

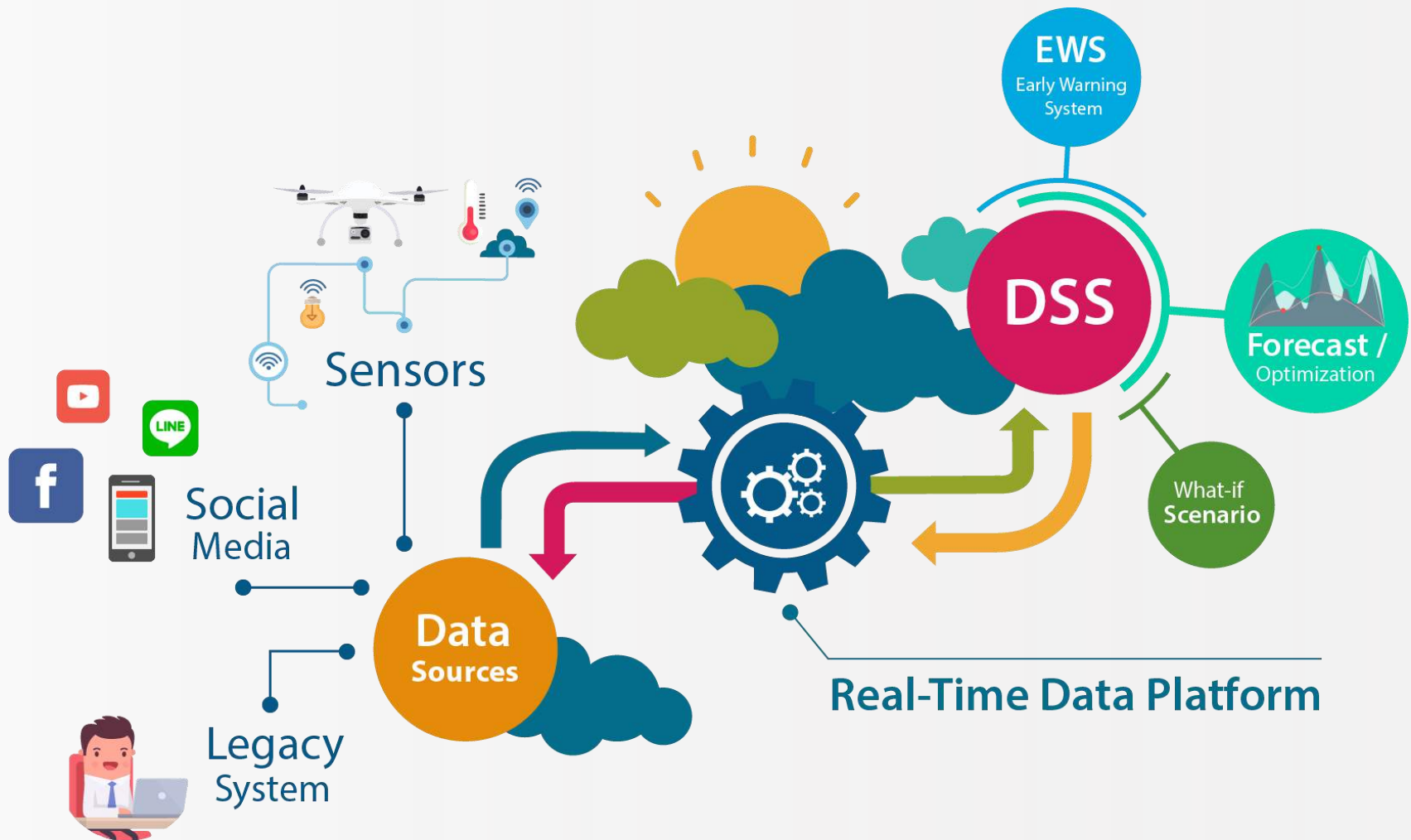
BigStream

แพลตฟอร์มสำหรับการบริหารจัดการข้อมูลแบบเรียลไทม์



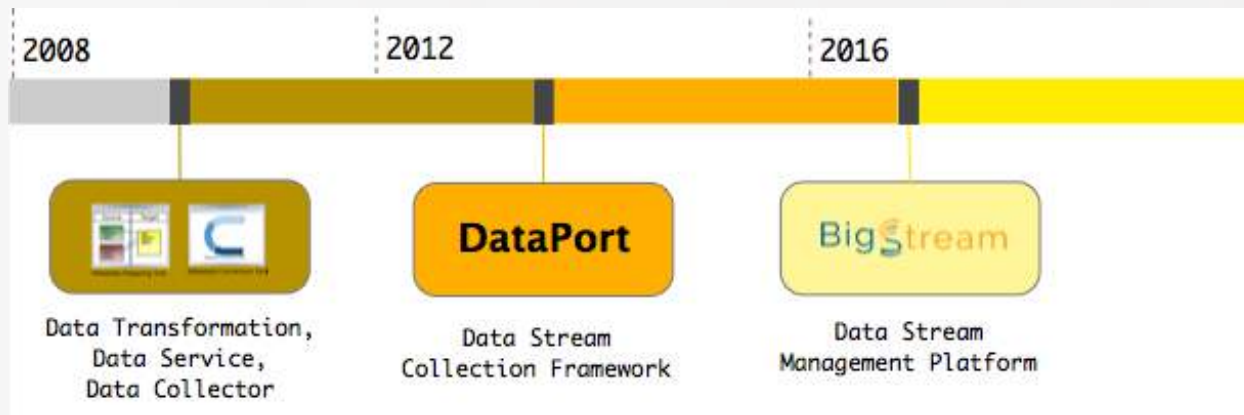
คำรณ อรุณเรื้อ
ผู้ช่วยวิจัย

ห้องปฏิบัติการวิจัยการจำลองขนาดใหญ่
หน่วยวิจัยการวิเคราะห์ข้อมูลและการคำนวณ
ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ

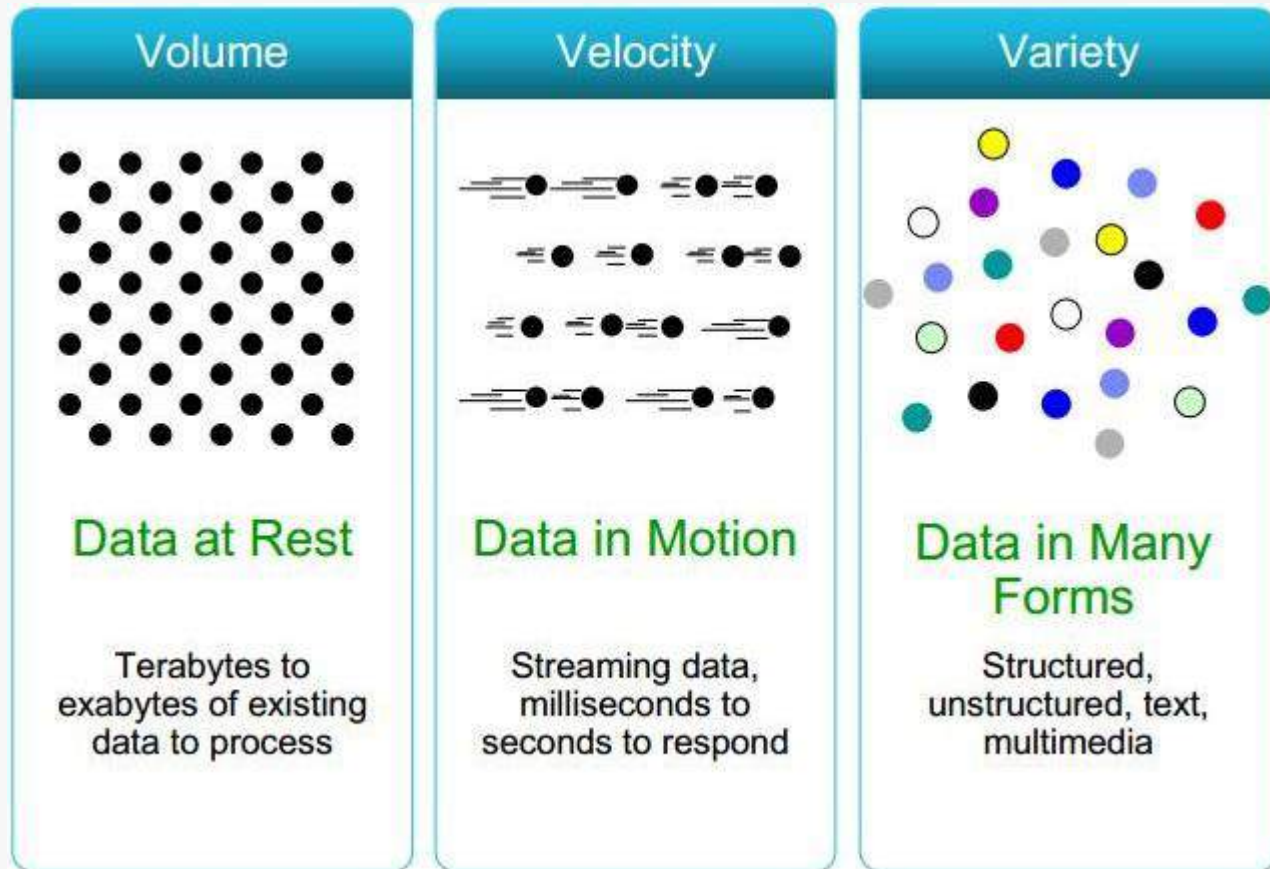


History of Data-Driven Platform

- Free style!
- Metadata Conversion Tool & Data Service
- DataPort
- BigStream



Characterization of Big Data



What is BigStream?

