

從前端資料擷取到後端文本/AI分析：
MATLAB與AIoT開發整合流程

Internet of Things

應用層

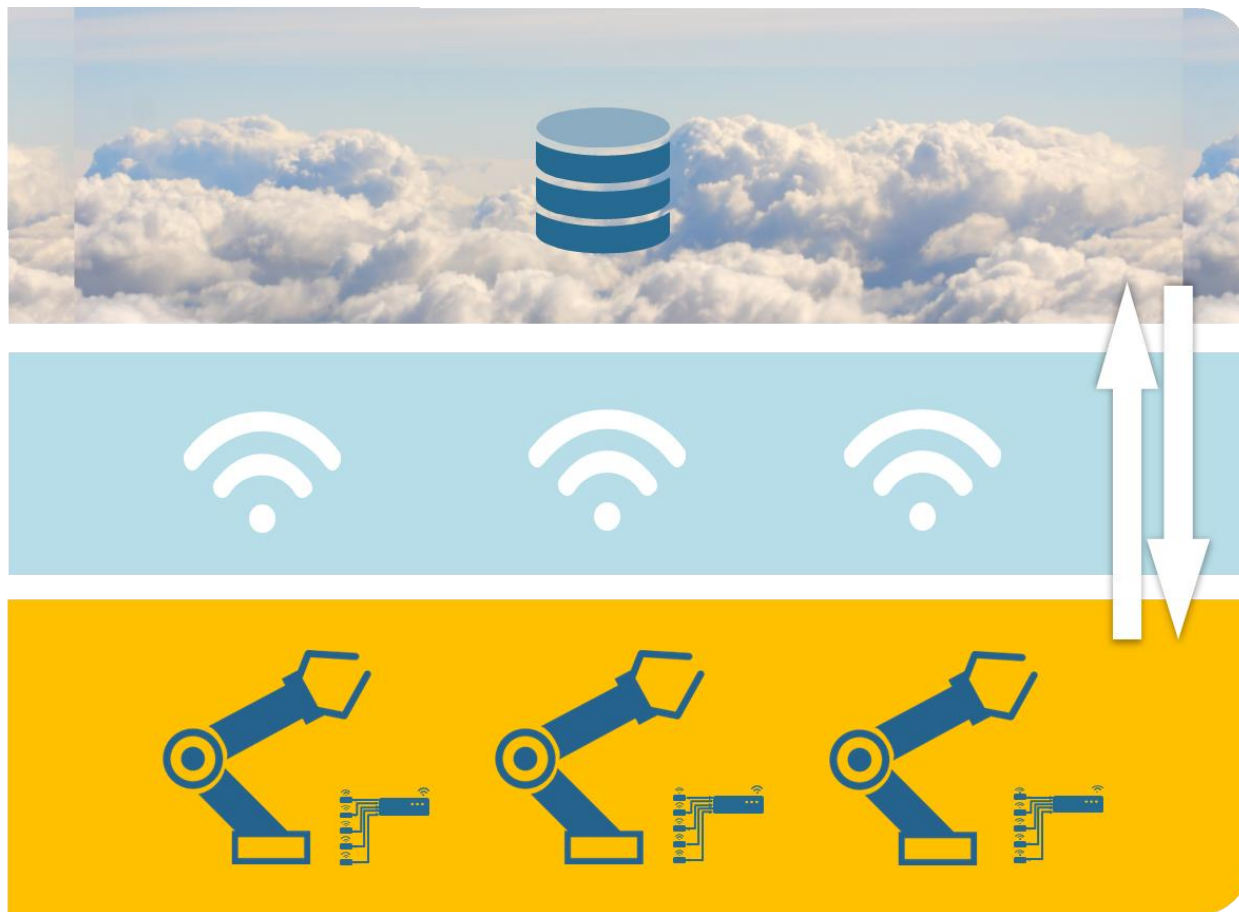
雲端主機接受網路層所傳遞資訊，藉由大數據分析與機器學習等演算法處理資料生成有用解決辦法。

網路層

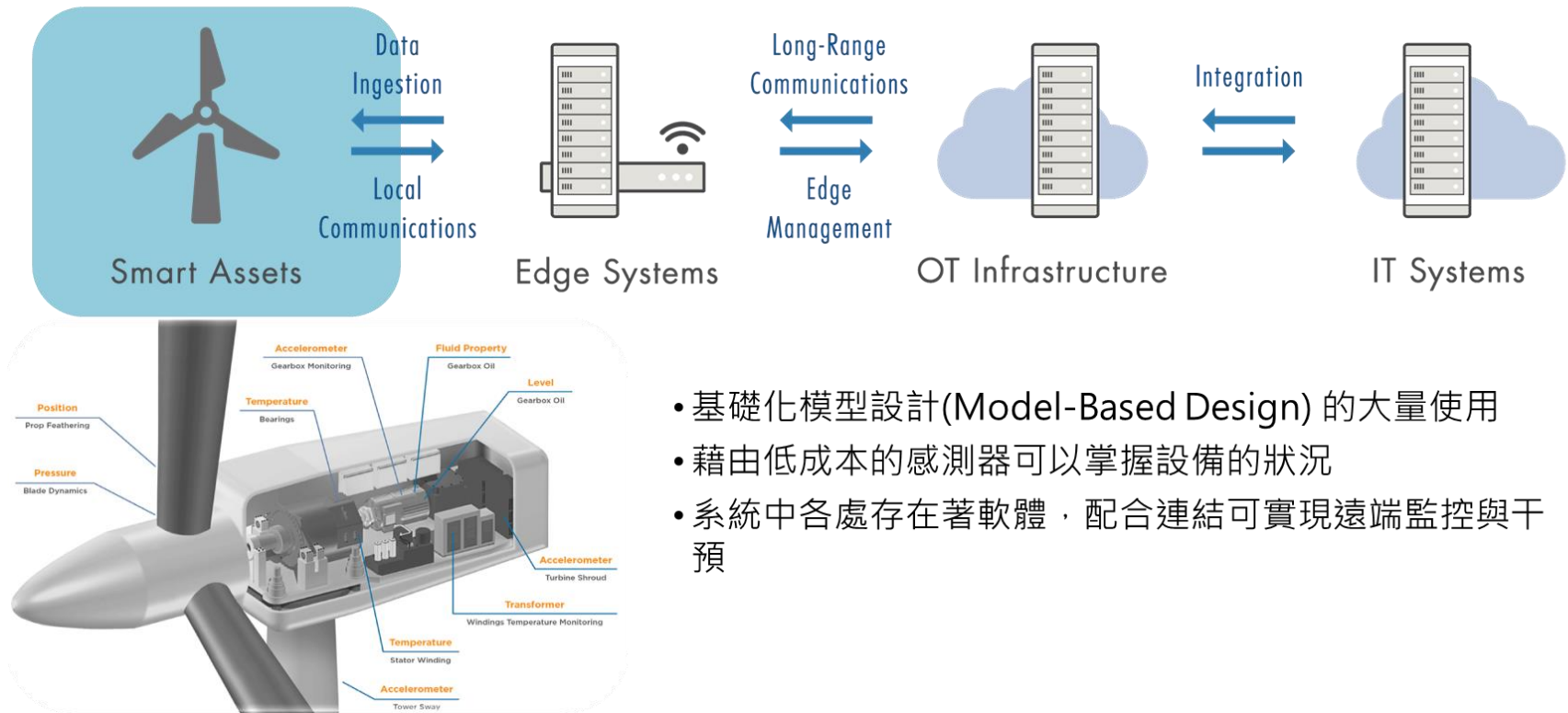
資訊透過有線或 3G、wi-fi 等無線通訊科技傳遞資訊。

感測層

透過壓力、扭力、濕度、加速規、鏡頭等各式感測計收集機台或是流程數據。

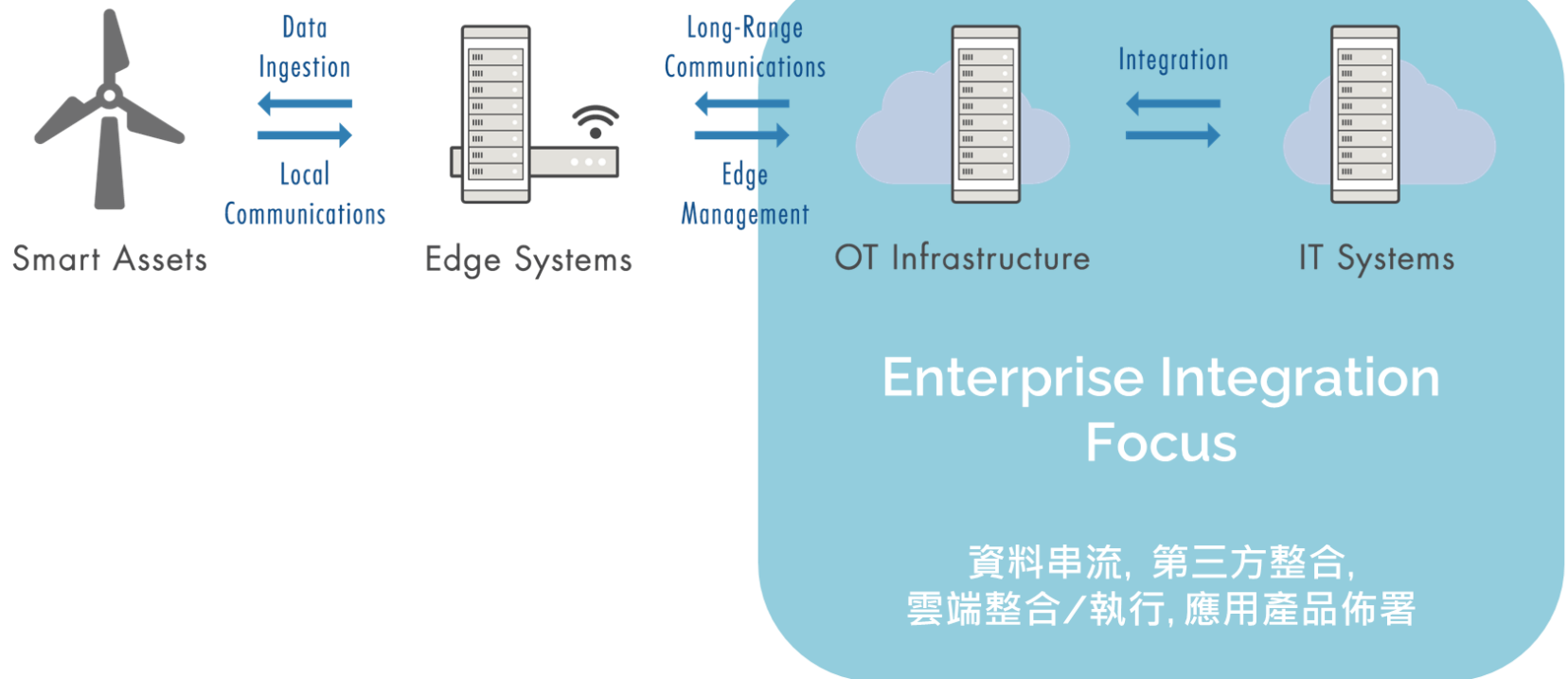


Industrial IoT Systems (IIoT)



- 基礎化模型設計(Model-Based Design) 的大量使用
- 藉由低成本的感測器可以掌握設備的狀況
- 系統中各處存在著軟體，配合連結可實現遠端監控與干預

Industrial IoT Systems





Edge Devices

- Signal Processing, Data Compression
- Embedded Systems (CPU, FPGA)



Data Aggregation

- Preprocessing, Synchronization
- Database Accesses

"A" IoT



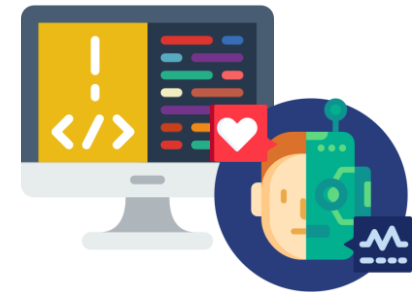
Edge Devices

- Signal Processing, Data Compression
- Embedded Systems (CPU, FPGA)



