เทคโนโลยีสารสนเทศสำหรับการจัดการข้อมูล

เจษฎา สกุลคู

ศูนย์ข้อมูลก๊าซเรือนกระจก

องค์การบริหารจัดการก๊าซเรือนกระจก (องค์การมหาชน)

เนื้อหาการบรรยาย

- เทคโนโลยีสารสนเทศ
- การจัดการข้อมูลและการดำเนินงานขององค์กร
- การพัฒนากระบวนงานเพื่อนำเทคโนโลยีสารเทศเข้าไปใช้ในสนับสนุนการดำเนินงานขององค์กร
- เทคโนโลยี
  - Database
  - Data Warehouse
  - Data Mining
  - Big Data

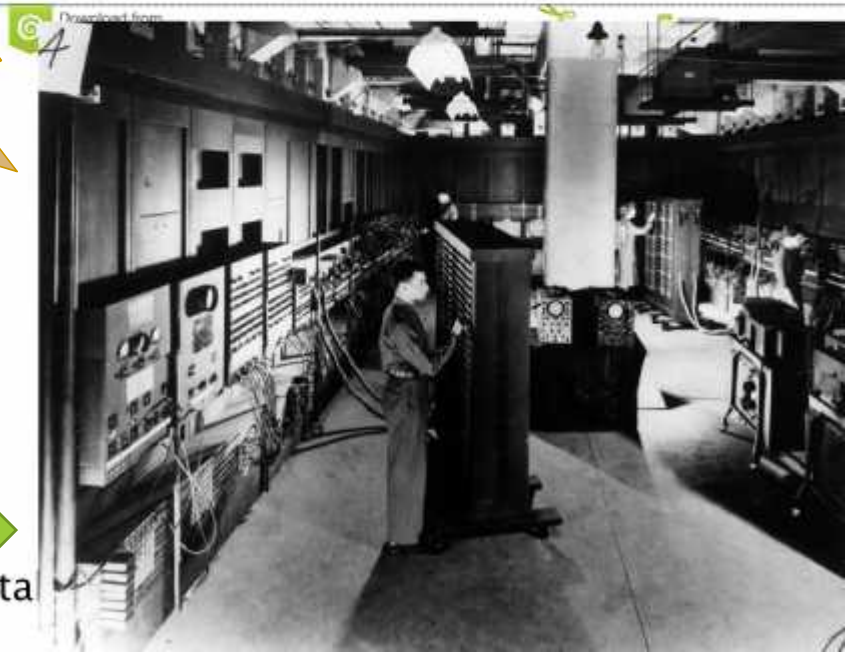
- Other Technology Trend
  - Internet of Thing (IoT)
  - Artificial Intelligent (AI)

For war purposes, constructive development after WWII

- การปฏิวัติอุตสาหกรรม
- การพัฒนาด้าน Operation Management

  เพื่อการพัฒนา–เพิ่มประสิทธิภาพด้านอุตสาหกรรมการผลิต

ENIAC
calculate artillery firing ta



การบริหารข้อมูล – เทคโนโลยีสารสนเทศ

Information technology ทำให้เกิดการเปลี่ยนแปลงขององค์กร
…from automation to paradigm shift…

1. Improve Operating Efficiency
2. Improve the Communication
3. Improve Decision Making Processes

Information System Components

Hardware
Software

การพัฒนากระบวนงานเพื่อนำเทคโนโลยีสารเทศเข้าไปใช้ในสนับสนุนการดำเนินงานขององค์กร

System analysis

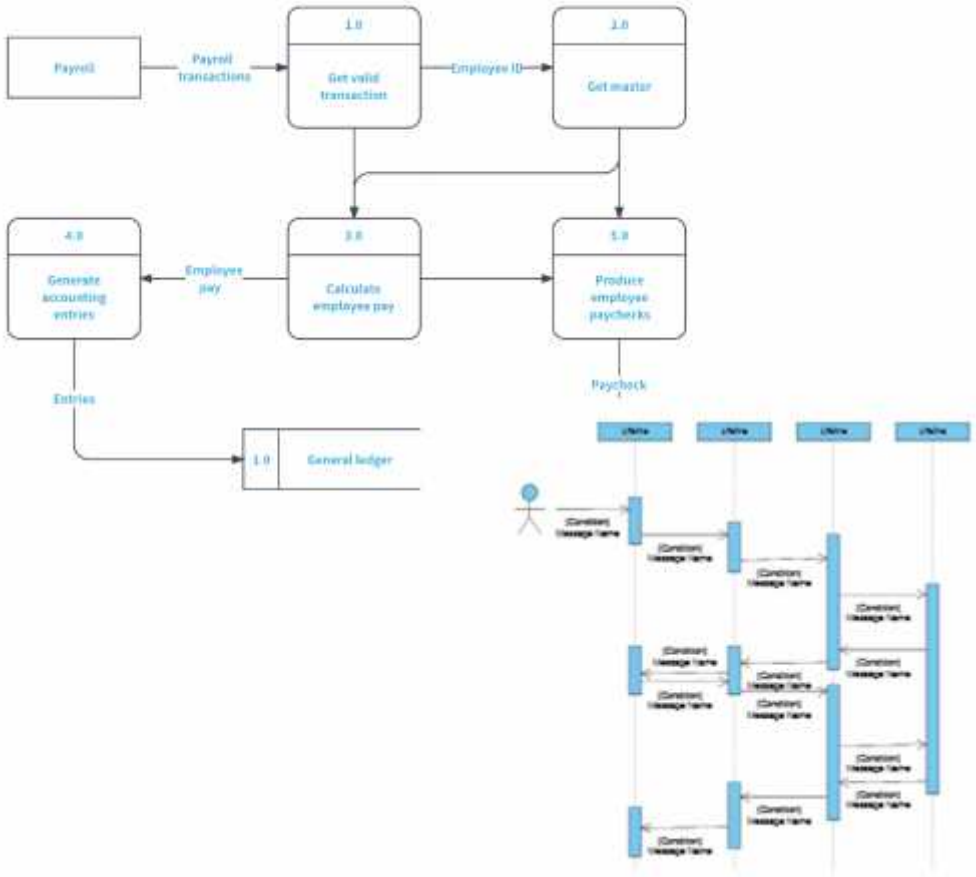
The development of a computer-based information system includes a system analysis phase. This helps produce the data model, a precursor to creating or enhancing a database. There are a number of different approaches to system analysis. When a computer-based information system is developed, system analysis (according to the Waterfall model) would constitute the following steps:

The development of a feasibility study: determining whether a project is economically, socially, technologically and organizationally feasible

Fact-finding measures, designed to ascertain the requirements of the system's end-users (typically involving interviews, questionnaires, or visual observations of work on the existing system)

Gauging how the end-users would operate the system (in terms of general experience in using computer hardware or software), what the system would be used for and so on

SYSTEM ANALYSIS & DESIGN

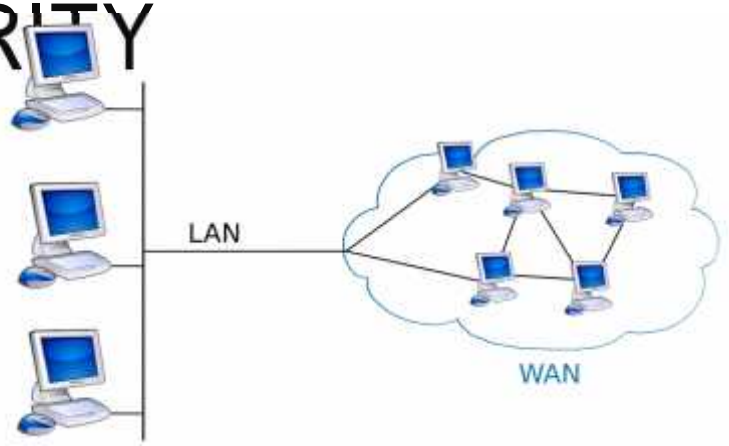


- The works are the flow of Data through processes.

แนวคิด : การทำงานคือ กระบวนการไหลของข้อมูลผ่านกระบวนงาน

- Issue : How process can be digitalized?
- Data Flow Diagram (DFD) : เครื่องมือในการออกแบบระบบงาน การพัฒนาโปรแกรม และการพัฒนาฐานข้อมูล
- UML : เครื่องมือในการออกแบบ Business Flow Process

COMMUNICATION & SECURITY

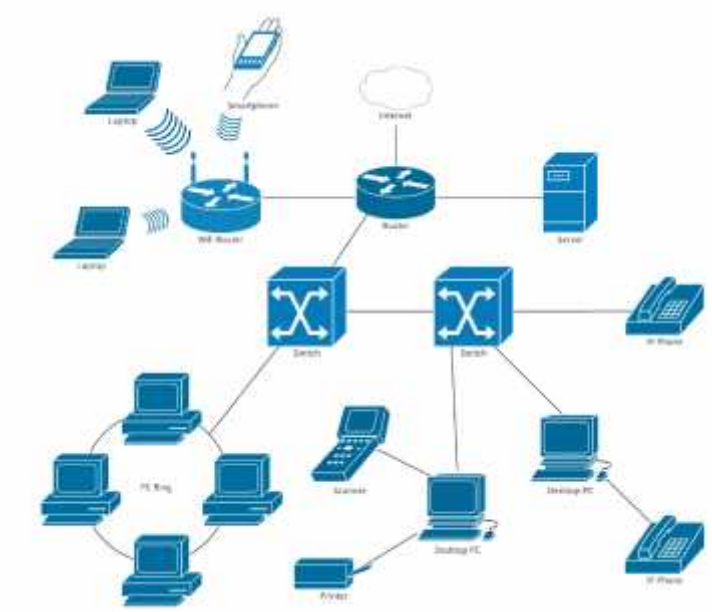


Data communication

- Local Area Network (LAN)
- Wide Area Network (WAN)
- Internet

Need to access and communicate

- Possible of intrusion
- Critical : access & control of the communication and data centric, even client machines
- Spoofing attack to the datagram over network



ABOUT THE DATA

Data *is a set of values of qualitative or quantitative variables.*

Data come in various types and forms

In general sense

1. Attribute Data vs Object Data
2. Structuring Data vs non-structuring data

Evolution of Database Technology

- 1960s:
  - Data collection, database creation, IMS and network DBMS
- 1970s:
  - Relational data model, relational DBMS implementation
- 1980s:
  - RDBMS, advanced data models (extended-relational, OO, deductive, etc.) and application-oriented DBMS (spatial, scientific, engineering, etc.)
- 1990s—2000s:
  - Data mining and data warehousing, multimedia databases, and Web databases

SOFTWARE FOR DATA MANAGEMENT
DATABASE SOFTWARE

1. เครื่องมือที่ใช้ จัดเก็บข้อมูล บริหารจัดการข้อมูล และค้นหาข้อมูล
2. จัดเก็บในรูปแบบ Tubular → 2 dimensions table
3. Table → Fields (Columns) and Records (Rows)

GHG_Id	GHG_Name	GHG	gwp_value	AR_Id
1	CARBON DIOXIDE	CO2	1	4
2	METHANE	CH4	25	4
3	NITROUS OXIDE	N2O	298	4
4	HFC-23	CHF3	14800	4
5	HFC-32	CH2F2	675	4
6	HFC-41	CH3F		4
7	HFC 43 10mee	CF3CHFCHFCF2CF3	1640	4
8	HFC-125	CHF2CF3	3500	4
9	HFC-134	CHF2CHF2		4
10	HFC-134a	CH2FCF3	1430	4
11	HFC-143	CHF2CH2F		4
12	HFC-143a	CF3CH3	4470	4