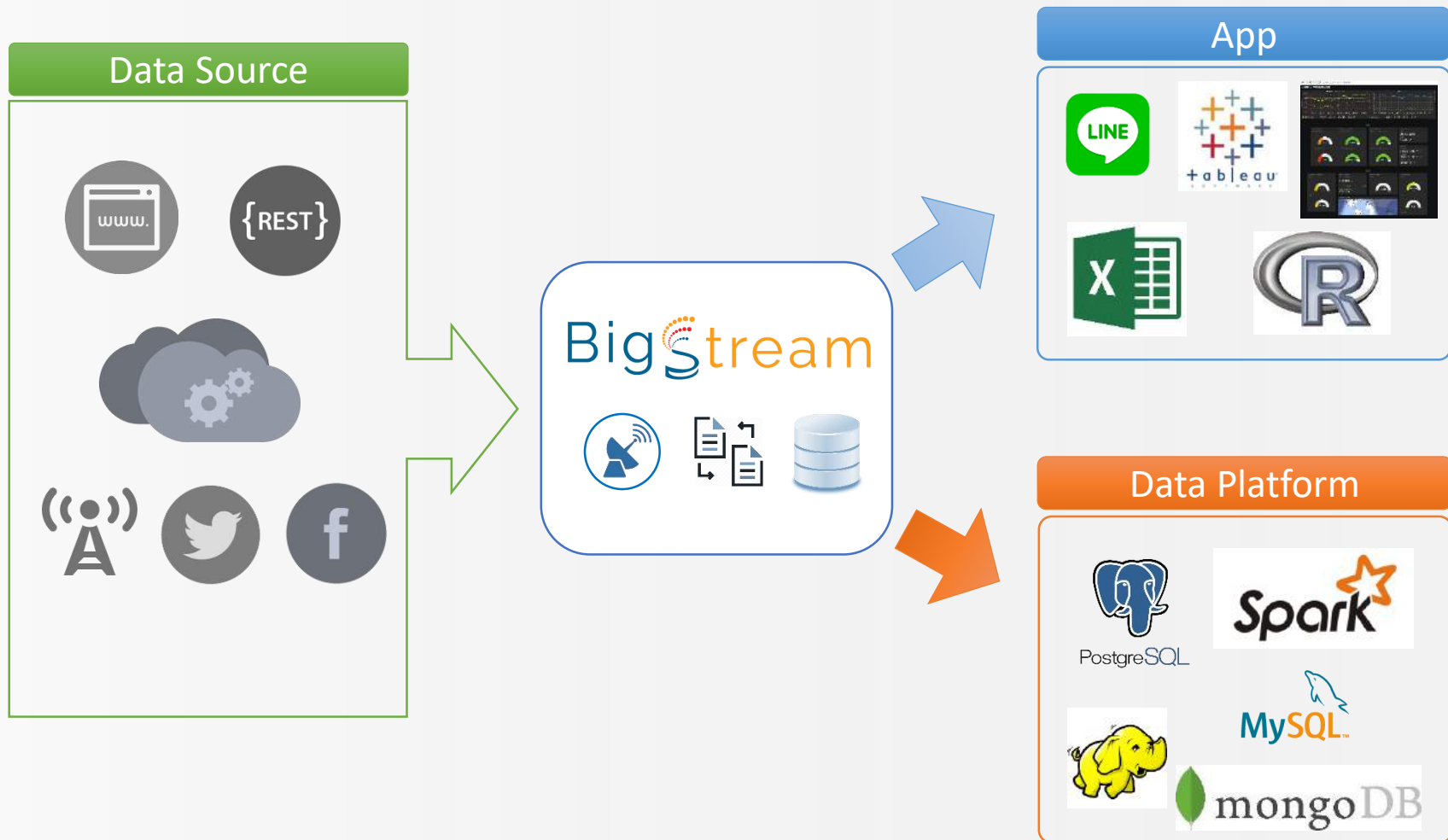
BigStream



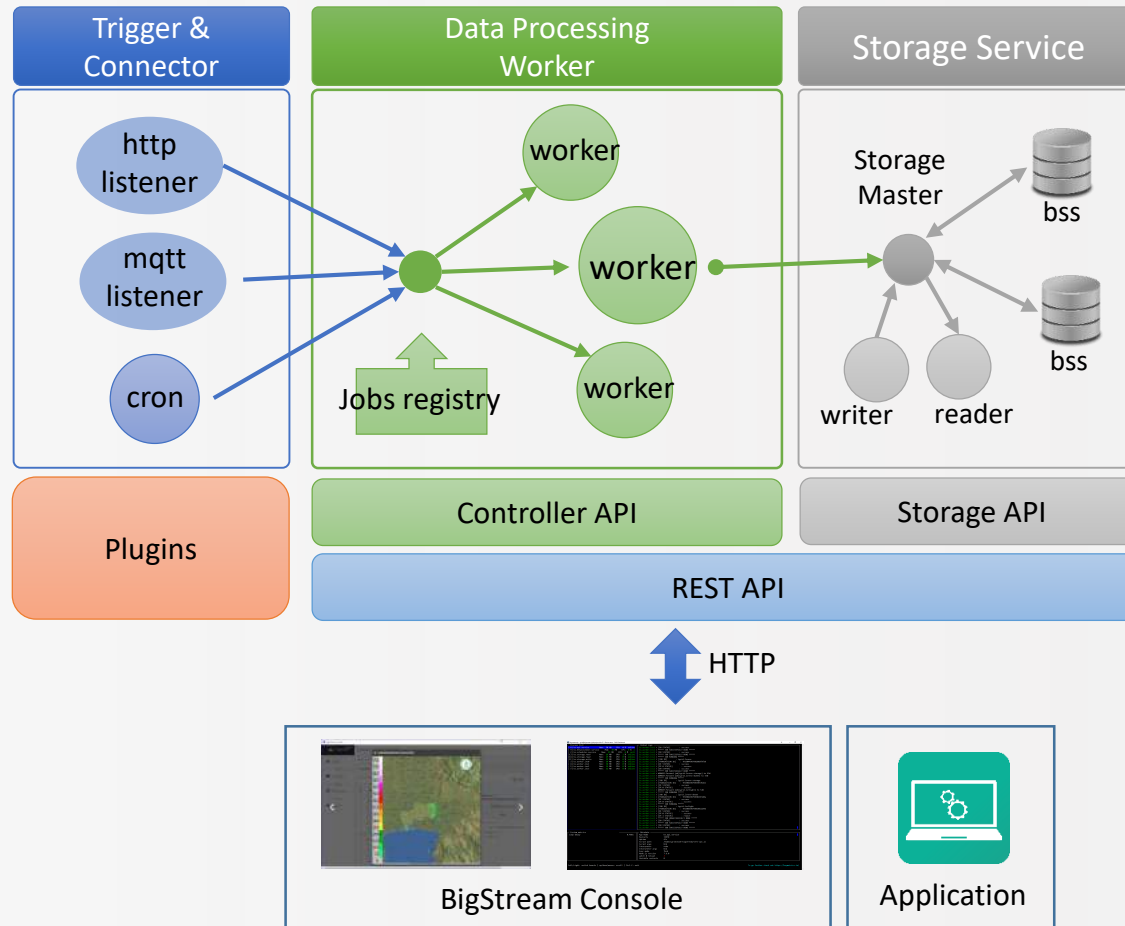
Real-time
Data
Management
Platform

- Data Stream Collector
- Various Data Types Support
- Distributed Data Processing
- Built-in Data Storage Service
- Extendibility & Scalability
- Multi-purpose Data App

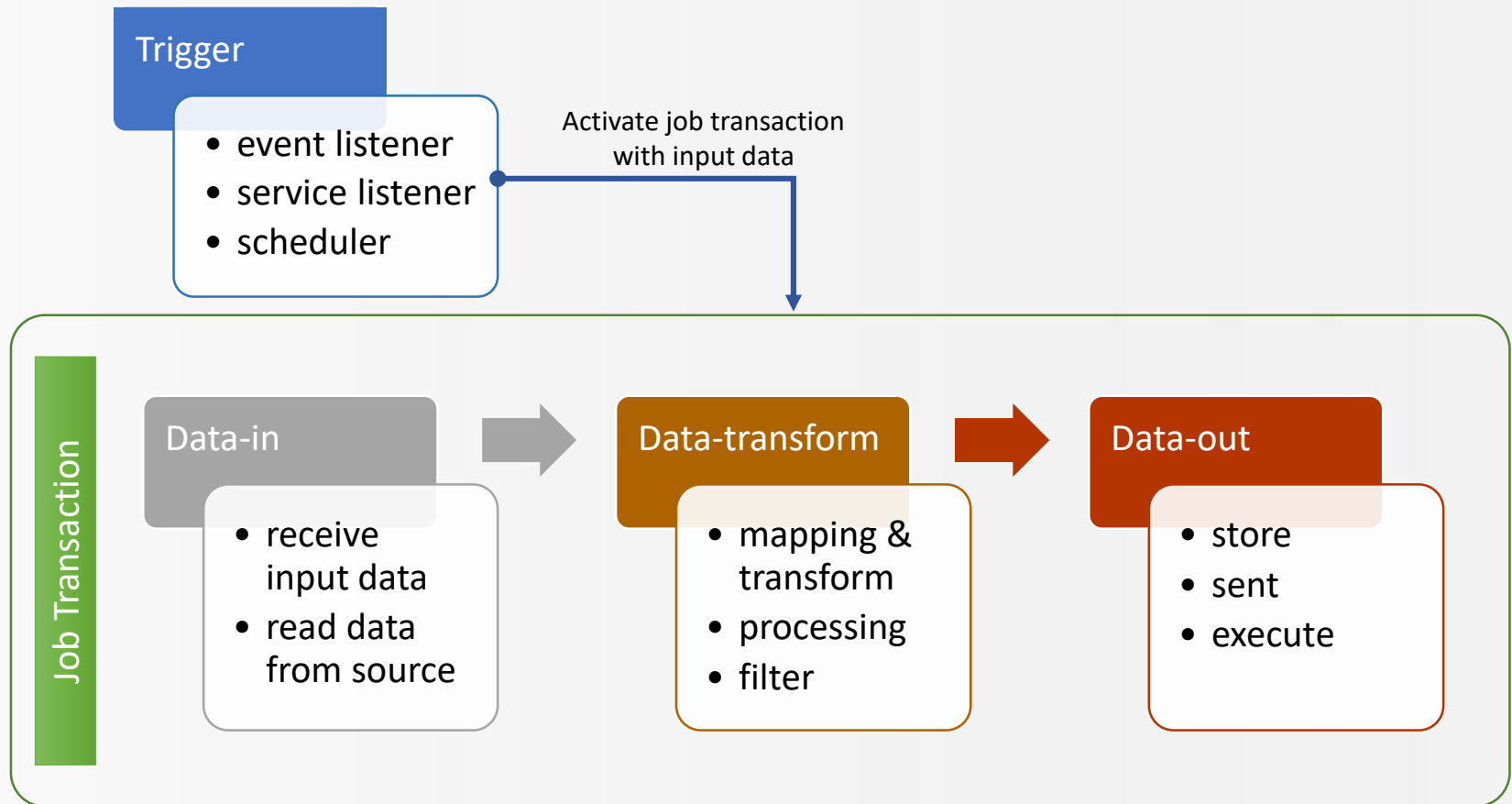
Where BigStream fits?