Data Aggregation

- Preprocessing, Synchronization
- Database Accesses



Data Analytics

- Trend Analysis, Anomaly Detection
- Data Visualization, Reporting

數位黃金



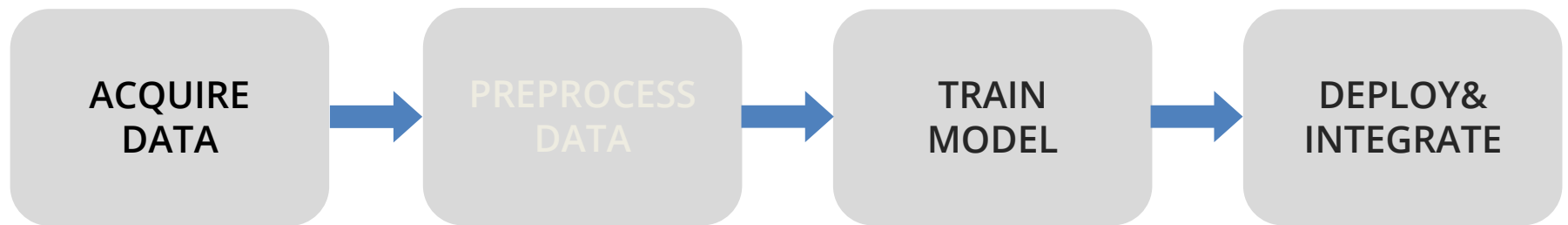
數位黃金

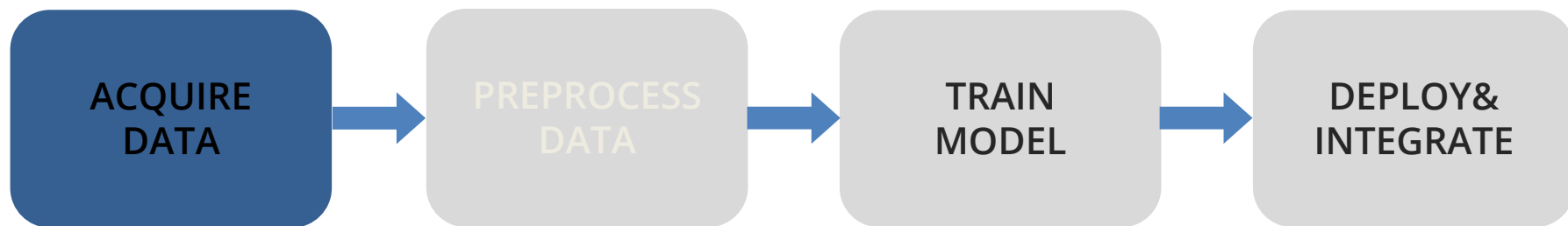


資料何處來 此物只應天上有？



Process workflow





資料收集

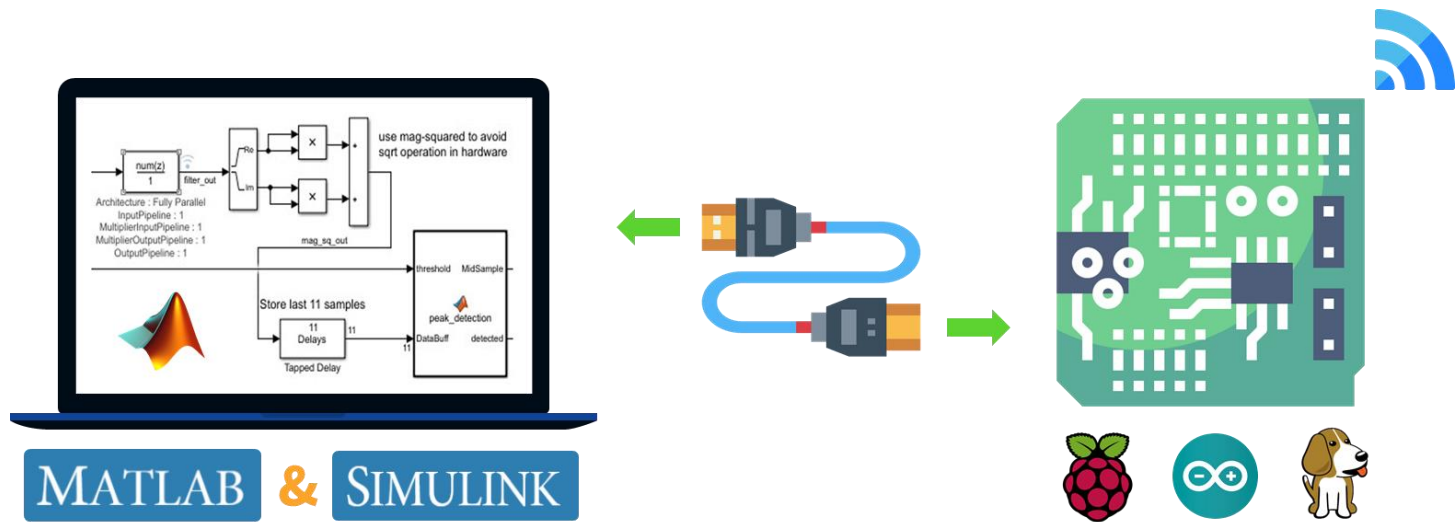
當你在 Messenger 群組貼了超爆笑的梗圖時



譯：好色龍

Source: <https://horndragon.blogspot.com/2018/04/974.html>

Edge Device



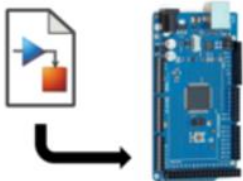
Hardware Library



MATLAB Support Package for Arduino Hardware

by MathWorks MATLAB Hardware Team **STAFF**

Acquire inputs and send outputs on Arduino boards



Simulink Support Package for Arduino Hardware

by MathWorks Simulink Team **STAFF**

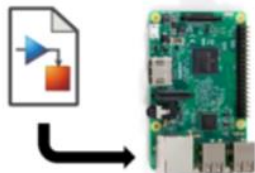
Run models on Arduino boards.



MATLAB Support Package for Raspberry Pi Hardware

by MathWorks MATLAB Hardware Team **STAFF**

Acquire sensor and image data from your Raspberry Pi.

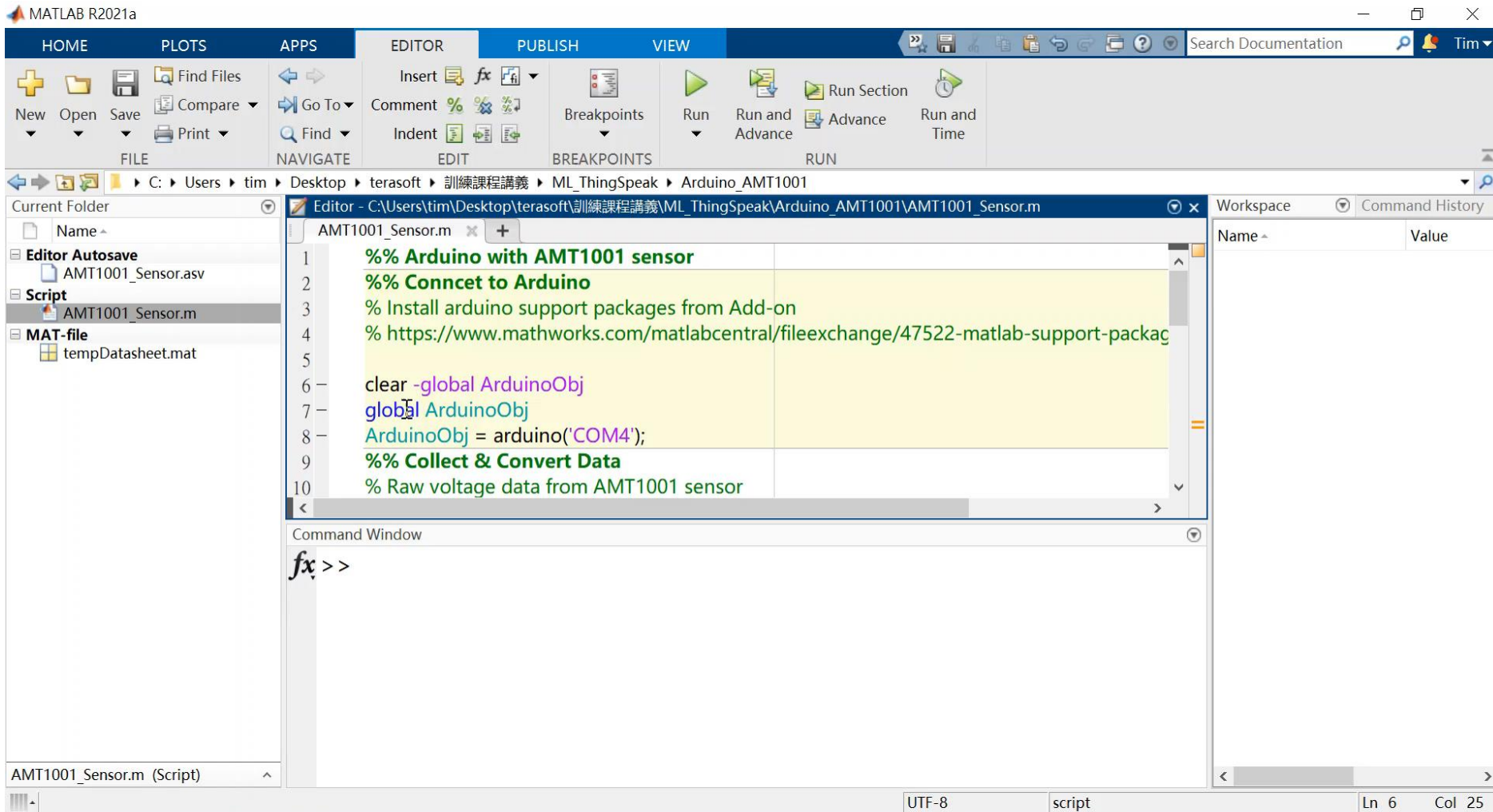


Simulink Support Package for Raspberry Pi Hardware

by MathWorks Simulink Team **STAFF**

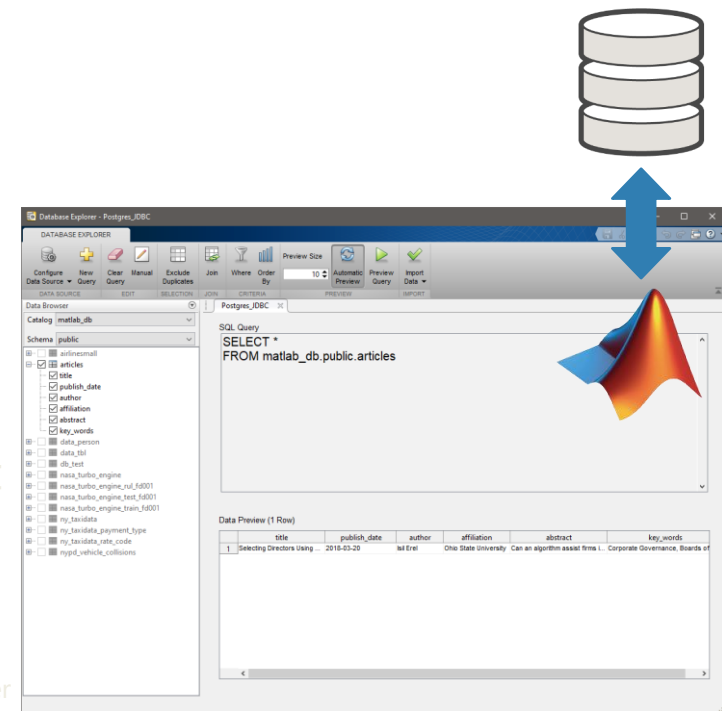
Run models on Raspberry Pi.