EXERCISE: CONVERT TABLE TO TUPLE AR FROM

Global warming potential (GWP) values relative to CO_2

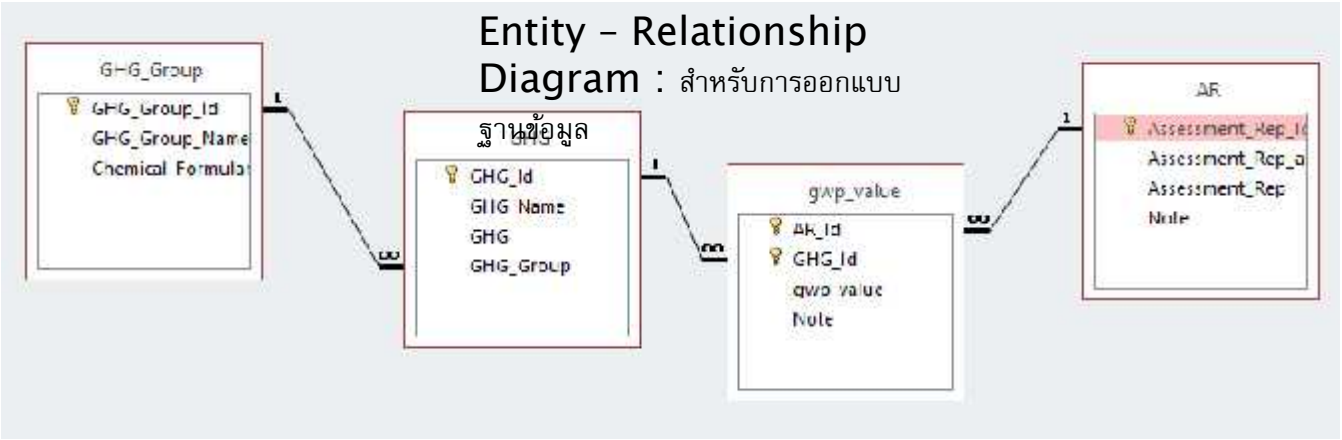
Industrial designation or common name	Chemical formula	GWP values for 100-year time horizon		
		Second Assessment Report (SAR)	Fourth Assessment Report (AR4)	Fifth Assessment Report (AR5)
Carbon dioxide	CO_2	1	1	1
Methane	CH_4	21	25	28
Nitrous oxide	N_2O	310	298	265
Substances controlled by the Montreal Protocol				
CFC-11	CCl_3F	3,800	4,750	4,660
CFC-12	CCl_2F_2	8,100	10,900	10,200
CFC-13	$CClF_3$		14,400	13,900
CFC-113	CCl_2FCClF_2	4,800	6,130	5,820
CFC-114	$CClF_2CClF_2$		10,000	8,590
CFC-115	$CClF_2CF_3$		7,370	7,670
Halon-1301	$CBrF_3$	5,400	7,140	6,290
Halon-1211	$CBrClF_2$		1,890	1,750
Halon-2402	$CBrF_2CBrF_2$		1,640	1,470
Carbon tetrachloride	CCl_4	1,400	1,400	1,730
Methyl bromide	CH_3Br		5	2
Methyl chloroform	CH_3CCl_3	100	146	160

GHG Name	Chemical Formula	GWP	Assessment Report	Protocol

DATABASE DESIGN : DATA MODEL

Relational DB Model

- Data store in tubular table
- Split dataset to different tables
- Prevent duplicate data in table
- Using data one or more value to distinctive data uniqueness → Key concept
- Table has relationship to other table in 1-to-many manner (Relational Function)



APPLICATION

- Custom development software
- Commercial proprietary software
- Open Source Software

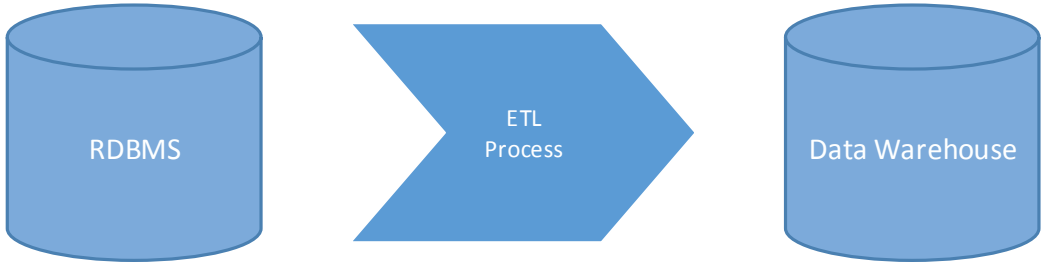
  - What's data to manage?

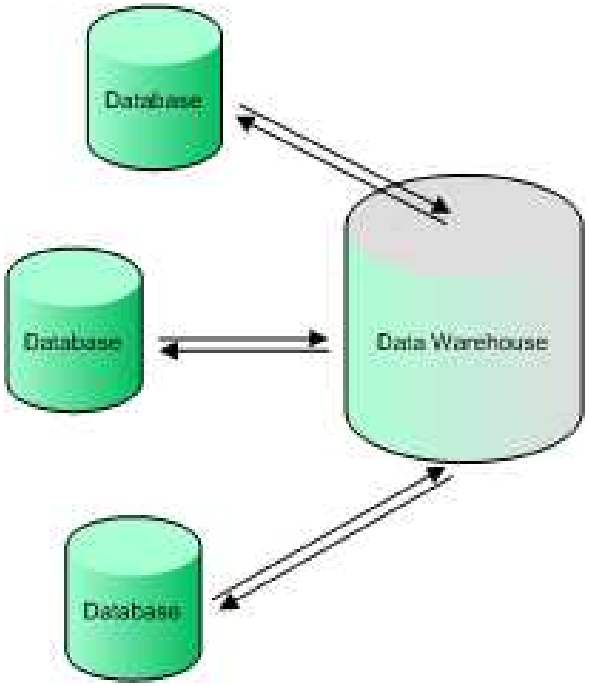
    Attribute data vs Object data
  - Data elementary level
  - Data atomicity in database

