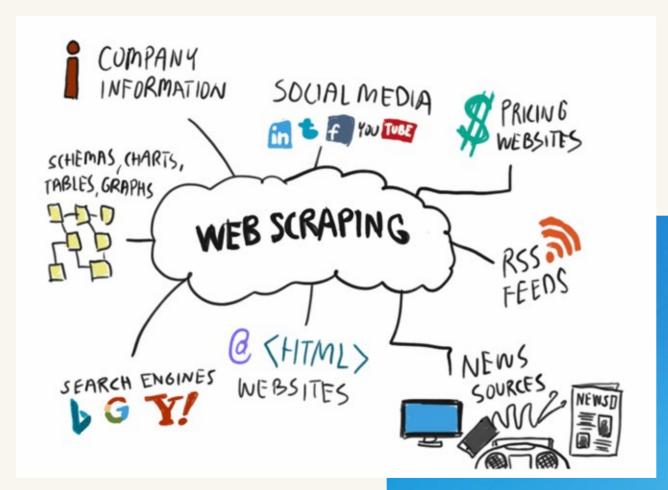


ส่วนที่ 1 ระบบ web scrapping



WEB SCRAPPING

แทคนิคการเข้าถึงข้อมูลจำนวนมหาศาล จากเว็บไซต์ต่างๆที่ต้องการ ไม่ว่าจะเป็น โซเชียล มีเดีย เว็บขายสินค้า เว็บข่าว ฯ ซึ่งเราสามารถเข้าถึงข้อมูลเหล่านั้นโดยใช้ ภาษาโปรแกรมมิ่งเป็นเครื่องมือ (ในที่นี้ใช้ภาษา Java)

