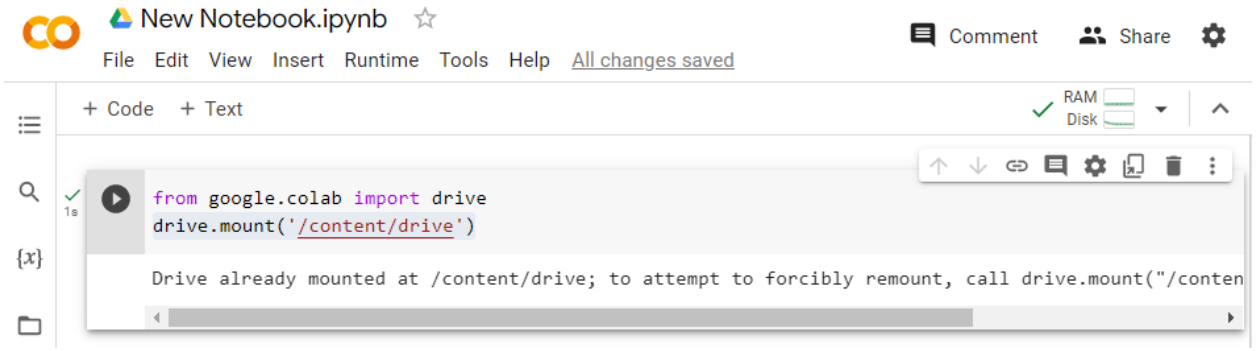
Question 1: Login into your Google Drive and start a Google Colab notebook file. Give Google Colab access to your Google Drive. Take a screenshot of it and copy and paste it.



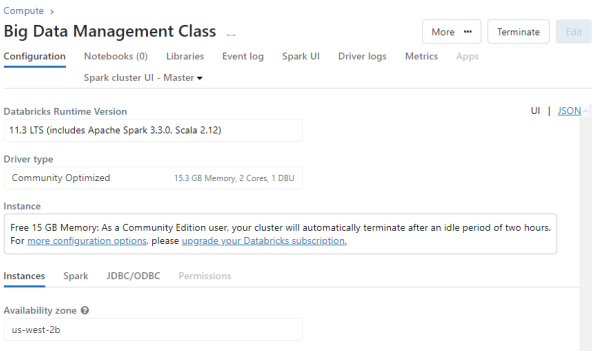
Step 1: I went to Google Colab website by visiting https://colab.research.google.com/
Step 2: Click on the "File" on the top-left of the screen and select "New notebook"
Step 3: After creating new notebook on Google Colab, I ran this code to get access to my Google Drive:

From google.colab import drive #import the drive module from google.colab package
drive.mount('/content/drive') #mount Google Drive to the specified path within the collab environment

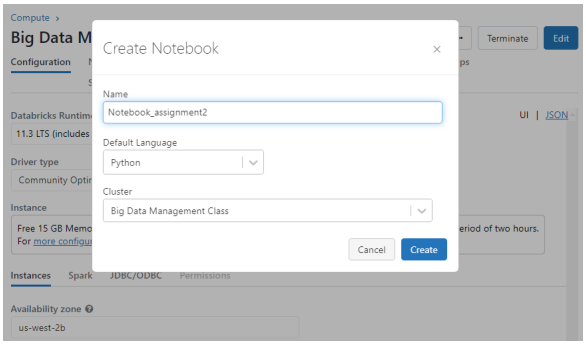
Step 4: After grant permissions that let Google Collab access the Google Drive, I can access Google Drive files directly. A message is appeared: "mounted at /content/drive"

Question 2: Start a Databricks account and then start a Databricks notebook file. Take a screenshot of it and copy and paste it.