BigStream Architecture



BigStream Job Framework



Plugins (job modules)

Trigger	Data-in	Data-transform	Data-out
• http-listener • cron • storage-event • netpie	• input • filereader • http-request • ftp • agritronics	• map • reduce • transform • filter-condition • filter-duplicate • windows-event • windows-time • noop	• bss • dir • storage • pgsql • mysql • hdfs • dweet • line-notify • netpie • call

Job configuration

```
{  
  ☐ job_id      :    unique id  
  ☐ active      :    true | false  
  ☐ job_timeout :    millisecond  
  ☐ trigger  
    ☐ type      :    trigger type  
    ☐ cmd       :    trigger command  
  ☐ data_in  
    ☐ type      :    plugin name  
    ☐ param     :    plugin parameter  
  ☐ data_transform  
    [  
      ☐ type      :    plugin name  
      ☐ param     :    plugin parameter  
    ]  
  ☐ data_out  
    ☐ type      :    plugin name  
    ☐ param     :    plugin parameter  
}
```

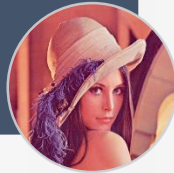
BigStream Storage

- Multi-purpose Stream Storage
- NoSQL Database (batch, real-time)
- Feature
 - large-scale data store (billions records per storage)
 - fast read/write, fast access
 - light weight & portable
 - various data types

Supported Data Types

- Image
- Wave Data
- Video Stream

Binary



- Web data
- Social network
- Encrypted text

Text



- Structured Data Set
- Tabular Data
- JSON

Object



BigStream

3rd Party Software & Tool

BigStream Console

The screenshot displays the BigStream Console web application. At the top, three summary cards show: 824 storages (blue), 868 Active jobs (green), and 868 Inactive jobs (red). Below these is a 'Console message' section for 'Mac Miller'. The left sidebar contains navigation links: Home, JOBS (with a 'Job editor' link), STORAGE (with a 'Storage lists' link), and SETTING (with a 'Setting' link). The main content area shows a table of messages for 'agritronics-MMSIIIV116-005-1'.

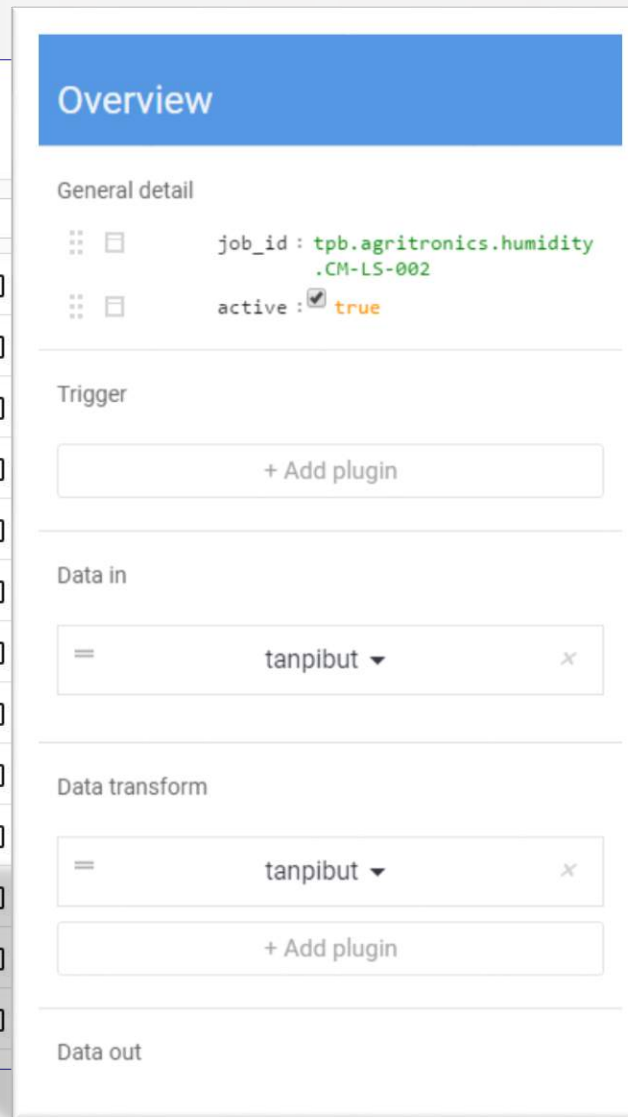
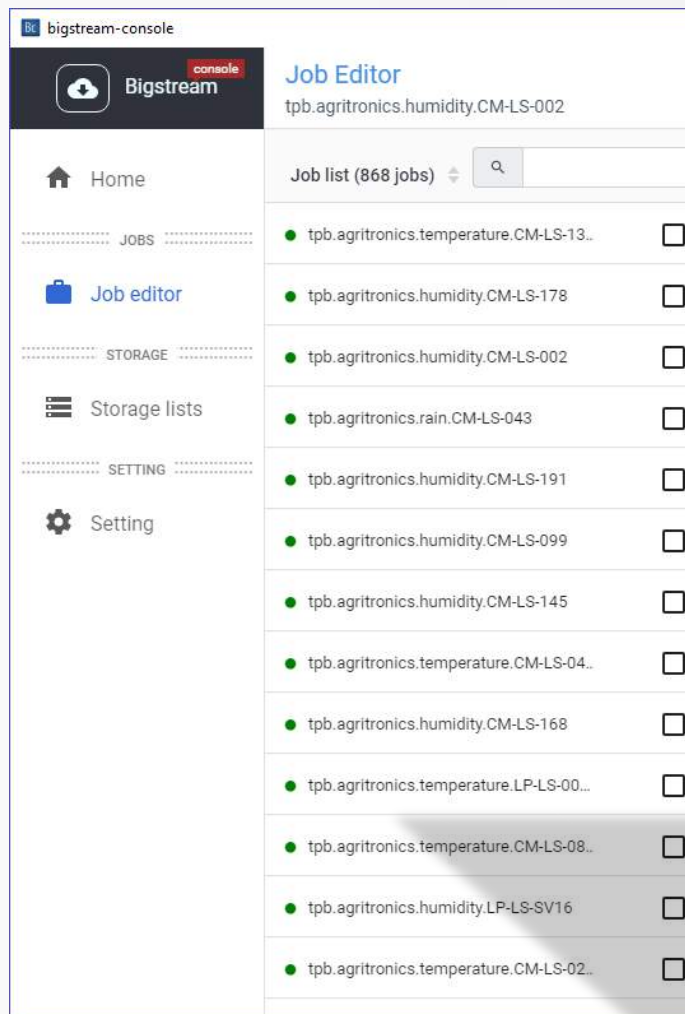
Last at	Messages
01-04-2017, 00:00:00	11 messages
01-04-2017, 00:00:00	23 messages
01-04-2017, 00:00:00	25 messages
01-04-2017, 00:00:00	25 messages
01-04-2017, 00:00:00	25 messages
01-04-2017, 00:00:00	25 messages

Job Editor

The screenshot displays the 'bigstream-console' interface. On the left is a sidebar with navigation links: Home, JOBS, Job editor (selected), STORAGE, Storage lists, SETTING, and Setting. The main area is titled 'Job Editor' and shows a 'Job list (868 jobs)' with a search bar and a refresh button. Below the list is a table of jobs, each with a green status dot and a checkbox. The selected job is 'tpb.agritronics.humidity.CM-LS-002'. To the right of the table is a detailed view of the job configuration, showing a JSON object with fields like 'job_id', 'active', 'job_timeout', 'data_in', 'profile', 'param', 'data_transform', and 'data_out'. The configuration is as follows:

```
1 {
2   "job_id": "tpb.agritronics.humidity.CM-LS-002",
3   "active": true,
4   "job_timeout": "600000",
5   "data_in": {
6     "type": "tanpibut",
7     "profile": {
8       "station_id": "CM-LS-002",
9       "latitude": "18.931601",
10      "longitude": "18.931601"
11    },
12    "param": {
13      "recover": false,
14      "url": "http://122.155.1.142/ws/get2.php",
15      "appkey": "0c5a295bd8c07a0806411b79e1fd73",
16      "station_id": "CM-LS-002",
17      "data_type": {
18        "type": "2",
19        "name": "Humidity",
20        "node_id": "4096"
21      },
22      "init_observed_date": "2017-08-01",
23      "init_observed_time": "00:00:00",
24      "limit": 20
25    }
26  },
27  "data_transform": {
28    "type": "tanpibut"
29  },
30  "data_out": {
31    "type": "storage",
32    "param": {
33      "storage_name": "tpb.agritronics.humidity.CM-LS-002"
34    }
35  }
36 }
```

Job Editor



Storage Viewer

bigstream-console

Bigstream console

Storage lists

Search

#	Name
0	tpb.agritronics.humidity.CM-LS-001
1	tpb.agritronics.humidity.CM-LS-002
2	tpb.agritronics.humidity.CM-LS-003
3	tpb.agritronics.humidity.CM-LS-004
4	tpb.agritronics.humidity.CM-LS-005
5	tpb.agritronics.humidity.CM-LS-007
6	tpb.agritronics.humidity.CM-LS-008
7	tpb.agritronics.humidity.CM-LS-009
8	tpb.agritronics.humidity.CM-LS-010
9	tpb.agritronics.humidity.CM-LS-011
10	tpb.agritronics.humidity.CM-LS-012
11	tpb.agritronics.humidity.CM-LS-013

Storage Viewer

The screenshot shows the Bigstream console interface for the 'igrid.env' storage. The left sidebar contains navigation links: Home, JOBS, Job editor, STORAGE, Storage lists, and SETTING. The main area displays a table of storage objects with columns for index, list ID, and creation time. A search bar and filter tabs are at the top. A detailed view of a selected object is shown on the right.

file name : igrid/env.bss | file size : 191555390 | 903581 rows

Last filter: From - limit filter | by ID | Date

Number of last index: [input field] [SEARCH]