นำข้อมูลที่ได้จากการทำ web scrapping มาแสดงบน แอพพลิเคชั่นค้นหาสินค้าลดราคา





ภาษาและเครื่องมือที่ใช้ในการพัฒนาระบบ

Frontend

Backend

Database

Server

- Angular
- Ionic

- Java

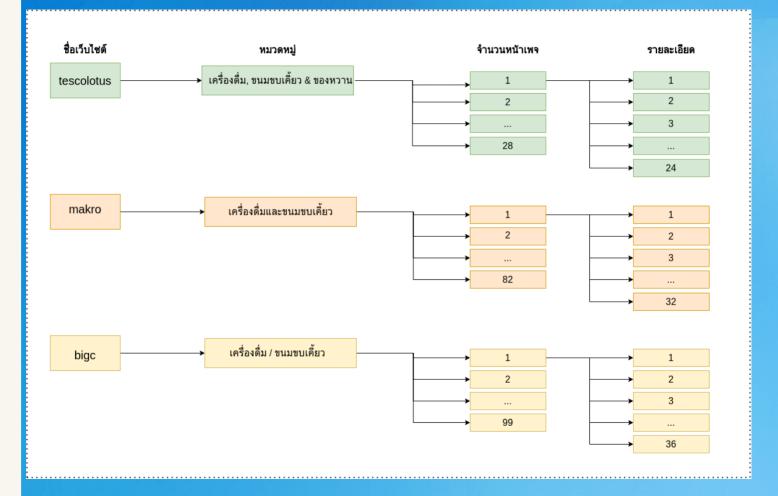
- Mariadb
- Elasticsearch
- Redis

- Ubuntu

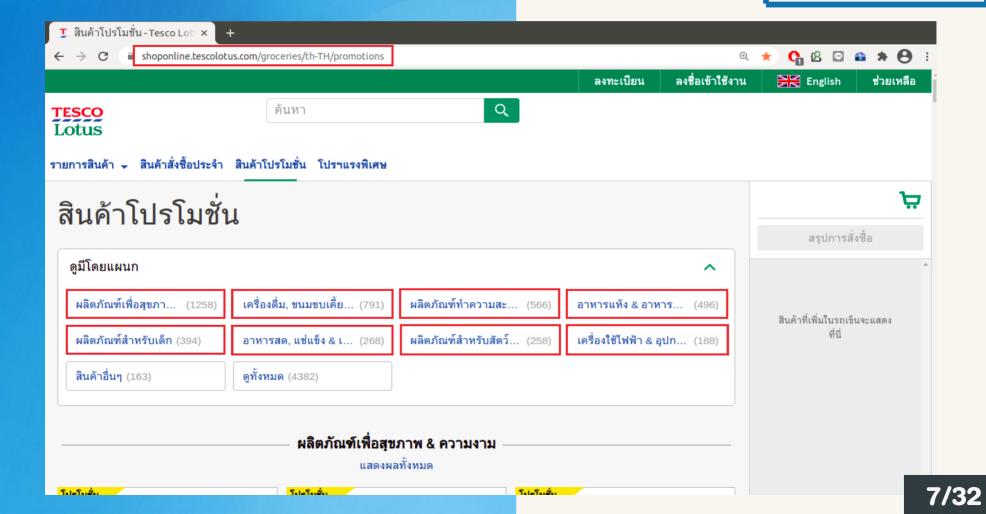
Server 20.04

LTS

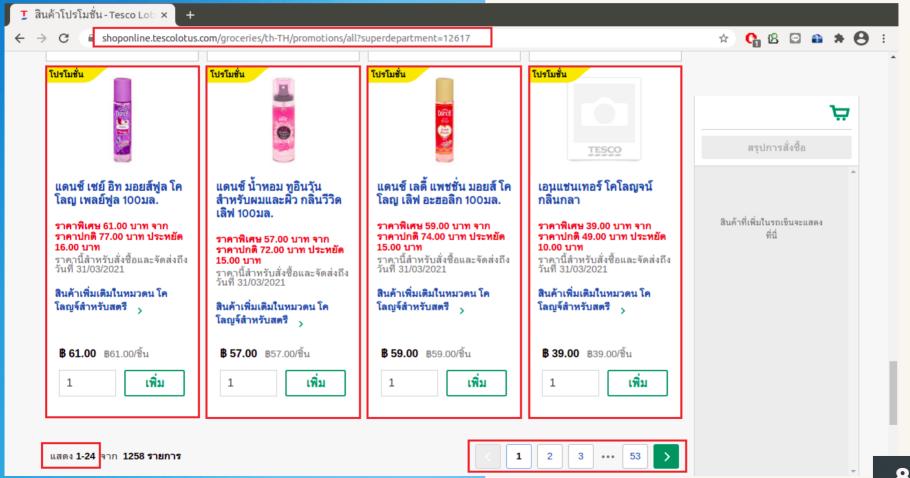
แผนผัง หน้าเว็บไซต์



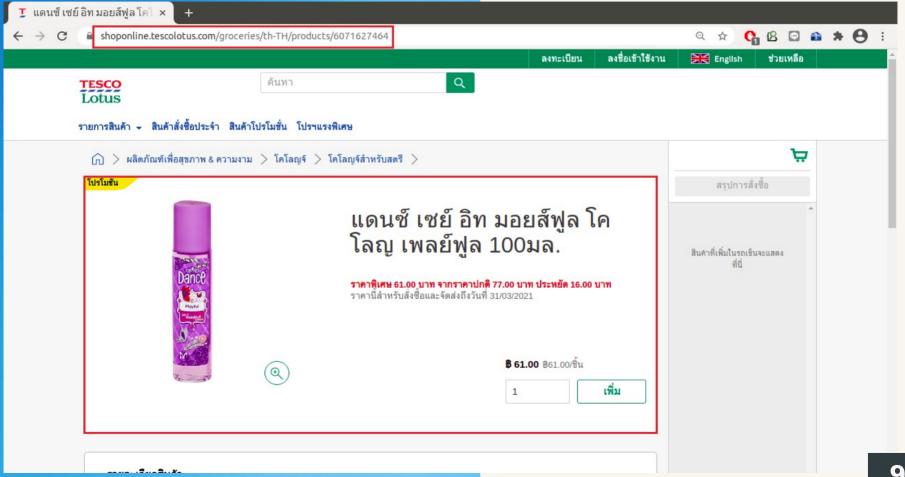
ตัวอย่างหน้าแรกของเว็บไซต์



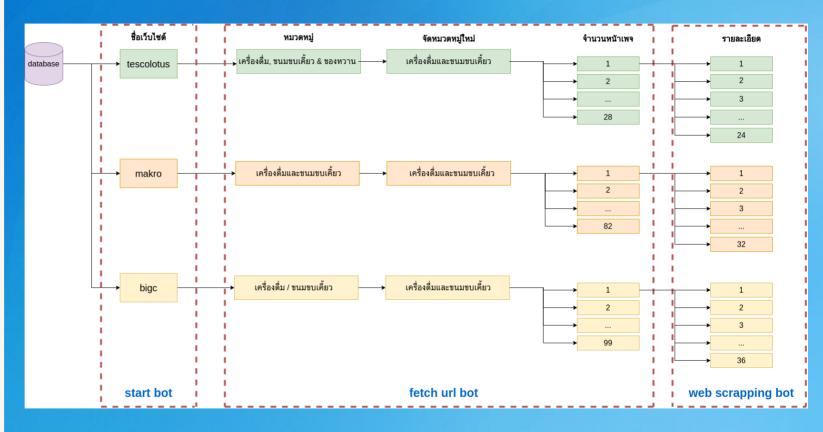
ตัวอย่างหน้าถัดไปเมื่อเลือกหมวดหมู่



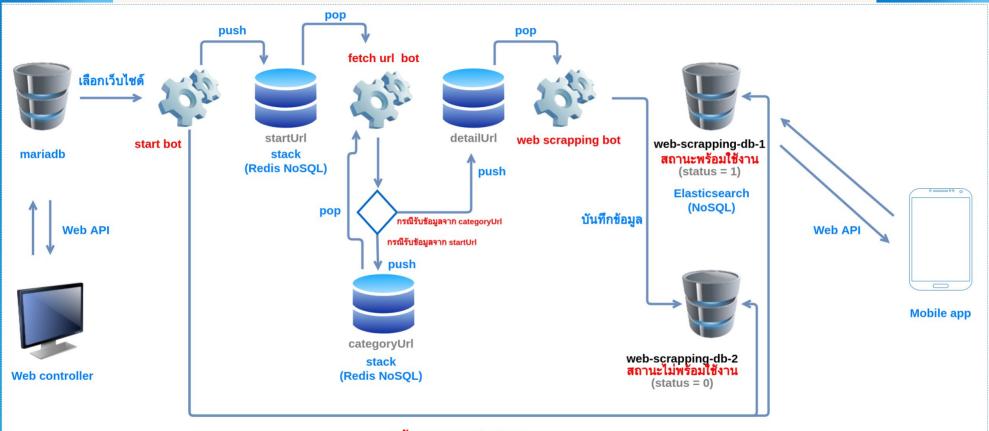
ตัวอย่างหน้าเว็บไซต์ที่ต้องการดึง**ข**้อมูล



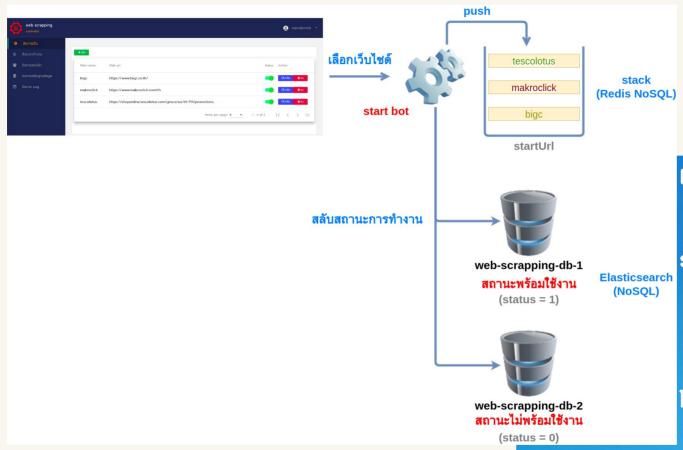
แผนผัง แบ่งการทำงานออก เป็นส่วนๆ



แผนผังการทำงานของ web scrapping



สลับสถานะของ database จากเดิม สถานะพร้อมใช้งาน เปลี่ยนเป็น สถานะไม่พร้อมใช้งาน จากเดิม สถานะไม่พร้อมใช้งาน เปลี่ยนเป็น สถานะพร้อมใช้งาน