DATA WAREHOUSE

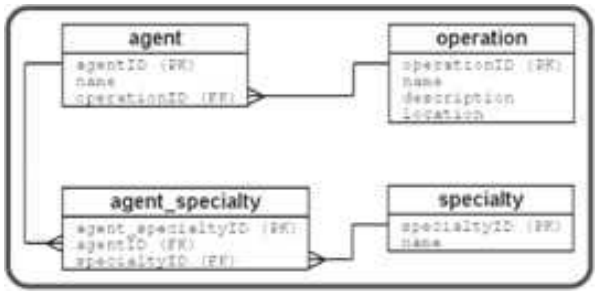
พัฒนาเพื่อวัตถุประสงค์ การเรียกคืนข้อมูล การสืบค้น และการประมวลผลในลักษณะหาคำตอบ มากกว่าใช้เพื่อรองรับการปฏิบัติงาน เนื่องจากเป็นการทำงานกับข้อมูลในปริมาณมากๆ จึงทำให้ต้องมีการพัฒนาแนวทางการออกแบบโครงสร้างข้อมูลเพื่อประสิทธิภาพในด้านที่ต้องการ

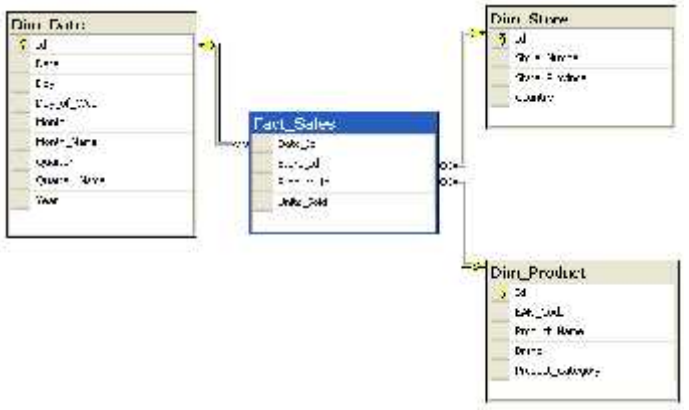
- ปรับโครงสร้างจาก Normal Form to Star Schemas