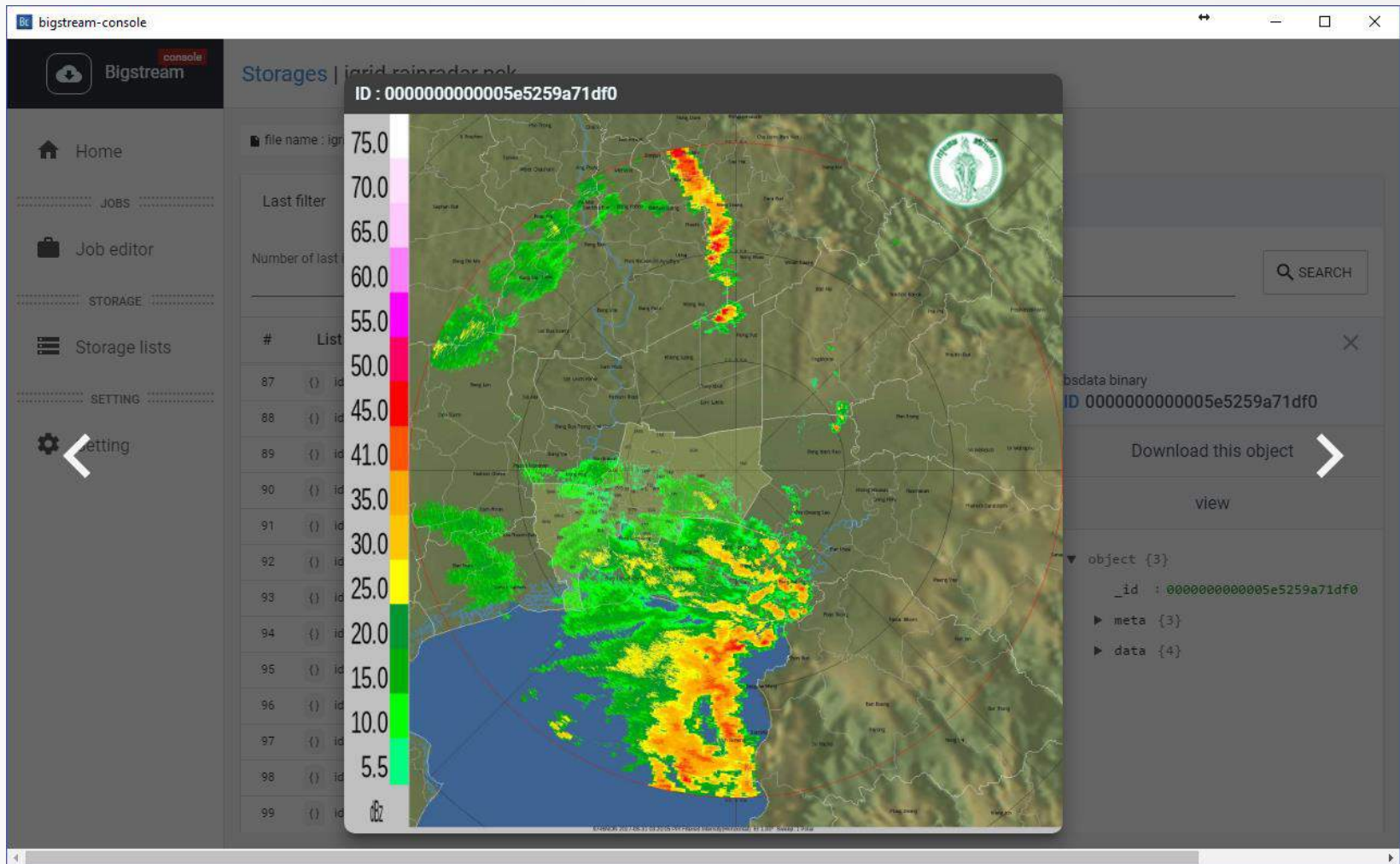
#	List	Create at
87	{ id : 01000000000dc99159a790f6	Thu Aug 31, 2017, 11:30:46
88	{ id : 01000000000dc99259a7910a	Thu Aug 31, 2017, 11:31:06
89	{ id : 01000000000dc99359a7911e	Thu Aug 31, 2017, 11:31:27
90	{ id : 01000000000dc99459a79133	Thu Aug 31, 2017, 11:31:47
91	{ id : 01000000000dc99559a79147	Thu Aug 31, 2017, 11:32:07
92	{ id : 01000000000dc99659a7915b	Thu Aug 31, 2017, 11:32:28
93	{ id : 01000000000dc99759a79170	Thu Aug 31, 2017, 11:32:48
94	{ id : 01000000000dc99859a79184	Thu Aug 31, 2017, 11:33:08
95	{ id : 01000000000dc99959a79198	Thu Aug 31, 2017, 11:33:29
96	{ id : 01000000000dc99a59a791ad	Thu Aug 31, 2017, 11:33:49
97	{ id : 01000000000dc99b59a791c1	Thu Aug 31, 2017, 11:34:09
98	{ id : 01000000000dc99c59a791d5	Thu Aug 31, 2017, 11:34:30
99	{ id : 01000000000dc99d59a791e9	Thu Aug 31, 2017, 11:34:50

ID 01000000000dc99d59a791e9

Download this object

```
object {3}
  _id : 01000000000dc99d59a791e9
  meta {3}
  data {4}
    temp : 23.6
    hum : 63.7
    light : 95
    ts : 1504154358
```

Storage Viewer



Monitoring Console

bigmaster.tlp - igrid@bigmaster.igridproject.info:22 - Bitvise xterm - PM2 Dashboard

Process list

	Mem:	CPU:	
[8] bs.api.service	30 MB	0 %	online
[9] bs.httplistener.service	30 MB	0 %	online
[2] bs.scheduler.service	29 MB	0 %	online
[13] bs.storage.read	29 MB	1 %	online
[14] bs.storage.read	29 MB	0 %	online
[11] bs.storage.write	29 MB	0 %	online
[3] bs.worker.ins0	34 MB	0 %	online
[4] bs.worker.ins1	38 MB	0 %	online
[5] bs.worker.ins2	36 MB	0 %	online
[7] bs.worker.ins3	32 MB	0 %	online

Global Logs

```
bs.worker.ins1 > [DO STATUS] : success
bs.worker.ins1 > ***** JOB SUCCESSFULLY DONE *****
bs.worker.ins2 > [DO STATUS] : success
bs.worker.ins2 > ***** JOB SUCCESSFULLY DONE *****
bs.worker.ins2 > ***** JOB RUNNING *****
bs.worker.ins2 > [JOB ID] : igrid.lsrenv
bs.worker.ins2 > [TRANSACTION ID] : TR15044987502602f8fa8
bs.worker.ins2 > [DI STATUS] : success
bs.worker.ins2 > [DT:0 STATUS] : success
bs.worker.ins2 > [DO STATUS] : success
bs.worker.ins2 > ***** JOB SUCCESSFULLY DONE *****
bs.worker.ins0 > WORKER:Forward job[igrid.lsrenv-storage] to SJW
bs.worker.ins3 > WORKER:Forward job[igrid.lsrenv-dweet] to SJW
bs.worker.ins3 > ***** JOB RUNNING *****
bs.worker.ins3 > [JOB ID] : igrid.lsrenv-storage
bs.worker.ins3 > [TRANSACTION ID] : TR15044987503507d5dee
bs.worker.ins3 > [DI STATUS] : success
bs.worker.ins3 > [DT:0 STATUS] : success
bs.worker.ins1 > WORKER:Forward job[igrid.lsrlight] to SJW
bs.worker.ins1 > ***** JOB RUNNING *****
bs.worker.ins1 > [JOB ID] : igrid.lsrenv-dweet
bs.worker.ins1 > [TRANSACTION ID] : TR15044987503816f420a
bs.worker.ins1 > [DI STATUS] : success
bs.worker.ins1 > [DT:0 STATUS] : success
bs.worker.ins2 > ***** JOB RUNNING *****
bs.worker.ins2 > [JOB ID] : igrid.lsrlight
bs.worker.ins2 > [TRANSACTION ID] : TR1504498750410bad29d
bs.worker.ins2 > [DI STATUS] : success
bs.worker.ins2 > [DT:0 STATUS] : success
bs.worker.ins2 > [DT:1 STATUS] : reject
bs.worker.ins2 > ***** JOB UNSUCCESSFULLY DONE *****
bs.worker.ins3 > [DO STATUS] : success
bs.worker.ins3 > ***** JOB SUCCESSFULLY DONE *****
bs.worker.ins1 > [DO STATUS] : success
bs.worker.ins1 > ***** JOB SUCCESSFULLY DONE *****
```

