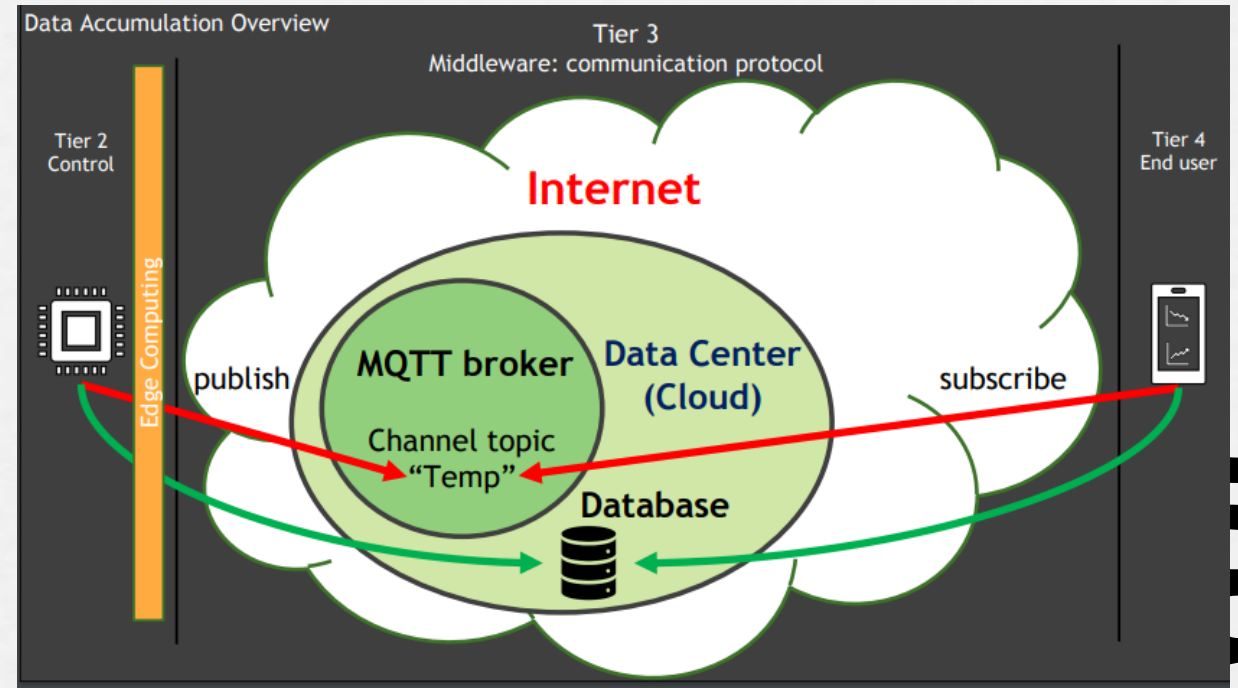


ดังนั้น data accumulation ในระบบ IOT คือการรวบรวม เปลี่ยนแปลง เก็บข้อมูลที่ถูกสร้างจากอุปกรณ์ IOT เช่น เซ็นเซอร์ที่ถูกส่งผ่านระบบเครือข่าย (computer network) ทั้งมีและไม่มีสายค่าจะไปเก็บไว้ใน DB ในระบบ cloud