start bot

หน้าที่ของ start bot

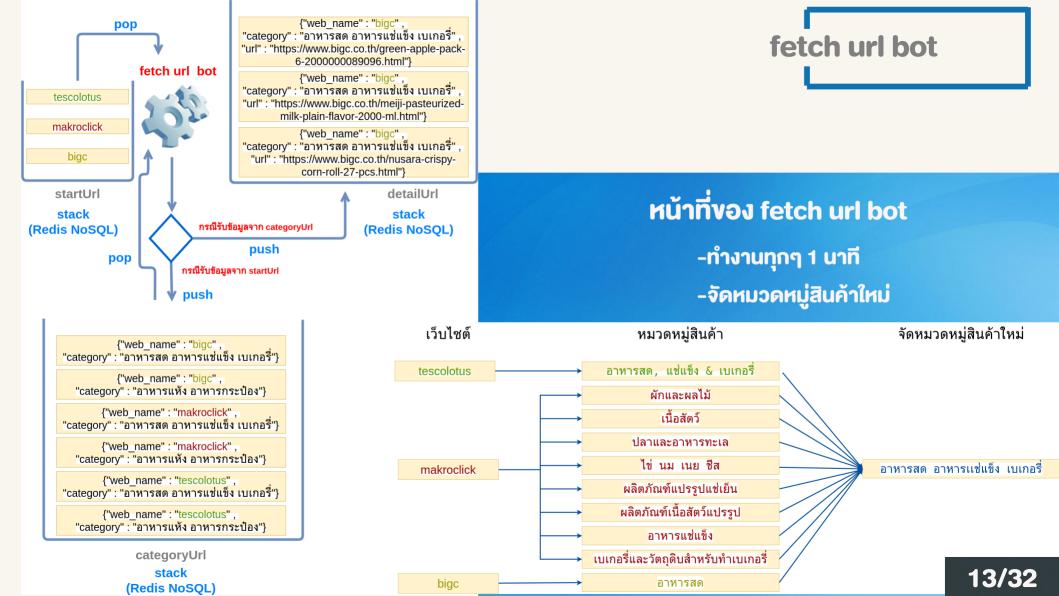
-ทำงานตามเวลาที่ได้ตั้งค่าไว้ใน

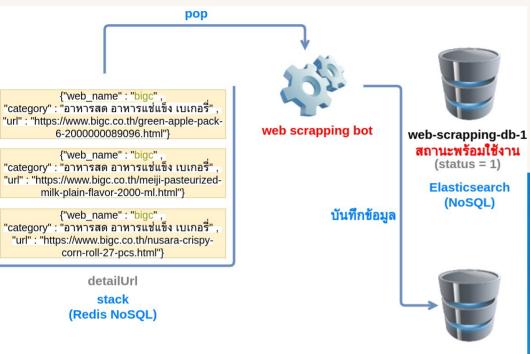
ระบบ

- -เลือกเว็บไซต์ที่ต้องการดึงข้อมูล
- -สลับฐานข้อมูล web-scrpping-db
- -ลบข้อมูลใน web-scrpping-db

โดยเลือกลบข้อมูลที่สถานะไม่พร้อมใช้งาน

-ลบข้อมูลใน stack (Redis NoSQL)





pop

{"web name": "bigc",

6-2000000089096.html"}

{"web_name" : "bigc" .

"category" : "อาหารสด อาหารแช่แข็ง เบเกอรี่"

milk-plain-flavor-2000-ml.html"}

"url": "https://www.bigc.co.th/nusara-crispy-

corn-roll-27-pcs.html"}

detailUrl

stack (Redis NoSOL) web scrapping bot

หน้าที่ของ web scrapping bot

-ทำงานทุกๆ 1 นาที

-เลือกเฉพาะสินค้าที่ลดราคา

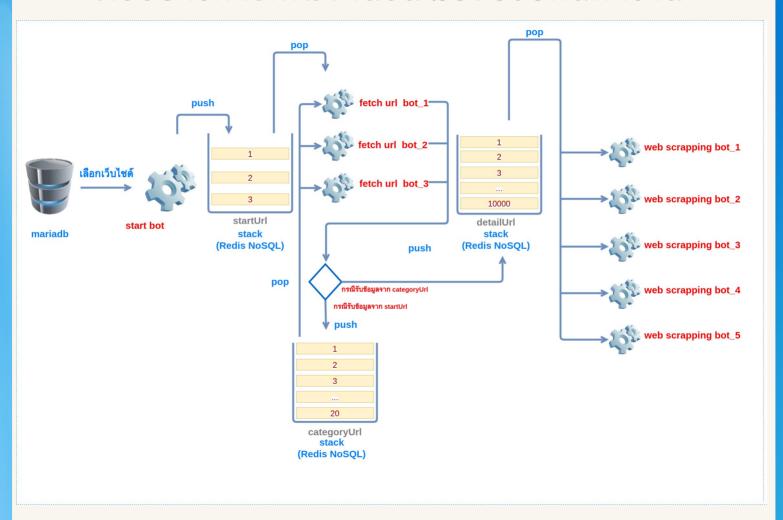
-จัดเก็บข้อมูลลงใน web-scrapping-db

โคยเลือกฐานข้อมูลที่สถานะไม่พร้อมใช้งาน

web-scrapping-db-2 สถานะไม่พร้อมใช้งาน (status = 0)

ลำดับ	วันที่	ช่วงเวลา	เว็บไซต์	สินค้าทั้งหมด	สินค้าลดราคา	สัดส่วนลดราคา	หมายเหตุ
1	08/06/64	06.00 - 09.00	makro	17265	1928	11.17%	
2	08/06/64	06.00 - 09.00	lotus	-	4463		ไม่สามารถทราบจำนวนสินค้าทั้งหมดได้ เนื่องจากดึงข้อมูลที่เมนูสินค้าโปรโมชั่น
3	08/06/64	06.00 - 09.00	bigc	23268	5641	24.24%	
4	08/06/64	10.00 - 13.00	makro	17265	1931	11.18%	
5	08/06/64	10.00 - 13.00	lotus	-	4467		ไม่สามารถทราบจำนวนสินค้าทั้งหมดได้ เนื่องจากดึงข้อมูลที่เมนูสินค้าโปรโมชั่น
6	08/06/64	10.00 - 13.00	bigc	23268	5641	24.24%	
7	08/06/64	14.00 - 17.00	makro	17263	1929	11.17%	
8	08/06/64	14.00 - 17.00	lotus	-	4467		ไม่สามารถทราบจำนวนสินค้าทั้งหมดได้ เนื่องจากดึงข้อมูลที่เมนูสินค้าโปรโมชั่น
9	08/06/64	14.00 - 17.00	bigc	23268	5640	24.24%	
10	08/06/64	18.00 - 21.00	makro	17259	1929	11.18%	
11	08/06/64	18.00 - 21.00	lotus	-	4467		ไม่สามารถทราบจำนวนสินค้าทั้งหมดได้ เนื่องจากดึงข้อมูลที่เมนูสินค้าโปรโมชั่น
12	08/06/64	18.00 - 21.00	bigc	23268	5640	24.24%	