Arduino + AMT1001

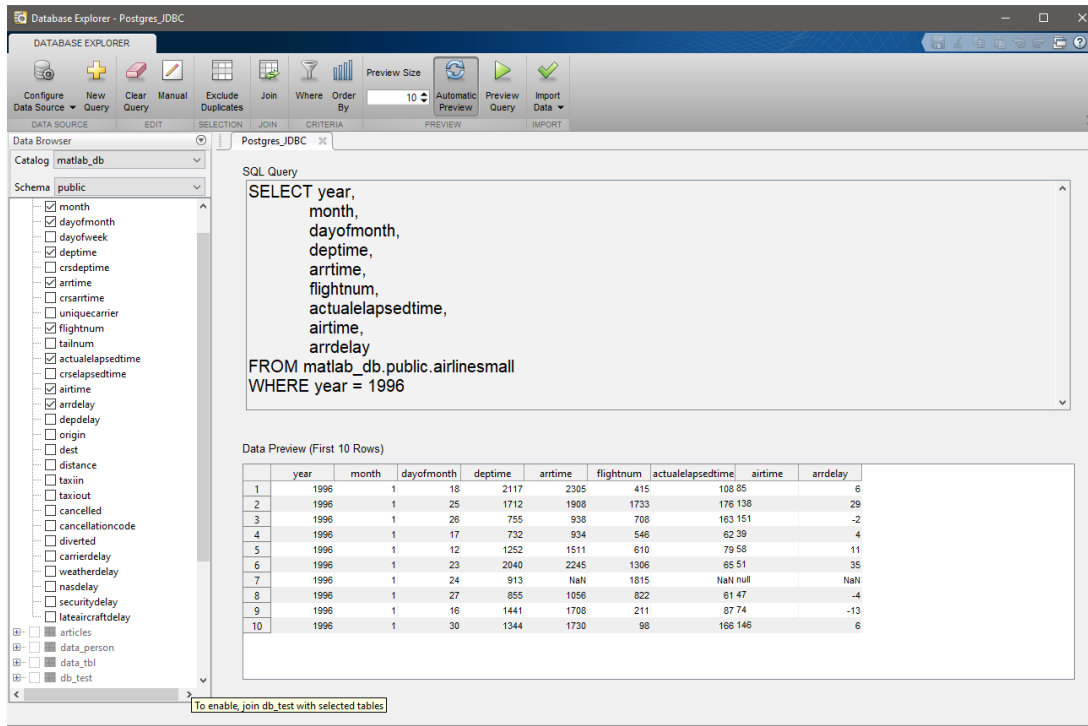


Database

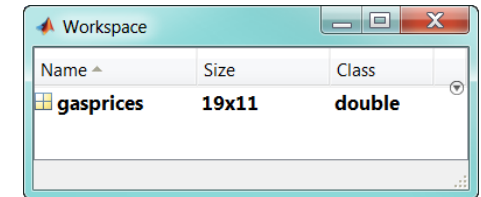
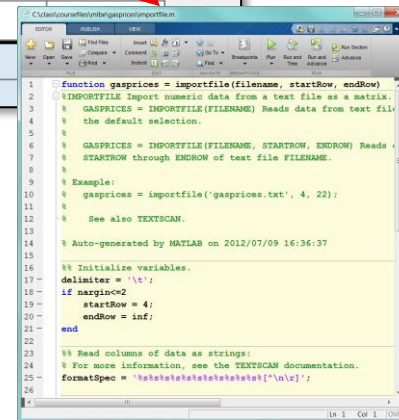
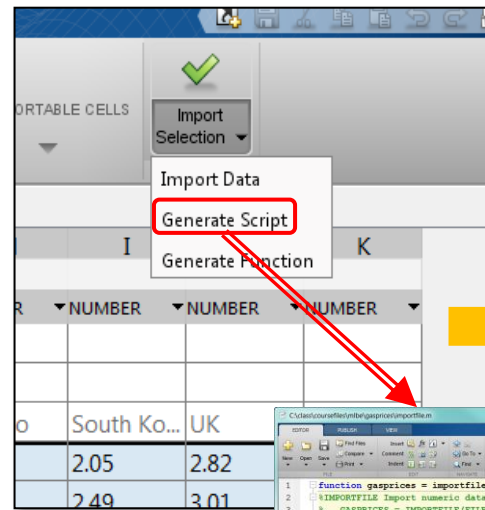
- Connection with relational DB (JDBC / ODBC compliant)
 - Simple command line without SQL
 - GUI access (Database Explorer app)
 - Connection with non-relational DB
 - Neo4j (Graph DB)
 - MongoDB (Document Oriented DB)
 - Apache Cassandra (column-oriented DB)
 - Reading and writing to SQLite files that do not require DB
 - Supports datastore for big data
 - SQL query partitioning, parallel access *
- ※ Required Parallel Computing Toolbox and MATLAB Distributed Computing Server



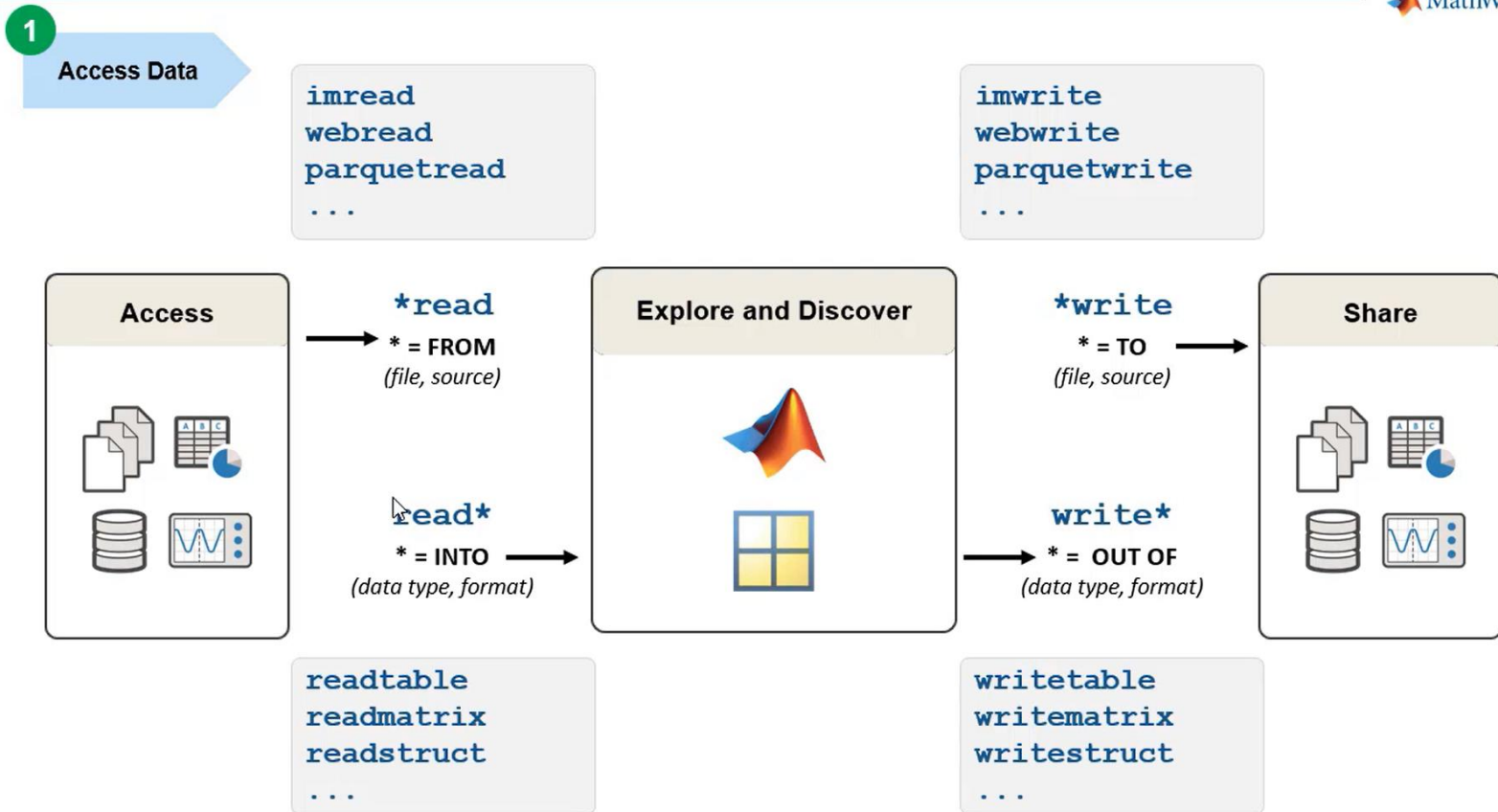
Database Explorer



可以在不了解SQL的情況下從GUI訪問和讀取數據
自動生成用於導入的SQL查詢語句和MATLAB腳本



Import Data



Web Crawling



Demo case – Dcard web crawler

 在美食板搜尋

所有看板

即時熱門看板

好物研究室

遊戲專區

即時熱門看板

感情

心情

閒聊

美妝

梗圖

女孩

美食

YouTuber

更多

Dcard 精選看板

 美食

熱門

最新

板規

追蹤

清水寺附近的親子丼

店名： ひさご親子丼，地址：京都市東山區下河原町484，在日本自助旅行時晃到了京都清水寺...

2

0

收藏

國立雲林科技大學 · 47 分鐘前

7-11 起司vs全家四種起司

最近這兩個實在有燒到我，因為這兩款比較類似 又不想踩雷，於是先上網查了評...

3

0

收藏

參阿甜 · 1 小時前

不甜不膩的白葡伯爵戚風蛋糕！

一直很想做這種清新風的蛋糕，剛好

4

0

收藏

參阿甜 · 1 小時前

不甜不膩的白葡伯爵戚風蛋糕！

一直很想做這種清新風的蛋糕，剛好

4

0

收藏

Web Data

RESTful Web Service

webread	Read data
webwrite	Write data
websave	Save data to file
weboptions	Specify options such as authentication and timeout

JSON

jsondecode	jsonencode
------------	------------

HTTP Messaging

Use the HTTP interface for more complex web communication: >>>

body = ...
matlab.net.http.MessageBody(x);

request = ...
matlab.net.http.RequestMessage ...
(method,header,body);

PLOTS

VARIABLE

VIEW

New from Selection

Open

Print

Rows

Columns

Insert

Delete

Transpose

Sort

VARIABLE

SELECTION

EDIT

dcardTable

330x10 table

	1	2	3	4	5	6	7	8	9
	id	title	excerpt	createdTime	updatedTime	gender	school	department	commentCount
1	235969282	清水寺附近的親子...	店名： ひさご親子丼，地...	2021年5月12日 ...	2021年5月12日 ...	"F"	"もちもち貝果"	"mochibagel"	0
2	235968985	"7-11 起司vs...	"最近這兩個實在有燒到我...	2021年5月12日 ...	2021年5月12日 ...	"F"	"國立雲林科技大..."	""	0
3	235968853	"不甜不膩的白葡...	"一直很想做這種清新風的蛋...	2021年5月12日 ...	2021年5月12日 ...	"F"	"參阿甜"	"abc12300901"	0
4	235968852	"不甜不膩的白葡...	"一直很想做這種清新風的蛋...	2021年5月12日 ...	2021年5月12日 ...	"F"	"參阿甜"	"abc12300901"	0
5	235968751	"#台中美食#手...	"哈囉！大家好~我們是怪...	2021年5月12日 ...	2021年5月12日 ...	"F"	"怪獸美食公司"	"monsterfood1..."	0
6	235968713	"火鍋料"	"想知道大家在火鍋裡面有什...	2021年5月12日 ...	2021年5月12日 ...	"F"	"seven"	"sevenseven777"	1
7	235968616	"初次手做可愛便...	"小班校外教學~這是0第一...	2021年5月12日 ...	2021年5月12日 ...	"F"	"Natalia"	"nataliaili"	0
8	235968605	"新北淡水 悠哉...	"淡江學生一定知道這家蛋...	2021年5月12日 ...	2021年5月12日 ...	"F"	"露露肚子餓"	"lu_801"	0
9	235968592	"都吃過的司康口..."	""	2021年5月12日 ...	2021年5月12日 ...	"M"	"國立雲林科技大學"	""	3

22

© Terasoft, Inc.

