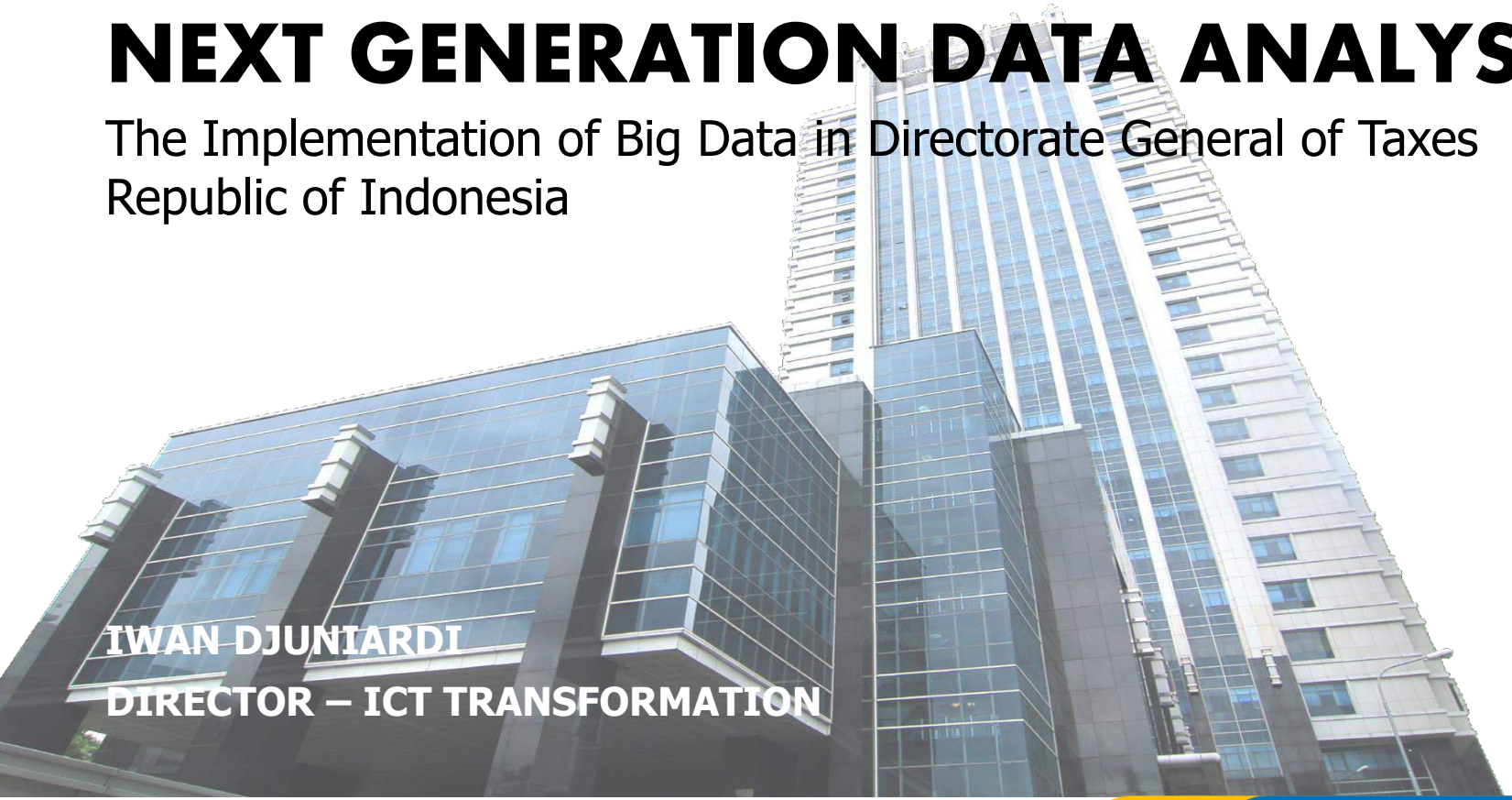
NEXT GENERATION DATA ANALYSIS

The Implementation of Big Data in Directorate General of Taxes Republic of Indonesia