ตัวอย่างการเพิ่มจำนวน bot ช่วยกันทำงาน



<mark>ทดสอบควา</mark>มครบถ้วนของข้อมูลจากเว็บไซต์ <u>เทสโก้ โลตัส</u>

วิธีที่ 1 ใช้ 1 bot ทำงานทั้งหมด

1-bot

99.98%

ลำดับ	วันที่	ช่วงเวลา	ใช้เวลา	start bot	fetch <u>url</u> bot	web scrapping bot	ความเร็วเน็ต(Mbps)	ข้อมูลสินค้าทั้งหมด	ข้อมูลสินค้าที่ได้	เปอร์เซ็นต์
1	03/05/64	02.00.00 - 04.34.57	02.34.57	00.00.02	00.05.34	02.29.21	56.18	4316	4315	99.98
2	03/05/64	08.00.00 - 10.20.05	02.20.05	00.00.02	00.05.00	02.15.03	28.62	4305	4303	99.95
3	03/05/64	11.00.00 - 13.25.04	02.25.04	00.00.02	00.04.42	02.20.20	37.79	4305	4304	99.98
4	03/05/64	14.00.00 - 16.50.11	02.50.11	00.00.02	00.05.17	02.44.52	22.48	4305	4305	100
5	03/05/64	17.00.00 - 19.49.32	02.49.32	00.00.02	00.05.19	02.44.11	14.66	4305	4304	99.98
6	03/05/64	20.00.00 - 23.32.51	03.32.51	00.00.05	00.08.35	03.24.11	6.52	4305	4304	99.98
7	04/05/64	00.00.00 - 02.34.33	02.34.33	00.00.02	00.04.50	02.29.41	17.3	4305	4304	99.98
8	04/05/64	11.00.00 - 13.45.24	02.45.24	00.00.02	00.04.39	02.40.43	17.37	4303	4303	100
9	04/05/64	16.00.00 - 18.43.27	02.43.27	00.00.03	00.05.21	02.38.03	25.04	4303	4302	99.98
10	04/05/64	20.00.00 - 23.38.56	03.38.56	00.00.04	00.07.59	03.30.53	2.38	4303	4303	100
		ค่าเฉลี่ย	02.49.31	00.00.03	00.05.44	02.43.44	22.83			99.98
		สัดส่วน	100%	0.03%	3.38%	96.59%				

วิธีที่ 2 ใช้ 10 bot โดยวิธีกระจายงาน

10-bot

99.95%

ลำดับ	วันที่	ช่วงเวลา	ใช้เวลา	start bot	fetch <u>url</u> bot	web scrapping bot	ความเร็วเน็ต(Mbps)	ข้อมูลสินค้าทั้งหมด	ข้อมูลสินค้าที่ได้	เปอร์เซ็นต์
1	10/05/64	10.40.00 - 11.00.27	00.20.27	00.00.05	00.03.09	00.18.27	36.3	4235	4230	99.88
2	10/05/64	11.10.00 - 11.30.21	00.20.21	00.00.02	00.03.12	00.18.21	30.78	4235	4235	100
3	10/05/64	11.40.00 - 12.00.41	00.20.41	00.00.03	00.02.17	00.18.41	27.9	4235	4235	100
4	10/05/64	12.10.00 - 12.33.03	00.23.03	00.00.02	00.03.38	00.21.03	29.93	4235	4235	100
5	10/05/64	12.40.00 - 13.02.09	00.22.09	00.00.03	00.03.53	00.20.09	18.37	4235	4223	99.71
6	10/05/64	13.10.00 - 13.32.11	00.22.11	00.00.05	00.03.31	00.20.11	22.06	4235	4235	100
7	10/05/64	13.40.00 - 14.03.49	00.23.49	00.00.02	00.02.44	00.21.49	8.25	4235	4234	99.98
8	10/05/64	14.10.00 - 14.32.43	00.22.43	00.00.02	00.03.43	00.20.43	18.26	4235	4235	100
9	10/05/64	14.40.00 - 15.03.43	00.23.43	00.00.02	00.04.52	00.21.43	17.61	4235	4232	99.93
10	10/05/64	15.10.00 - 15.34.31	00.24.31	00.00.03	00.03.53	00.22.31	23.21	4235	4235	100
		ค่าเฉลี่ย	00.22.22	00.00.03	00.03.29	00.20.22	23.27			99.95
		สัดส่วน	pipeline	0.22%	15.57%	91.06%				