Demo case – Google comments crawler

← 所有評論

5

4

3

2

1

4.3

★★★★★

6,464 篇評論

撰寫評論

全部 景觀 565 下午茶 275 干貝 228 舒芙蕾 203 夜景 184 窗 180 電梯 147 羊排 98

Yvonne

在地嚮導 · 46 則評論

★★★★★ 3 週前

位在信義微風46樓的吃到飽餐廳
餐餐果然名不虛傳
假日下午茶一個人1290\$/晚餐1890\$
價格雖然很高 但用料實在
將自助餐分為約六區塊

Helper Function

```
61 function info = DataFormatting(row)
62     info.username = '';
63     info.day = '';
64     info.comment = '';
65     info.rate = '';
66
67     if ~isempty(row{1}{2})
68         info.username = row{1}{2};
```

搜尋關鍵字為: 饕饕

現在位置為: 24.8227918°N, 121.62299415°E

爬取搜尋內容: INPARADISE 饕饕

****開始爬取****

爬取第0-10篇貼文
爬取第10-20篇貼文
爬取第20-30篇貼文
爬取第30-40篇貼文
爬取第40-50篇貼文
爬取第50-60篇貼文
爬取第60-70篇貼文
爬取第70-80篇貼文
爬取第80-90篇貼文
爬取第90-100篇貼文
爬取第100-110篇貼文

****爬取結束****

爬取成功! 共爬取110篇貼文

ans = 8x4 table

	username	day	comm...	rate
1	'Yvonne'	'3 週前'	'位在信義微...	4
2	'Jimmy Wang'	'3 週前'	'人山人海 吃...	2
3	'Drodma Tharl...	'2 個月前'	'饕饕果然名...	4
4	'Austin Wu'	'1 個月前'	'食材種類非...	4
5	'Sean Huang'	'1 週前'	'餐點選擇多...	5
6	'Sunny Farmer'	'1 個月前'	'個人覺得是...	5
7	'曼蒂'	'1 個月前'	'2個月前就要...	5
8	'KaiDir'	'2 個月前'	'週四晚上六...	5

資料的類型



有點像專有名詞與量詞

屬性資料 (Qualitative Data)

不以數值來表示，以類別來區分資料，又稱為『類別資料』(Categorical Data)。如性別、學歷、公司、國家等等...

屬量資料 (Quantitative Data)

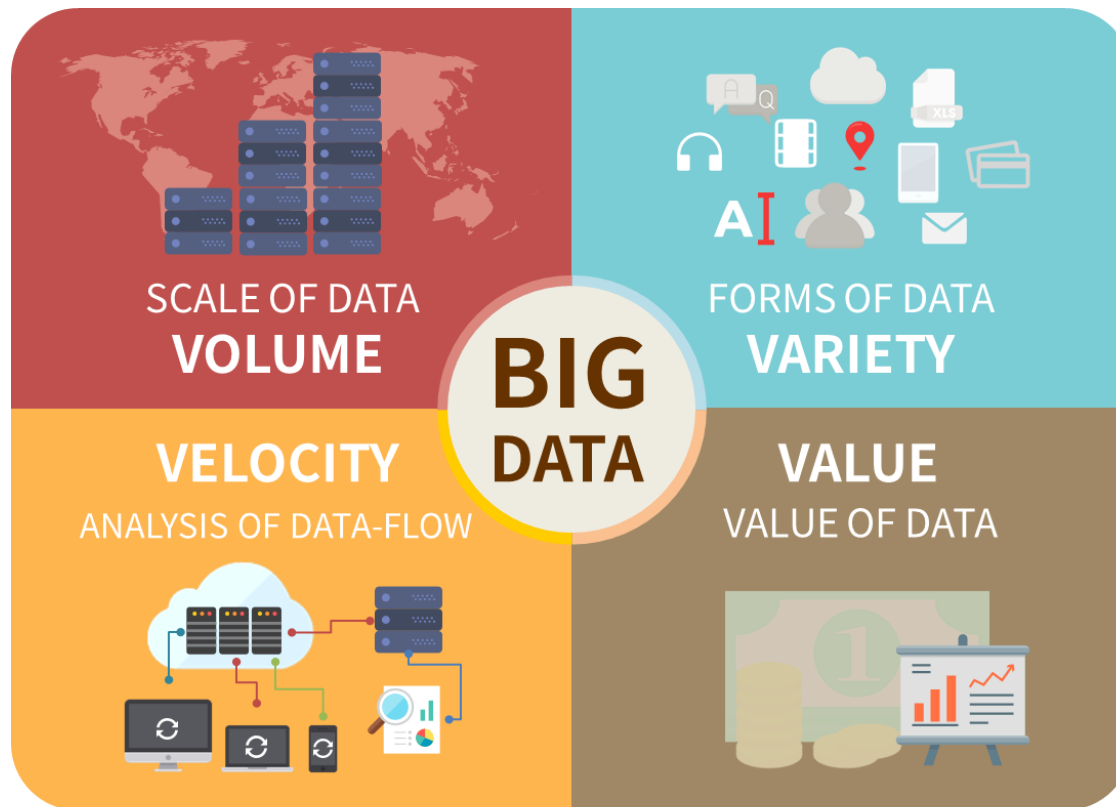
離散資料 (Discrete Data)

屬於可數的屬量資料，
例如：人數，機器台數等

連續資料 (Continuous Data)

在任兩數值間可插入無限多數值之資料一般，
例如：身高、體重等

資料的量級也跟著擴展



https://www.iconfinder.com/icons/1889193/analytics_chart_finance_financial_financial_report_flipchart_report_icon

Datastore

- Datastore 如何存取資料? 如何處理?
 - Datastore 如何存取資料? 如何處理?



Datastore

- Datastore 如何存取資料? 如何處理?
 - Datastore 如何存取資料? 如何處理?



Read portion



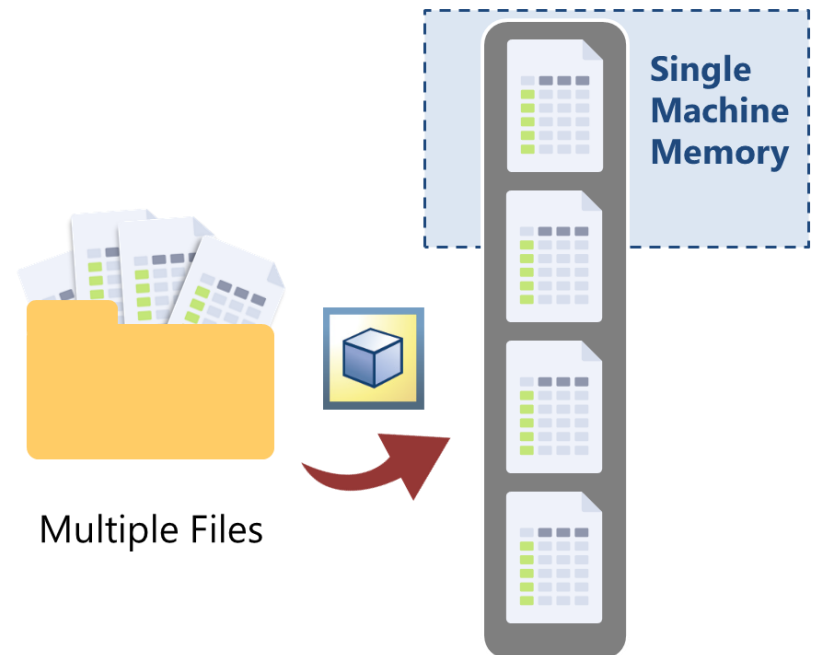
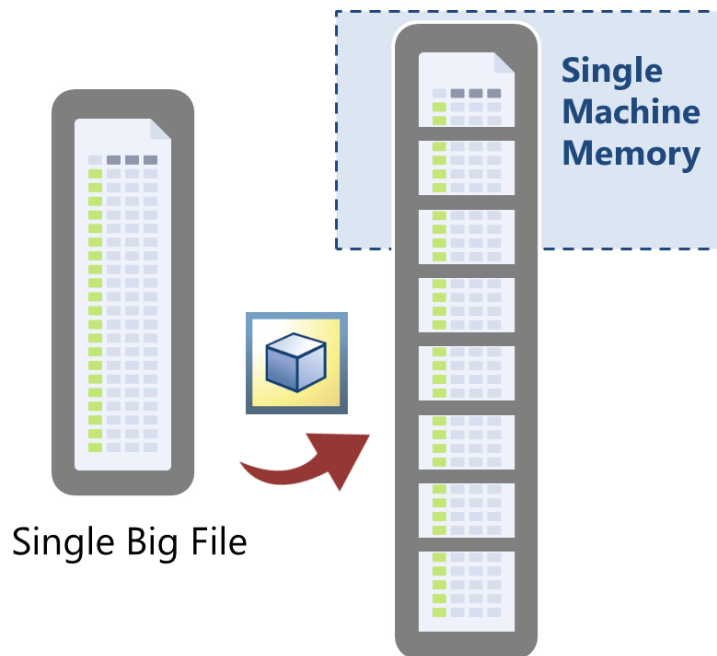
Block processing



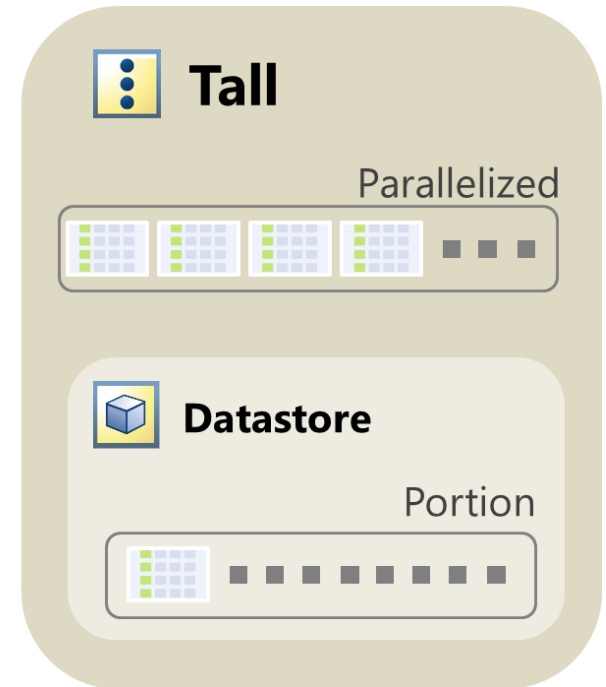
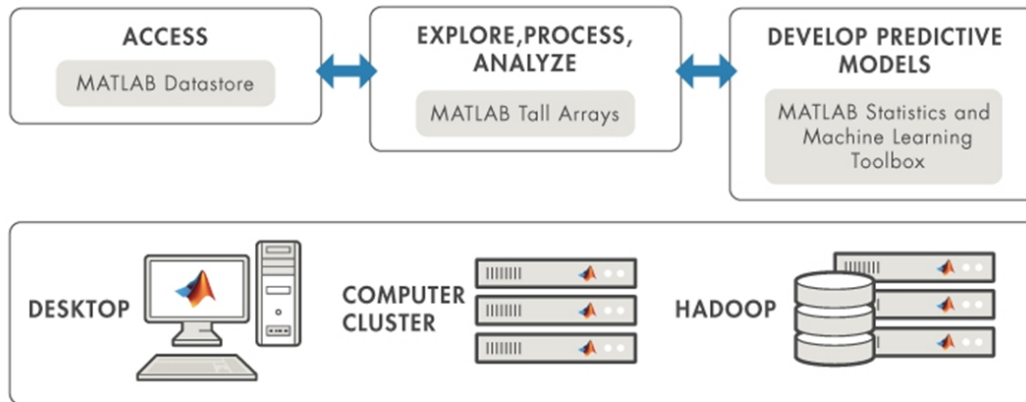
Stream processing

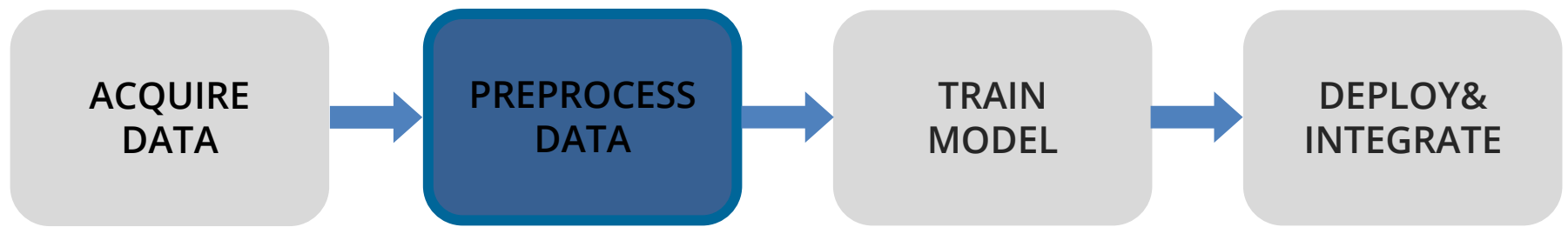
<http://www.gnu.org/licenses/gpl-3.0.html>