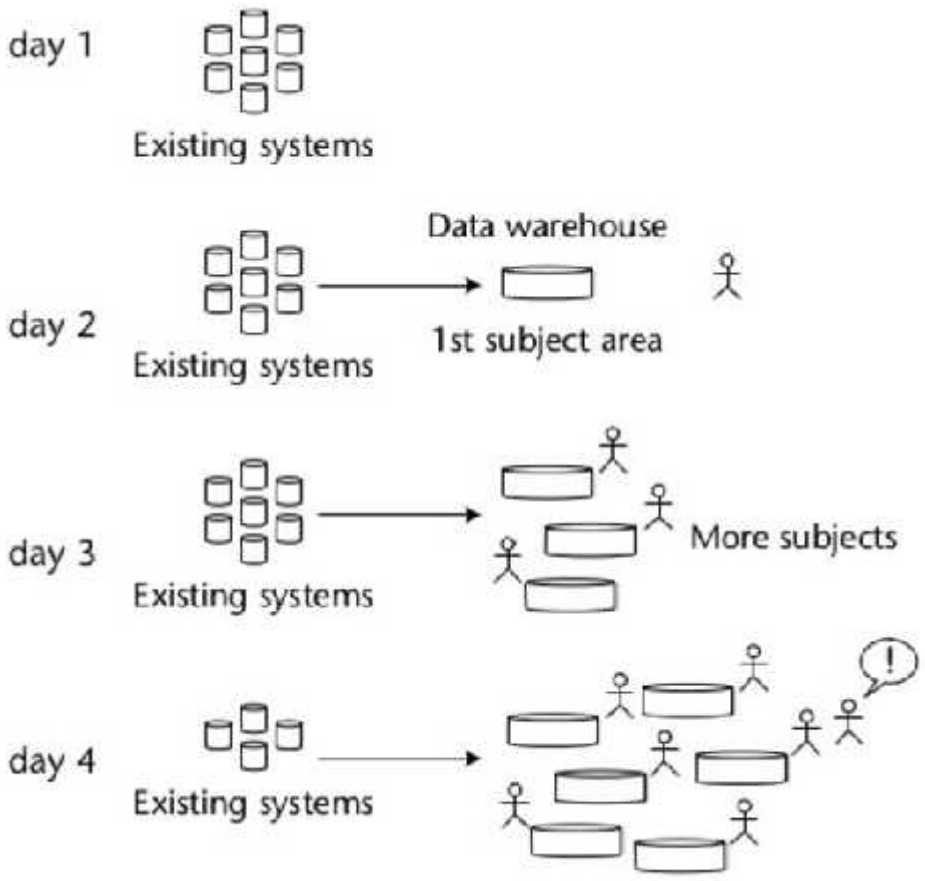


EXTRACT TRANSFORM LOAD (ETL) PROCESS





DAY 1 TO DAY N PHENOMENON



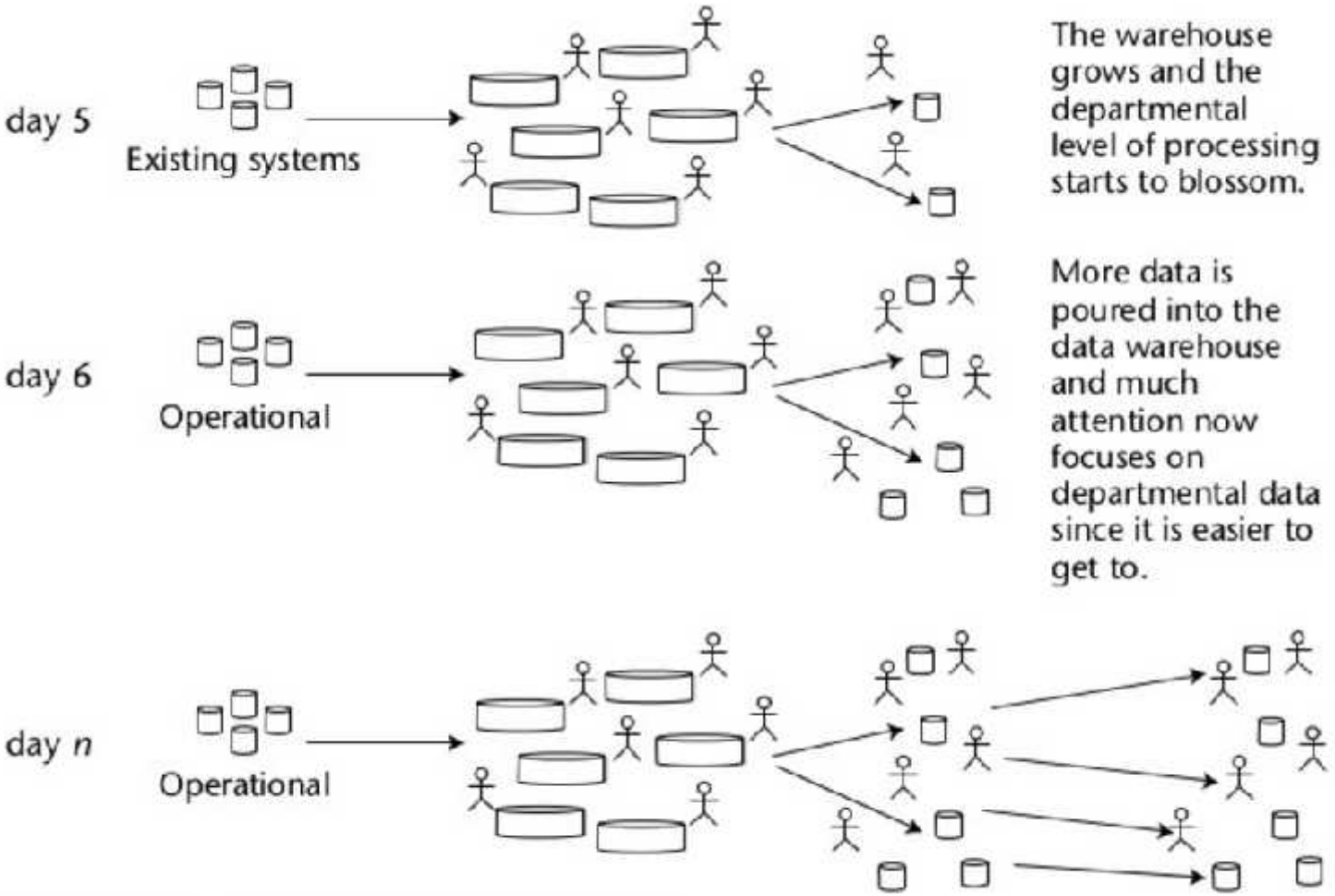
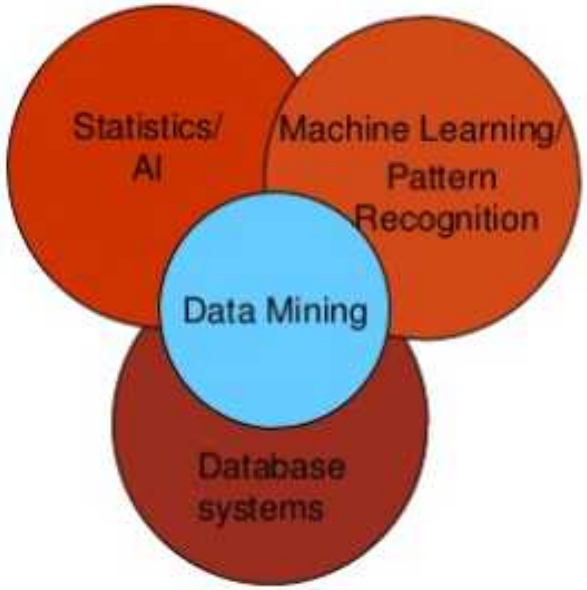


Figure 2-10 Day 1 to day *n* phenomenon.

DATA MINING KNOWLEDGE DISCOVERY IN DATABASES

- What is not Data Mining?
  - Look up phone number in phone directory
  - Query a Web search engine for information about "Amazon"

- What is Data Mining?
  - Certain names are more prevalent in certain US locations (O'Brien, O'Rurke, O'Reilly… in Boston area)
  - Group together similar documents returned by search engine according to their context (e.g. Amazon rainforest, Amazon.com,)



DATA MINING TASKS

- Prediction Tasks
  - Use some variables to predict unknown of future values of other variables.
- Description Tasks
  - Find human-interpretable patterns that describe the data.

- Common data mining tasks
  - Classification → (Predictive)
  - Clustering → (Descriptive)
  - Association Rule Discovery → (Descriptive)
  - Sequential Pattern Discovery → (Descriptive)
  - Regression → (Predictive)
  - Deviation Detection → (Predictive)

BIG DATA

Big data is a term that describes the large volume of data – both structured and unstructured – that inundates a business on a day-to-day basis. But it's not the amount of data that's important. It's what organizations do with the data that matters.

Big data can be analyzed for insights that lead to better decisions and strategic business moves.

BIG DATA PARADIGM

Tradition Form of Data

Structured Data

- Transaction data
- Legacy databases

Non-structuring Data

- Report
- Study
- Map
- Archive Image

- Time consume for processing

- To find the relationship between 2 groups of data
- Result data will able to predict behavior
- Behavior data will able to predict the result