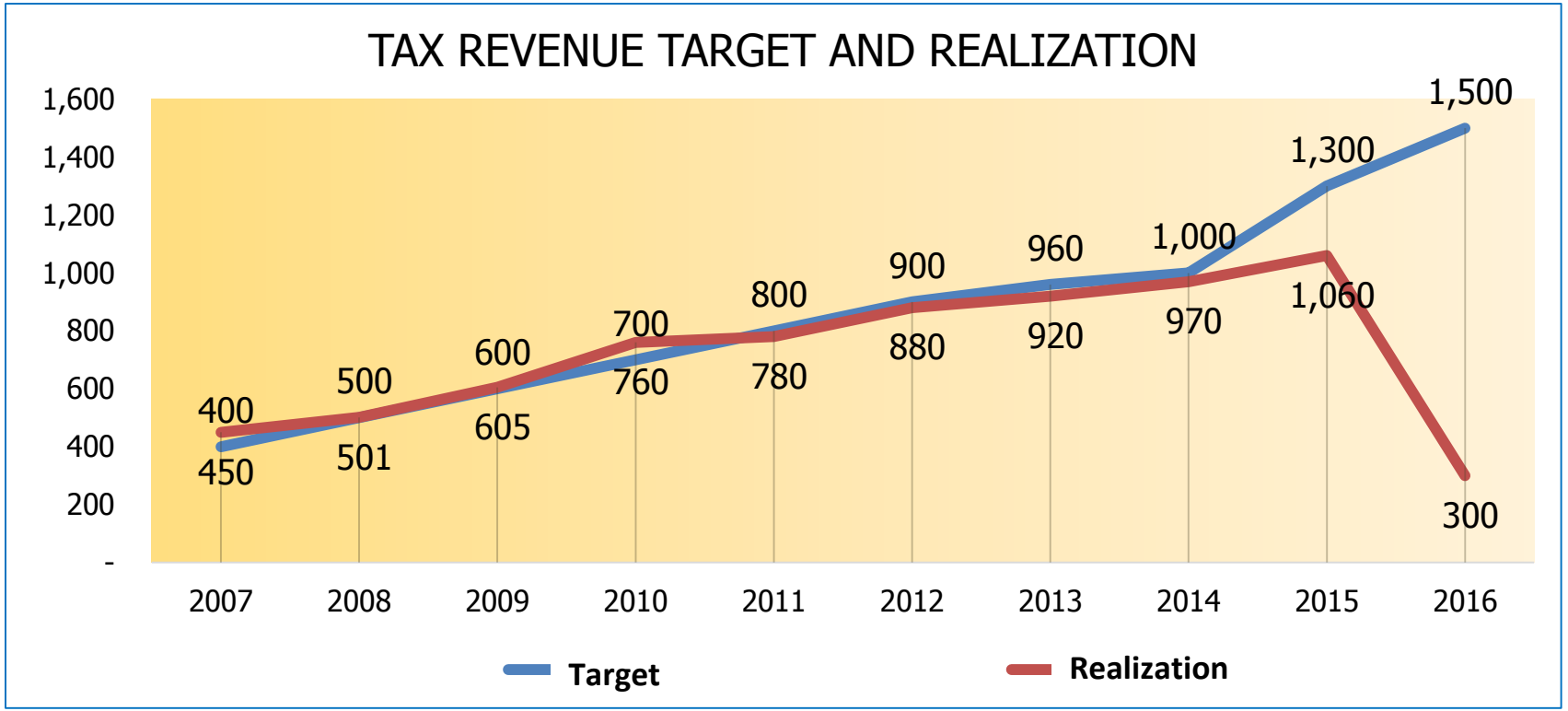
IWAN DJUNIARDI

DIRECTOR – ICT TRANSFORMATION