處理不同的狀況



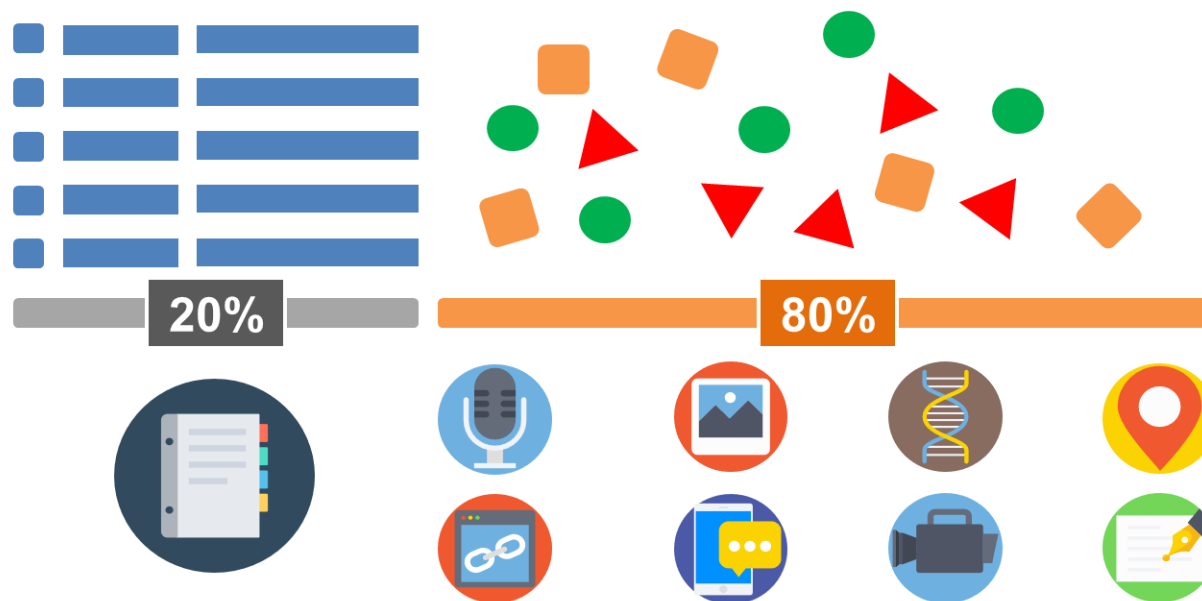
Tall – 結合平行運算的 Datastore





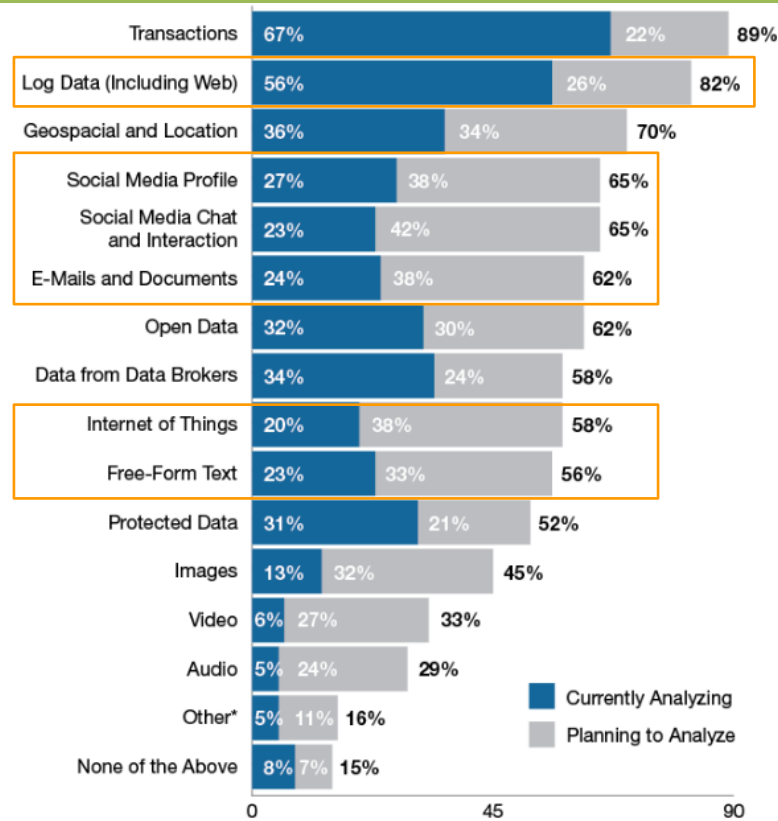
資料爆炸的時代中資料的儲存格式

- 20% 結構性資料，80% 半/非結構性資料



[href="https://www.flaticon.com/authors/prosymbols"](https://www.flaticon.com/authors/prosymbols)

Data Sources



n=195
 Source: Nick Heudecker, "Best Practice for Designing Your Data Lake," Gartner, 19 October 2016.
 *Other includes genomic, meteorological, traffic, security, and customer data.
 © 2016 Gartner, Inc. and/or its affiliates. All rights reserved.
 Gartner is a registered trademark of Gartner, Inc. or its affiliates.

Gartner

然後他就結構了



批踢踢實業坊 > 看板 Gossiping

看板 精華區 最新 (上頁) (下頁) 最新

搜尋文章...

- 4 [問卦] 有沒有義大利政局不穩的八卦? IslamicState 8/21 ...
- 2 [問卦] 遇到盜獵集團該如何自保? cychine 8/21 ...
- 4 [問卦] 小鳥被吹傷該怎麼辦? 在線等 ilikebrandy 8/21 ...
- [問卦] 肥宅我會不會練的太狀啊 malaman 8/21 ...
- [問卦] 請問除了嘴癢 喉嚨痛的舌苔那裏買 wiwi0526 8/21 ...

PTTCrawl.mlx | Final_table = 13x4 t...

	Reply	Date	Author	Title
1	" 4 "	"8/21"	"IslamicState"	" [問卦] 有...
2	" 2 "	"8/21"	"cychine"	" [問卦] 遇...
3	" 6 "	"8/21"	"ilikebrandy"	" [問卦] 小...
4	" "	"8/21"	"malaman"	" [問卦] 肥...
5	" "	"8/21"	"wiwi0526"	" [問卦] 讀...
6	" "	"8/21"	"go190214"	" [新聞] 公...
7	" "	"8/21"	"roanmaus"	" [問卦] 夠...
8	" "	"8/21"	"Scion"	" Re: [新聞]...
9	" "	"8/21"	"bengowa"	" [問卦] 違...

[半/非結構性資料]

非結構的純文字或是半結構的網頁原始碼

[結構性資料]

整理過後，欄位定義明確的表格資料

Data preprocessing with Live Editor tasks

The screenshot displays the MATLAB Live Editor interface for a project titled "Data Analytics - Load Forecasting Case Study". The main workspace shows a script with the following code:

```
load LETdata.mat
head(nyiso)
```

A "Missing Data" task window is open, showing the "Clean Missing Data" section. The "Input data" is set to "nyiso". The "Cleaning method" is "Fill missing". A dropdown menu for "Method to fill missing entries" is open, listing several options: Constant value, Previous value, Next value, Nearest value, Linear interpolation (selected), Spline interpolation, Shape-preserving cubic interpolation (PCHIP), Modified Akima cubic interpolation, Moving median, Moving mean, and Linear interpolation. The "Visualize results" section has checkboxes for "Cleaned data" and "Filled missing entries", both of which are checked.

On the right side of the interface, the output of the script is displayed as an 8x11 timetable:

	Date	CAPITL	CENT
1	05/01/2007 ...	981.9000	1.571
2	05/01/2007 ...	991.8000	1.568
3	05/01/2007 ...	950.1000	1.560
4	05/01/2007 ...	968.9000	1.560
5	05/01/2007 ...	968.5000	1.555
6	05/01/2007 ...	949.2000	1.564
7	05/01/2007 ...	941.6000	1.538
8	05/01/2007 ...	939.4000	1.557

Below the table, a line plot titled "Number of filled missing entries: 4" shows the "Cleaned data" (blue line) and "Filled missing entries" (red dots) over time from July 2007 to April 2008. The y-axis ranges from 200 to 1600.

Data preprocessing in Text

`"Items are occasionally getting stuck in the scanner spools."`

tokenizedDocument

`10 tokens: Items are occasionally getting stuck in the scanner spools`

removeStopWords

`7 tokens: Items occasionally getting stuck scanner spools .`

normalizeWords

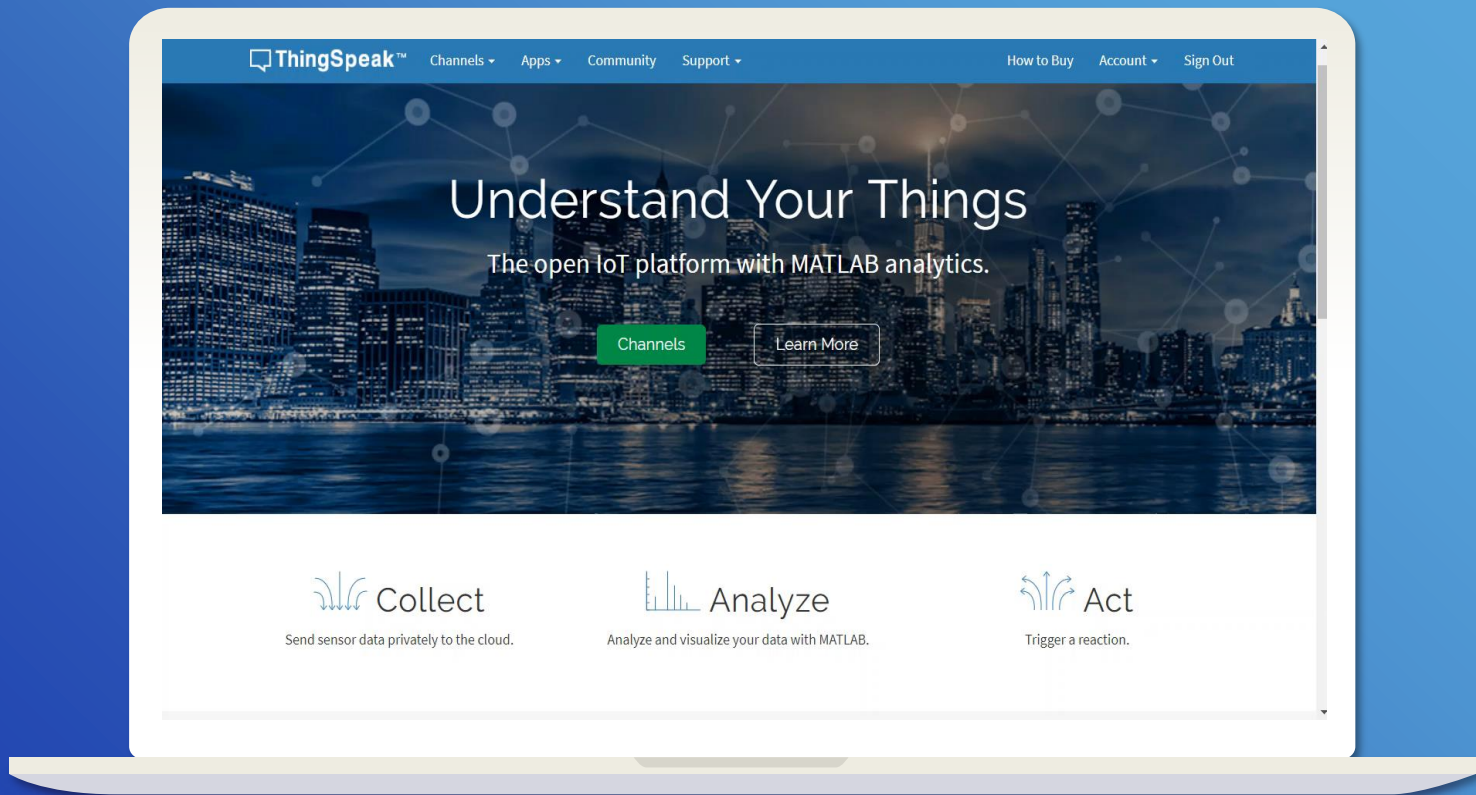
`7 tokens: items occasionally get stuck scanner spool .`