ในปัจจุบันได้ในระบบ IOT ที่เป็นการรวบรวมข้อมูลจาก tier1 ไป 4 จะมี 2 รูปแบบ คือ

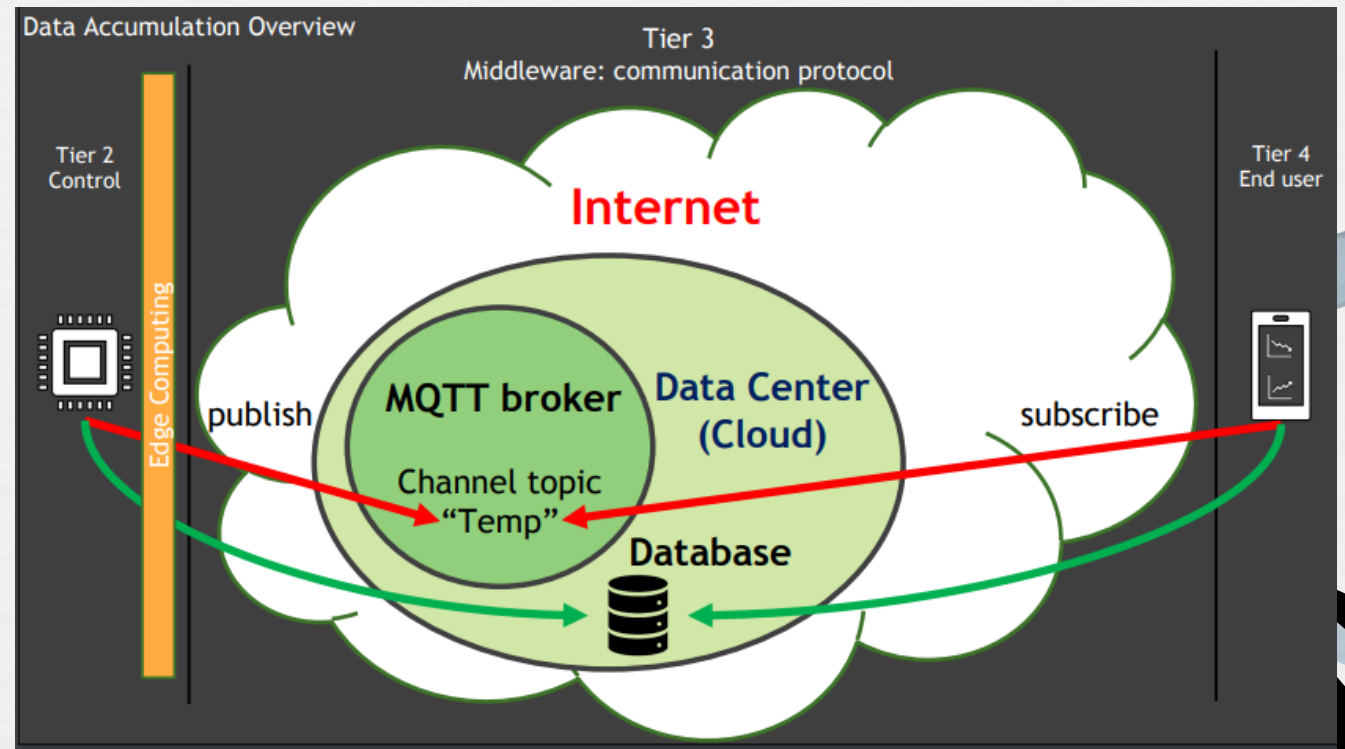
1. การเก็บข้อมูลลงฐานข้อมูลเพื่อการเรียกขอเพื่อนำไปใช้
2. การนำข้อมูลไปแสดงทันทีโดยที่ไม่ต้องมีการเก็บข้อมูล



Data Accumulation Diagram



จากรูปจะเห็นได้ว่าลูกครีแสดงจะเป็นการนำค่ามาเก็บไว้ใน cloud โดยใช้โปรโตคอล MQTT ซึ่งจะไม่มีการบันทึกข้อมูลลงฐานข้อมูลดังแสดงด้วยลูกครีแดง ในขณะที่ลูกครีเขียวแสดงถึงการบันทึกข้อมูลลงฐานข้อมูลเพื่อที่จะรอการร้องขอ (request) จาก tier4 ในการเลือกใช้งาน 2 สองแบบนี้ขึ้นอยู่กับความต้องการของระบบและจะมีโปรโตคอลให้เลือกใช้งานที่หลากหลาย