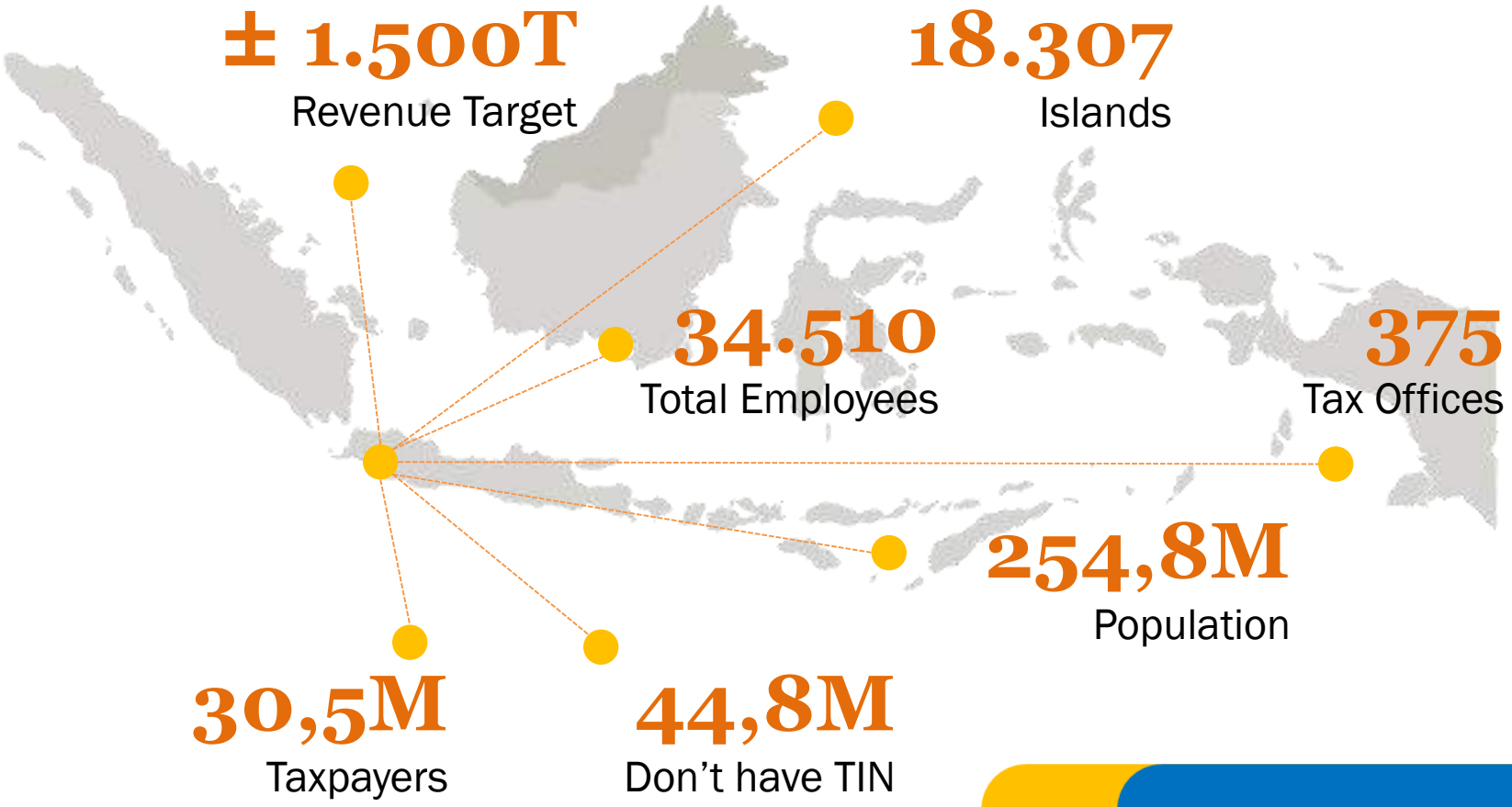
Directorate General of Taxes, Republic of Indonesia
2016