erasePunctuation

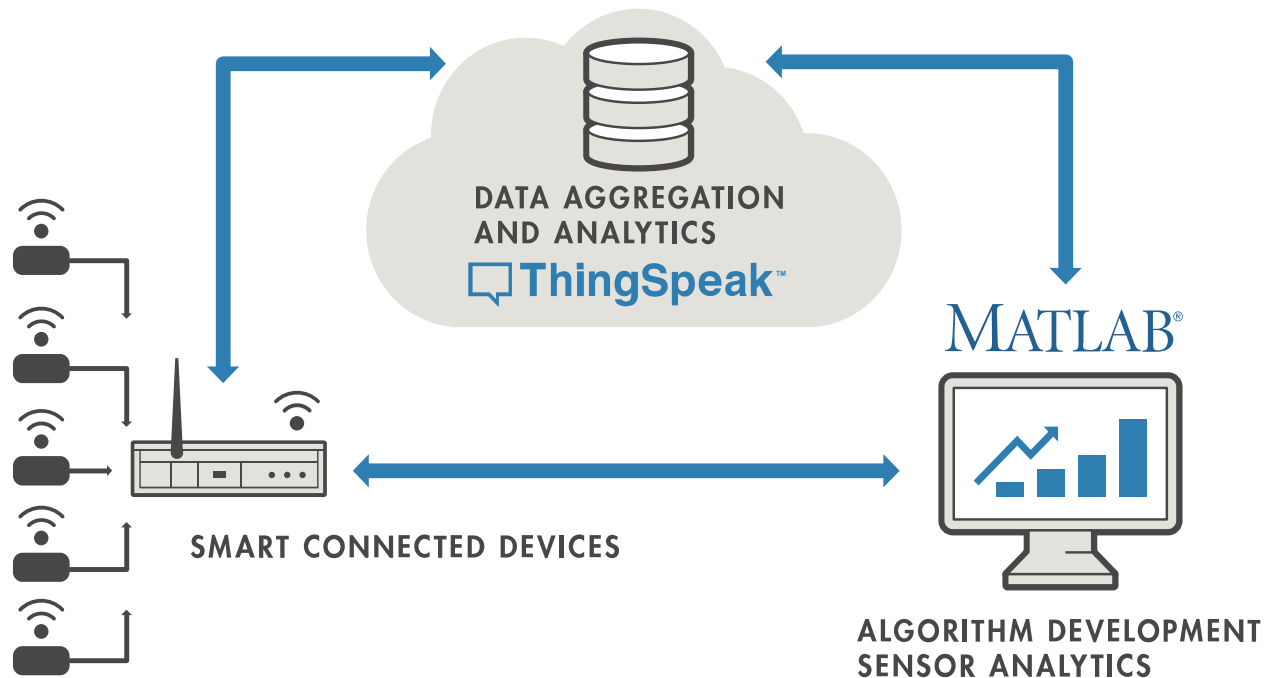
`6 tokens: items occasionally get stuck scanner spool`

雲端儲存平台-ThingSpeak

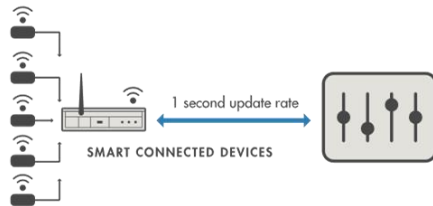
ThingSpeak



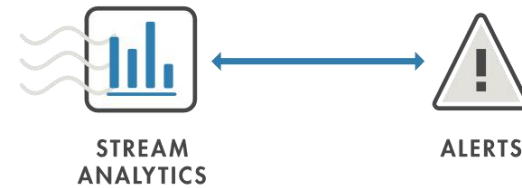
AIoT Environment in MATLAB



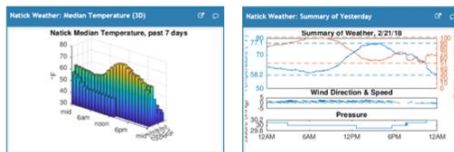
ThingSpeak Capability



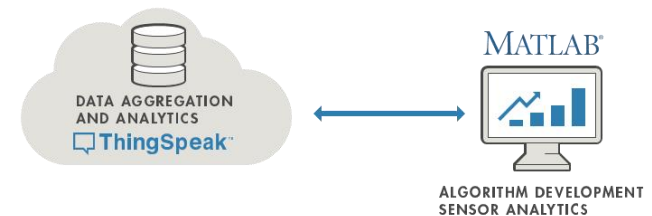
- 最高可更新速度 1 筆/秒
- 可同時更新多個 Channel



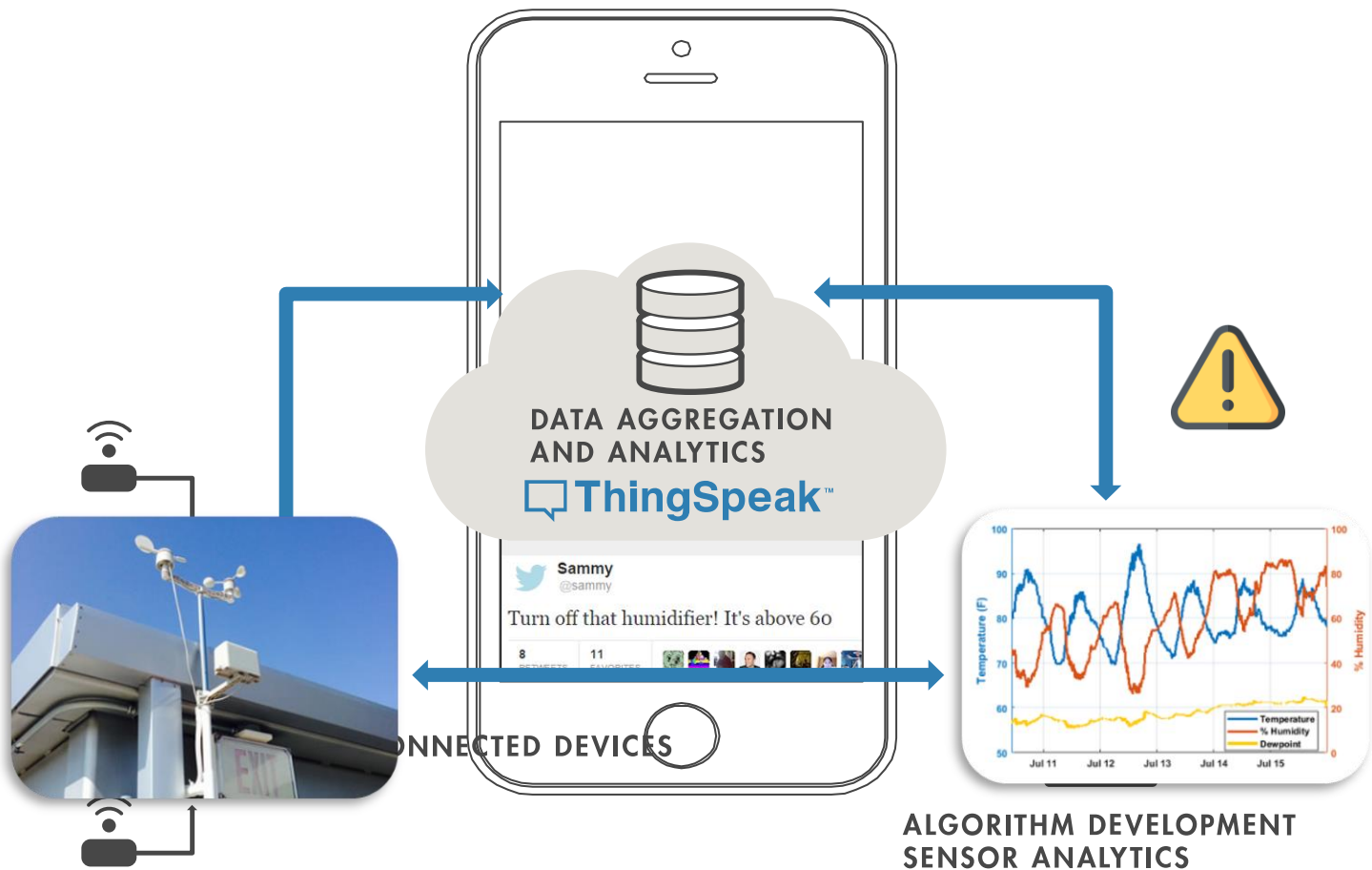
- 結合第三方應用程式推播



- 以MATLAB語法撰寫視覺化呈現

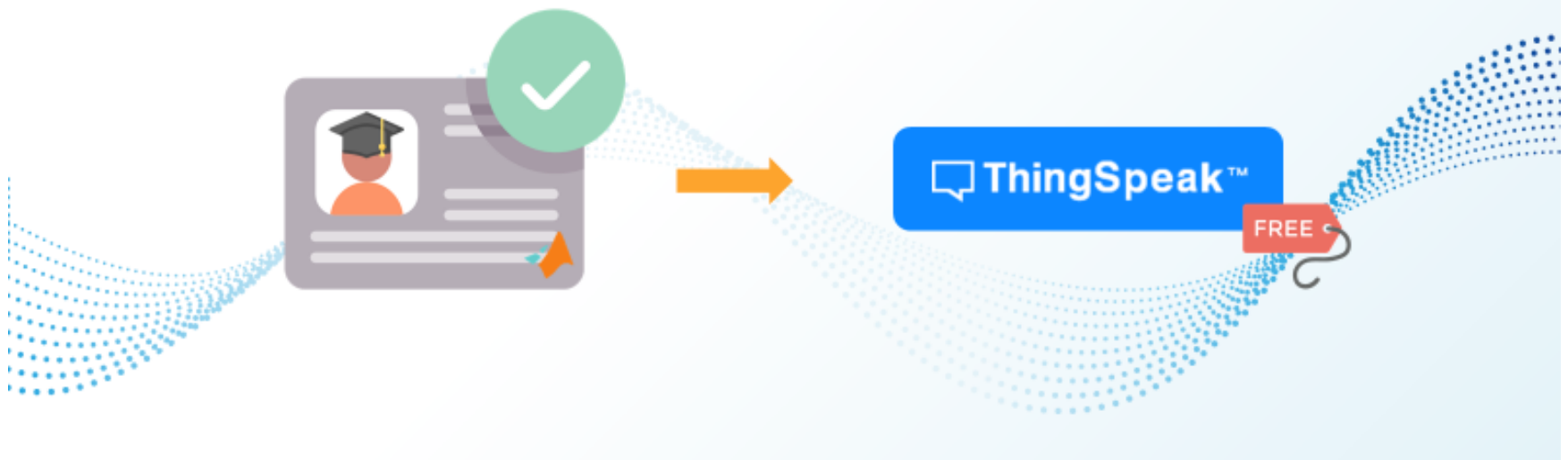


- 上傳或下載成 CSV · JSON 等檔案格式
- 在平台上以 MATLAB 撰寫資料分析演算法

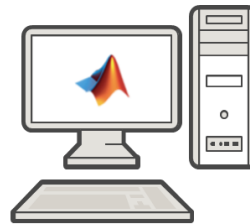
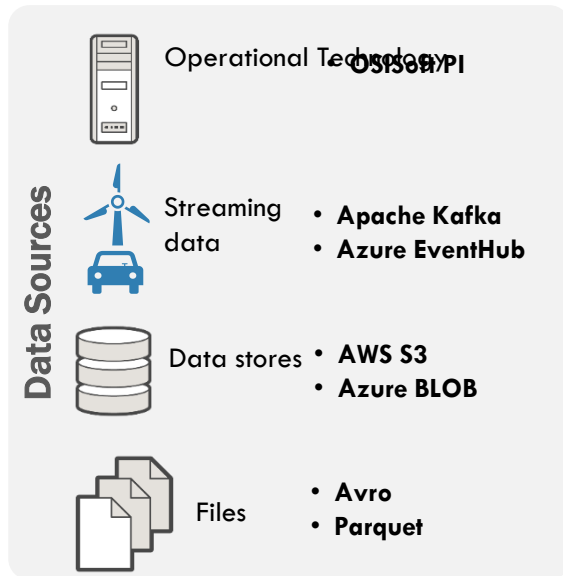