MQTT : Message Queuing Telemetry Transport



ความหมาย การส่งข้อความระยะไกลแบบตามลำดับ



Message = ข้อความ



Queuing = การเข้าคิว การทำตามลำดับ



Telemetry = การตรวจวัดจากระยะไกล



Transport = การขนส่ง



เป็นโปรโตคอลที่สร้างขึ้นเพื่อ IOT โดยเฉพาะเป็นการรับส่งข้อมูลแบบ publish/subscribe เป็นการเชื่อมต่อแบบ M2M (Machine to Machine) คือ อุปกรณ์กับอุปกรณ์



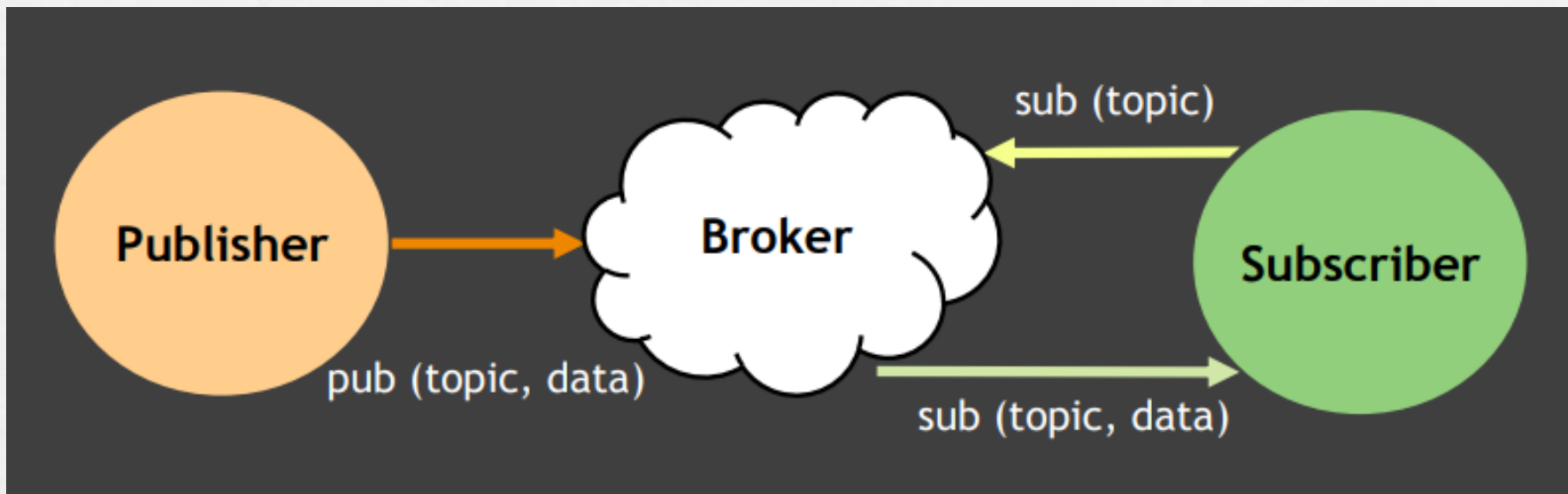
ออกแบบมาสำหรับอุปกรณ์ที่มีทรัพยากรจำกัด เช่น โหนดเซ็นเซอร์ที่ใช้พลังงานจากแบตเตอรี่และมีการรับ-ส่งข้อมูลในปริมาณที่น้อย โดยตัว MQTT จะใช้พลังงานน้อย มีการเชื่อมต่ออยู่บนโปรโตคอลรับส่งแบบ TCP/IP ส่งผลให้ข้อมูลนั้นได้รับการตรวจสอบอย่างครบถ้วน

MQTT architecture






ประกอบด้วย 3 ส่วน



1. Broker
2. Publisher
3. Subscriber





MQTT Publisher : pub

-  ทำหน้าที่ในการส่งข้อมูลซึ่งโดยปกติแล้วจะเป็น microcontroller ทำหน้าที่ส่งข้อมูลไปยังระบบ cloud
-  ระบบที่อยู่บน cloud คือ broker ในลักษณะนี้มักจะเป็นระบบที่เป็นการ monitor สภาพแวดล้อม เช่น ดูปุณภูมิและความชื้นสัมพัทธ์ในอากาศของสวนผัก เป็นต้น
-  user tier4 ก็เป็น publisher ได้ เช่น ระบบเปิด/ปิดหลอดไฟผ่านโทรศัพท์มือถือ เป็นต้น

