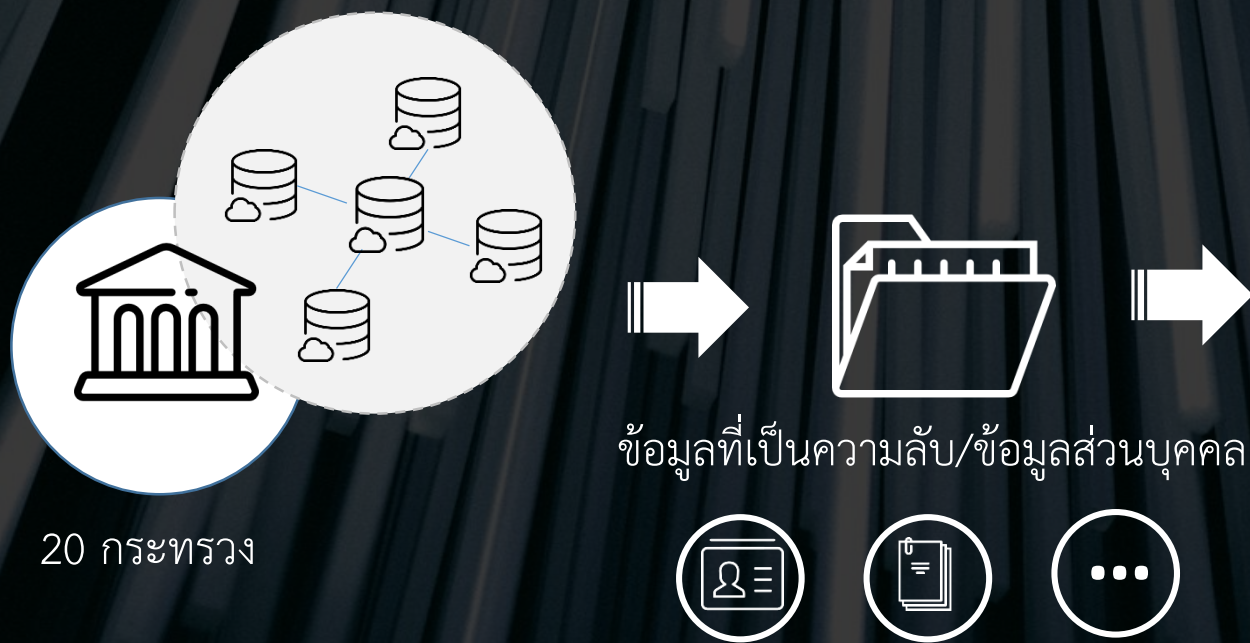


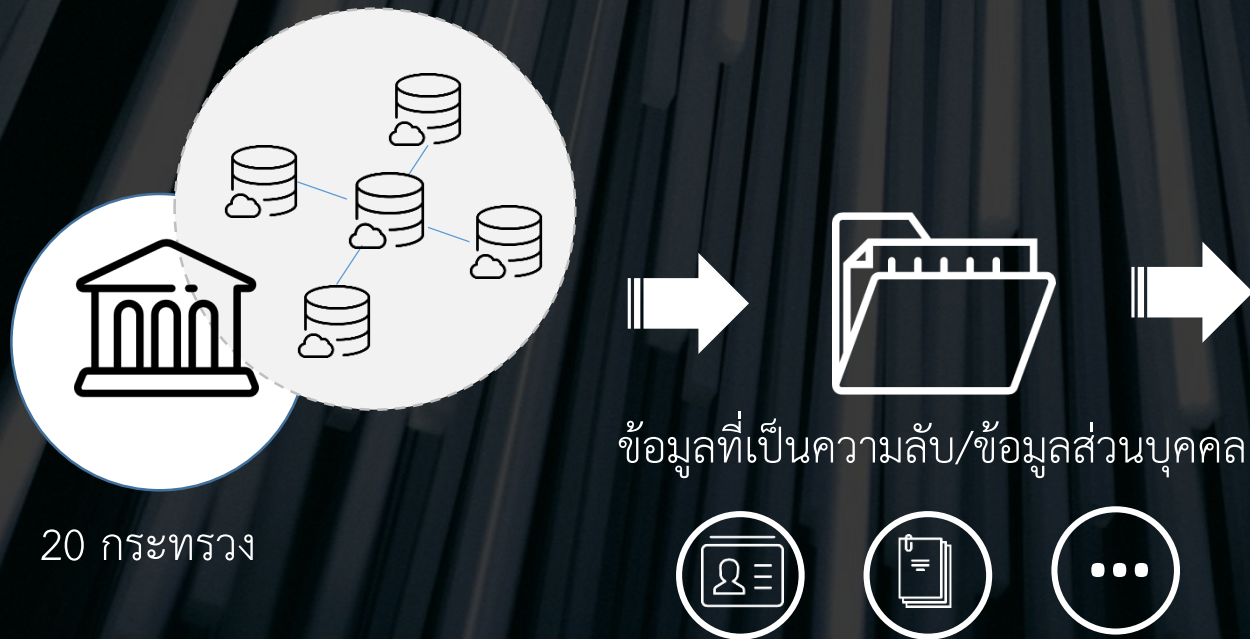
DATA PROCESSING IN BIG DATA PROJECT

การดำเนินงานด้านการจัดการข้อมูลขนาดใหญ่ภาครัฐ
(Government Big Data)



PRIVACY PROTECTION PROCESS

กระบวนการปกป้องข้อมูลที่เป็นส่วนบุคคล



PRIVACY PROTECTION PROCESS

กระบวนการปกป้องข้อมูลที่เป็นส่วนบุคคล

Procedure and Architecture

Hashing Citizen ID

Anonymization

PRIVACY PROTECTION PROCESS

กระบวนการปกป้องข้อมูลที่เป็นส่วนบุคคล

Procedure and Architecture

Procedure and Architecture

■ หน่วยงานดำเนินการ ■ สรอ. ดำเนินการ



Data Format Preparation

Step **1**



จัดเตรียมรูปแบบข้อมูล
(Data Format Preparation)

เช่น เลขบัตรประชาชน
รวมถึงข้อมูลที่ได้รับตามโจทย์

หน่วยงานดำเนินการ

เลขบัตรประชาชน	0000000077770 0 0000 00088 88 0 0-0000-00099-99-0	→	Data Format Preparation	0000000077770 0000000088880 0000000099990
เพศ	ช. หรือ หญิง. ชาย หรือ หญิง Male หรือ Female	→		M หรือ F
วันเกิด (คำนวณอายุ)	04/06/2521 [dd/mm/yyyy] 05/06/22 [dd/mm/yy] 12/31/2523 [mm/dd/yyyy] 2524/06/01 [yyyy/mm/dd] → 2525-06-01 [yyyy-mm-dd]	→		2521 2522 2523 2524 2524
จำนวนเงิน	1900 30000 77700 100050 →	→		1-10000 10001-50000 500001-100000 100050-500000

PRIVACY PROTECTION PROCESS

กระบวนการปกป้องข้อมูลที่เป็นส่วนบุคคล

Hashing Citizen ID

Hashing Citizen ID

How to Hashing ?