校園授權免費使用

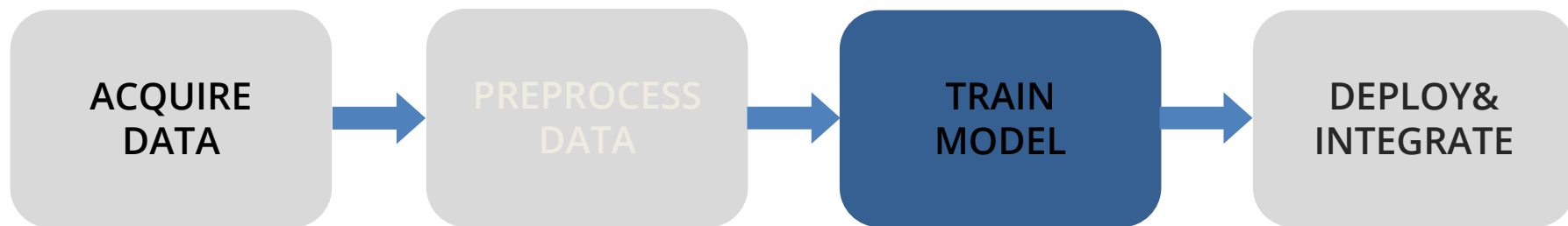
**Free Use with
Campus Wide License
(CWL)**



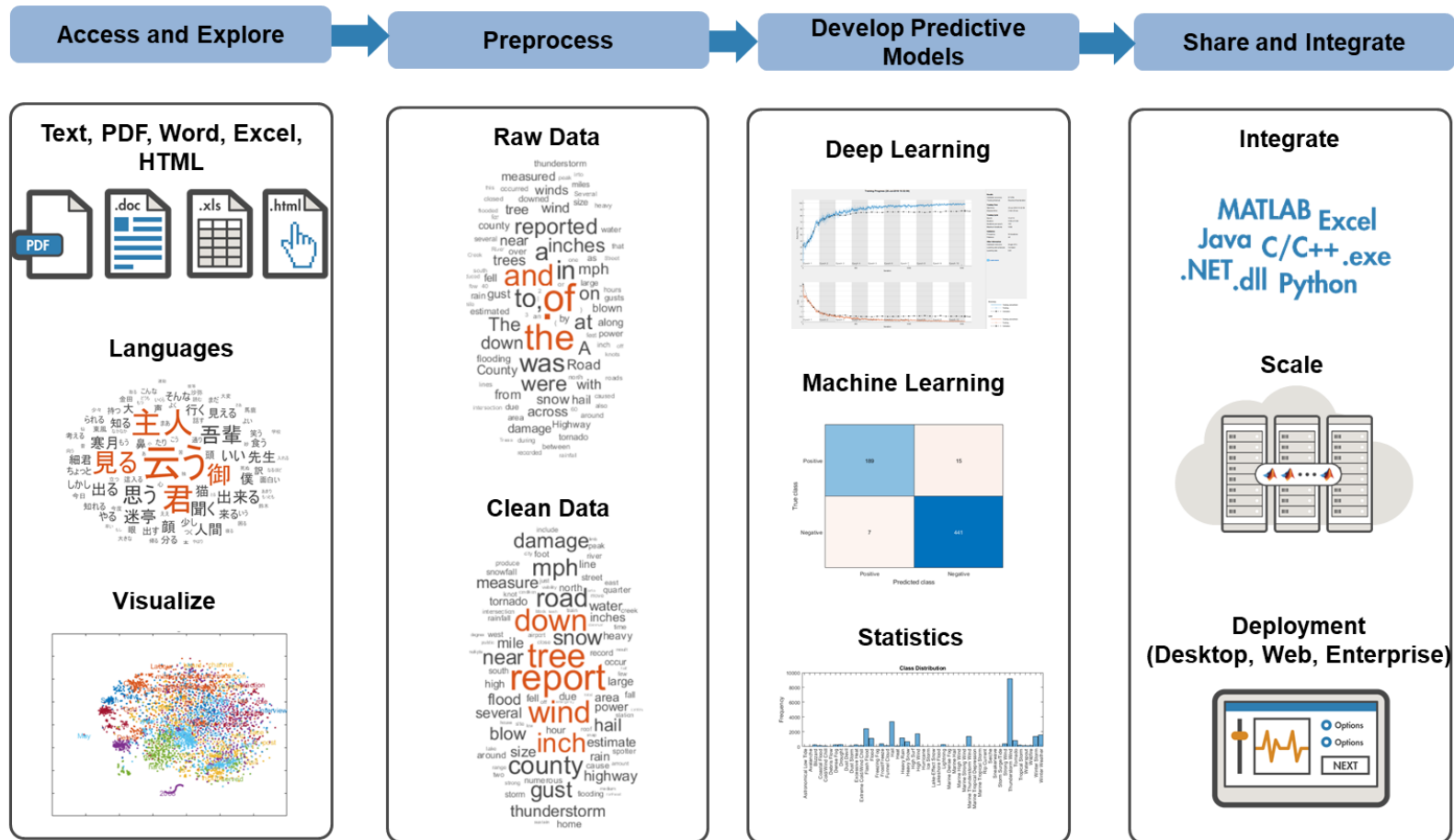
Public Cloud Storage



Scale to high-performance clusters using **MATLAB Parallel Server™**, and incorporate **MATLAB analytics** into applications with **MATLAB Production Server™**.



Text Analytics Workflow



Text Analytics Applications

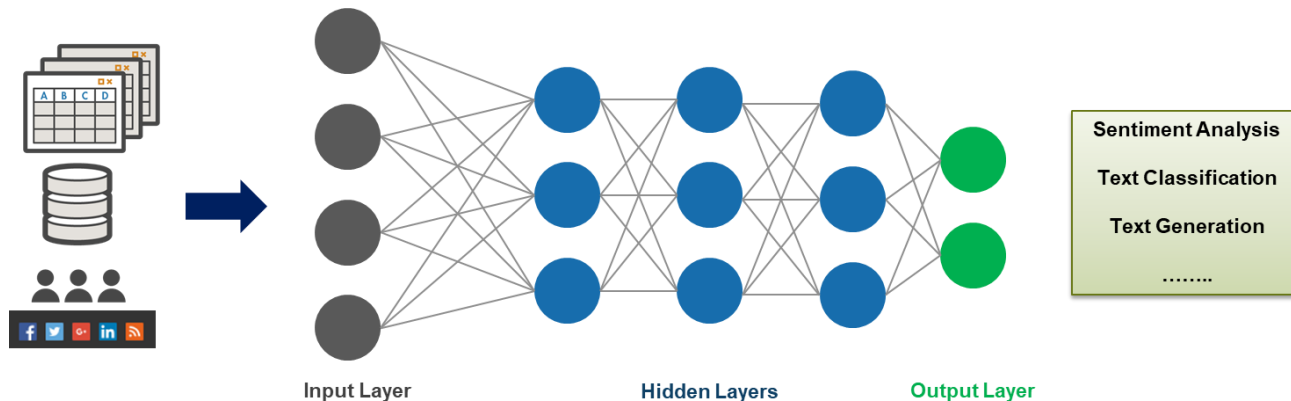


Text Analytics or Text Mining



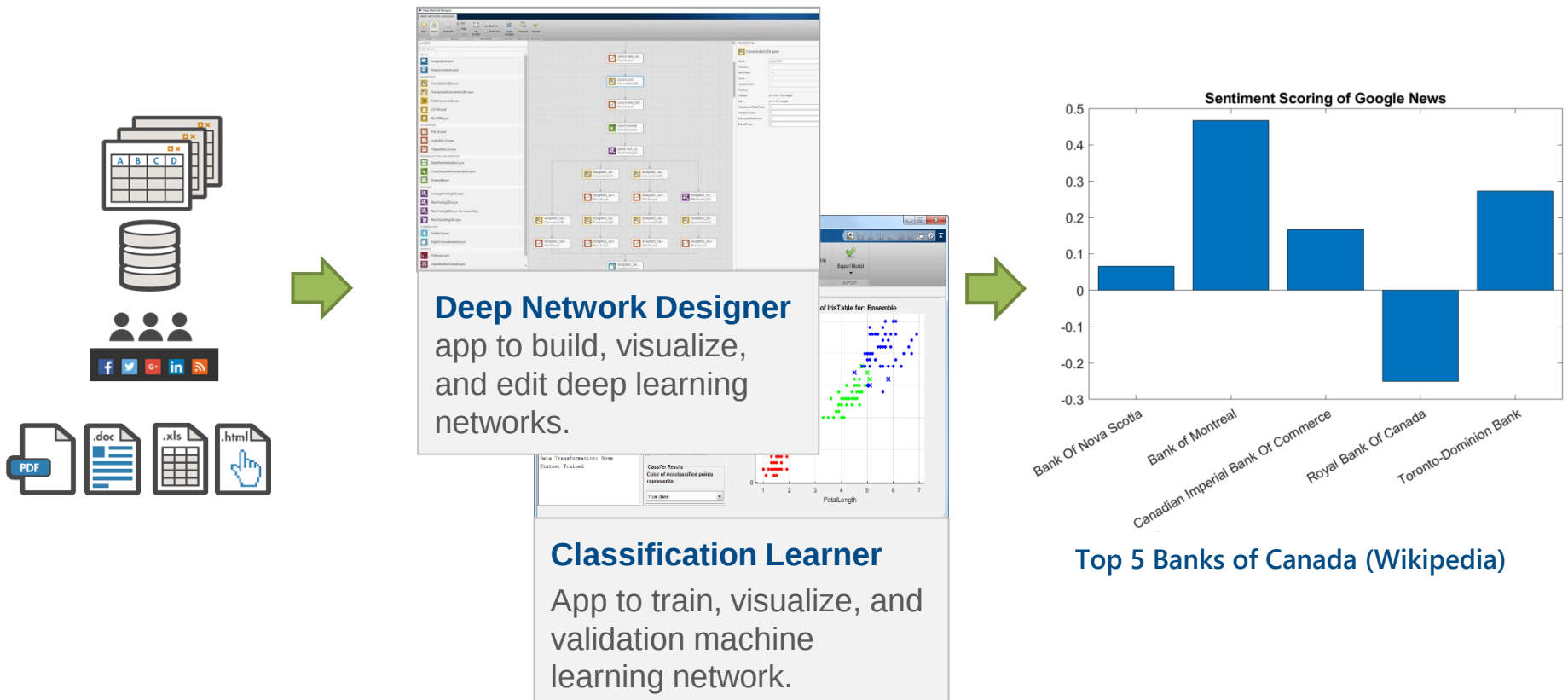
= Natural Language Processing + Machine Learning

Deep Learning and Text Analytics



More than 3 layers (including input & output) → Deep Learning

Train with APP



CONTENTS

- « Documentation Home
- « Examples
- « Text Analytics Toolbox

Category

- Get Started with Text Analytics Toolbox 6
- Text Data Preparation 9
- Modeling and Prediction 23**
- Display and Presentation 2
- Language Support 2

All Examples Functions

Modeling and Prediction — Examples

Training Data

Classify Text Data Using Deep Learning

Classify text data using a deep learning long short-term memory (LSTM) network.

[Open Live Script](#)

Text Data: Preprocessed Bigrams

Analyze Text Data Using Multiword Phrases

Analyze text using n-gram frequency counts.

[Open Live Script](#)

Topic 1

Topic 2

Topic 3

Topic 4

Analyze Text Data Using Topic Models

Use the Latent Dirichlet Allocation (LDA) topic model to analyze text data.

[Open Live Script](#)

Raw Data

Cleaned Data

Create Simple Preprocessing Function

Create a function which cleans and preprocesses text data for analysis.