MQTT Subscriber : sub

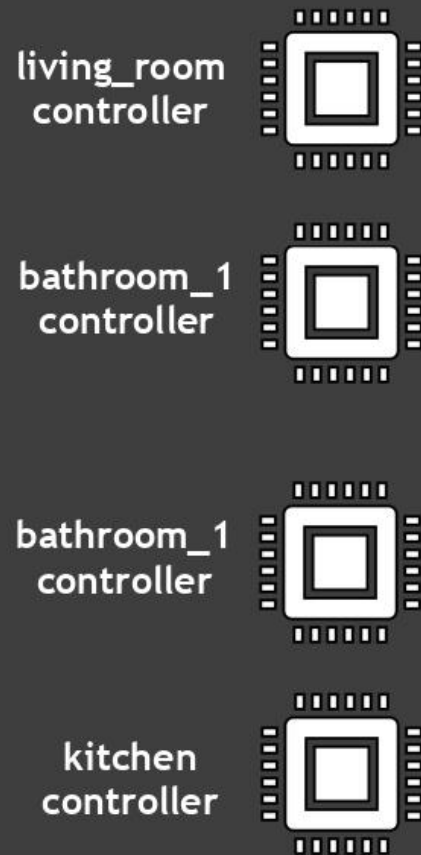
-  เป็นอุปกรณ์ที่อยู่ฝั่งตรงข้ามกับ pub เช่น ถ้า pub เป็น tier2 ทางฝั่ง sub จะเป็น tier4 และในทางกลับกันจะสลับกัน
-  ทั้ง 2 จะทำการ pub/sub ผ่านตัวกลางคือ broker

MQTT Broker

-  ทำหน้าที่เป็นตัวกลางระหว่าง pub และ sub โดยผู้พัฒนาจะทำการตั้งค่า topic เพื่อให้ทั้ง 2 ฝ่ายสื่อสารกันได้ การติดตั้ง broker นั้นสามารถติดตั้งบน cloud ที่อยู่บน server ใดๆ โดยปกติแล้วจะไม่มี การบันทึกข้อมูลลงในฐานข้อมูล
-  broker จะทำการบันทึกค่าล่าสุดที่ pub ได้ทำการ publish ขึ้นมา และเมื่อมีค่าใหม่ถูกส่งเข้ามา ก็จะมาแสดงแทนค่าเดิม ดังนั้นทางฝั่งของ sub เมื่อทำการ subscribe เข้ามาที่ topic ก็จะเห็นค่าล่าสุดที่มีใน topic บน broker ไปซึ่งถือว่าเป็นโปรโตคอลที่แสดงผลแบบ real-time

MQTT Diagram (tier-4 to tier-2)

Subscriber



subscribe

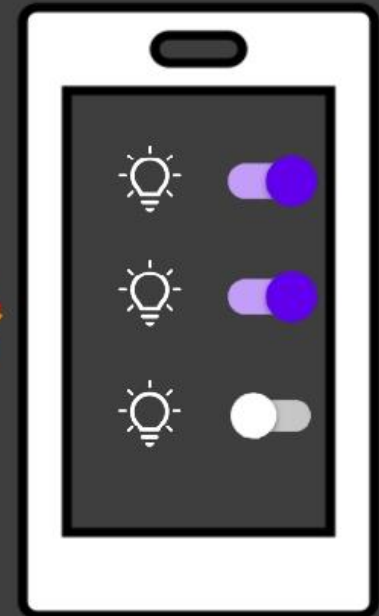
MQTT Broker

Topics



Publisher

publish



MQTT Diagram (tier-4 to tier-2)

Subscriber

subscribe



MQTT Broker

Topics

"water_level"
value = "80"

"air_temp"
value = "32.65"

"water_temp"
value = "25.81"