Custom metrics

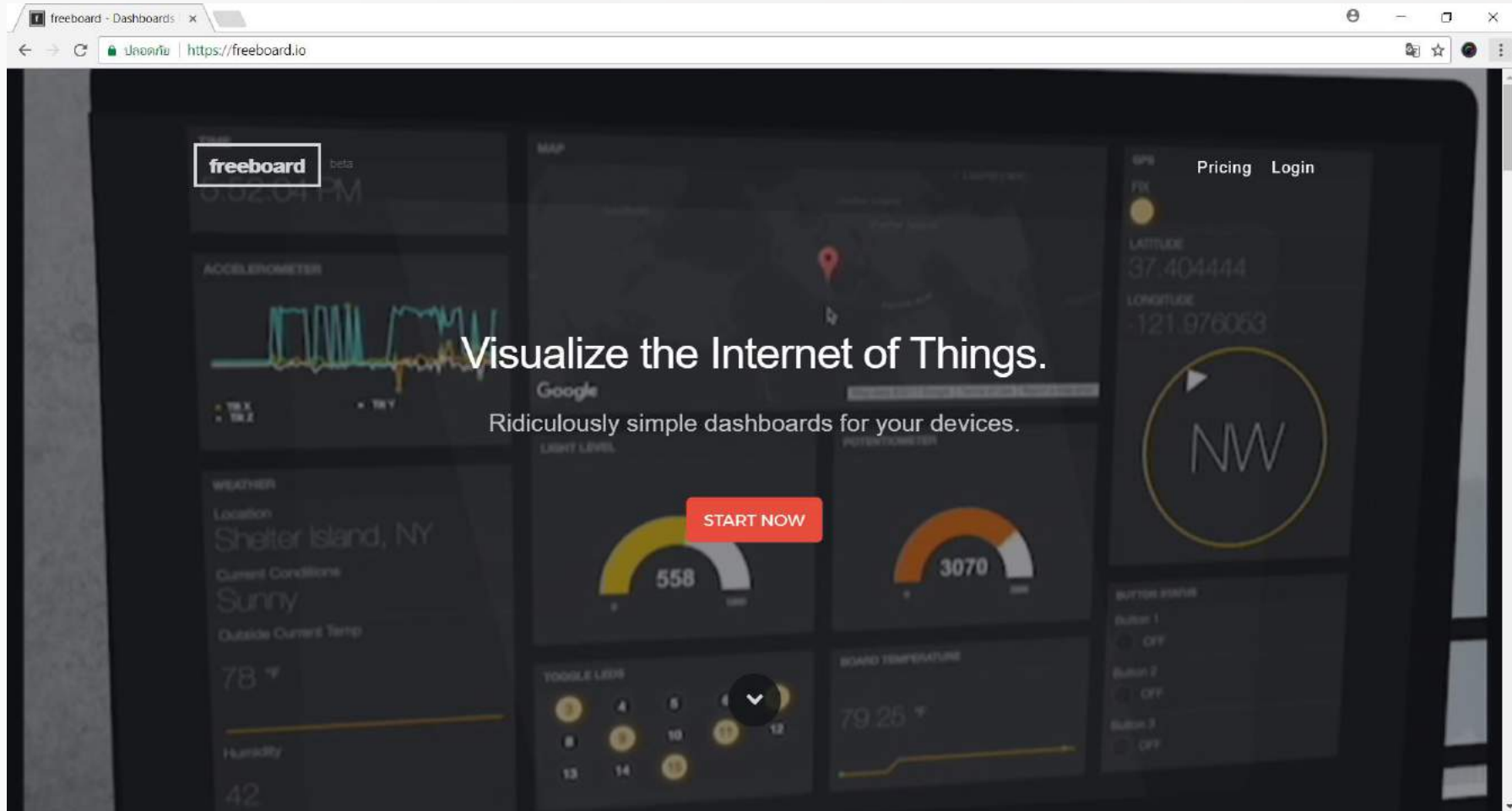
Loop delay 0.48ms

Metadata

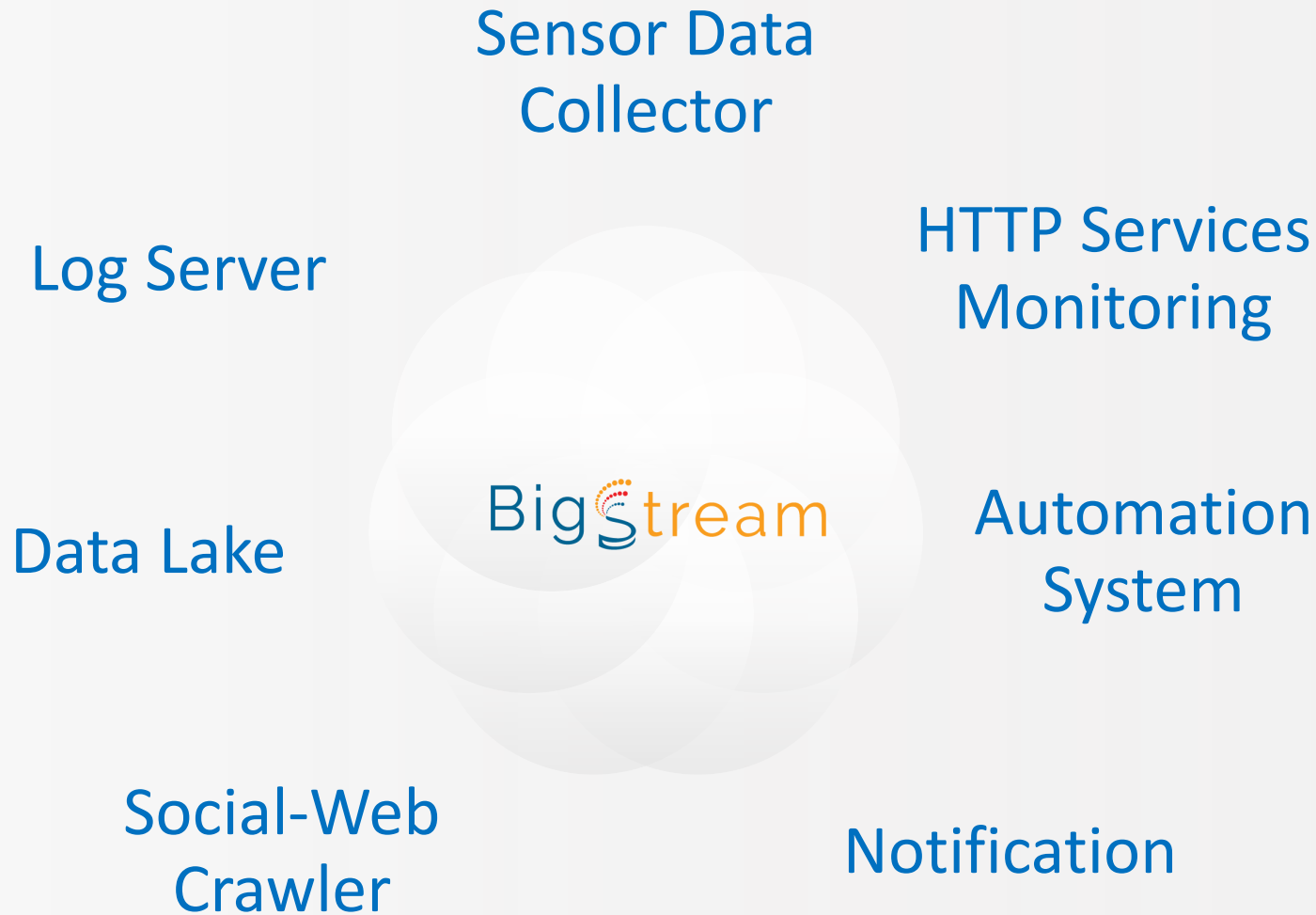
App Name	bs.api.service
Restarts	74970
Uptime	47s
Script path	/home/igrid/node-bigstream/serv-api.js
Script args	N/A
Interpreter	node
Interpreter args	N/A
Exec mode	fork
Node.js version	7.6.0
watch & reload	x
Unstable restarts	0

left/right: switch boards | up/down/mouse: scroll | Ctrl-C: exit

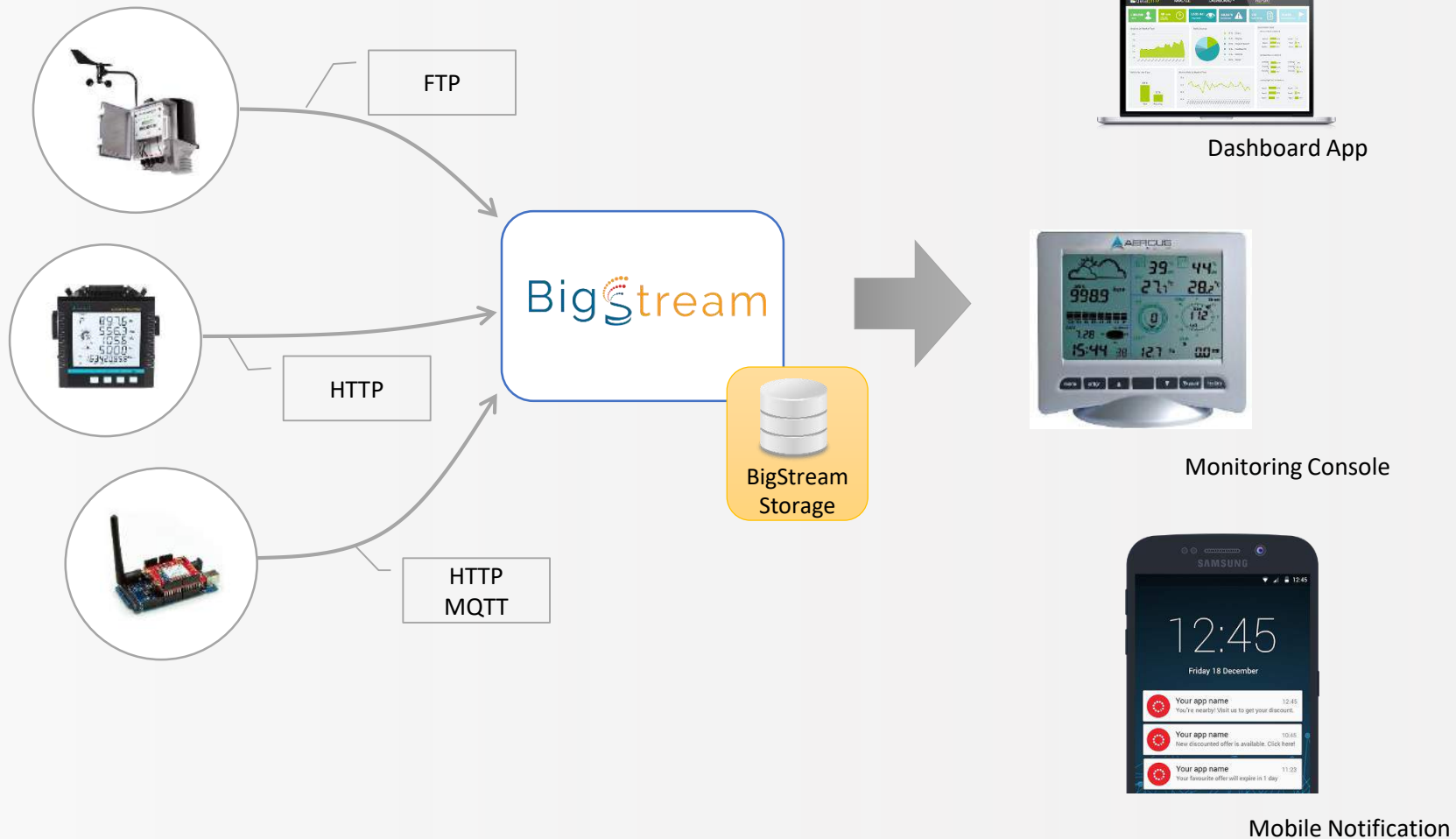
To go further check out <https://keymetrics.io/>