- Duplicate a Table

Step 2



Hashing CID

แปลงข้อมูล (Data) ที่ระบุตัวตน

	IDCARD	FIRSTNAME	LASTNAME	BIRTHDATE	AGE	SALARY
1	1-1111-11111-11-1	มนัส	ศิริศักดิ์	28-DEC-17	49	47069
2	2-2222-22222-22-2	วรศาสตร์	เกตุณัฐ	12-DEC-17	29	27115
3	3-3333-33333-33-3	อุทยานนท์	เชตุนเดช	19-DEC-17	31	35561
4	4-4444-44444-44-4	เมทินทร์	พนาสีผ่อง	23-DEC-17	32	38174
5	5-5555-55555-55-5	สมดวงฉิน	อุสาศาสตร์	16-DEC-17	33	30047
6	6-6666-66666-66-6	เมษา	วรภูมิ	11-DEC-17	43	41270
7	7-7777-77777-77-7	พีชร์	ชจรศาสตร์	29-DEC-17	22	22000

หน่วยงานดำเนินการ

Hashing Citizen ID

How to Hashing ?

- Duplicate a Table
- Cleansing CID No.

Step 2



Hashing CID

แปลงข้อมูล (Data) ที่ระบุตัวตน

1. UPDATE cloned_table
SET *idcard* = REPLACE(*idcard*, '-', '');
2. UPDATE cloned_table
SET *idcard* = REPLACE(*idcard*, ' ', '');

IDCARD	FIRSTNAME	LASTNAME	BIRTHDATE	AGE	SALARY
111111111111	มนัส	ศิริศักดิ์	28-DEC-17	49	47069
222222222222	วรศาสตร์	เกตุรัฐ	12-DEC-17	29	27115
333333333333	อุษานนท์	เชตบุตร	19-DEC-17	31	35561
444444444444	เมทินทร์	พนาสีวงศ์	23-DEC-17	32	38174
555555555555	สมดวงณี	อุสาศาสตร์	16-DEC-17	33	30047
666666666666	เมษา	วรภูมิ	11-DEC-17	43	41270
777777777777	พีช	ขจรศาสตร์	29-DEC-17	22	22000

หน่วยงานดำเนินการ

Hashing Citizen ID

How to Hashing ?

- Duplicate a Table
- Cleansing ID No.
- Verifier Records

Step 2



Hashing CID

แปลงข้อมูล (Data) ที่ระบุตัวตน

IDCARD	FIRSTNAME	LASTNAME	BIRTHDATE	AGE	SALARY
1111111111111	มนัส	ศิริศักดิ์	28-DEC-17	49	47069
2222222222222	วรศาสตร์	เกตุรัฐ	12-DEC-17	29	27115
3333333333333	อุทยานนท์	เชตเดช	19-DEC-17	31	35561
4444444444444	เมคินทร์	พนาลีพ้อง	23-DEC-17	32	38174
5555555555555	สมดวงนิณ	อุสาศาสตร์	16-DEC-17	33	30047
6666666666666	เมษา	วรภูมิ	11-DEC-17	43	41270
7777777777777	พีชร์	ชจรศาสตร์	29-DEC-17	22	22000
8888888888888	name1	name1	29-DEC-17	26	27893
9999999999999	name2	name2	29-DEC-17	32	32456
0000000000000	name3	name3	29-DEC-17	38	42900