"water_salinity"
value = "8.2"

publish

Publisher

water tank
controller



water level
sensor



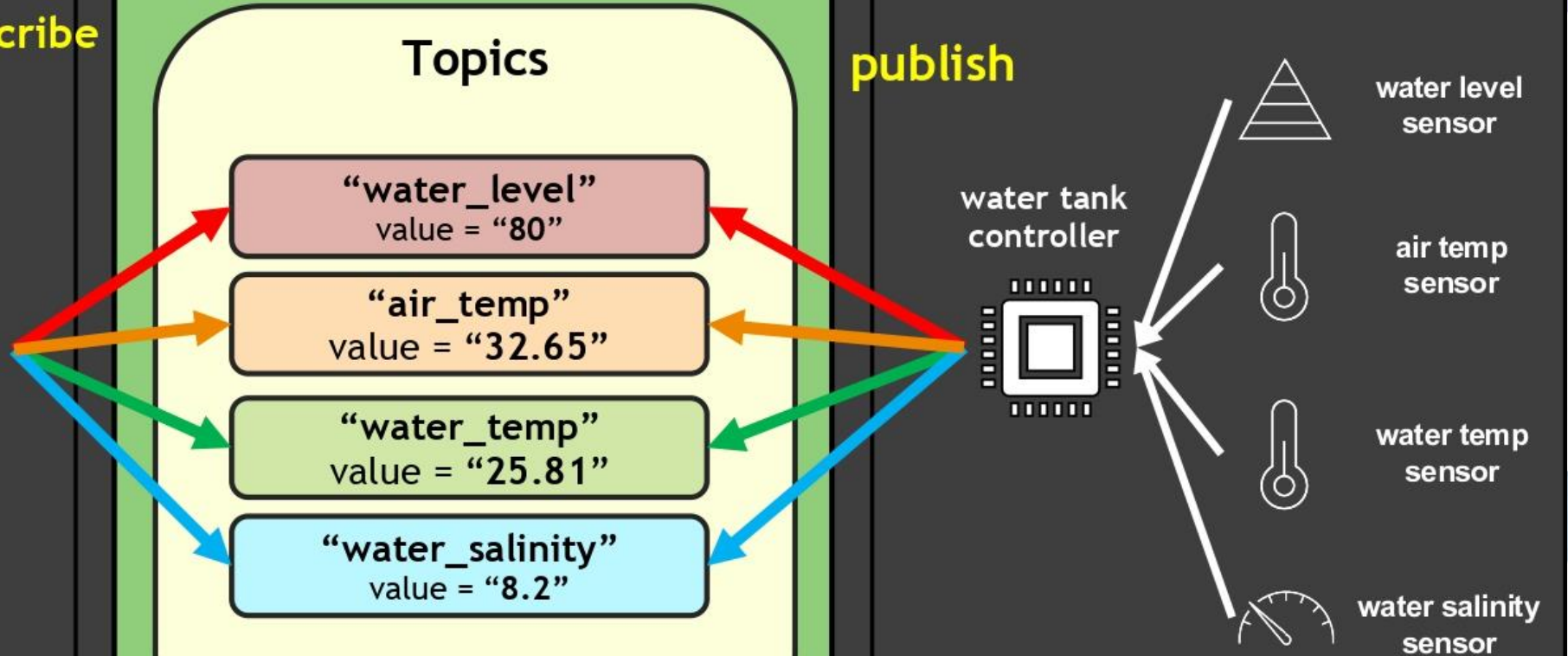
air temp
sensor



water temp
sensor



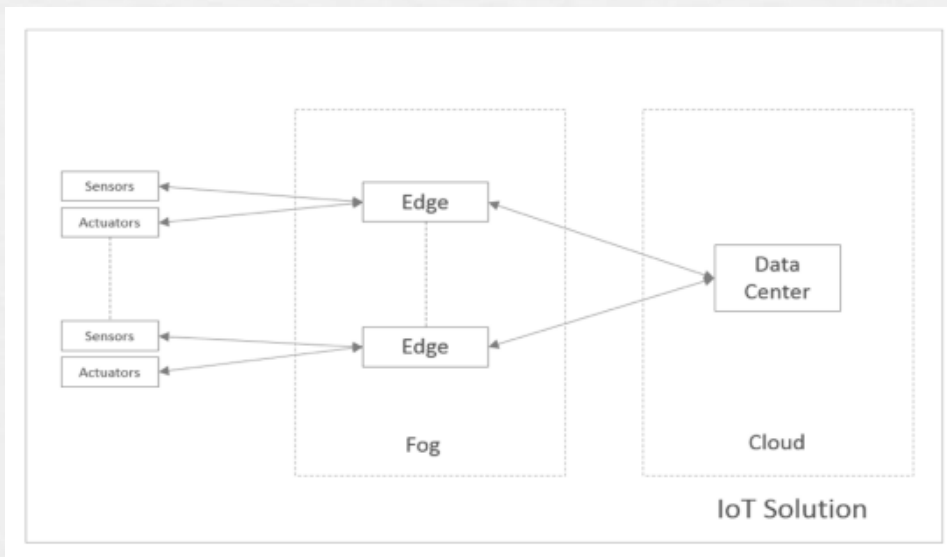
water salinity
sensor



การส่งข้อมูลในกรณีที่มี edge



เมื่อ tier-2 ส่งข้อมูลออกไปผ่านเครือข่ายไร้สายแล้วจะถูกนำไปพักไว้ที่ edge แล้วรวบรวมส่งต่อไปที่ main data center ในคลาวด์โดยความถี่ในการส่งจาก edge ไปยัง main data center จะขึ้นอยู่กับผู้ดูแลระบบเป็นผู้กำหนด



การส่งข้อมูลแบบบันทึกไว้ที่ฐานข้อมูล



ระบบ IoT ที่มีระบบฐานข้อมูลโดยจะมีการบันทึกข้อมูลที่ถูกส่งจาก tier2 หรือ tier4 ก็ได้โดยสามารถใช้ระบบฐานข้อมูลใดก็ได้เช่น relational database (ต.ย. SQL) หรือ real-time database (ต.ย. firebase) ซึ่งการสร้างระบบฐานข้อมูลนั้นสามารถสร้างได้ด้วยวิธีการสร้างฐานข้อมูลแบบปกติตามทฤษฎีและวิธีการของฐานข้อมูลนั้นๆ

Database



คือ ที่เก็บข้อมูลที่ถูกส่งมาจาก microcontroller ใน tier2 หรือ 4 ก็ได้ข้อมูลจะถูกบันทึกไว้ในฐานข้อมูลได้เรื่อยๆ ขึ้นอยู่กับขนาดของพื้นที่เก็บ (harddisk)