Step 1: Create a cluster



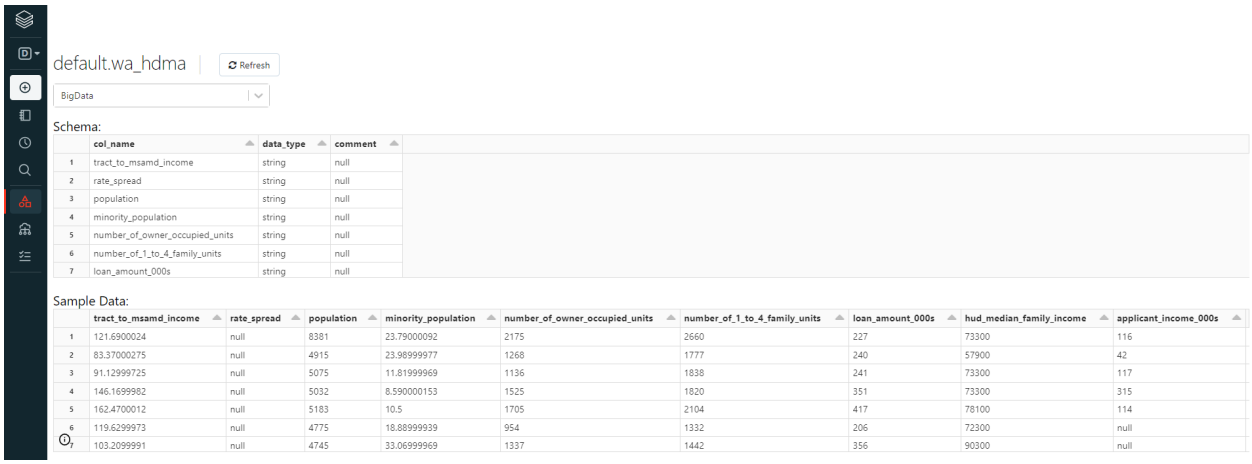
Step 2: Create a notebook



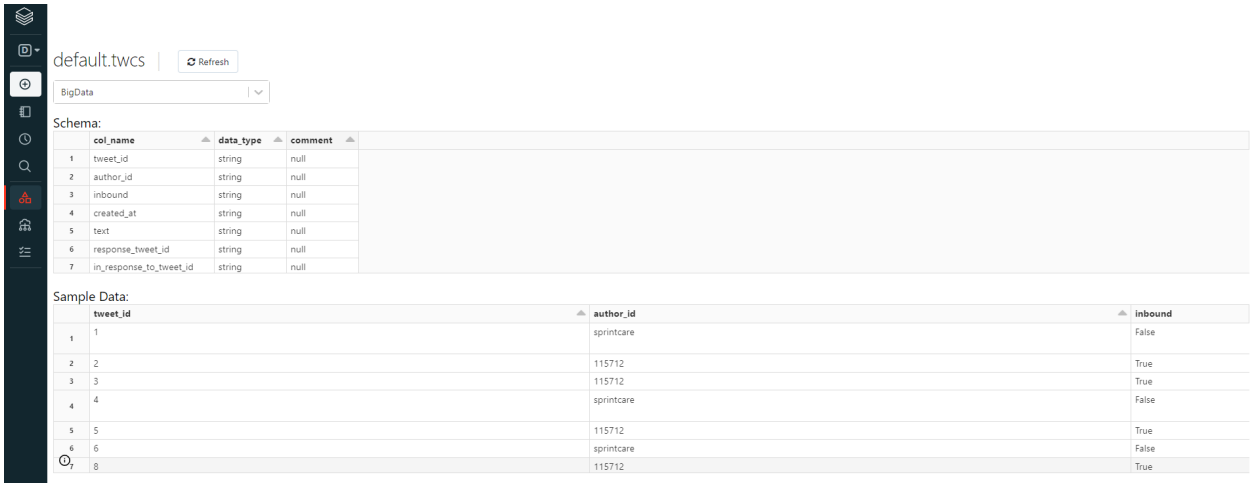
New notebook is created

Question 3: Upload the following datasets onto Databricks and create the tables and name them as follows. Take a screenshot of these tables on Databricks.