BACKGROUND

There were continuous increases on the revenue target

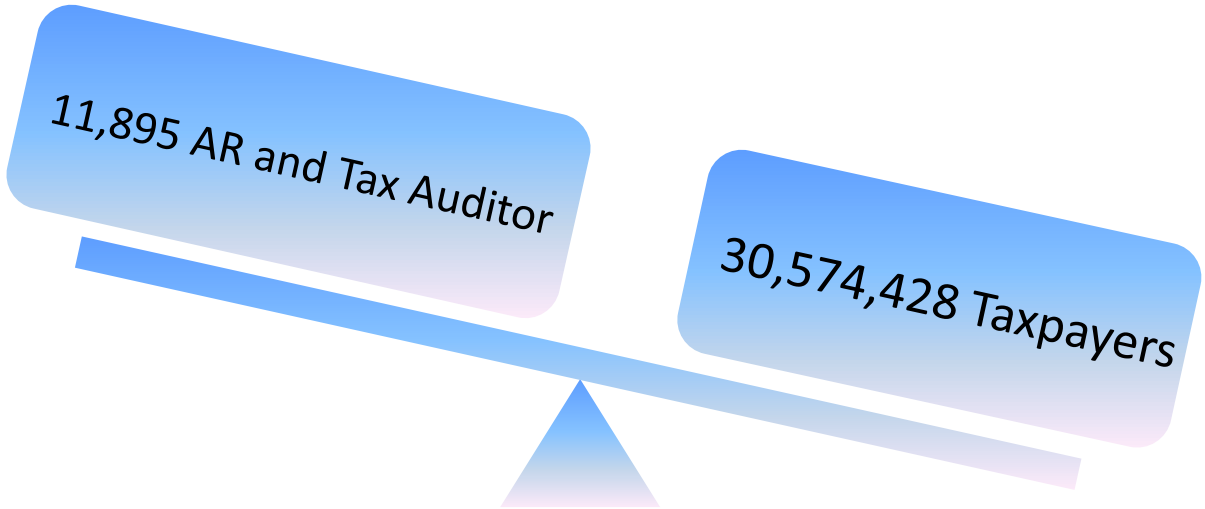


BACKGROUND


± 1.500T
Revenue Target
18.307
Islands
34.510
Total Employees
375
Tax Offices
254,8M
Population
30,5M
Taxpayers
44,8M
Don’t have TIN

BACKGROUND

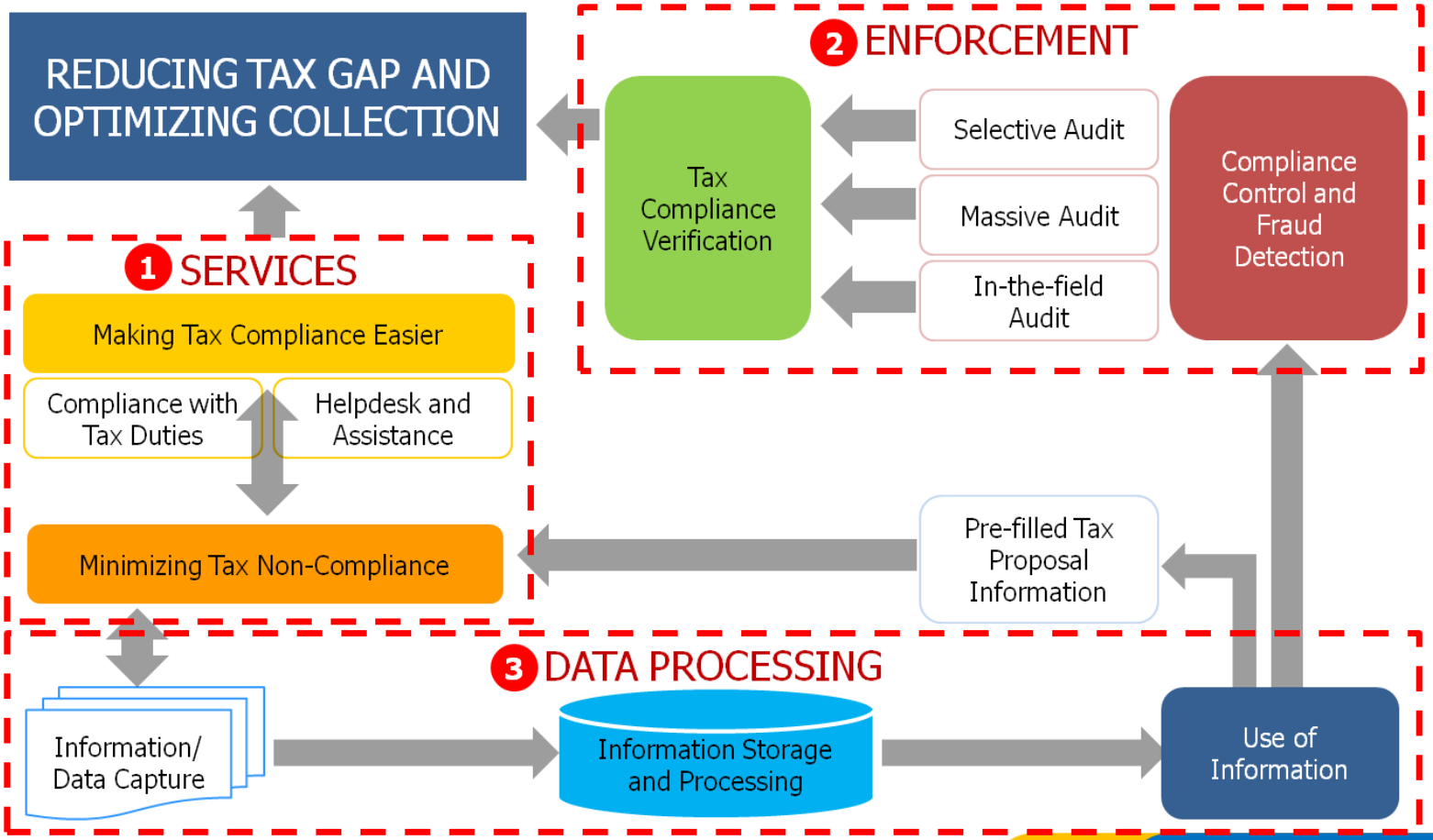
The lack number of employees (in DGT) to look after the taxpayers