Behavior Data

- from sensor technology
- both structured and non-structuring

- fast detecting
- may relate or not relate to the focus topic

EXAMPLE : GOOGLE SERVICE BY USING BIG DATA

Google provide traffic data and people visit of tourism spot

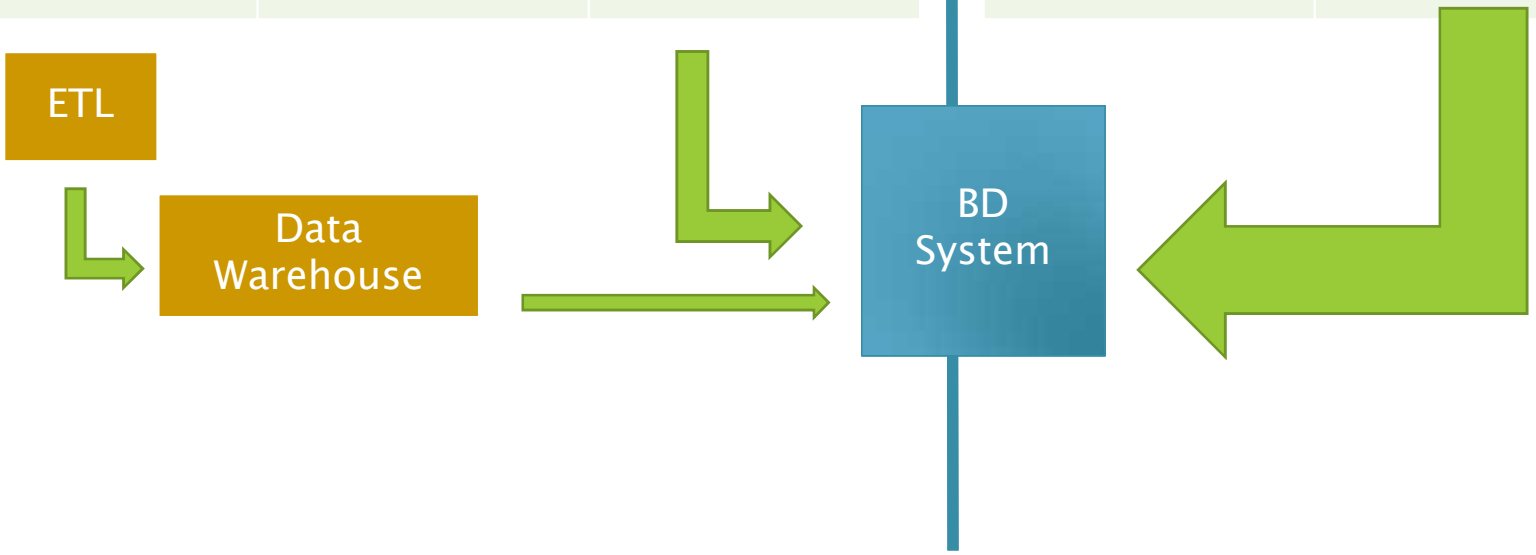
- Basis data are; map and geolocation
- Behavior data;
  - Detect by android smartphone user "open locator"
  - Processing to estimate result not exactly reality
  - Can help to make decision but cannot quantify the result no.

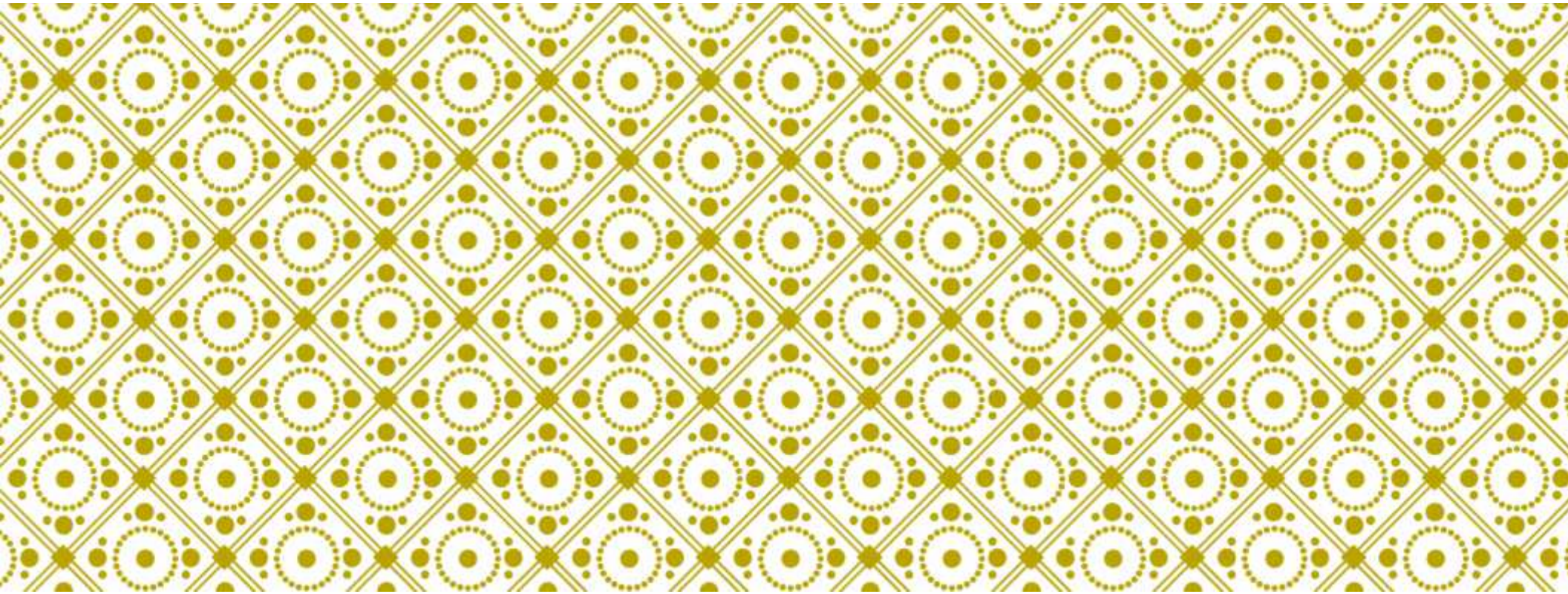
Result Data

Major Data	Dimension	Acquisition Technique
ปริมาณการผลิต		Transaction Processing
ปริมาณการขาย		
ปริมาณการสูญเสีย		

Behavior Data

Major Data	Dimension	Acquisition Technique
Power Plant PLC		Metering/Sensor
Distribution		
Demand Load		