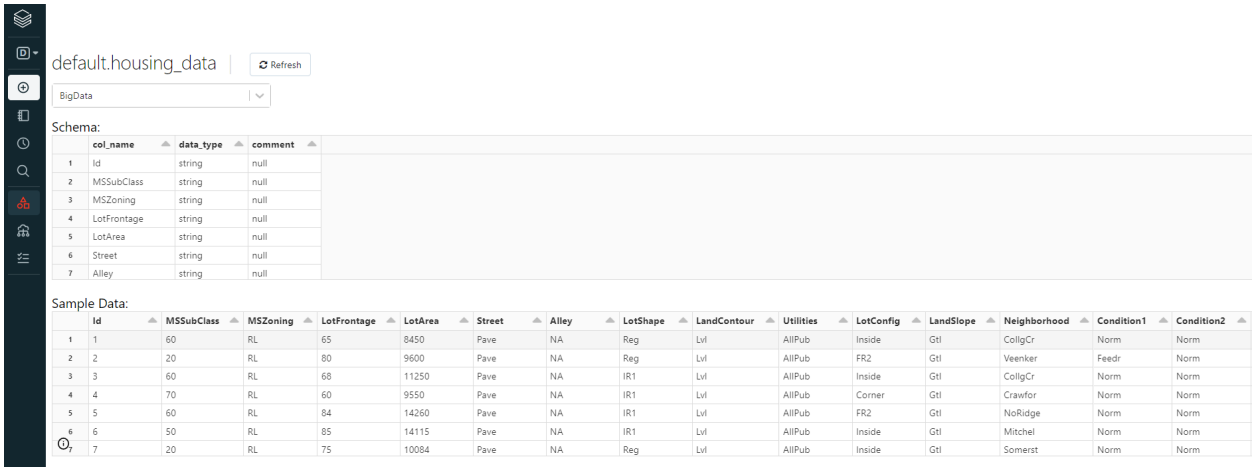
Wa_hdma table



Twcs table



Housing_data table



default.housing_data | Refresh

BigData

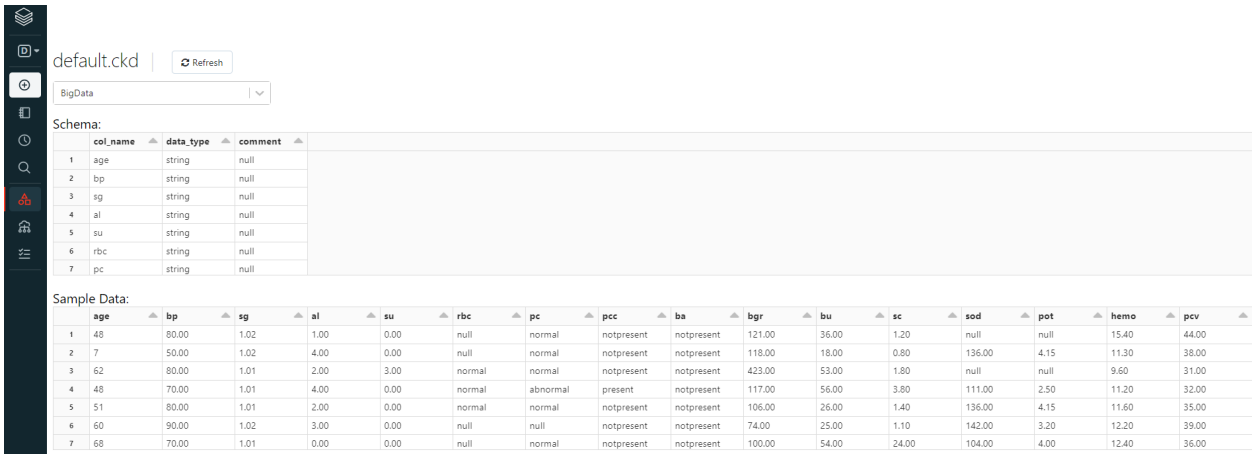
Schema:

	col_name	data_type	comment
1	Id	string	null
2	MSSubClass	string	null
3	MSZoning	string	null
4	LotFrontage	string	null
5	LotArea	string	null
6	Street	string	null
7	Alley	string	null

Sample Data:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	LotConfig	LandSlope	Neighborhood	Condition1	Condition2
1	1	60	RL	65	8450	Pave	NA	Reg	Lvl	AllPub	Inside	Gtl	CollgCr	Norm	Norm
2	2	20	RL	80	9600	Pave	NA	Reg	Lvl	AllPub	FR2	Gtl	Veenker	Feedr	Norm
3	3	60	RL	68	11250	Pave	NA	IR1	Lvl	AllPub	Inside	Gtl	CollgCr	Norm	Norm
4	4	70	RL	60	9550	Pave	NA	IR1	Lvl	AllPub	Corner	Gtl	Crawfor	Norm	Norm
5	5	60	RL	84	14260	Pave	NA	IR1	Lvl	AllPub	FR2	Gtl	NoRidge	Norm	Norm
6	6	50	RL	85	14115	Pave	NA	IR1	Lvl	AllPub	Inside	Gtl	Mitchel	Norm	Norm
7	7	20	RL	75	10084	Pave	NA	Reg	Lvl	AllPub	Inside	Gtl	Somerst	Norm	Norm