Potential Applications of BigStream



Sensor Data Collector



ESP8266 web client

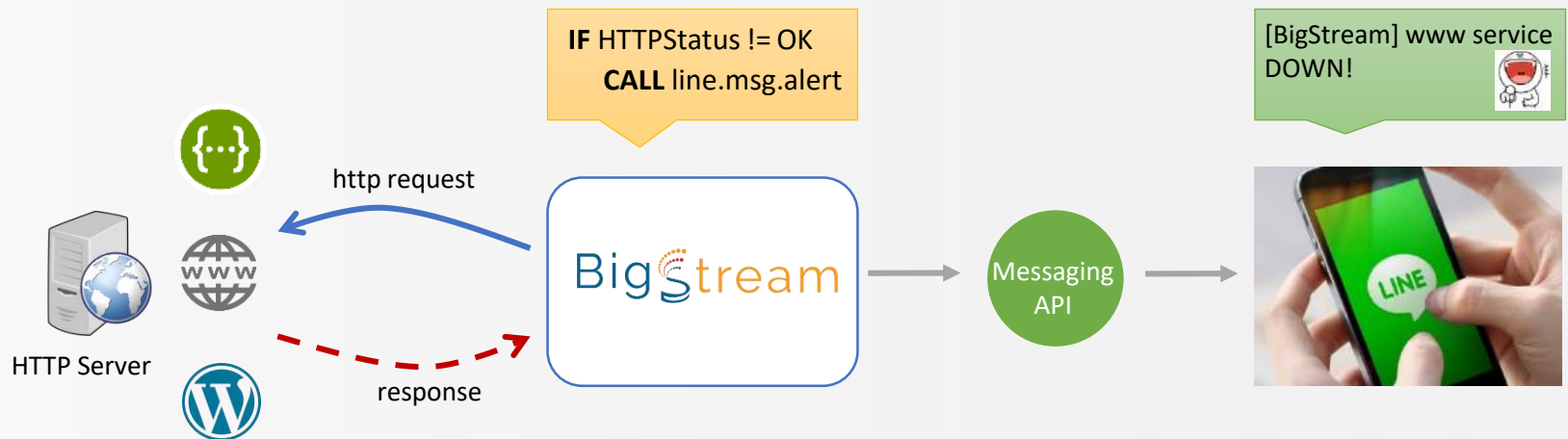
ARDUINO SKETCH

```
construct HTTP request to be sent
String cmd;
cmd = "GET /appkey?data=value&...";
cmd += temp;
cmd += "\r\n";
send the length of the command
espSerial.print("AT+CIPSEND=");
espSerial.println(cmd.length());

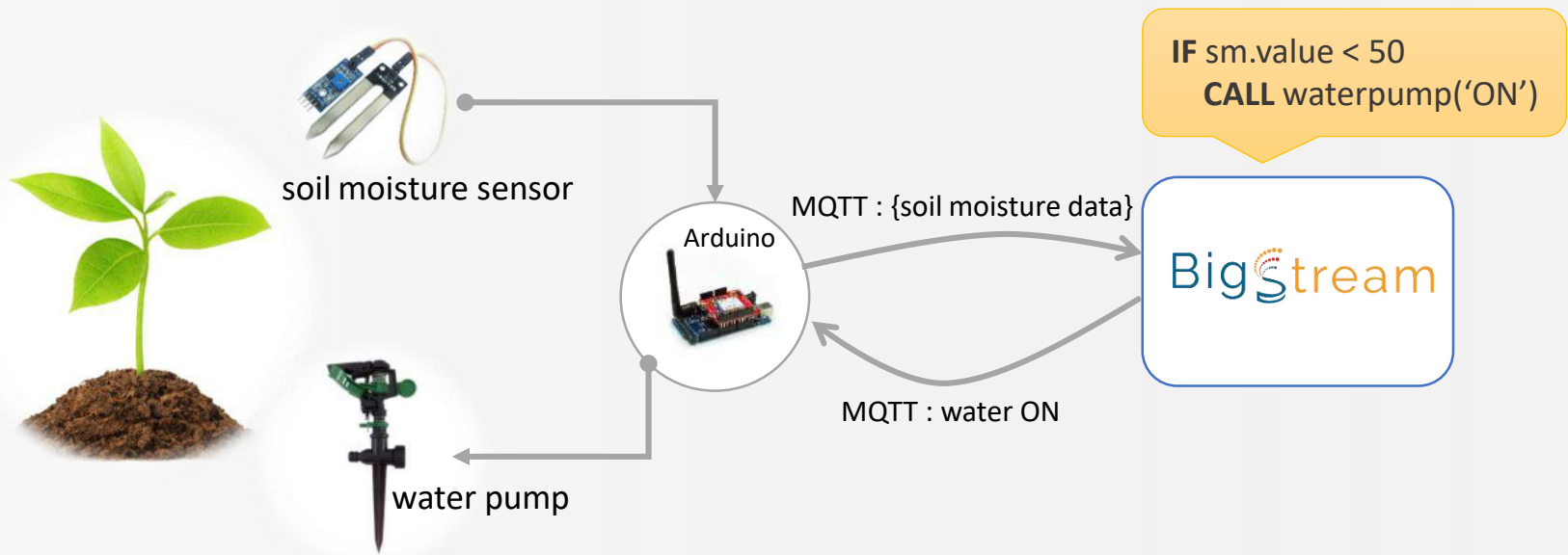
wait for ">" with espSerial.find(">")
espSerial.println(cmd);
decode the reply with espSerial.find("SEND OK")
```



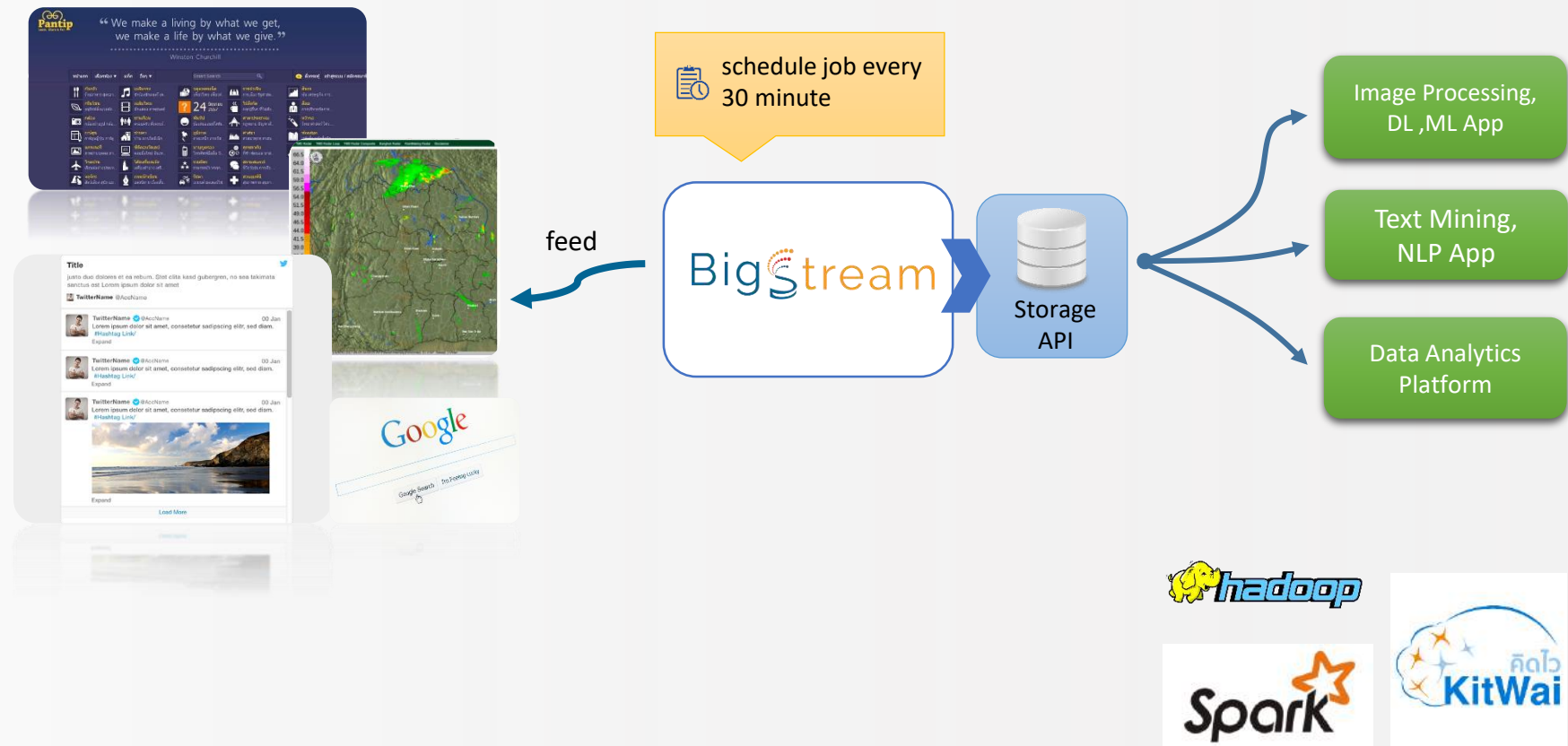
HTTP Service Monitoring



Automation System



Social Network - Web Crawler



LOCAL TIME

2:49:36 PM

BIGSTREAM MONITOR

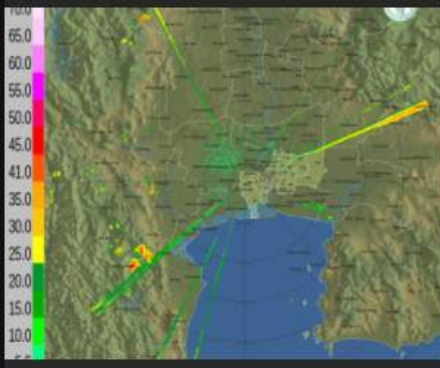
Agritrionics Data

110898 objects

Gistda-water

255919 objects

RADAR



KLONG LUANG

Temp

33 °C

Humidity

63 %



LSR ENVIRONMENT

Room Light

ON [brightness = 95%]

Temp

23.6 °C

Humidity



LAAN MAN PARKING

Total

299 cars



TANRABAD SURVEY

Service Status

UP [8/31/2017, 2:49:33 PM]