วิธีที่ 1 ใช้ 1 bot ทำงานทั้งหมด

1-bot

วิธีที่ 2 ใช้ 10 bot โดยวิธีกระจายงาน

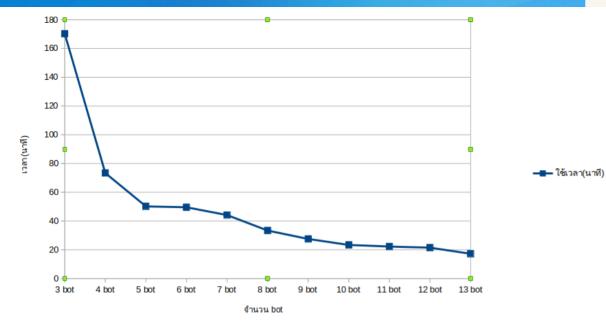
10-bot

ใช้ 10 bot ทำงานได้เร็วกว่า 1 bot ประมาณ 7.7 เท่า

ทดสอบดึงข้อมูลจากทั้ง 3 เว็บไซต์

ลำดับ	วันที่	ช่วงเวลา	ใช้เวลา	start bot	fetch <u>url</u> bot	web scrapping bot	ความเร็วเน็ต(Mbps)	ข้อมูลสินค้าที่ได้
1	07/05/64	08.00.00 - 11.06.33	03.06.33	00.00.02	00.39.54	02.26.37	21.02	10628
2	07/05/64	12.00.00 - 15.10.25	03.10.25	00.00.02	00.29.46	02.40.37	23.84	12598
3	07/05/64	16.00.00 - 19.19.57	03.19.57	00.00.02	00.34.32	02.45.23	7.46	11138
4	07/05/64	20.00.00 - 00.23.44	04.23.44	00.00.06	00.35.55	03.47.43	9.52	12596
5	08/05/64	01.00.00 - 04.36.46	03.36.46	00.00.04	00.50.33	02.46.09	23.95	12611
6	08/05/64	10.00.00 - 13.35.42	03.35.42	00.00.03	00.54.56	02.40.43	14.01	12607
7	08/05/64	16.10.00 - 19.59.28	03.49.28	00.00.03	00.46.53	03.02.32	11.24	10467
8	08/05/64	20.10.00 - 00.14.18	04.04.18	00.00.06	00.30.44	03.33.28	2.29	12608
9	09/05/64	06.10.00 - 09.21.42	03.21.42	00.00.02	00.46.16	02.35.24	29.89	12596
10	09/05/64	10.10.00 - 13.10.43	03.00.43	00.00.02	00.22.19	02.38.22	28.65	12605
		ค่าเฉลี่ย	03.32.56	00.00.03	00.39.11	02.53.42	17.19	
		สัดส่วน	100%	81.58%	18.40%	0.02%		

ลำดับ	วันที่	ช่วงเวลา	ใช้เวลา	start bot	fetch <u>url</u> bot	web scrapping bot	ความเร็วเน็ต(Mbps)	ข้อมูลสินค้าที่ได้
1	10/05/64	19.00.00 - 19.30.44	00.30.44	00.00.02	00.26.24	00.28.44	12.34	12387
2	10/05/64	20.00.00 - 20.47.11	00.47.11	00.00.06	00.21.27	00.45.11	1.55	12391
3	11/05/64	00.00.00 - 00.24.42	00.24.42	00.00.02	00.11.20	00.22.42	20.87	12374
4	11/05/64	00.40.00 - 01.05.15	00.25.15	00.00.02	00.09.44	00.23.15	44.65	12367
5	11/05/64	01.20.00 - 01.44.22	00.24.22	00.00.02	00.23.03	00.22.22	36.31	12392
6	11/05/64	02.10.00 - 02.34.33	00.24.33	00.00.02	00.10.32	00.22.33	36.98	12395
7	11/05/64	02.40.00 - 03.02.42	00.22.42	00.00.02	00.09.45	00.20.42	52.49	12396
8	11/05/64	03.10.00 - 03.39.00	00.29.00	00.00.02	00.27.22	00.27.00	48.85	12417
9	11/05/64	08.30.00 - 08.56.01	00.26.01	00.00.02	00.24.25	00.24.01	33.92	12418
10	11/05/64	09.10.00 - 09.31.59	00.21.59	00.00.02	00.10.54	00.19.59	29.88	12417
		ค่าเฉลี่ย	00.27.39	00.00.02	00.17.30	00.25.39	31.78	
		สัดส่วน	pipeline	0.12%	63.29%	92.77%		17/32



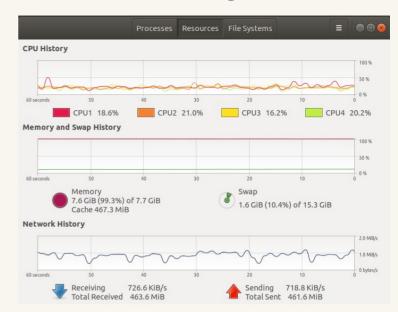
ทดสอบการทำงานของ bot โดย วิธีกระจายงานเป็นส่วนๆ

ลำดับ	จำนวน bot	วันที่	ช่วงเวลา	ใช้เวลา(นาที)	ใช้แรม(gb)	ความเร็วเน็ต(Mbps)	ข้อมูลที่ได้(รายการ)	รายละเอียดของ bot
1	3 bot	11/03/64	13.00.00 - 15.50.24	170.24	2.53	29.46	12156	start bot = 1 , fetch url bot = 1 , web scrapping bot = 1
2	4 bot	11/03/64	16.00.00 - 17.13.38	73.38	2.91	20.61	11792	start bot = 1 , fetch url bot = 1 , web scrapping bot = 2
3	5 bot	11/03/64	17.30.00 - 18.20.16	50.16	3.52	8.63	12424	start bot = 1 , fetch url bot = 1 , web scrapping bot = 3
4	6 bot	11/03/64	18.40.00 - 19.29.55	49.55	4.45	8.88	12426	start bot = 1 , fetch url bot = 1 , web scrapping bot = 4
5	7 bot	12/03/64	00.00.00 - 00.44.14	44.14	4.81	17.98	12370	start bot = 1 , fetch url bot = 2 , web scrapping bot = 4
6	8 bot	12/03/64	01.00.00 - 01.33.36	33.36	5.17	25.81	12366	start bot = 1 , fetch url bot = 2 , web scrapping bot = 5
7	9 bot	12/03/64	01.50.00 - 02.17.51	27.51	5.63	28.82	12373	start bot = 1 , fetch url bot = 2 , web scrapping bot = 6
8	10 bot	12/03/64	02.30.00 - 02.53.34	23.34	6.19	44.46	12372	start bot = 1 , fetch url bot = 2 , web scrapping bot = 7
9	11 bot	12/03/64	03.00.00 - 03.22.24	22.24	7.55	51.71	12406	start bot = 1 , fetch url bot = 3 , web scrapping bot = 7
10	12 bot	12/03/64	03.40.00 - 04.01.47	21.47	7.67	37.53	12402	start bot = 1 , fetch url bot = 3 , web scrapping bot = 8
11	13 bot	12/03/64	04.20.00 - 04.37.21	17.21	8.45	48.56	12405	start bot = 1 , fetch url bot = 3 , web scrapping bot = 9