Ckd table



default.ckd | Refresh

BigData

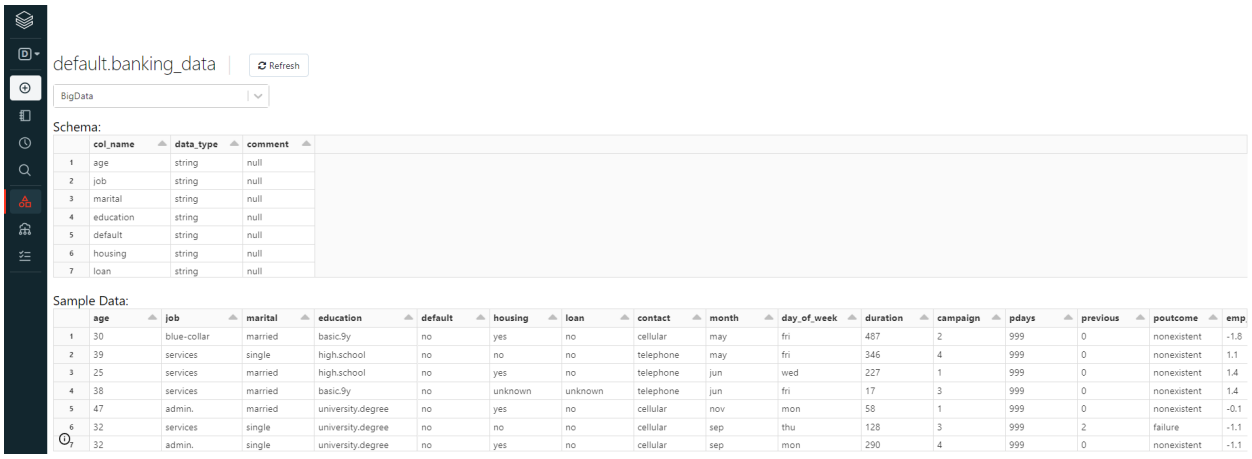
Schema:

	col_name	data_type	comment
1	age	string	null
2	bp	string	null
3	sg	string	null
4	al	string	null
5	su	string	null
6	rbc	string	null
7	pc	string	null

Sample Data:

	age	bp	sg	al	su	rbc	pc	pcc	ba	bgr	bu	sc	sod	pot	hemo	pcv
1	48	80.00	1.02	1.00	0.00	null	normal	notpresent	notpresent	121.00	36.00	1.20	null	null	15.40	44.00
2	7	50.00	1.02	4.00	0.00	null	normal	notpresent	notpresent	118.00	18.00	0.80	136.00	4.15	11.30	38.00
3	62	80.00	1.01	2.00	3.00	normal	normal	notpresent	notpresent	423.00	53.00	1.80	null	null	9.60	31.00
4	48	70.00	1.01	4.00	0.00	normal	abnormal	present	notpresent	117.00	56.00	3.80	111.00	2.50	11.20	32.00
5	51	80.00	1.01	2.00	0.00	normal	normal	notpresent	notpresent	106.00	26.00	1.40	136.00	4.15	11.60	35.00
6	60	90.00	1.02	3.00	0.00	null	null	notpresent	notpresent	74.00	25.00	1.10	142.00	3.20	12.20	39.00
7	68	70.00	1.01	0.00	0.00	null	normal	notpresent	notpresent	100.00	54.00	24.00	104.00	4.00	12.40	36.00