Total

42034 surveys

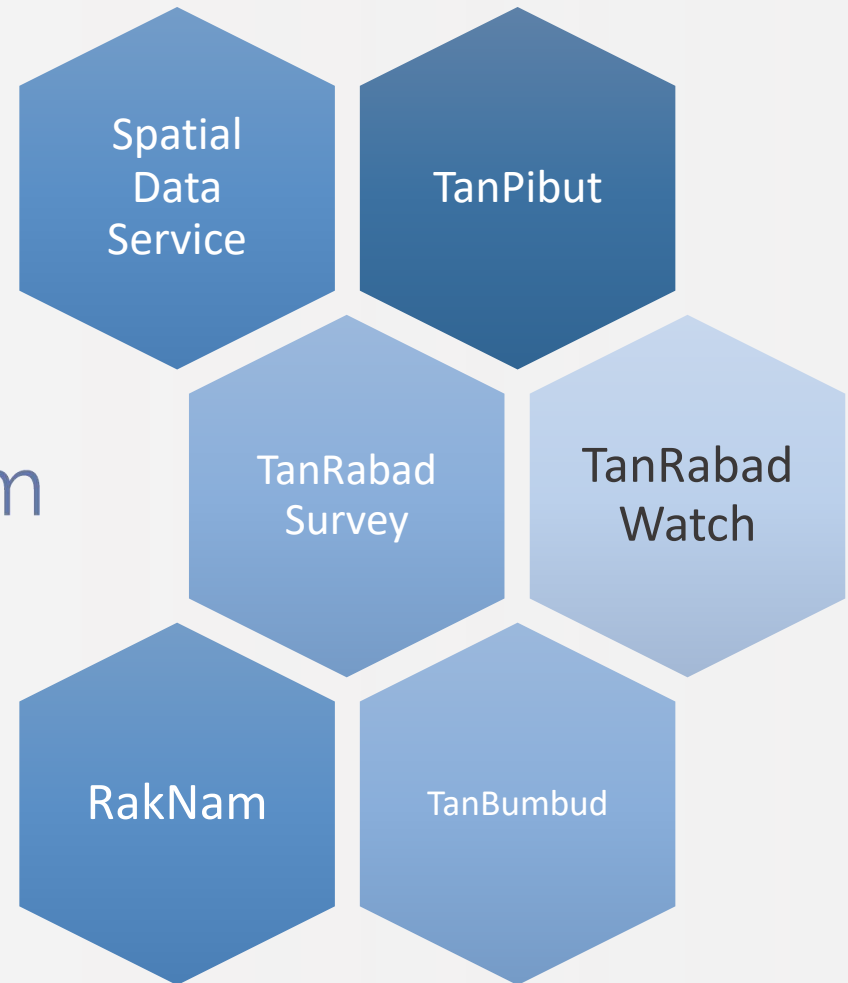
Last Survey

Surveyor : jidapa

Time : 8/31/2017, 11:08:49 AM



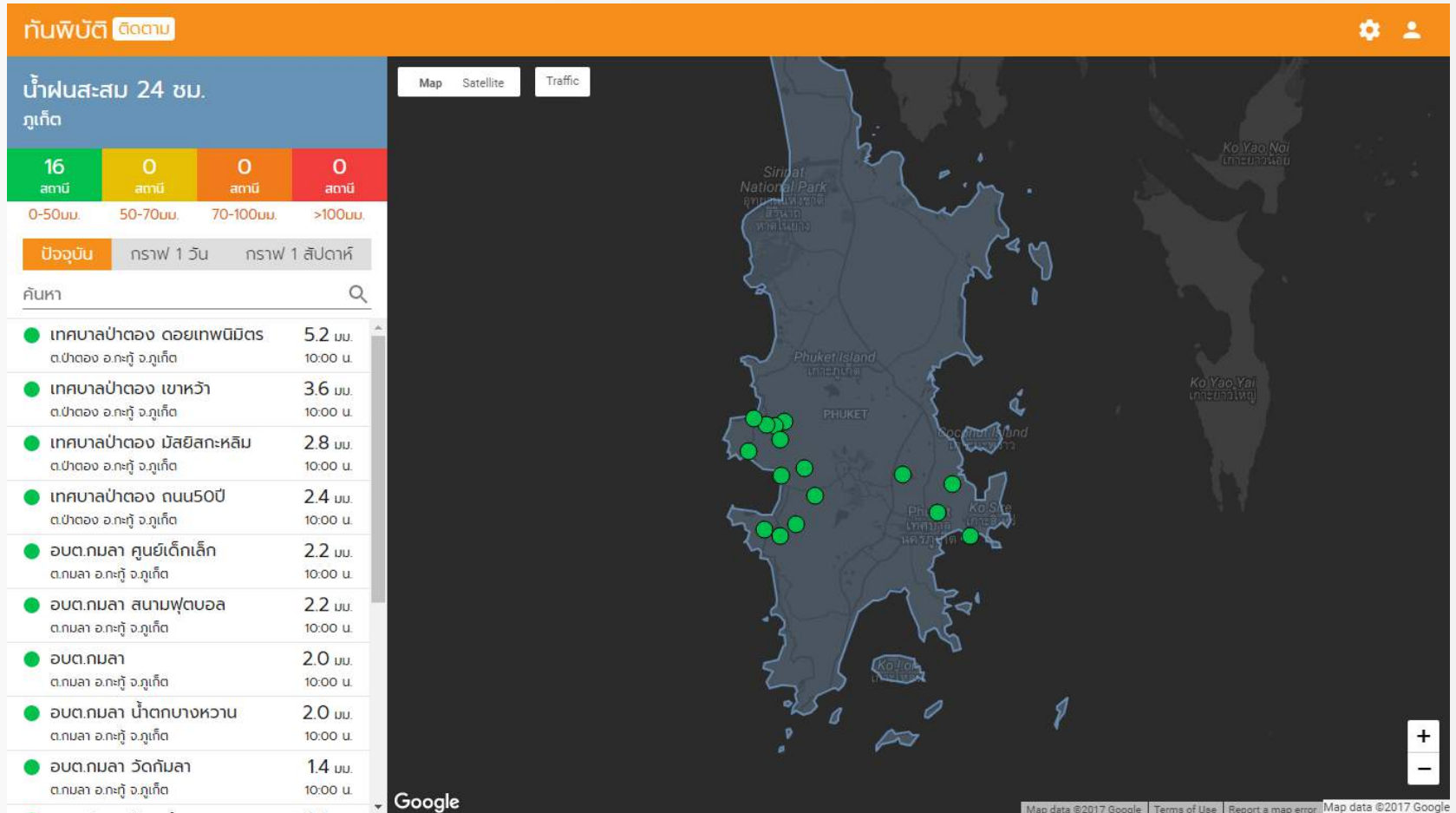
Projects with BigStream



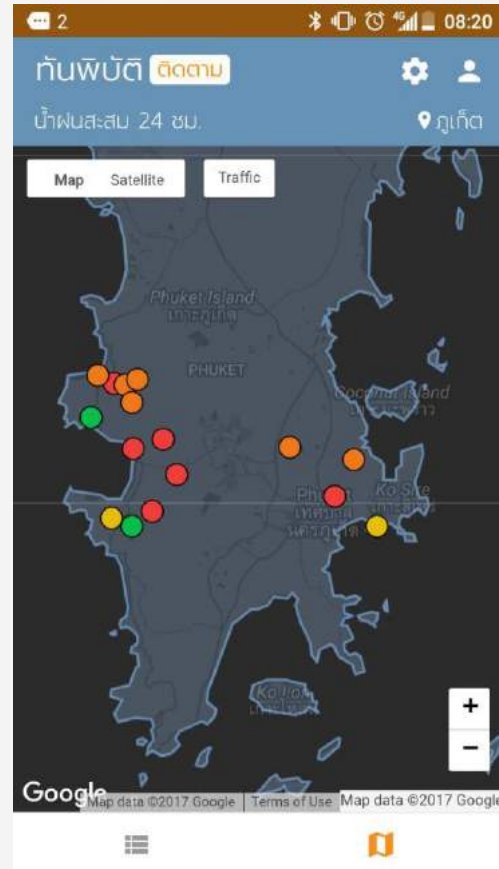
TanPibut (ทันพิบัติ)



TanPibut (ทันพิบัติ)



TanPibut (ทันพิบัติ)



TanRabad (ทันระบาด)

← สำรอง บันทึก

บ้านเลขที่ 20/1
ชุมชนอยู่เจริญสุขมาก
จำนวนผู้อยู่อาศัย ๒ คน

ภายในอาคาร	ภายนอกอาคาร
น้ำใช้ ทั้งหมด ☐ พว ☐	น้ำใช้ ทั้งหมด ☐ พว ☐
น้ำดื่ม ทั้งหมด ☐ พว ☐	น้ำดื่ม ทั้งหมด ☐ พว ☐
แฉก ทั้งหมด ☐ พว ☐	แฉก ทั้งหมด ☐ พว ☐
เครื่องกันมด ทั้งหมด ☐ พว ☐	เครื่องกันมด ทั้งหมด ☐ พว ☐
จานรองกระถาง ทั้งหมด ☐ พว ☐	จานรองกระถาง ทั้งหมด ☐ พว ☐
อ่างบัว/ใบไม้ ทั้งหมด ☐ พว ☐	อ่างบัว/ใบไม้ ทั้งหมด ☐ พว ☐
ยางรถยนต์เก่า ทั้งหมด ☐ พว ☐	ยางรถยนต์เก่า ทั้งหมด ☐ พว ☐