หน่วยงานดำเนินการ

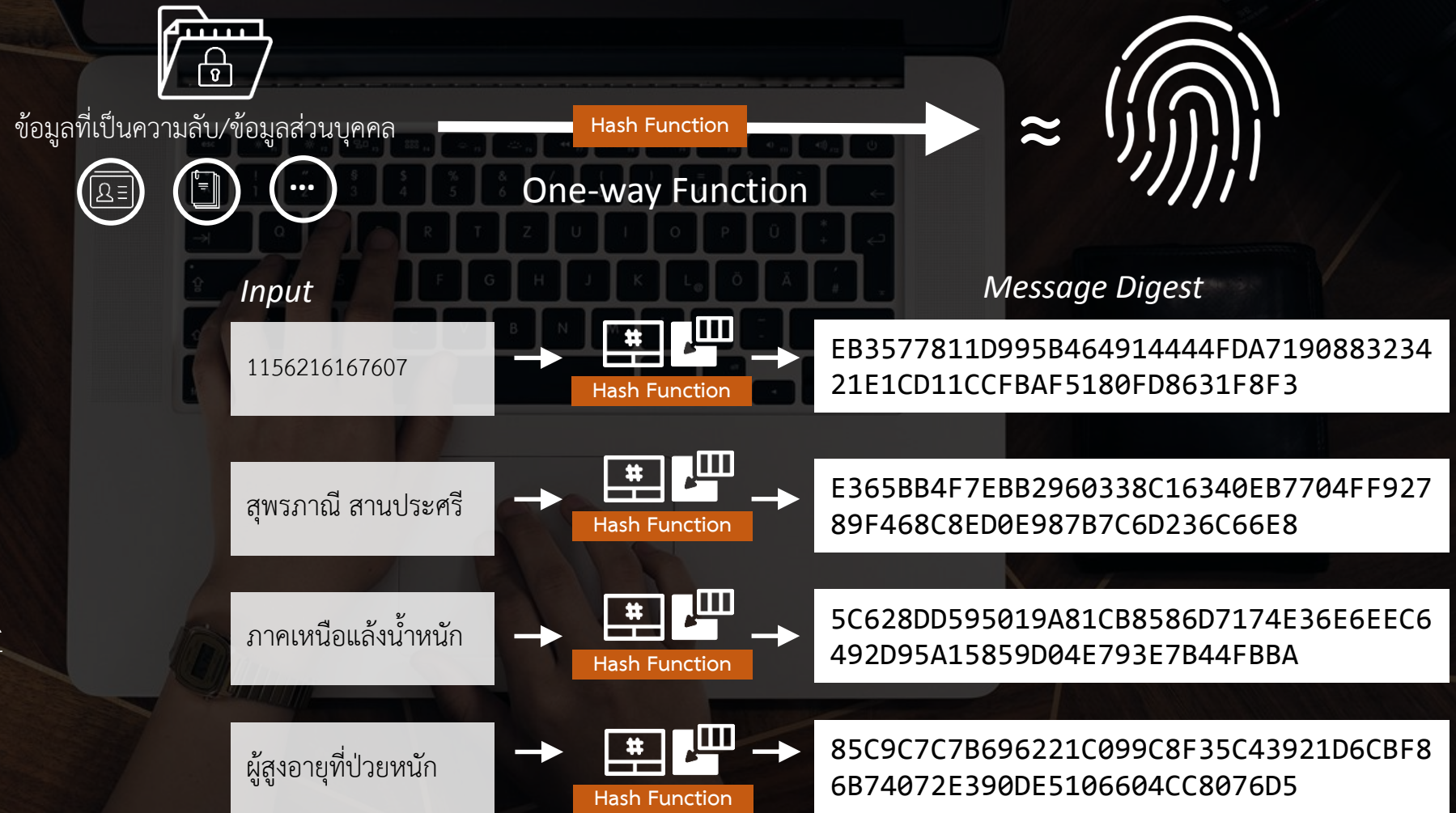
Step 2



Hashing CID

แปลงข้อมูล (Data) ที่ระบุตัวตน

หน่วยงานดำเนินการ



Hashing Citizen ID

Intro

- Collision Problem

Step 2



Hashing CID

แปลงข้อมูล (Data) ที่ระบุตัวตน



Hash
Function

สมรัฐ

ภูมिरพี

สมยศ

John

0000001

0000002

0000003

Hashing Citizen ID

Intro

- Collision Problem
- Hash Algorithms

Step 2



Hashing CID

แปลงข้อมูล (Data) ที่ระบุตัวตน

Algorithm and variant		Output size (bits)	Internal state size (bits)	Block size (bits)	Max message size (bits)	Rounds	Operations	Security bits (Info)
MD5 (as reference)		128	128 (4 × 32)	512	Unlimited ^[2]	64	And, Xor, Rot, Add (mod 2 ³²), Or	<64 (collisions found)
SHA-0		160	160 (5 × 32)	512	2 ⁶⁴ − 1	80	And, Xor, Rot, Add (mod 2 ³²), Or	<34 (collisions found)
SHA-1								<63 (collisions found ^[3])
SHA-2	SHA-224 SHA-256	224 <u>256</u>	256 (8 × 32)	512	2 ⁶⁴ − 1	64	And, Xor, Rot, Add (mod 2 ³²), Or, Shr	112 128
	SHA-384 SHA-512	384 512	512 (8 × 64)	1024	2 ¹²⁸ − 1	80	And, Xor, Rot, Add (mod 2 ⁶⁴), Or, Shr	192 256
	SHA-512/224 SHA-512/256	224 256						112 128

Hashing Citizen ID

Intro

- Collision Problem
- Hash Algorithms
- Hash Algorithms – SHA-256

Step 2



Hashing CID

แปลงข้อมูล (Data) ที่ระบุตัวตน



ข้อมูลที่เป็นความลับ/ข้อมูลส่วนบุคคล



SHA-256

Hash Function

Hash Value

85C9C7C7B696221C099C8F35
C43921D6CBF86B74072E390D
E5106604CC8076D5

Hashing Citizen ID

Intro

- Collision Problem
- Hash Algorithms
- Hash Algorithms – SHA256
- SHA-256 with Salt

Step 2



Hashing CID

แปลงข้อมูล (Data) ที่ระบุตัวตน

ข้อมูลที่เป็นความลับ/ข้อมูลส่วนบุคคล



SHA-256 w/ Salt#1

Hash Function

Hash Value

311F8DA0C3F0FDCA2770BEC7
AD241C53F3DB4A2ED4CC99C2
9BAA204B73AC0B9C

Hashing Citizen ID

How to SHA256 ?

- Collision Problem
- Hash Algorithms
- Hash Algorithms – SHA256
- SHA-256 with Salt
- Generate Hash

Step **2**



Hashing CID

แปลงข้อมูล (Data) ที่ระบุตัวตน

```
ALTER TABLE cloned_table ADD hashed varchar2(64);
```

```
Define salt = 'BigData7Projects';
```

```
UPDATE cloned_table
```

```
SET hashed = STANDARD_HASH(idcard || '&salt','SHA256');
```

IDCARD	FIRSTNAME	LASTNAME	BIRTHDATE	AGE	SALARY	HASHED
111111111111	มนัส	ศิริศักดิ์	28-DEC-17	49	47069	
222222222222	วรศาสตร์	เกตุรัฐ	12-DEC-17	29	27115	
333333333333	อุทยานนท์	เชตุนเดช	19-DEC-17	31	35561	
444444444444	เมทินทร์	พนาสีผ่อง	23-DEC-17	32	38174	
555555555555	สมดวงกนิ	อุสาศาสตร์	16-DEC-17	33	30047	
666666666666	เมษา	วราภูมิ	11-DEC-17	43	41270	
777777777777	พีชร์	ขจรศาสตร์	29-DEC-17	22	22000	
888888888888	name1	name1	29-DEC-17	26	27893	
999999999999	name2	name2	29-DEC-17	32	32456	
000000000000	name3	name3	29-DEC-17	38	42900	

หน่วยงานดำเนินการ

Hashing Citizen ID

How to SHA256 ?

- Collision Problem
- Hash Algorithms
- Hash Algorithms – SHA256
- SHA-256 with Salt
- Generate Hash

Step 2



Hashing CID

แปลงข้อมูล (Data) ที่ระบุตัวตน

```
ALTER TABLE cloned_table ADD hashed varchar2(64);
```

```
Define salt = 'BigData7Projects';
```

```
UPDATE cloned_table
```

```
SET hashed = STANDARD_HASH(idcard || '&salt', 'SHA256');
```

IDCARD	FIRSTNAME	LASTNAME	BIRTHDATE	AGE	SALARY	HASHED
111111111111	มนัส	ศิริศักดิ์	28-DEC-17	49	47069	7AE93289F8682756DC57CF5CF55D7E7D84898E89CD8D81C230EDFB33090D7C1D
222222222222	วรศาสตร์	เกตุณัฐ	12-DEC-17	29	27115	5C443954B201ECEBF9A341588E32AE2D506E800EB7ADF094F6F7B57536915BEA
333333333333	อุยานนท์	เชตุนเดช	19-DEC-17	31	35561	3AB16F2004B0A0294F033DB3255F54F2CC02D122DEDEC513523C6D372B01B281
444444444444	เมทินทร์	พนาสีผ่อง	23-DEC-17	32	38174	E9445BD4FAF0C4E13431B8B5713D58CD9FC75873648F4F78566E325DCD0D3028
555555555555	สมดวงกนิ	อุสาศาสตร์	16-DEC-17	33	30047	F3A9FC03725BDD82036DCB0FEF6ADFD27A6551B14648048DEC368C62C290F549
666666666666	เมษา	วราภูมิ	11-DEC-17	43	41270	329A99A260D19701BF897A8B9A4972EA37709536BA50C82B684B620C46A1990A
777777777777	พีชร์	ชจรศาสตร์	29-DEC-17	22	22000	C6B57C4641C46F467B424168A5F1133C42A8F8C1D93A75B5392CB3C96D058570
888888888888	name1	name1	29-DEC-17	26	27893	8A6F383B4BC570A0445C5E5052E02A1EE0488EC72868A453DE292BDC4C5EAD65
999999999999	name2	name2	29-DEC-17	32	32456	BFDEA38A8F5B65D465B24029065FB855A9C1049BAA2B82CCDBCDD486618F9C0
000000000000	name3	name3	29-DEC-17	38	42900	E32AE7EA65FF12E48250E2E2486813973BC3E9A8F42579BA9A5BF7C064F7B6FB