The huge amount of data used on the law enforcement activity and tax potential excavation

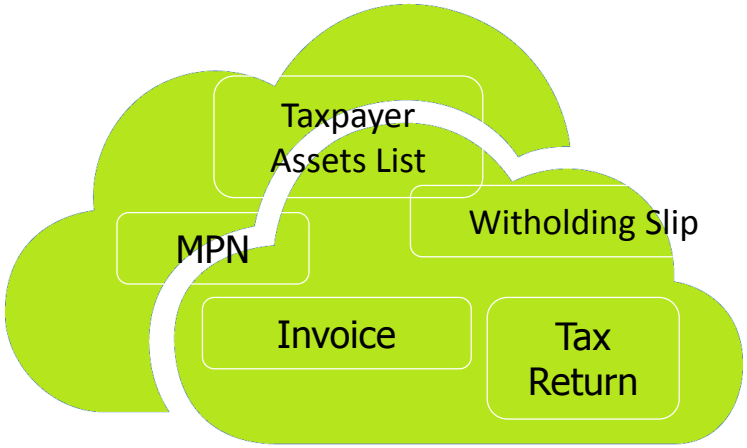
- 1,1 Billions Records of Internal Data per year
- 63 Different Agencies and Social Media Data
- unstandardized data, no single ID

TAX ADMINISTRATION BUSINESS MODEL

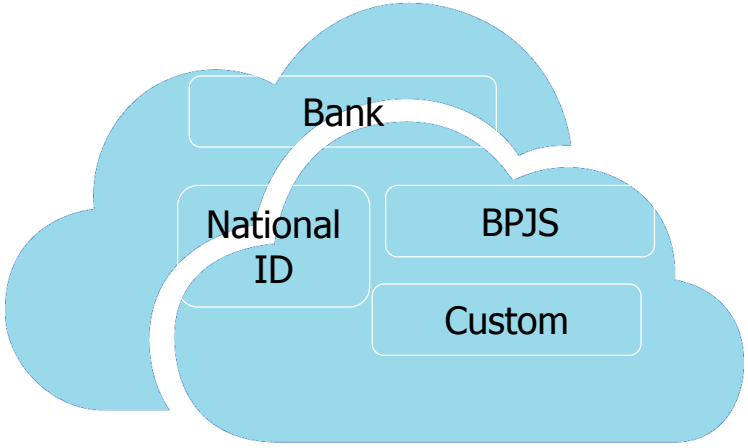


Implementation Purpose

- Collecting data from various sources, both internal and external
- Identification, investigation, and escalate the tax revenue potential
- To implement a kind of management information system that has a single case management and a system of predictive analytics based on risk assessment