Banking_data table



default.banking_data | Refresh

BigData

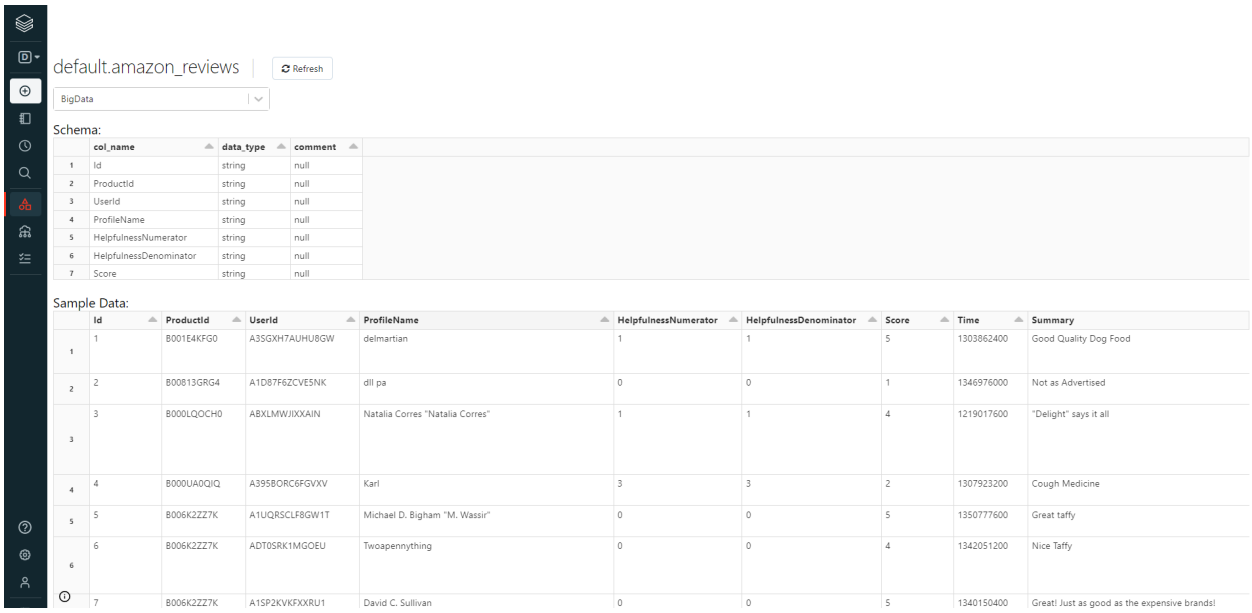
Schema:

	col_name	data_type	comment
1	age	string	null
2	job	string	null
3	marital	string	null
4	education	string	null
5	default	string	null
6	housing	string	null
7	loan	string	null

Sample Data:

	age	job	marital	education	default	housing	loan	contact	month	day_of_week	duration	campaign	pdays	previous	poutcome	emp
1	30	blue-collar	married	basic.9y	no	yes	no	cellular	may	fri	487	2	999	0	nonexistent	-1.8
2	39	services	single	high.school	no	no	no	telephone	may	fri	346	4	999	0	nonexistent	1.1
3	25	services	married	high.school	no	yes	no	telephone	jun	wed	227	1	999	0	nonexistent	1.4
4	38	services	married	basic.9y	no	unknown	unknown	telephone	jun	fri	17	3	999	0	nonexistent	1.4
5	47	admin.	married	university.degree	no	yes	no	cellular	nov	mon	58	1	999	0	nonexistent	-0.1
6	32	services	single	university.degree	no	no	no	cellular	sep	thu	128	3	999	2	failure	-1.1
7	32	admin.	single	university.degree	no	yes	no	cellular	sep	mon	290	4	999	0	nonexistent	-1.1

Amazon_review table



Question 4: Compare and contrast in-house (on-premise or also known as the private cloud) data warehousing and analytics platforms with those provided through cloud services (fully cloud or also known as the public cloud).

The differences between ***in-house (on-premise or private cloud)*** and ***cloud-based services*** (public cloud) data warehousing and analytics platforms are explained in following factors:

Factors	Private cloud	Public cloud
Deployment	• Hosted on the local servers and infrastructure. • Manage and maintained by the in-house teams	• Hosted on the cloud service provider ‘s infrastructure. • Managed and maintained by the cloud service provider team.
Scalability	• Scale take time, money and human resource and can't change the plan quickly	• Easy to scale based on the demand and can change the plan quickly
Security	• Has full control: security measurement, intrusion detection and access control which help keep the security in the highest level.	• Rely on the cloud service provider ‘s security practices and has limitations in security measurement.
Management	• Has full control and flexibility to customize the infrastructure	• Has less control and customized options. Can’t

	based on the demands.	deploy customization immediately cause the organizations need to work with the cloud service provider for any customizations.
Cost	● Higher upfront cost for infrastructure setup. ● Higher cost for management and maintenance because organizations need a dedicated IT team to take care. ● Higher cost when scale the infrastructure.	● Lower upfront cost for infrastructure setup. ● Cost for management and maintenance is flexible based on demands. However, do not need a dedicated IT team to take care, instead, some headcounts will work closely with the cloud service provider IT team. ● Lower cost and time when scale because the service provider has available plans for clients' demands.