หน่วยงานดำเนินการ

Hashing Citizen ID

Verifier Records

- Collision Problem
- Hash Algorithms
- Hash Algorithms – SHA256
- SHA-256 with Salt
- Generate Hash
- Verify the Verifier Records

Step 2



Hashing CID

แปลงข้อมูล (Data) ที่ระบุตัวตน

```
SELECT idcard, hashed
FROM cloned_table
WHERE (idcard = '888888888888' AND hashed =
'8A6F383B4BC570A0445C5E5052E02A1EE0488EC72868A453DE292BDC4C5EAD65')
OR (idcard = '999999999999' AND hashed =
'BFDEA38A8F5B65D465B24029065FB855A9C1049BAA2B82CCDBCBD486618F9C0')
OR (idcard = '000000000000' AND hashed =
'E32AE7EA65FF12E48250E2E2486813973BC3E9A8F42579BA9A5BF7C064F7B6FB')
;
```

IDCARD	HASHED
888888888888	8A6F383B4BC570A0445C5E5052E02A1EE0488EC72868A453DE292BDC4C5EAD65
999999999999	BFDEA38A8F5B65D465B24029065FB855A9C1049BAA2B82CCDBCBD486618F9C0
000000000000	E32AE7EA65FF12E48250E2E2486813973BC3E9A8F42579BA9A5BF7C064F7B6FB

หน่วยงานดำเนินการ

PRIVACY PROTECTION PROCESS

กระบวนการปกป้องข้อมูลที่เป็นส่วนบุคคล

Anonymization

Pseudonymization is a method to substitute identifiable data with a reversible, consistent value. **Anonymization** is the destruction of the identifiable data.

```
ALTER TABLE cloned_table ADD age_range varchar2(20);
```

```
UPDATE cloned_table  
SET num_range =  
CASE
```

```
WHEN age <= 20 THEN '0-20'  
WHEN age BETWEEN 21 AND 30 THEN '21-30'  
WHEN age BETWEEN 31 AND 40 THEN '31-40'  
WHEN age BETWEEN 41 AND 50 THEN '41-50'  
WHEN age BETWEEN 51 AND 60 THEN '51-60'  
WHEN age > 60 THEN '>60'
```

```
END
```

```
WHERE age_range IS NOT NULL;
```

IDCARD	FIRSTNAME	LASTNAME	BIRTHDATE	AGE	SALARY	HASHED
111111111111	มนัส	ศิริศักดิ์	28-DEC-17	49	47069	7AE932...
222222222222	วรศาสตร์	เกตุรัฐ	12-DEC-17	29	27115	5C4439...
333333333333	อุทยานนท์	เชตนะเดช	19-DEC-17	31	35561	3AB16F...
444444444444	เมทินท์	พนาลีผ่อง	23-DEC-17	32	38174	E9445B...
555555555555	สมดวงณิน	อุสาศาสตร์	16-DEC-17	33	30047	F3A9FC...
666666666666	เมษา	วรกิจ	11-DEC-17	43	41270	329A99...
777777777777	พีชร์	ชจรศาสตร์	29-DEC-17	22	22000	C6B57C...
888888888888	name1	name1	29-DEC-17	26	27893	8A6F38...
999999999999	name2	name2	29-DEC-17	32	32456	BFDEA3...
000000000000	name3	name3	29-DEC-17	38	42900	E32AE7...

```
ALTER TABLE cloned_table ADD age_range varchar2(20);
```

```
UPDATE cloned_table
SET num_range =
CASE
```

```
  WHEN age <= 20 THEN '0-20'
  WHEN age BETWEEN 21 AND 30 THEN '21-30'
  WHEN age BETWEEN 31 AND 40 THEN '31-40'
  WHEN age BETWEEN 41 AND 50 THEN '41-50'
  WHEN age BETWEEN 51 AND 60 THEN '51-60'
  WHEN age > 60 THEN '>60'
```

```
END
```

```
WHERE age_range IS NOT NULL;
```

IDCARD	FIRSTNAME	LASTNAME	BIRTHDATE	AGE	SALARY	HASHED	AGE_RANGE
111111111111	มนัส	ศิริศักดิ์	28-DEC-17	49	47069	7AE932...	
222222222222	วรศาสตร์	เกตุรัฐ	12-DEC-17	29	27115	5C4439...	
333333333333	อุษานนท์	เชตนะเดช	19-DEC-17	31	35561	3AB16F...	
444444444444	เมทินทร์	พนาลีผ่อง	23-DEC-17	32	38174	E9445B...	
555555555555	สมดวงณิน	อุสาศาสตร์	16-DEC-17	33	30047	F3A9FC...	
666666666666	เมษา	วรกิจ	11-DEC-17	43	41270	329A99...	
777777777777	พีชร์	ชจรศาสตร์	29-DEC-17	22	22000	C6B57C...	
888888888888	name1	name1	29-DEC-17	26	27893	8A6F38...	
999999999999	name2	name2	29-DEC-17	32	32456	BFDEA3...	
000000000000	name3	name3	29-DEC-17	38	42900	E32AE7...	

```
ALTER TABLE cloned_table ADD age_range varchar2(20);
```

```
UPDATE cloned_table  
SET num_range =  
CASE
```

```
  WHEN age <= 20 THEN '0-20'  
  WHEN age BETWEEN 21 AND 30 THEN '21-30'  
  WHEN age BETWEEN 31 AND 40 THEN '31-40'  
  WHEN age BETWEEN 41 AND 50 THEN '41-50'  
  WHEN age BETWEEN 51 AND 60 THEN '51-60'  
  WHEN age > 60 THEN '>60'
```

```
END
```

```
WHERE age_range IS NOT NULL;
```