INTERNAL



EXTERNAL

BIG DATA IMPLEMENTATION

Starting Point

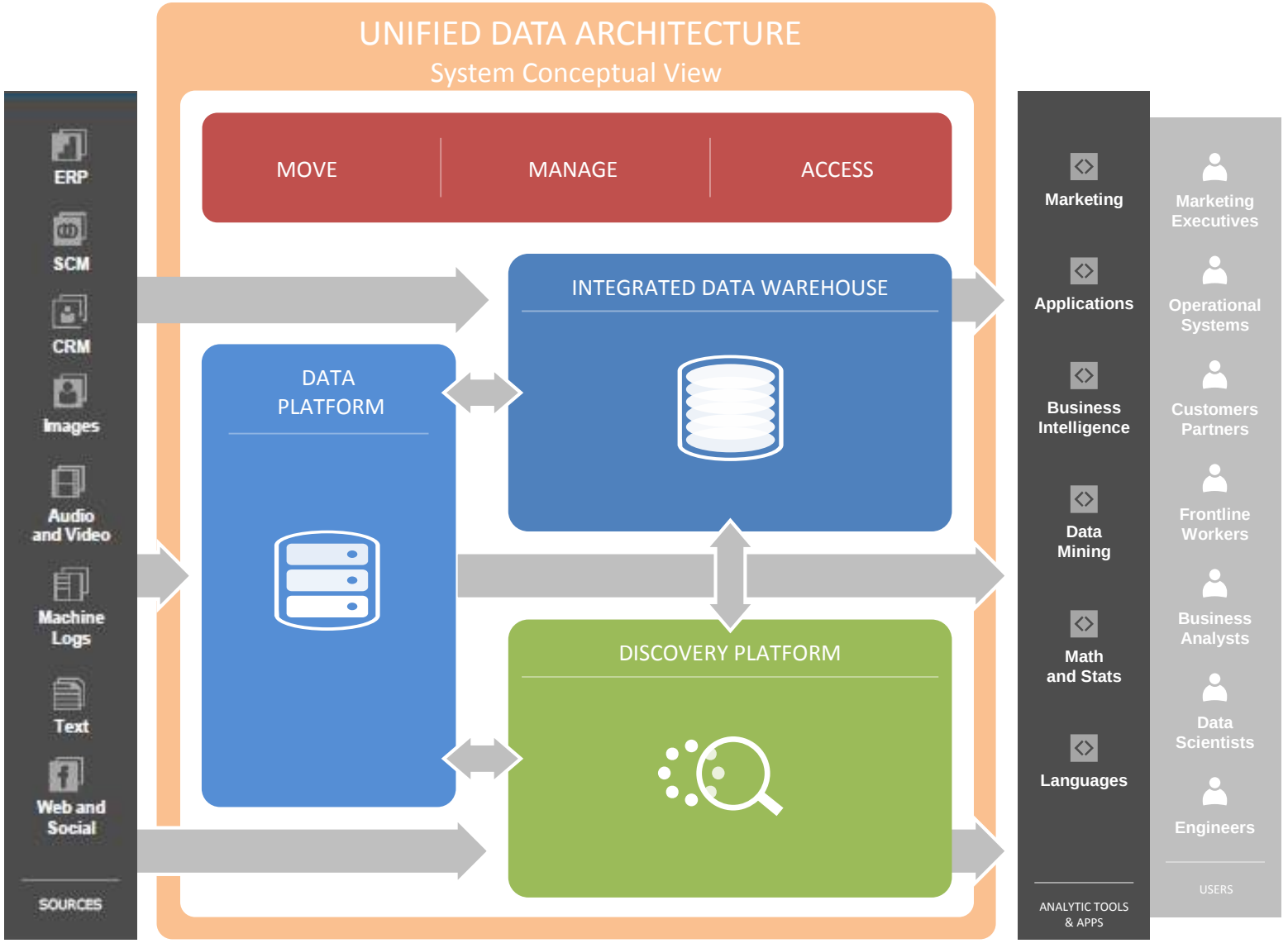
- 10 personal Computer with i7 intel processor, each has initial storage capacity of 1 TB and 8 GB RAM (10 TB storage and 80 GB RAM in total)
- OS : open source
- Hadoop version : open source
- RDBMS : open source

a small step for of a better larger data management

BIG DATA IMPLEMENTATION

Current Stage

- DGT currently operates Enterprise Data Warehouse which has 2000 GB RAM and storage capacity of 500 TB used (Hadoop 300 TB and RDBMS 200 TB)
- Big Data, has a very important role for the DGT especially in data virtualization process and throughout the preparation of consolidated report
- DGT currently administers approximately
  - 7 billion records of Tax Return data (total)
  - 337 million records of Electronic Tax Invoice data (2014-2016)
  - 5 million records of Tax Letter (SKP/STP) data (total)
  - 255 million records of National Identity Card data (2015-2016)
  - 51 million records of Custom data (2006-2016)
  - 35 million records of Credit Card data (per year)
  - 30.5 million records of other banking data (2015-2016)


UNIFIED DATA ARCHITECTURE
System Conceptual View
MOVE
MANAGE
ACCESS
INTEGRATED DATA WAREHOUSE
DATA PLATFORM
DISCOVERY PLATFORM
ERP
SCM
CRM
Images
Audio and Video
Machine Logs
Text
Web and Social
SOURCES
Marketing
Applications
Business Intelligence
Data Mining
Math and Stats
Languages
ANALYTIC TOOLS & APPS
Marketing Executives
Operational Systems
Customers Partners
Frontline Workers
Business Analysts
Data Scientists
Engineers
USERS