เครื่องที่ทำการทดสอบ

- -Intel core i5-2430M 2.40 GHz
- -Ram 8 gb
- -SSD 500 gb

ทคสอบใช้ 13 bot คึงข้อมูลจากทั้ง 3 เว็บไซต์



ลำดับ	วันที่	ค่อมออก	ใช้เวลา	Memory(Gb)			CPU		ความเร็วเน็ต(Mbps)	ข้อมูลสินค้าทั้งหมด	
ถาทบ	านท	ช่วงเวลา	เอเงสา	Ram	Swap	cpu1	cpu2	cpu3	cpu4	и листии(морь)	ขอมูลเลเนท เทงทมท
1	25/05/64	10.00.00 - 10.18.02	0.18.02	7.26	1.57	18.20%	21.00%	16.20%	20.20%	26.1	12760
2	25/05/64	10.30.00 - 10.41.33	0.11.33	7.17	1.56	20.80%	14.10%	19.80%	16.50%	29.01	12758
3	25/05/64	11.00.00 - 11.11.31	0.11.31	7.26	1.4	17.60%	22.70%	21.80%	14.00%	23.41	12760
4	25/05/64	11.30.00 - 11.42.06	0.12.06	7.25	2.36	27.30%	20.20%	9.50%	17.80%	34.02	12761
5	25/05/64	12.00.00 - 12.12.28	0.12.28	7.25	1.37	19.20%	17.80%	23.20%	16.30%	35.58	12761
	·	ค่าเฉลี่ย	0.13.08	7.24	1.65	20.62%	19.16%	18.10%	16.96%	29.62	



	Test-load.jmx (/home/m	nandarin/apache-jmete	er-5.3/bin/Test-load.jmx)	- Apache JMeter (5.3)	
Eile Edit Search Bun Option					
📑 🍪 🚔 🖫 📈	📵 🟮 + - 🤣	 > & ® •	🖏 👹 🙉 🦫	00:00:00	<u>∧</u> 0 0/0 ↔ 🤽
A Test Plan Start Plan Stummary Report Constant Timer HTTP Request Defa	Thread Group Name: 270 Comments: Action to be taken after a Si © continue © Start Net Thread Properties Number of Threads (users), Ramp-up period (seconds). Loop Count: Infinite Ø same user on each ter Delay Thread creation Second Start Net Startup delay (seconds).		7 Thread Stop Test	Stop Test Now	

Load testing

ลำดับ	วันที่	จำนวน user ต่อ 1 วินาที	ใช้ cpu server(%)	error (%)	error message
1	08/06/64	20	20.03	0.00	
2	08/06/64	40	28.57	0.00	
3	08/06/64	60	39.18	0.00	
4	08/06/64	80	40.82	0.00	
5	08/06/64	100	47.24	0.00	
6	08/06/64	120	50.51	0.00	
7	08/06/64	140	53.81	0.00	
8	08/06/64	160	57.22	0.00	
9	08/06/64	180	64.4	0.00	
10	08/06/64	200	72.28	0.00	
11	08/06/64	270	77.16	0.00	
12	08/06/64	271	79.8	7.01	Connection timed out

จากผลการทดลอง Load testing จากโปรแกรม Apache Jmeter ผลลัพธ์ที่ได้ ระบบ สามารถรองรับการใช้งานจาก user ได้จำนวน 270 user ต่อ วินาที

ลำดับ	จำนวน user ต่อ 1 วินาที	ครั้งที่ 1 error (%)	ครั้งที่ 2 error (%)	ครั้งที่ 3 error (%)	ค่าเฉลี่ย error (%)
1	300	1.00	2.67	5.00	2.89
2	400	28.75	28.50	26.50	27.92
3	500	47.00	38.00	44.20	43.07
4	540	43.70	44.44	48.89	45.68
5	600	49.17	50.83	52.67	50.89
6	700	55.00	56.43	58.71	56.71
7	800	59.25	60.25	62.88	60.79
8	810	60.62	64.69	62.10	62.47
9	900	67.67	65.56	67.89	67.04
10	1000	70.70	70.40	70.30	70.47

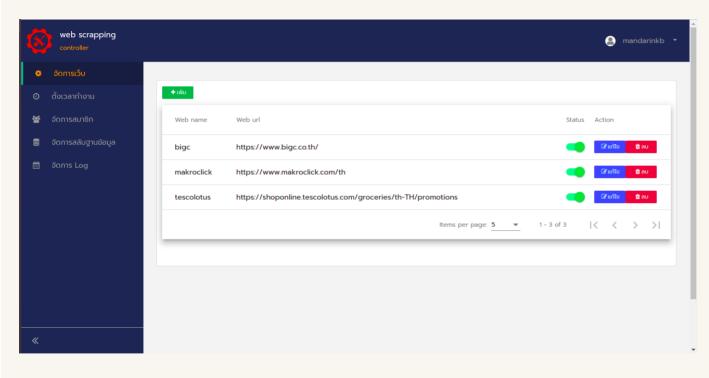
ลำดับ	จำนวน server	รองรับการใช้งาน (จำนวน user)	จำนวนเงินค่าserver (บาท)/เดือน
1	1	270	642.47
2	2	540	1284.94
3	3	810	1927.41
4	4	1080	2569.88