IDCARD	FIRSTNAME	LASTNAME	BIRTHDATE	AGE	SALARY	HASHED	AGE_RANGE
111111111111	มนัส	ศิริศักดิ์	28-DEC-17	49	47069	7AE932...	41-50
222222222222	วรศาสตร์	เกตนุรัฐ	12-DEC-17	29	27115	5C4439...	21-30
333333333333	อุทยานนท์	เชตนะเดช	19-DEC-17	31	35561	3AB16F...	31-40
444444444444	เมทินทร์	พนาลีผ่อง	23-DEC-17	32	38174	E9445B...	31-40
555555555555	สมดวงณิน	อุสาศาสตร์	16-DEC-17	33	30047	F3A9FC...	31-40
666666666666	เมษา	วรภูมิ	11-DEC-17	43	41270	329A99...	41-50
777777777777	พีชร์	ชจรศาสตร์	29-DEC-17	22	22000	C6B57C...	21-30
888888888888	name1	name1	29-DEC-17	26	27893	8A6F38...	21-30
999999999999	name2	name2	29-DEC-17	32	32456	BFDEA3...	31-40
000000000000	name3	name3	29-DEC-17	38	42900	E32AE7...	31-40

Step 3



Anonymization



Data for anonymization

Import



Export



CSV

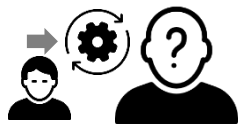
Anonymized data

หน่วยงานดำเนินการ

Anonymization

Anonymized Data

Step 3



Anonymization

หน่วยงานดำเนินการ

Input

		AGE	SALARY	HASHED
1	✓	29	27115	5C443954B201ECEBF9A341588E32AE2D506E800EB7ADF094F6F7B57536915BEA
2	✓	22	22000	C6B57C4641C46F467B424168A5F1133C42A8F8C1D93A75B5392CB3C96D058570
3	✓	26	27893	8A6F383B4BC570A0445C5E5052E02A1EE0488EC72868A453DE292BDC4C5EAD65
4	✓	31	35561	3AB16F2004B0A0294F033DB3255F54F2CC02D122DEDEC513523C6D372B01B281
5	✓	32	38174	E9445BD4FAF0C4E13431B8B5713D58CD9FC75873648F4F78566E325DCD0D3028
6	✓	33	30047	F3A9FC03725BDD82036DCB0FEF6ADFD27A6551B14648048DEC368C62C290F549
7	✓	32	32456	BFDEA38A8F5B65D465B24029065FB855A9C1049BAA2B82CCDBCBD486618F9C0
8	✓	38	42900	E32AE7EA65FF12E48250E2E2486813973BC3E9A8F42579BA9A5BF7C064F7B6FB
9	✓	49	47069	7AE93289F8682756DC57CF5CF55D7E7D84898E89CD8D81C230EDFB33090D7C1D
10	✓	43	41270	329A99A260D19701BF897A8B9A4972EA37709536BA50C82B684B620C46A1990A

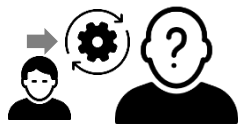
Output

		AGE	SALARY	HASHED
1	✓	[20, 60]	[20000, 30000]	5C443954B201ECEBF9A341588E32AE2D506E800EB7ADF094F6F7B57536915BEA
2	✓	[20, 60]	[20000, 30000]	C6B57C4641C46F467B424168A5F1133C42A8F8C1D93A75B5392CB3C96D058570
3	✓	[20, 60]	[20000, 30000]	8A6F383B4BC570A0445C5E5052E02A1EE0488EC72868A453DE292BDC4C5EAD65
4	✓	[20, 60]	[30000, 40000]	3AB16F2004B0A0294F033DB3255F54F2CC02D122DEDEC513523C6D372B01B281
5	✓	[20, 60]	[30000, 40000]	E9445BD4FAF0C4E13431B8B5713D58CD9FC75873648F4F78566E325DCD0D3028
6	✓	[20, 60]	[30000, 40000]	F3A9FC03725BDD82036DCB0FEF6ADFD27A6551B14648048DEC368C62C290F549
7	✓	[20, 60]	[30000, 40000]	BFDEA38A8F5B65D465B24029065FB855A9C1049BAA2B82CCDBCBD486618F9C0
8	✓	[20, 60]	[40000, 50000]	E32AE7EA65FF12E48250E2E2486813973BC3E9A8F42579BA9A5BF7C064F7B6FB
9	✓	[20, 60]	[40000, 50000]	7AE93289F8682756DC57CF5CF55D7E7D84898E89CD8D81C230EDFB33090D7C1D
10	✓	[20, 60]	[40000, 50000]	329A99A260D19701BF897A8B9A4972EA37709536BA50C82B684B620C46A1990A

Anonymization

Export Data to CSV

Step 3



Anonymization

หน่วยงานดำเนินการ

Export Wizard - Step 1 of 3

Source/Destination

Connection: SYSTEM

☐ Export DDL

☒ Pretty Print ☒ Show Schema ☐ Grants ☒ Add BYTE keyword

☒ Terminator ☒ Add Force to Views ☐ Drops ☐ Cascade Drops

☒ Storage ☐ Dependents

☒ Export Data

Format: csv ☒ Header

Delimiter: , Line Terminator: environment default

Left Enclosure: " Right Enclosure: "

Save As: Single File ☐ Compressed Encoding: UTF-8

File: C:\Users\Administrator\done_table.csv Browse...

☐ Proceed to summary.

Help < Back Next > Finish Cancel

Export Wizard - Step 2 of 3

Specify Data

Source/Destination Specify Data Export Summary

Data Columns

Columns

☒ All

☐ IDCARD

☐ FIRSTNAME

☐ LASTNAME

☐ BIRTHDATE

☒ AGE

☒ SALARY

☒ HASHED

Help OK

Global Where:

Help < Back Next > Finish Cancel

PRIVACY PROTECTION PROCESS

กระบวนการปกป้องข้อมูลที่เป็นส่วนบุคคล

สรุปขั้นตอนสำคัญที่หน่วยงานต้องดำเนินการ



แต่ละหน่วยงานเป็นผู้จัดทำพร้อมก่อนนำเข้าสู่ระบบ

Step 1



จัดเตรียมรูปแบบข้อมูล
(Data Format Preparation)