TECHNOLOGY TREND

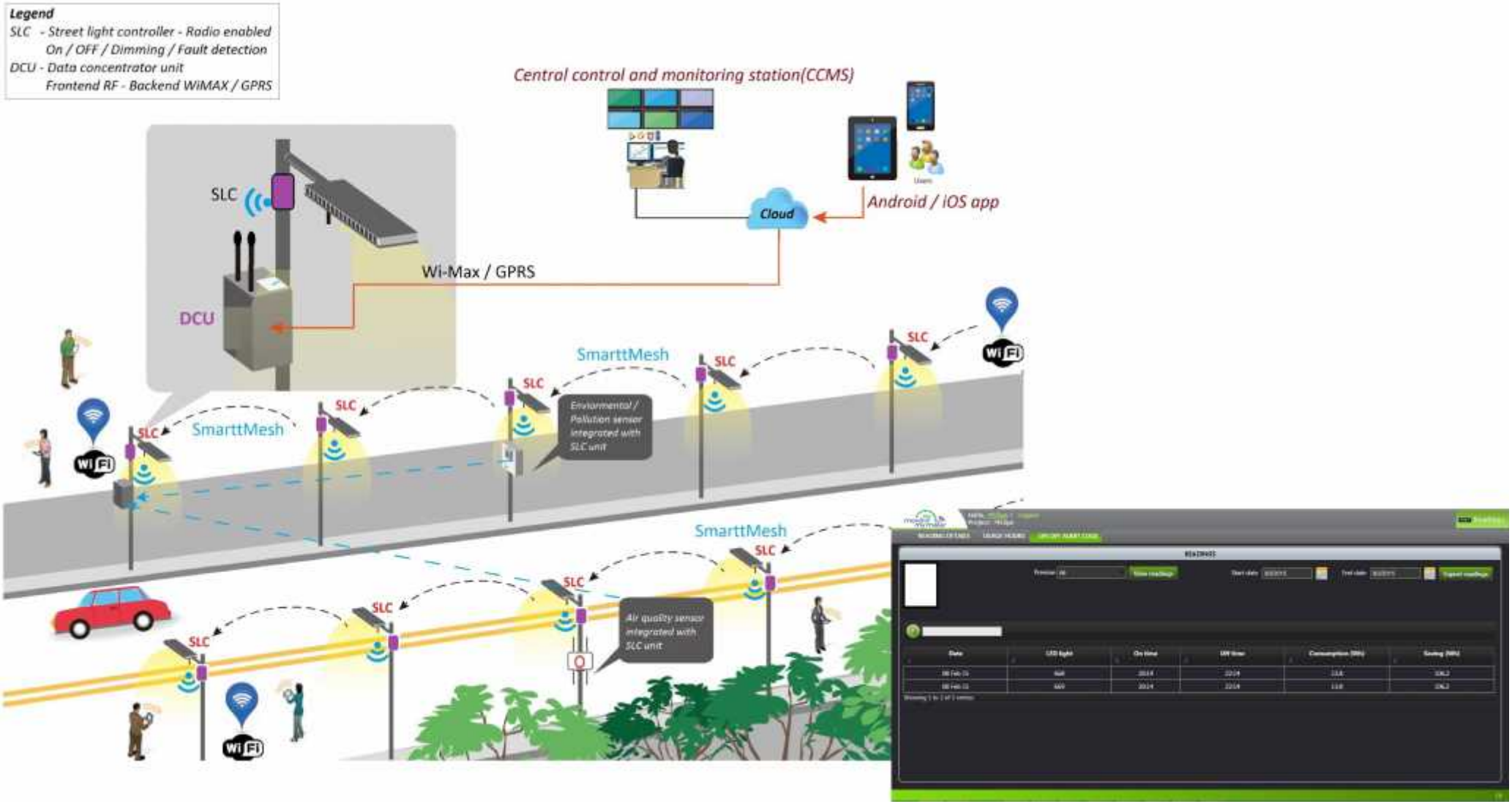
INTERNET OF THINGS (IOT)

Use Case : IoT for city street light

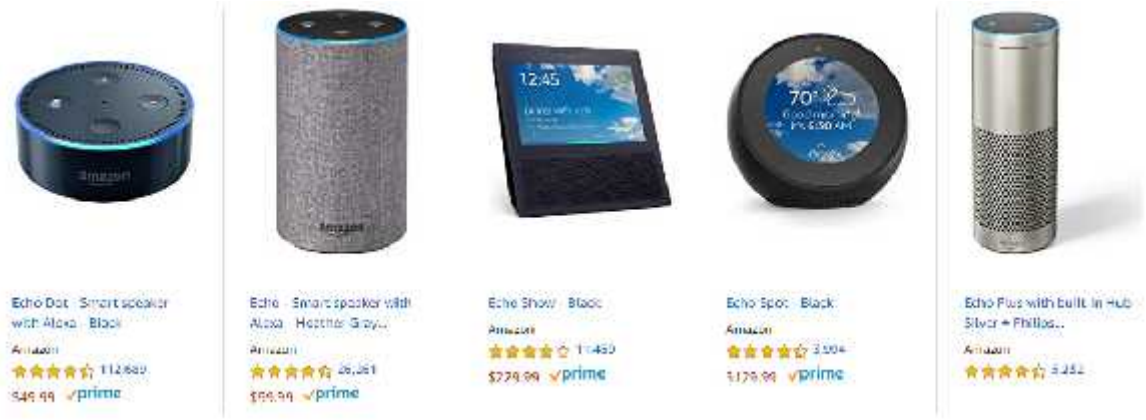
CityTouch Ready luminaires

CityTouch connectivity devices

Cabinet-based group connectivity


Legend
SLC - Street light controller - Radio enabled
On / OFF / Dimming / Fault detection
DCU - Data concentrator unit
Frontend RF - Backend WiMAX / GPRS
Central control and monitoring station(CCMS)
Cloud
Android / iOS app
Wi-Max / GPRS
SLC
DCU
SmarttMesh
Environmental / Pollution sensor integrated with SLC unit
Air quality sensor integrated with SLC unit
WiFi

IOT @ HOME


Echo Dot - Smart speaker with Alexa - Black
Amazon
$49.99 ✓prime
Echo - Smart speaker with Alexa - Heather Gray...
Amazon
$99.99 ✓prime
Echo Show - Black
Amazon
$229.99 ✓prime
Echo Spot - Black
Amazon
$129.99 ✓prime
Echo Plus with built-in Hub Silver + Philips...
Amazon


Ok Google, what does "circumlocution" mean?
Circumlocution: the use of many words where fewer would do, especially in a deliberate attempt to be vague or evasive

ARTIFICIAL INTELLIGENT (AI)

Artificial intelligence (AI) is an area of computer science that emphasizes the creation of intelligent machines that work and react like humans. Some of the activities computers with artificial intelligence are designed for include:

- Speech recognition
- Learning
- Planning
- Problem solving

AI USE CASE : PLANT IDENTIFY TOOL

About PlantSnap

PlantSnap is a mobile app that helps you identify plants, flowers, trees, cacti and mushrooms. Simply snap a photo of the plant, and PlantSnap tells you what it is!