Load testing 1000 user

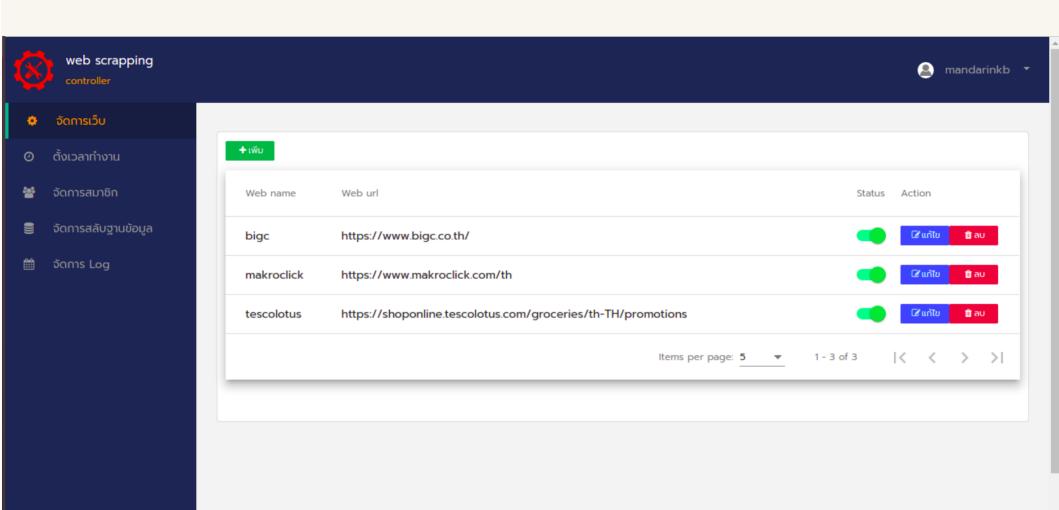
ปัจจุบันใช้งาน server 1
server ในอนาคตกรณีมีผู้ใช้งาน
1000 user ต่อ วินาที ต้องขยาย
server เพื่อรองรับการใช้งาน
เป็นจำนวน 4 server คิดเป็นเงิน
2569.88 บาท/เดือน
ต้นทุนต่อผู้ใช้งานคิดเป็น
2.4 บาท / คน / เดือน