[Open Live Script](#)

Class Distribution

Create Simple Text Model for Classification

Text Data: Preprocessed Bigrams

Analyze Text Data Using Multiword Phrases

Topic 1

Topic 2

Topic 3

Topic 4

Analyze Text Data Using Topic Models

Validation Accuracy

Choose Number of Topics for LDA Model

GitHub interface showing the repository **matlab-deep-learning / transformer-models**. The repository is described as "Deep Learning Transformer models in MATLAB" and has 42 stars and 13 forks.

The repository structure includes the following files and folders:

File/Folder	Description	Last Commit
+bert	slightly better fix, use mustBeNumericOrLogical and convert to logica...	last month
+finbert	finbert src	2 months ago
+gpt2	merge common changes	
+sampling	First commit	
+transformer/+layer	fix typo, closes #9	
.circleci	ok try using the default - suppose 21	
test	slightly better fix, use mustBeNumericOrLogical and convert to logica...	
.gitignore	add gitignore	
ClassifyTextDataUsingBERT.m	readme and example updates	
FineTuneBERT.m	readme and example updates	
PredictMaskedTokensUsingBERT.m	updates to readme and examples	
PredictMaskedTokensUsingFinBERT.m	updates to readme and examples	

The repository is currently on the **master** branch, with 2 branches and 3 tags. The last commit was made 18 days ago by **ea9fed4**, with 53 commits in total.

Repository statistics:

- Contributors: 2
- Issues: 0
- Stars: 42
- Forks: 13

The repository was last updated 2 months ago.

<https://github.com/matlab-deep-learning/transformer-models>

Demo case: Summarize from BBC news

C:\Users\tim\Desktop\ArticleSummary.mlx

LIVE EDITOR INSERT VIEW

New Open Save Print Export FILE

Go To Find Bookmark NAVIGATE

Text B I U M TEXT

Code Control Task CODE

Refactor Run Section Run and Advance Run Run to End SECTION

Run Step Stop RUN

Summarize BBC News Using GPT-2 and Extract Keywords Using RAKE Model

BBC Sign In Home News Sport Reel Worklife Trz

NEWS

Home Coronavirus Video World Asia UK Business Tech Science Stories Entertainment & Arts

Business Market Data New Economy New Tech Economy Companies Entrepreneurship Technology o Economy Global Car Industry

Cryptocurrency: Musk's SpaceX to launch dogecoin moon mission

3 days ago Comments



從網頁讀取HTML純文字檔

url = 'https://www.bbc.com/news/business-57052811':

```
source =
16x1 htmlTree:

<P><B class="ssrcss-hmf8q1-BoldText e5tfeyi3">Elon Musk has said his rocket company SpaceX will now accep
<P>In a tweet he said the commercial space exploration firm will launch the "DOGE-1 Mission to the Moon"
<P>It came as dogecoin continued to fall after the technology billionaire said on television that it was
<P>Last month, he tweeted that SpaceX was going to put a "literal Dogecoin on the literal moon".</P>
<P>Canada's Geometric Energy Corporation announced the dogecoin-funded mission earlier on Sunday, althoug
<P>"This mission will demonstrate the application of cryptocurrency beyond Earth orbit and set the founda
<P lang="en" dir="ltr">SpaceX launching satellite Doge-1 to the moon next year<BR/><BR/>♦♦ Mission paid
<P>Mr Musk's tweets in recent months have helped to turn the once-obscure digital currency, which was sta
<P>It has soared by more than 800% over the last month, with its total market value rising above $70bn, a
<P>However, dogecoin lost more than a third of its price on Sunday, after Mr Musk called it a "hustle" wh
<P>In February, electric carmaker Tesla, where Mr Musk is chief executive, said it bought $1.5bn worth of
<P>The announcement was seen as another major step towards the mainstream acceptance of bitcoin and helpe
<P>Last month, Tesla announced that its profits for the first three months of the year were $438m, up fro
<P>Mr Musk also revealed on Saturday Night Live that <A href="https://www.bbc.co.uk/news/world-us-canada-
<P>People with Asperger's interpret the environment around them differently to other people.</P>
<P>It is thought to be the first time Mr Musk has spoken about his condition.</P>

str = 16x1 string
Rows 3:12
"It came as dogecoin continu...
>Last month, he tweeted that...
>Canada's Geometric Energy C...
>"This mission will demonstr...
>SpaceX launching satellite ...
>Mr Musk's tweets in recent ...
>It has soared by more than ...
>However, dogecoin lost more...
>In February, electric carma...
>The announcement was seen a...

inputText =
'Elon Musk has said his rocket company SpaceX will now accept the meme-inspired cryptocurrency dogecoin a

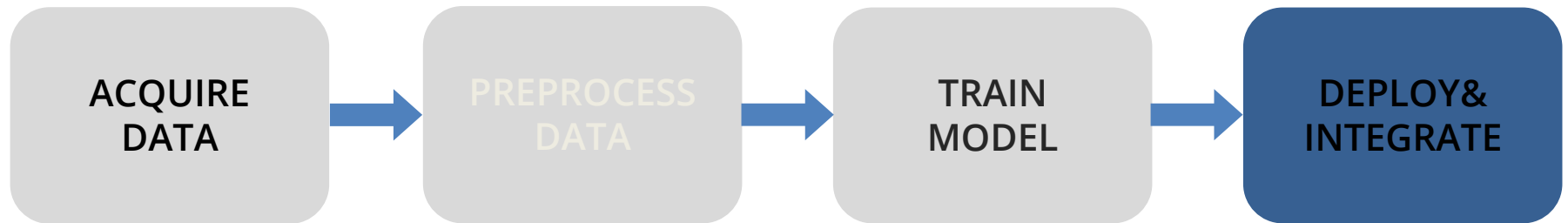
♦♦ Mission paid for in Doge
♦♦ 1st crypto in space
♦♦ 1st meme in space

To the moooooon!! Mr Musk's tweets in recent months have helped to turn the once-obscure digital curren

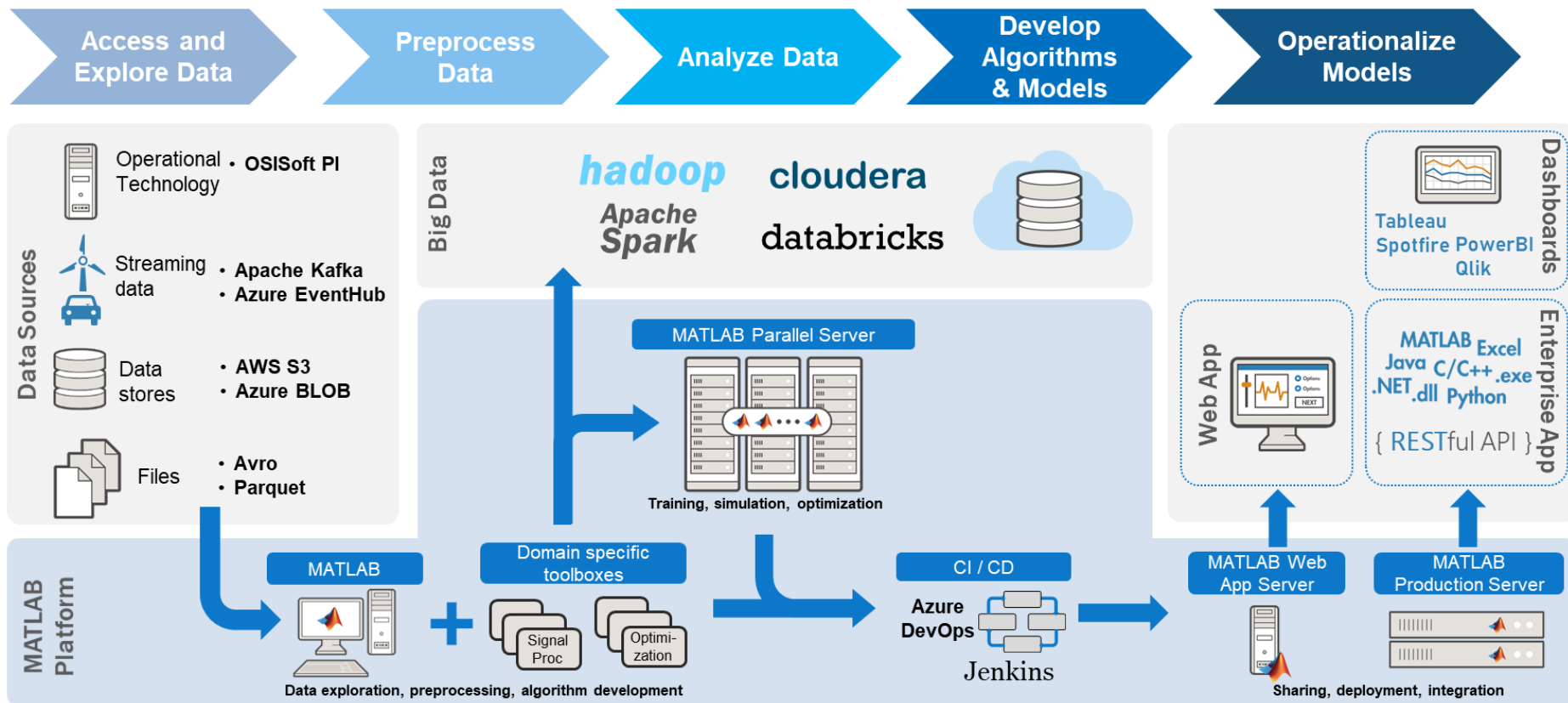
summary = ' : Dogecoin, the cryptocurrency that has become the darling of the internet, is now accepted by the
```

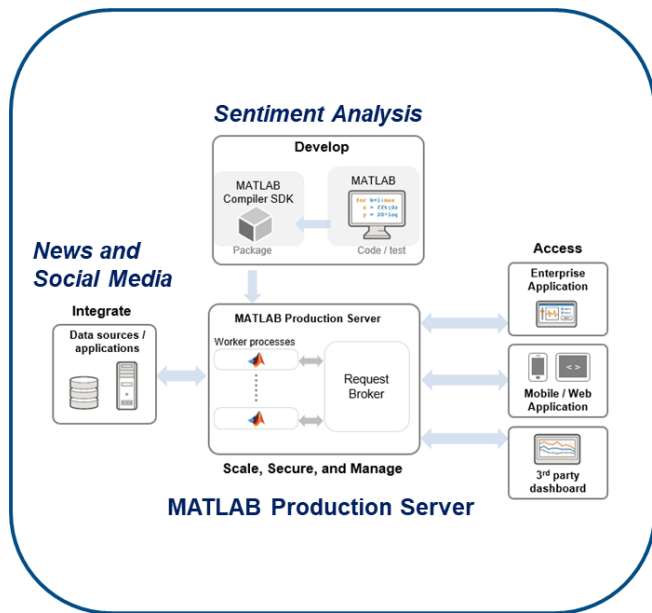
UTF-8 LF script

Process workflow



MathWorks end-to-end solution

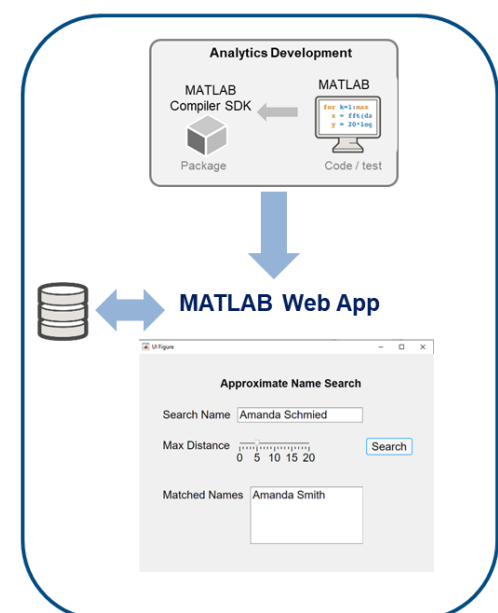




Identify Potential Issues



Web Activity Monitoring



Approximate Name Matching



最專業的MATLAB技術支援及服務團隊－鈦思科技

台北：台北市忠孝東路六段21號8樓之3 Tel：(02)2788-9300

新竹：竹北市復興一街251號13樓之 6 Tel：(03)550-5590