เช่น เลขบัตรประชาชน
รวมถึงข้อมูลที่ได้รับตามโจทย์

Step 2

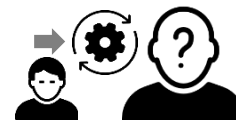


Hashing CID

แปลงข้อมูล (Data) ที่ระบุตัวตน

SHA-256 w/ Salt#1

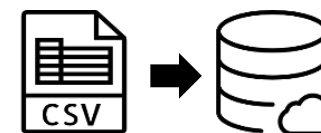
Step 3



Anonymization

Step

4

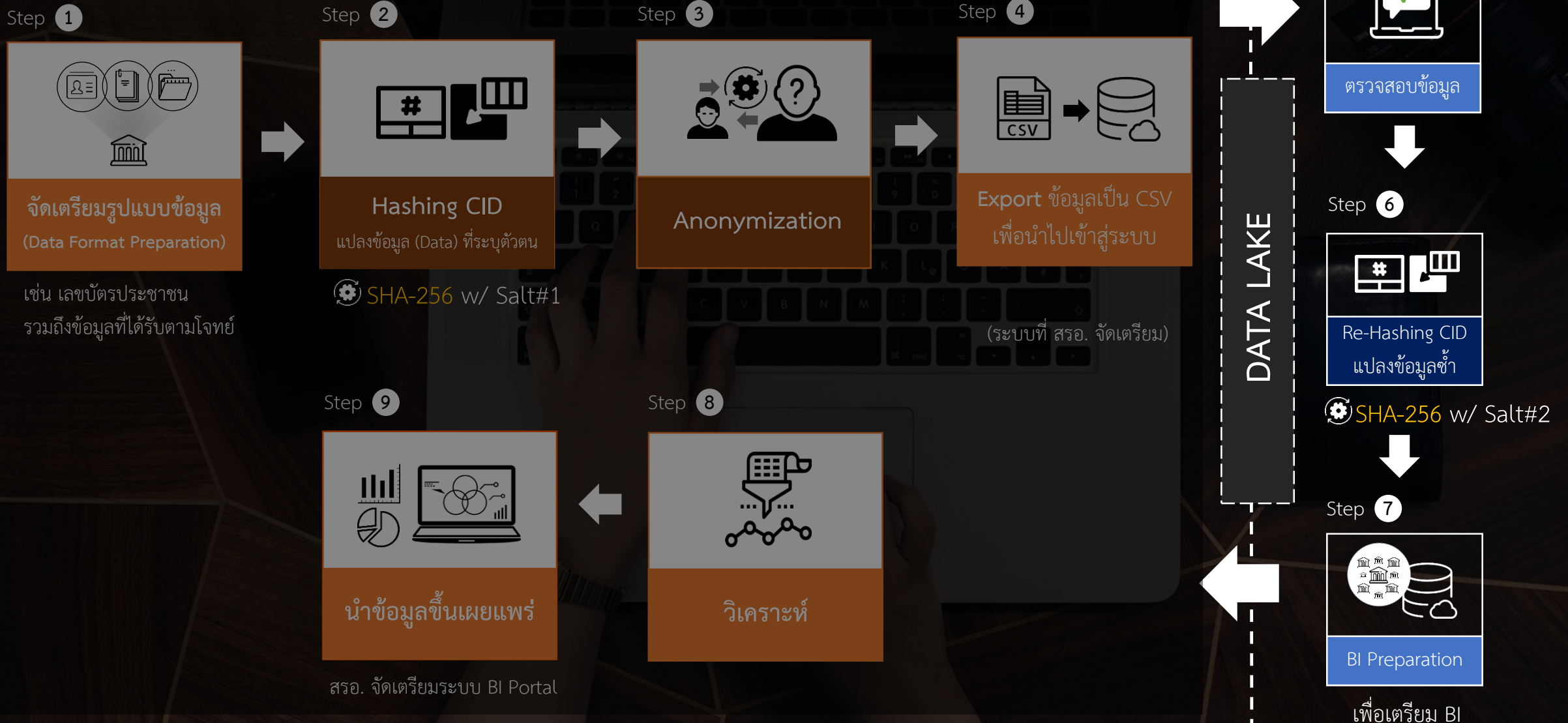


Export ข้อมูลเป็น CSV เพื่อ
นำไปเข้าสู่ระบบ

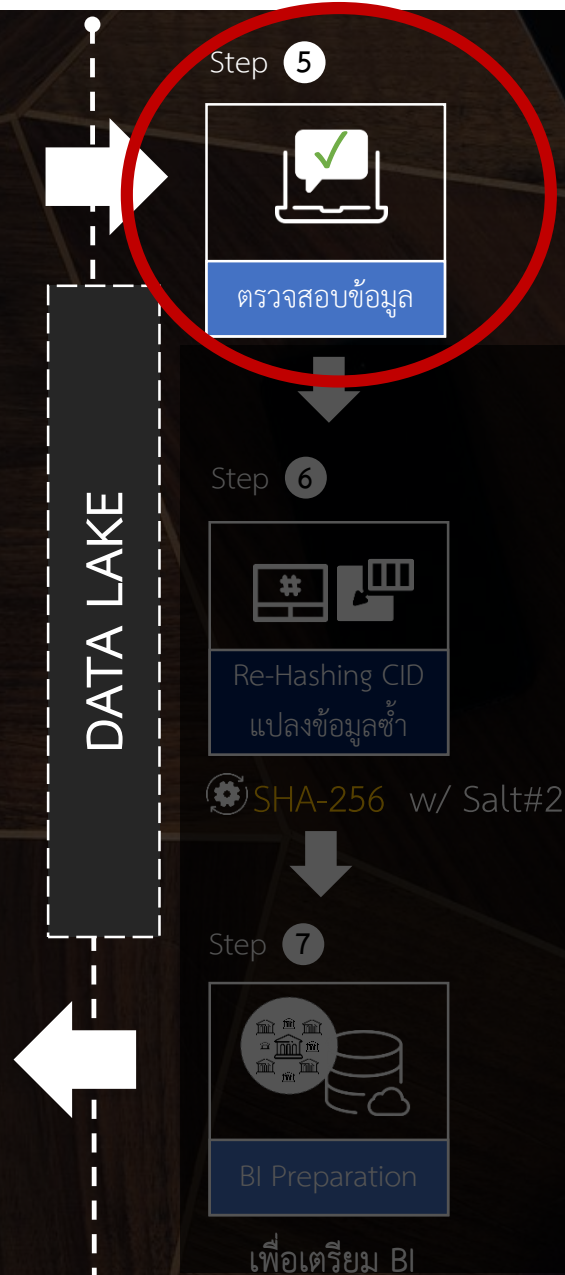
*** ในไฟล์ที่นำมาใช้ต้องไม่มีข้อมูลที่สามารถระบุถึงตัวตนของบุคคล ***