Web Socket

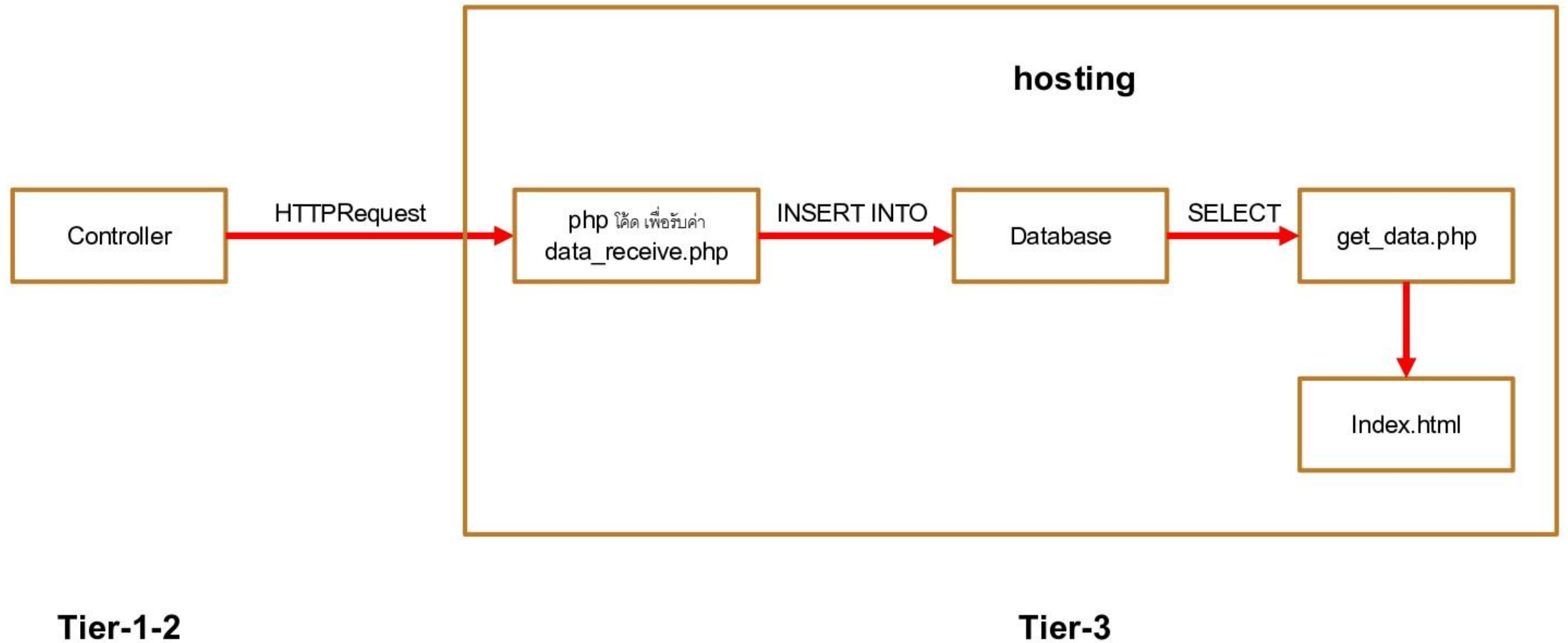


วิธีการหนึ่งที่สามารถส่งข้อมูลจาก tier2 ไปยังฐานข้อมูล SQL คือการใช้ Web socket protocol โดยจะต้องทำการสร้างไฟล์ขึ้นมาบน server เพื่อรอรับค่าจาก controller หลังจากนั้นในไฟล์นี้จะทำการ insert ข้อมูลลงในตารางฐานข้อมูล



ในฝั่งของ controller จะใช้คำสั่ง “HTTPRequest” เพื่อเชื่อมต่อไปยังไฟล์ใน server ที่รอรับข้อมูล ดังแผนภาพในสไลด์ถัดไป

- ถ้ามี database ต้องใช้ HTTPRequest
- แต่ถ้าไม่มี database ให้ใช้ MQTT ก็ได้





Big Data



แปลว่า ข้อมูลขนาดใหญ่



ปัจจุบันระบบคอมพิวเตอร์หลายระบบสามารถขับเคลื่อนได้ด้วยข้อมูล เช่นระบบพยากรณ์ระบบแนะนำ ซึ่งหากต้องการสร้างระบบที่มีประสิทธิภาพ ให้ผลที่แม่นยำและเที่ยงตรงจะต้องอาศัยข้อมูลที่มีมาก ยิ่งเยอะยิ่งดีแต่ด้วยบางครั้งข้อมูลที่มีอยู่มากมายนั้นอาจมีการจัดกระจาย อาจมีข้อมูลเสีย ซึ่งการนำข้อมูลมาวิเคราะห์นั้นคือขั้นตอนของการวิเคราะห์ข้อมูล (data analysis) อีกทั้งการเข้าถึงข้อมูลมหาศาลเหล่านั้นและใช้ข้อมูลเหล่านั้นให้เป็นประโยชน์ถือเป็นสิ่งท้าทายและถือเป็นการแข่งขันในการทำธุรกิจเป็นอย่างยิ่ง

ใหญ่ขนาดไหน?