Based on the factors compared between private and public cloud, organizations can make decisions based on their requirements. Both cloud infrastructures have pros and cons that organizations consider to maximize their advantages and profits.

Question 5: Elaborate on what you think business leaders should consider when choosing on-premise, public, or hybrid cloud infrastructures?

Hybrid cloud infrastructure is the combination between both private and public cloud services in which the use of private and public cloud services is dependent on the organization's demands and requirements. The hybrid cloud infrastructure is the mix of private and public cloud, there are some pros and cons when choosing this plan:

- Pros:
    - Businesses using hybrid cloud service are flexible to scale infrastructure and control over critical resources.
    - Take advantage of private cloud and prevent the limitation of public cloud on security.
- Cons:
    - Increase the workload while organizations and cloud service providers have to collaborate consistently, leading to the complexity in management and increase of upfront cost, management cost and data transfer cost.

Combined with the comparison on question 4, each private, public and hybrid cloud service has its advantages and disadvantages which business leaders should consider. These factors are: security, cost, scalability, integration, deployment, and management. Organizations can leverage the advantages of each infrastructure when making the right decision based on specific needs and situations of business.

Question 6: What are the primary cloud service providers? Briefly compare and contrast them.

There are 3 primary cloud service providers: Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP). Here are some factors when comparing them:

Factors	AWS	Microsoft Azure	GCP
Services	Storage, compute, networking, security, analytics	AI, ML, computing, hybrid cloud	Analytics, ML, infrastructure as a service
Features	EC2 (compute), S3 (storage), RDS (relational database), Lamda (serverless computing)	Virtual Machines (compute), Azure storage (storage), Azure SQL Database (relational database), Azure Functions (serverless computing)	Compute Engine (compute), Cloud Storage (storage), Cloud SQL (relational database), Cloud Functions (serverless computing)
User types	Multiple business types in various industries	Large enterprise in various industries	Companies with technology focus
Infrastructure	Worldwide infrastructure with highest number of data centers	Worldwide infrastructure with second-largest number of data centers	Worldwide infrastructure with third-largest number of data centers
Open-source support	Open-source support but not so focused	Open-source support but not so focused	Focus on open-source more than Microsoft Azure and AWS

Question 7: Elaborate on the security issues and challenges that come with cloud computing.

According to IBM: “Cloud computing is on-demand access, via the internet, to computing resources—applications, servers (physical servers and virtual servers), data storage, development

tools, networking capabilities, and more—hosted at a remote data center managed by a cloud services provider (or CSP). ” Cloud computing is a data storage service, therefore, security issues regarding this service include: data loss, access control, infrastructure security, and data breaches. When the systems of cloud service providers have vulnerabilities, malicious actors can attack to steal data used for negative purposes. The vulnerabilities in systems can also cause data loss, severely impacting organizations when leading to legal issues, business downtime and financial loss. To avoid these security problems, organizations and cloud services providers have to ensure having robust security measures to avoid risks in third-party providers, infrastructure and access control.

Question 8: Elaborate on the link between IoT and Cloud Computing and Big Data.

According to Oracle: “The Internet of Things (IoT) describes the network of physical objects—“things”—that are embedded with sensors, software, and other technologies for the purpose of connecting and exchanging data with other devices and systems over the internet.” Therefore, IoT, Cloud Computing and Big Data has mutual relationship, in which:

- **IoT helps create more data:** Sensors, software, and other technologies are essential parts in generating data.
- **Big Data and Cloud Computing help increase the competency and capability of IoT:** Cloud Computing plays an important part in managing the IoT ecosystem when providing a centralized management platform. In addition, Big Data offers analytics capabilities to process and analyze data in real-time and near-time, helping extract data from IoI devices. In addition, the advantages from data analytics can be applied to enhance the power of IoT devices, helping generate more data used for development of many industries and organizations.