PlantSnap can currently recognize 90% of all known species of plants and trees, which covers most of the species you will encounter in every country on Earth.

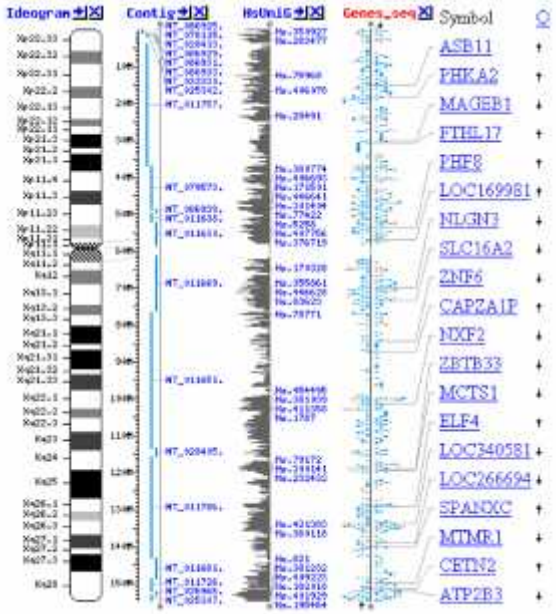
USE CASE : MIX OF TECHNOLOGY

DATA WAREHOUSE, DATAMINING, BIG DATA, AI

Bio informatics

- Blast service of base code in gene
- Collect genome, molecular biology research data
- Complex algorithm in bio-data

```
5'ATGACGTGGGGA3'
3'TACTGCACCCCT5'
```




```
A5ASC3.1    14 SIKLWPPSQTTRLLLVERMANNLST..PSIFTRK..YGSLSKEEARENAKQIEEVACSTANQ.....HYEKEPDGDGGSAVQLYAKECSKLILEVLK 101
B4F917.1    13 SIKLWPPSESTRIMLVDRMTNNLST..ESIFSRK..YRLLGKQEAHENAKTIEELCFALADE.....HFREEPDGDGSSAVQLYAKETSKMMLEVLK 100
A9S1V2.1    23 VFKLWPPSQGTREAVRQKMALKLSS..ACFESQS..FARIELADAQEHARAIEEVAFGAAQE......ADSGGDKTGSAVVMVYAKHASKLMLETLR 109
B9GSN7.1    13 SVKLWPPGQSTRLMLVERMTKNFIT..PSFISRK..YGLLSKEEAEEDAKKIEEVAFAAANQ.....HYEKQPDGDGSSAVQIYAKESSRLMLEVLK 100
Q8H056.1    30 SFSIWPPTQRTRDAVVRRLVDTLGS..DTILCKR..YGAVPAADAEPAARGIEAEAFDAAAA..SGEAAATASVEEGIKALQLYSKEVSRRLLDFVK 120
Q0D4Z3.2    44 SLSIWPPSQRTRDAVVRRLVQTLVA..PSILSQR..YGAVPEAEAGRAAAAVEAEAYAAVTES.SSAAAAPASVEDGIEVLQAYSKEVSRRLLELAK 135
B9MVW8.1    56 SFSIWPPTQRTRDAIISRLIETLST..TSVLSKR..YGTIPKEEASEASRRIEEEAFSGAST.......VASSEKDGLEVLQLYSKEISKRMLETVK 141
Q0IYC5.1    29 SFAVWPPTRRTRDAVVRRLVAVLSGDTTTALRKRYRYGAVPAADAERAARAVEAQAFDAASA....SSSSSSSVEDGIETLQLYSREVSNRLLAFVR 121
A9NW46.1    13 SIKLWPPSESTRLMLVERMTDNLSS..VSFFSRK..YGLLSKEEAAENAKRIEETAFLAAND.....HEAKEPNLDDSSVVQFYAREASKLMLEALK 100
Q9C500.1    57 SLRIWPPTQKTRDAVLNRLIETLST..ESILSKR..YGTLKSDDATTVAKLIEEEAYGVASN.......AVSSDDDGIKILELYSKEISKRMLESVK 142
Q2HRI7.1    25 NYSIWPPKQRTRDAVKNRLIETLST..PSVLTKR..YGTMSADEASAAAIQIEDEAFSVANA.......SSSTSNDNVTILEVYSKEISKRMIETVK 110
Q9M7N3.1    28 SFKIWPPTQRTREAVVRRLVETLTS..QSVLSKR..YGVIPEEDATSAARIIEEEAFSVASV.ASAASTGGRPEDEWIEVLHIYSQEIXQRVVESAK 119
Q9M7N6.1    25 SFSIWPPTQRTRDAVINRLIESLST..PSILSKR..YGTLPQDEASETARLIEEEAFAAAGS.......TASDADDGIEILQVYSKEISKRMIDTVK 110
Q9LE82.1    14 SVKMWPPSKSTRLMLVERMTKNITT..PSIFSRK..YGLLSVEEAEQDAKRIEDLAFATANK.....HFQNEPDGDGTSAVHVYAKESSKLMLDVIK 101
Q9M651.2    13 SIKLWPPSLPTRKALIERITNNFSS..KTIFTEK..YGSLTKDQATENAKRIEDIAFSTANQ.....QFEREPDGDGGSAVQLYAKECSKLILEVLK 100
B9R748.1    48 SLSIWPPTQRTRDAVITRLIETLSS..PSVLSKR..YGTISHDEAESAARRIEDEAFGVANT.......ATSAEDDGLEILQLYSKEISRRMLDTVK 133
```