หากถามว่า ข้อมูลปริมาณแค่ไหนหรือใหญ่มากแค่ไหนถึงจะเข้าข่ายที่จะเรียกว่า big data ได้โดยทั่วไปแล้วจะมองในมุมของจำนวนข้อมูลมากกว่าขนาดของข้อมูล (ขนาดของไฟล์) เช่น ในหนึ่งวันมีลูกค้าเติมน้ำมันในปั้มน้ำมันยี่ห้อ A ทั่วประเทศเฉลี่ยจำนวน 1 ล้านใบเสร็จ ภายใน 30 วันก็จะมีประมาณ 30 ล้านใบเสร็จ ภายใน 1 ปีก็ประมาณ 365 ล้านใบเสร็จ



การสังเคราะห์ข้อมูลที่มีนัยสำคัญจากจำนวนข้อมูลในปริมาณมากขนาดนี้จะเกิดจากการวิเคราะห์ที่เรียกว่า big data analysis

ตัวอย่างหน่วยงานที่มีการวิเคราะห์bigdata



ดังนั้นเราสามารถสรุปได้ว่า หากมีปริมาณข้อมูลมากในหลักหมื่น records ขึ้นไปก็อาจจะมองว่าเป็นการจัดการกับข้อมูลขนาดใหญ่ได้ในบางระบบอาจมีประเด็นของขนาดข้อมูลเข้ามามีบทบาทมากในกรณีที่ข้อมูลแต่ละชิ้นมีขนาดใหญ่



บริษัทที่มีการบริหารจัดการข้อมูลขนาดใหญ่ เช่น

1. Facebook, YouTube, Netflix มีการวิเคราะห์พฤติกรรมของผู้ใช้การยิงโฆษณา พฤติกรรมการดูวิดีโอ เป็นต้น
2. 7-11, ปตท., Major Cineplex มีการวิเคราะห์พฤติกรรมการซื้อสินค้า วิเคราะห์ยอดขายในแต่ละพื้นที่สินค้าขึ้นไหนขายดีในช่วงเวลาใดของปีที่สินค้าขึ้นไหนขายดี
3. มหาวิทยาลัยสามารถวิเคราะห์พฤติกรรมของผู้เรียน นักเรียนที่สมัครเข้ามาเรียน บันทึกที่สำเร็จการศึกษาและภาวการณ์ปฏิบัติงาน
4. เจ้าของโรงงาน สวนผลไม้ฟาร์มกุ้ง สามารถวิเคราะห์ผลผลิต สภาพแวดล้อมที่มีผลต่อผลผลิตและยอดขายในช่วงต่างๆของปีได้เพื่อวางแผนการดำเนินธุรกิจในช่วงเวลานั้นๆของปีถัดๆไป

Bigdata เกี่ยวข้องยังไงกับ IoT?



ระบบ IoT บางระบบมีเครือข่ายเซ็นเซอร์ที่มีการบันทึกข้อมูลเป็นประจำ เช่น สวณผลไม้มีเซ็นเซอร์ทั้งหมด 100 ตัว ส่งข้อมูลวันละ 3 ครั้ง ครั้งละ 5 ค่า รวมใน 1 วันจะมีจำนวนข้อมูลถูกบันทึกไว้ในฐานข้อมูลจำนวน 1,500 records โดยประมาณ ใน 1 สัปดาห์จะมีเท่ากับ 10,500 records ใน 1 เดือนมี 45,000 records และใน 1 ปีจะมีจำนวน 547,500 records ซึ่งถือว่าอยู่ในปริมาณที่มากและค่อนข้างยากที่จะดำเนินการในขั้นตอน data cleansing และกระบวนการต่างๆเพื่อให้ได้ข้อมูลที่ใช้ได้เนื่องจากในระหว่างการเก็บข้อมูลนั้นอาจมีข้อมูลที่ไม่ต้องการหรือช่วงขาดตอน เช่น น้ำท่วม ไฟดับ ฝนตก ซึ่งเหตุการณ์เหล่านี้ถือเป็นปัจจัยที่จะส่งผลต่อประสิทธิภาพของการวิเคราะห์ข้อมูลโดยรวม