Procedure and Architecture

■ หน่วยงานดำเนินการ ■ สรอ. ดำเนินการ



Procedure and Architecture

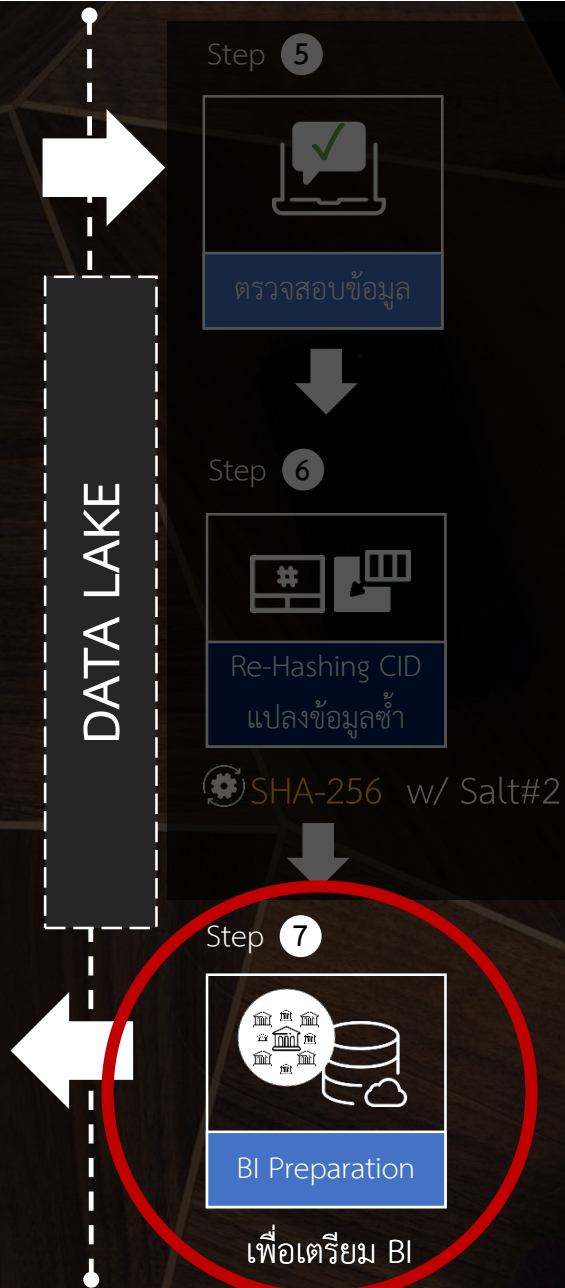


การตรวจสอบ	Detail	หน่วยงาน	EGA ตรวจสอบ
ตรวจสอบจำนวนข้อมูล	จำนวนข้อมูลครบตามที่หน่วยงานแจ้งหรือไม่	แจ้ง 1,000,000 record	ยืนยันจำนวนเท่ากัน
Hashing	ข้อมูลถูกต้องหรือไม่	A00000	ยืนยันข้อมูลจาก verifier records
การเชื่อมโยงข้อมูลเบื้องต้น	ทดลองหาความสัมพันธ์ข้อมูลระหว่างหน่วยงานเพื่อตรวจสอบว่าข้อมูลที่ hash มานั้นถูกต้อง	-	นำข้อมูลของหน่วยงาน A มา join กับข้อมูลของหน่วยงาน B
ข้อมูลซ้ำ	ตรวจสอบข้อมูลซ้ำเช่น เลขบัตรประชาชนที่ hash แล้ว	-	ตรวจสอบข้อมูลมี duplicate หรือไม่

Procedure and Architecture



Procedure and Architecture



เตรียม	Detail	EGA	หน่วยงาน
กำหนดสิทธิ์	กำหนดสิทธิ์การเข้าถึงข้อมูลของแต่ละหน่วยงาน ตามที่ตกลงกันระหว่างหน่วยงาน	เพิ่ม users และสิทธิ์ในการเข้าถึงข้อมูล	เข้าถึงข้อมูล
ทดสอบความสัมพันธ์ข้อมูลหลังจาก hash ครั้งที่ 2	ข้อมูลทดสอบต้องมีจำนวน record เท่ากันกับที่ hash มาในรอบแรก	ทดสอบโดยการนำข้อมูลของหน่วยงาน A มา join หน่วยงาน B	-

Procedure and Architecture

■ หน่วยงานดำเนินการ ■ สรอ. ดำเนินการ

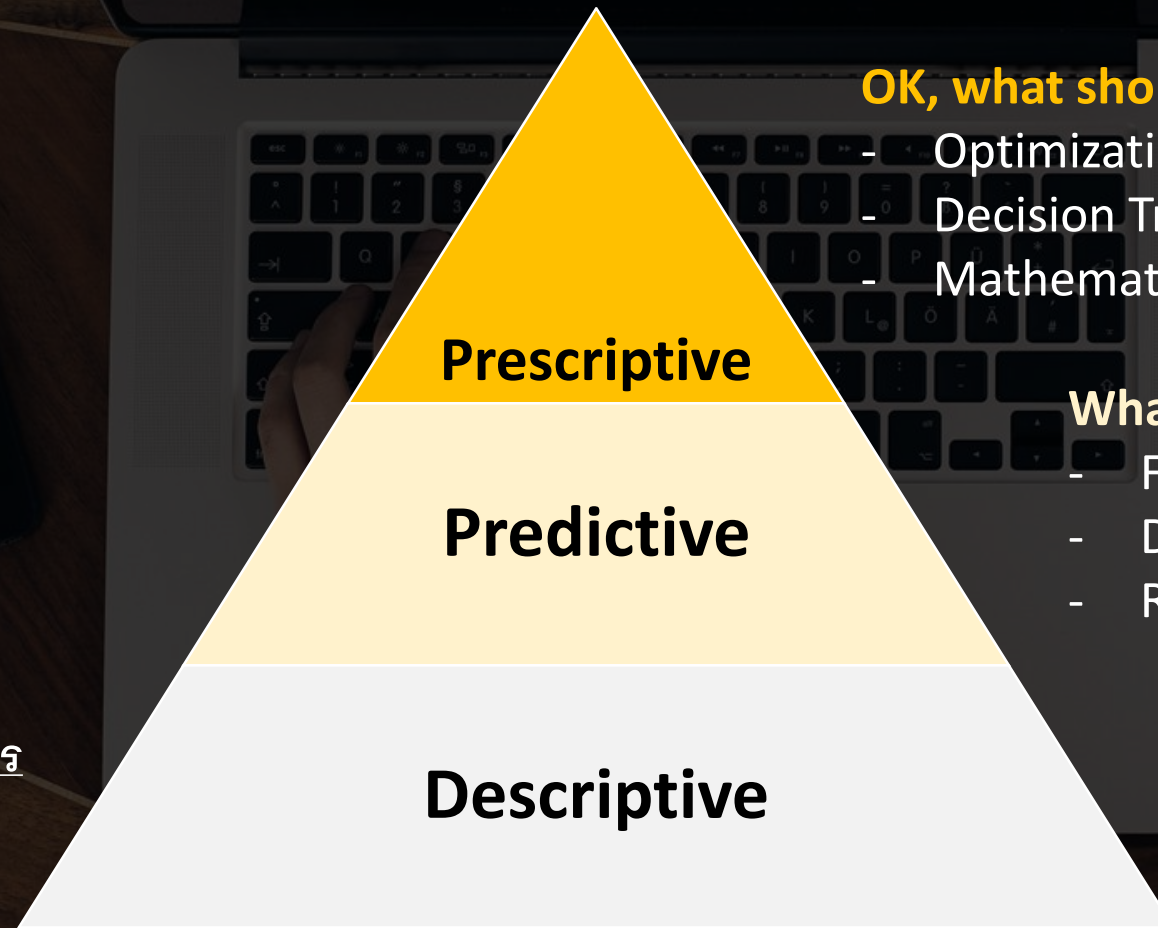


Step 8



วิเคราะห์

หน่วยงานดำเนินการ



OK, what should we do ?