sync ข้อมูล

ผลการสำรวจ
1 มี.ค. 2559

จัดสรร
 ชุมชนหมู่บ้านโพธิ์
แขวงบางนา เขตบางนา กรุงเทพมหานคร

HI 19.15% CI 11.86% BI 29.79

บ้านที่สำรวจทั้งหมด 47

- พบลูกน้ำ 9
- ไม่พบลูกน้ำ 31
- ไม่มีภาชนะชั่งน้ำ 7
- สำรวจซ้ำ 0 ครั้ง

ภาชนะสำรวจทั้งหมด 118 พว 14

ภาชนะที่พบลูกน้ำยุ่งลายมากที่สุด

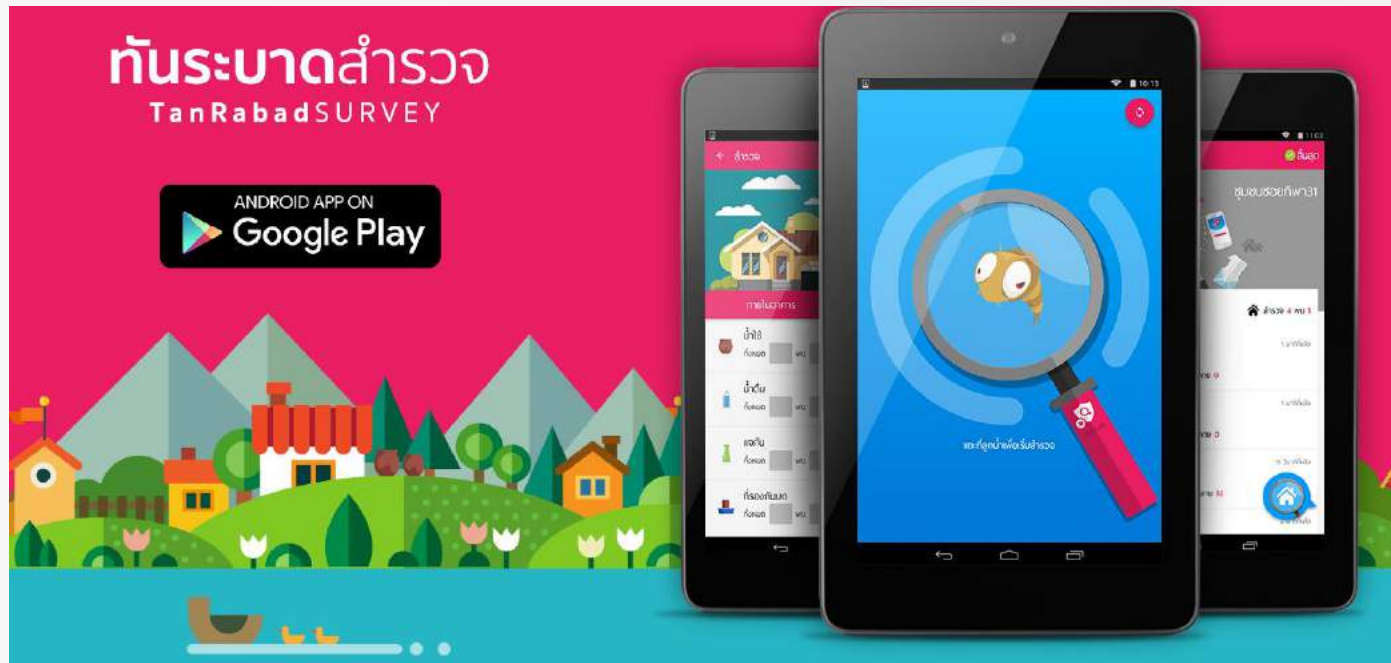
ภายในอาคาร	ภายนอกอาคาร
น้ำใช้	ภาชนะที่ไม่ใช่
อื่นๆ (ถ้าใช่ประโยชน์)	แฉก

ประมวลผลเมื่อ 1 ชั่วโมงที่แล้ว

รับทราบ



TanRabad-SURVEY



- collect larval survey data via on-line & off-line modes
- specify geo-location of places with GPS via Google Maps
- automatically transfer larval survey data to cloud
- real-time report entomology indices once the survey is completed

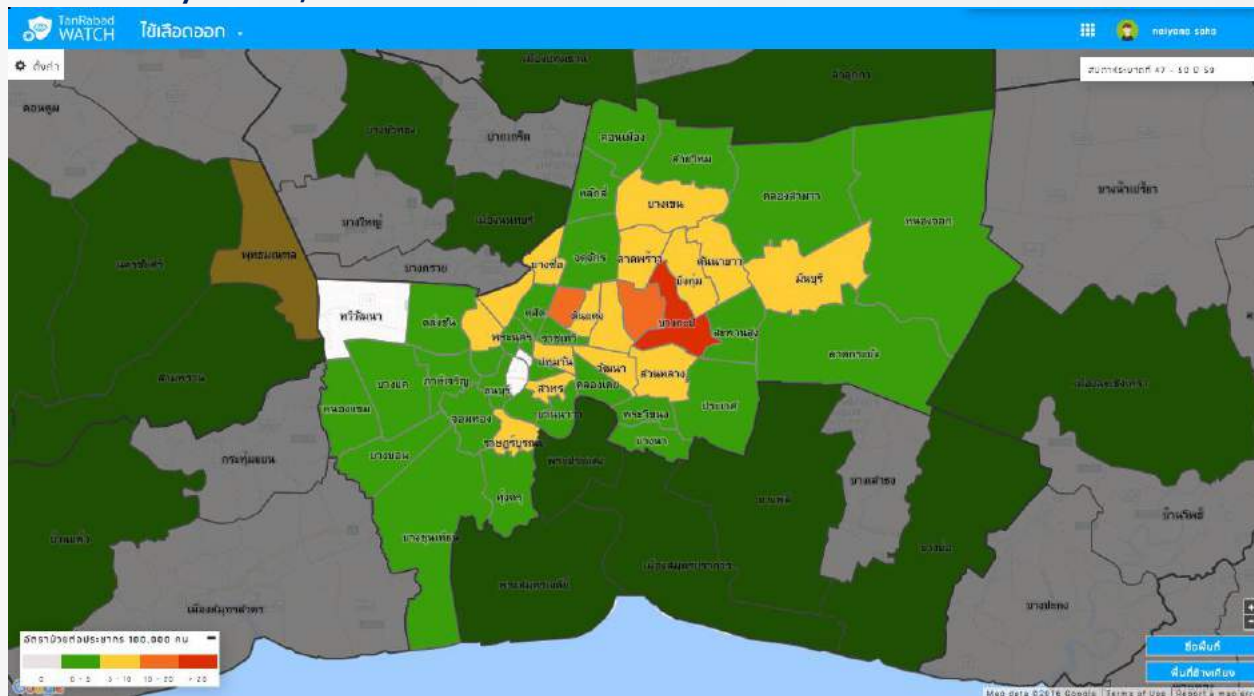
Real-time Epidemic Monitoring



TanRabad-WATCH



Morbidity Rate, Number of Patients with Median

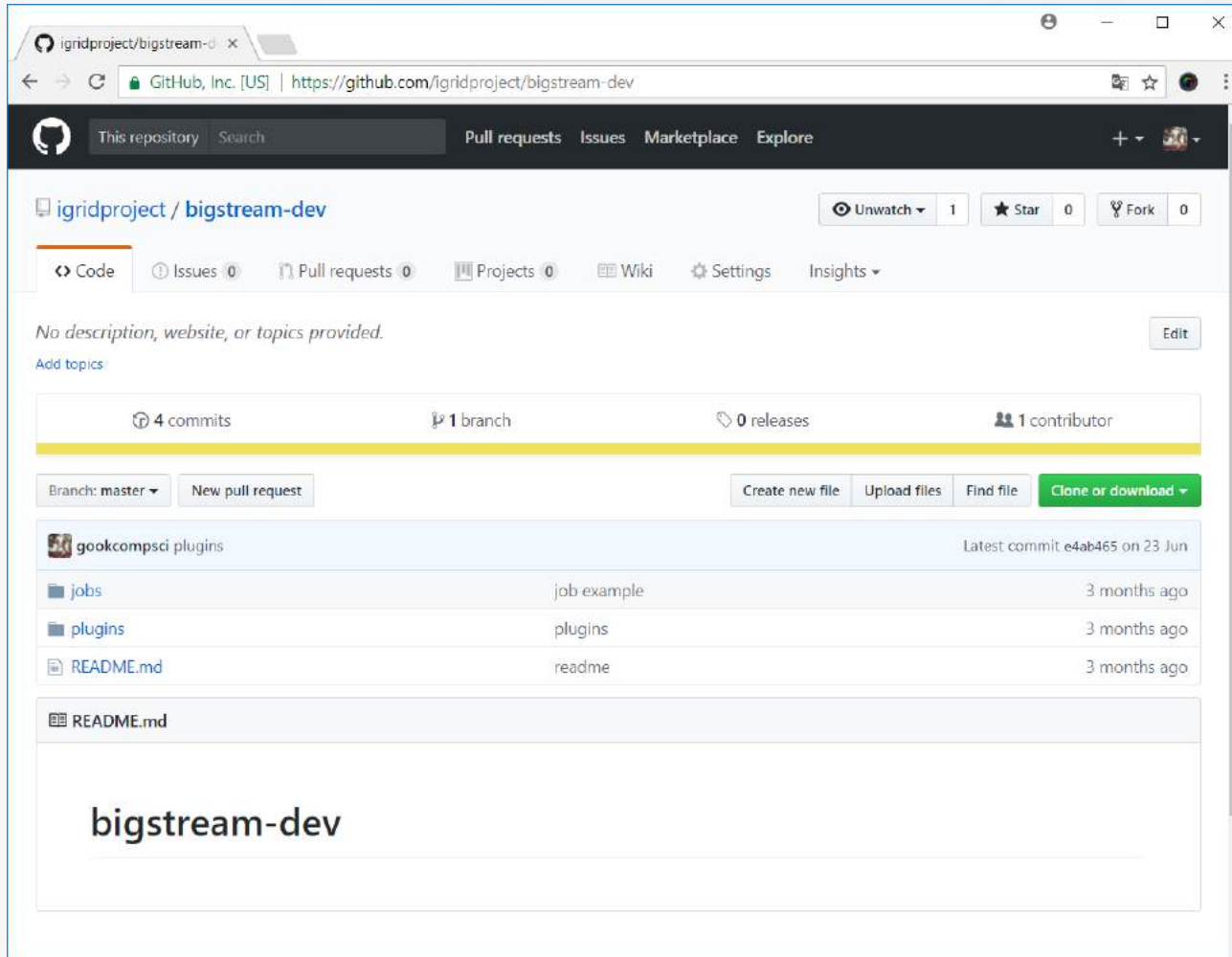


Try BigStream

The screenshot shows the Docker Hub page for the repository `igridproject/bigstream`. The page is titled "PUBLIC REPOSITORY" and includes a search bar with "bigstream" entered. The repository is marked as "Last pushed: 2 months ago". The "Repo Info" tab is selected, showing a "Short Description" of "Realtime Data Platform". The "Full Description" section contains the text "BigStream Docker", "Realtime Data Platform", and a link to the GitHub repository: <https://github.com/igridproject/node-bigstream>. Below this, the "Run Container" section lists instructions: "เปิด port 19080,19180,19980,5672" and "mount volume ของ bigstream storage data ซึ่งเก็บอยู่ที่ /var/bigstream/data". The "Docker Pull Command" section shows the command `docker pull igridproject/bigstream`. The "Owner" section shows the user `igridproject`.

<https://hub.docker.com/r/igridproject/bigstream/>

BigStream Dev



<https://github.com/igridproject/bigstream-dev>

“ งานวิจัยใช้ได้จริง สปริงบอร์ดนวัตกรรม ”



สนใจข้อมูลเพิ่มเติม กรุณาติดต่อ

ทีมวิจัย

คำรณ อรุณเรือ

kamron.aroonrua@nectec.or.th



Group
“NECTEC BigStream”

ดร.นัยนา สหเวชชภัณฑ์

naiyana.sahavechaphan@nectec.or.th

Session 7 “Dynamic Data Driven Application Systems (DDAS) เพื่อสนับสนุนการบริหารจัดการสิ่งแวดล้อมในยุค Big Data”
วันจันทร์ที่ 11 กันยายน 2560 | Meeting room 3 ศูนย์ประชุมแห่งชาติสิริกิติ์