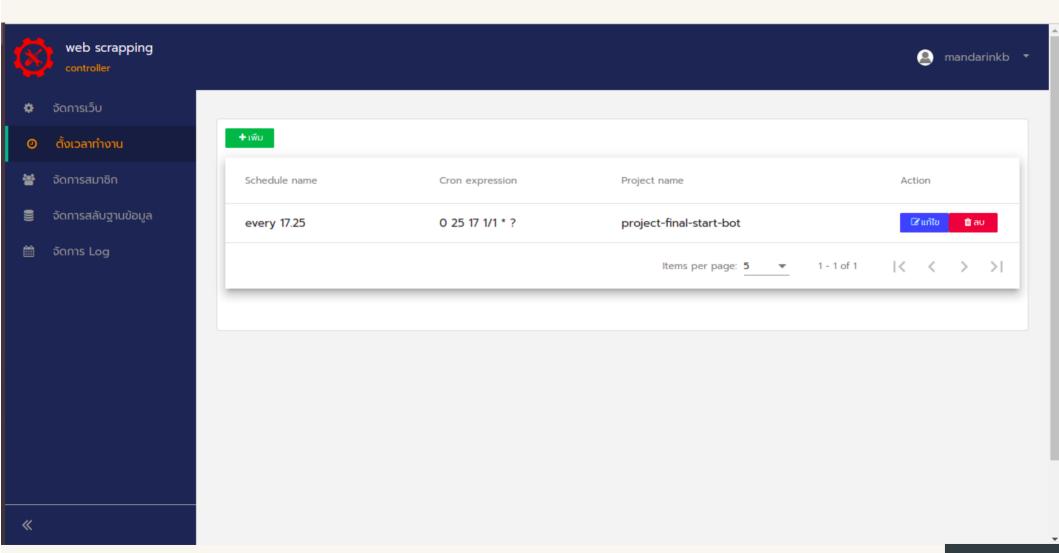
ส่วนที่ 2 แอพพลิเคชั่น

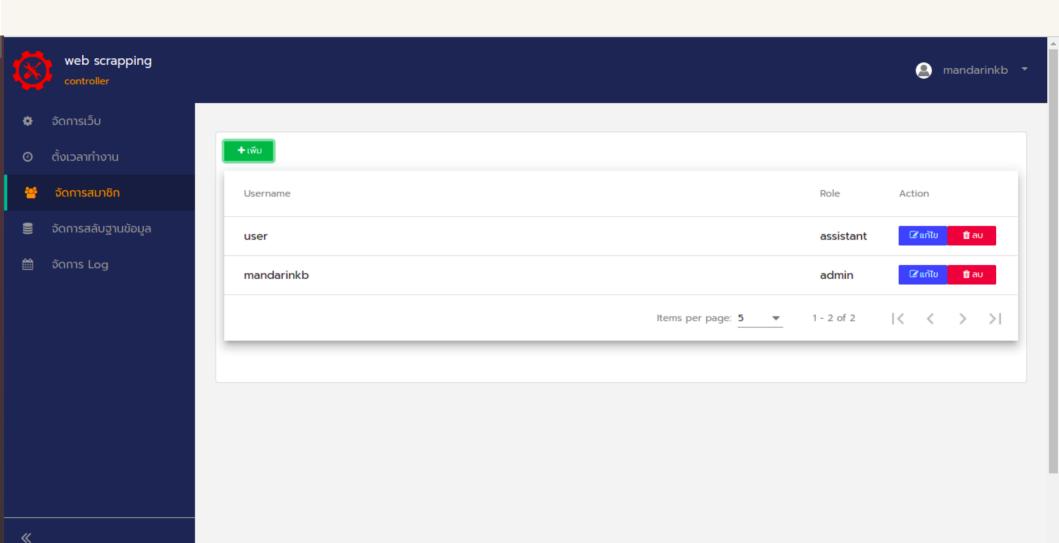
ตัวอย่าง หน้าเว็บและแอพพลิเคชั่น

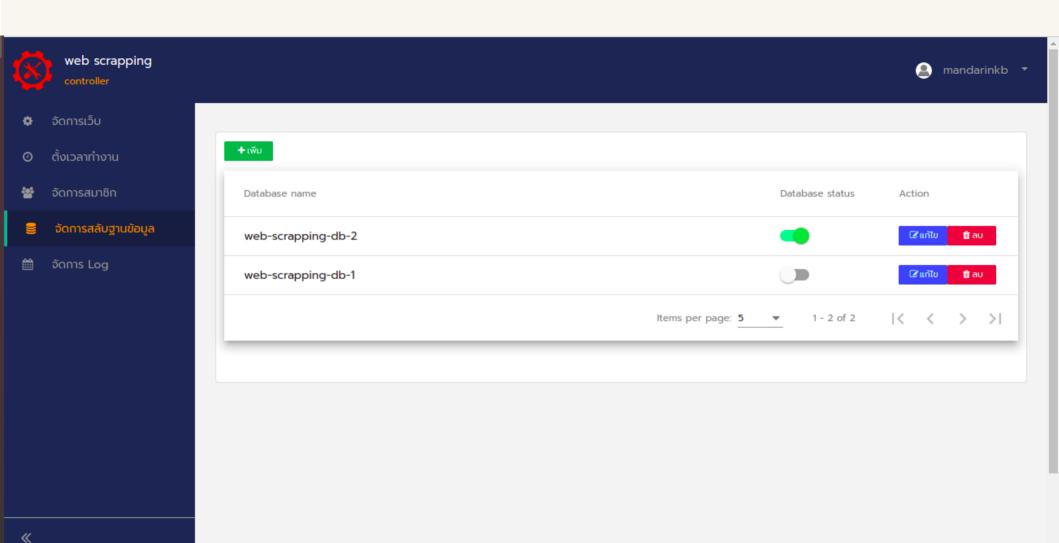


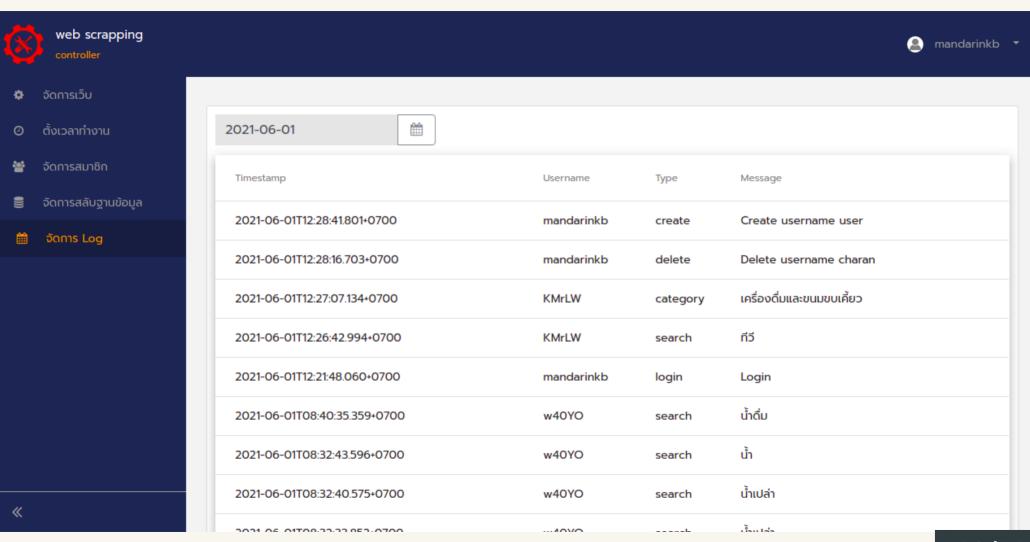








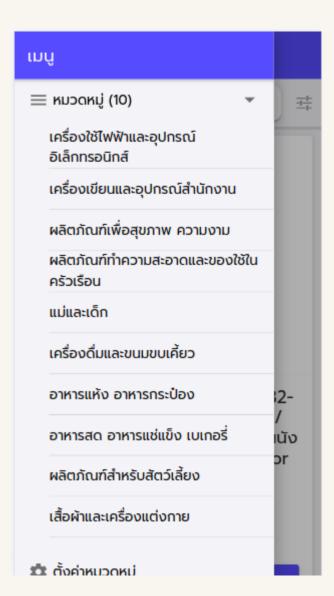


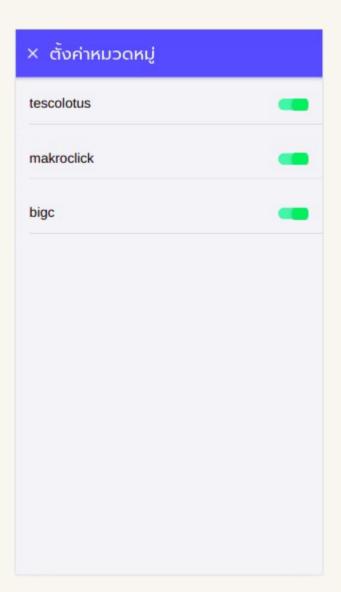














ลำดับ	รายการแบบประเมิน	ค่าเฉลี่ยมุมมองผู้ใช้	ค่าเฉลี่ยมุมมองนักพัฒนาระบบ
1	ความสวยงาม ทันสมัย ความน่าสนใจของแอพพลิเคชั่น	4.2	3.92
2	การจัดรูปแบบแอพพลิเคชั่นง่ายต่อการอ่านและการใช้งาน	4.47	4.47
3	ขนาดตัวอักษรและรูปแบบตัวอักษร อ่านง่ายและสวยงาม	4.2	4.47
4	การออกแบบให้ใช้งานง่ายเมนูไม่ชับซ้อน	4.2	4.47
5	แอพพลิเคชั่นทำงานได้ถูกต้องมีประสิทธิภาพ	4.2	3.67
6	การออกแบบแอพพลิเคชั่นสอดคล้องกับความต้องการของผู้ใช้	4.47	4.73
7	รูปแบบการใช้งานแอพพลิเคชั่น ความยาก - ง่าย	4.2	4.73
8	ความรวดเร็วในการใช้งานแอพพลิเคชั่น	3.67	3.4
9	ความพึงพอใจในการใช้งาน	4.2	3.92
10	ความสามารถของแอพพลิเคชั่น ในการนำไปใช้ประโยชน์	4.2	4.2
	ค่าเฉลี่ยทั้งหมด	4.201	4.198

ทดสอบจากผู้ใช้งาน 10 คน -ผู้ใช้งานทั่วไป 5 คน -นักพัฒนาระบบ 5 คน

ได้ทดลองให้ผู้ใช้และนักพัฒนาทดลองใช้แล้วให้ความคิดเห็น สรุป ได้ว่า ผู้ใช้มีความพึงพอใจอยู่ในระดับมาก (ค่าเฉลี่ย 4.201) ส่วนผู้ พัฒนามีความพึงพอใจในมุมมองการออกแบบและฟังก์ชันการใช้ งานอยู่ในระดับมาก (ค่าเฉลี่ย 4.198)

31/32