- Optimization
- Decision Trees
- Mathematical Programming

What going to happen ?

- Forecasting
- Data Mining
- Regressions

What's going on ?

- Business Intelligence (BI)
- Dashboard

Step

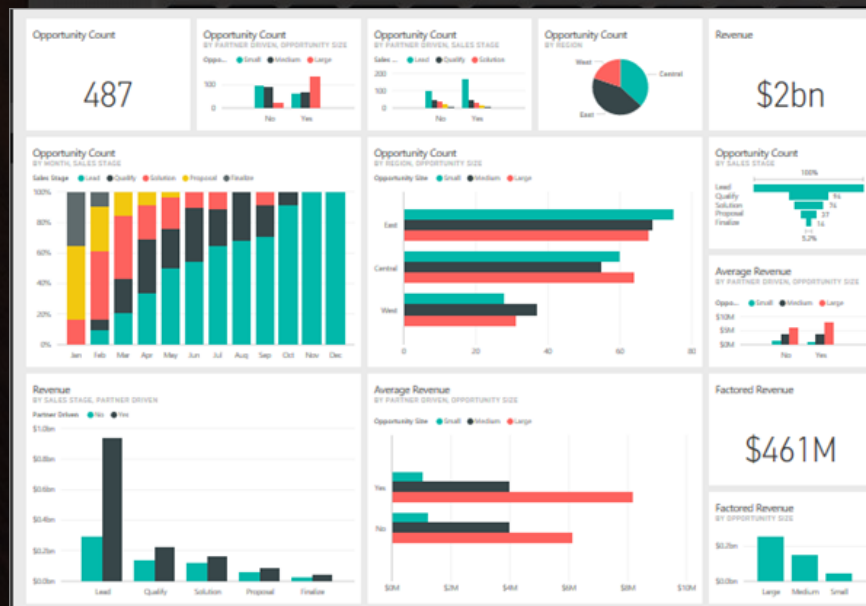
8



วิเคราะห์

หน่วยงานดำเนินการ

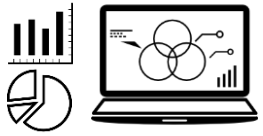
Business Intelligence (BI) and Dashboard



Business Intelligence หรือ BI คือ เทคโนโลยีที่ช่วยในการสรุปภาพรวมของข้อมูลทางธุรกิจในหลายมิติ เพื่อสะท้อนผลการดำเนินงาน และสนับสนุนทางเลือกในการตัดสินใจและช่วยในการวางแผนทางธุรกิจ หรือนโยบายขององค์กร

Step

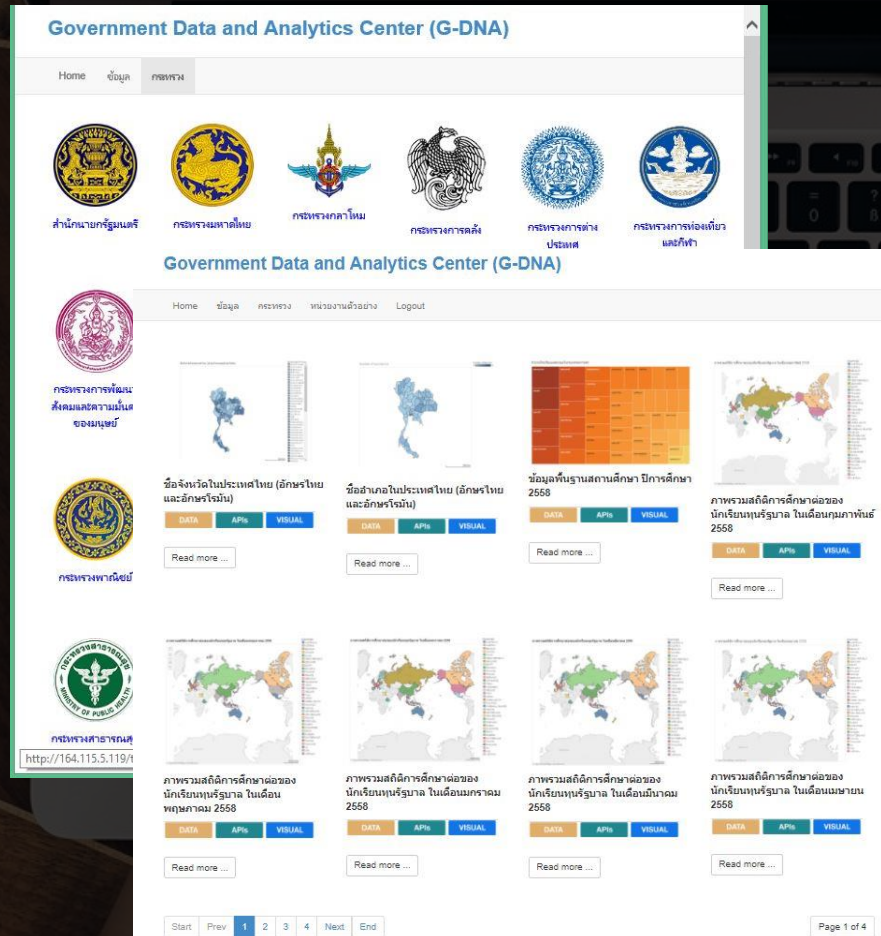
9



นำข้อมูลขึ้นเผยแพร่

สรอ. จัดเตรียมระบบ BI Portal

หน่วยงานดำเนินการ



ข้อมูลที่ทำการวิเคราะห์กระทั่งอยู่ในลักษณะของ BI, dashboard ต่างๆ แล้วหน่วยงานสามารถนำไปใช้ประโยชน์ ไม่ว่าจะเป็นการนำเสนอผู้บริหาร เพื่อใช้พิจารณาประกอบการตัดสินใจในการออกนโยบายต่างๆ รวมถึง publish บน website ของหน่วยงานเพื่อเผยแพร่ข้อมูลให้ประชาชนรับทราบได้

THANK YOU

DATA PROCESSING IN BIG DATA PROJECT

การดำเนินงานด้านการจัดการข้อมูลขนาดใหญ่ภาครัฐ

(Government Big Data)