BIG DATA IMPLEMENTATION

Next Stage

- Integrate all of the existing applications into an integrated Information Systems which serves as tax intelligence center
- Tax service will be made wider and easier so that the data collection activities can be held more effectively
- A system of predictive analytics based on risk assessment and self-service business intelligence will continue to be developed which in turn is expected to increase the tax revenue

Case study

Several question that can be answered by Big Data Analytics

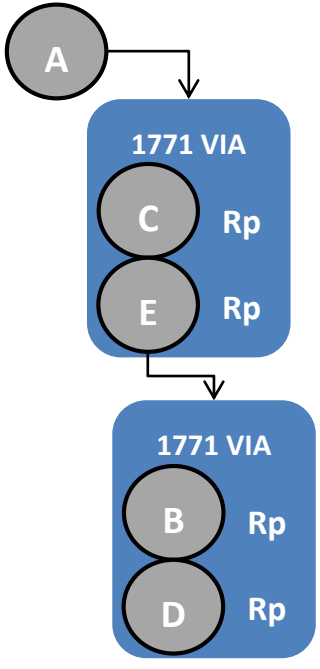
What would DGT do if knew

Company ownership structure ?

Identification – Company ownership structure

Company can own a share of other company which recorded in form PPH1771 VIA (Corporate Income Tax Return)

DGT found **1296** *company in sample data who have ownership in other company*



From Reporting perspective, can be identified how many company recorded in PPH 1771 VIA

What insight that can be discovered further ?

Discovery – Company ownership structure

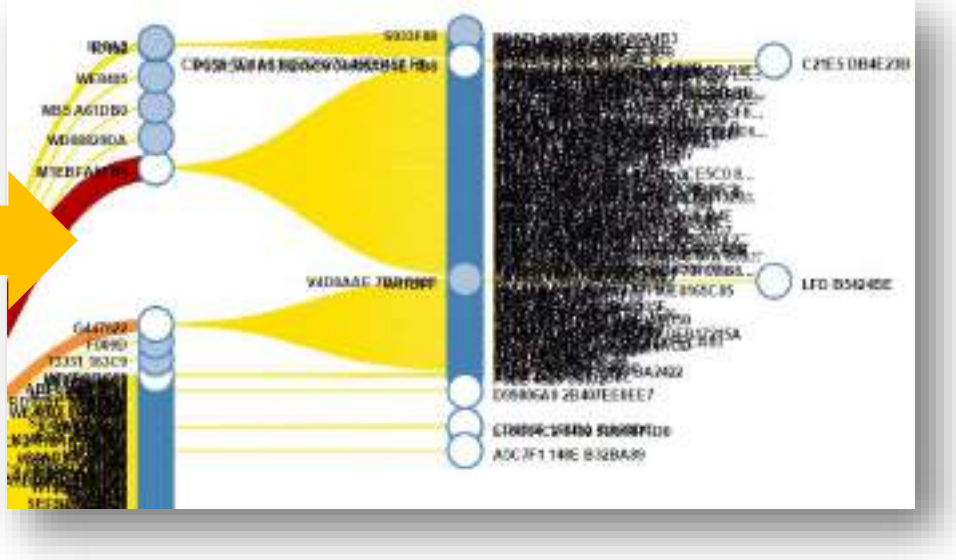
```
select * from ntree(
on pajak_sri.pph1771va_cleaned
        partition by 1
        root_node(nama_pemegang_saham !='')
        node_id(nama)
        parent_id(nama_pemegang_saham)
        mode('down')
        allow_cycles('true')
        starts_with(nama!='')
        output('all')
        result(path(nama) as path, level(*))
)
where level >1;
```

Big Data Analytics can show company ownership structure more than 1 level (multilevel)

Results :

No	Path
1	ID6A8 --> A92681 --> M507F760AC088D 536A1520A F3C6
2	ID6A8 --> A92681 --> A2A06E707F
3	ID6A8 --> A92681 --> GF58513E8656D
4	ID6A8 --> A92681 --> P5314DF6C41B1D6121
5	ID6A8 --> A92681 --> A3AF1C3198120
6	IC152 --> S93F88 --> TF18BF405FABA 6CDAE
7	IC152 --> S93F88 --> IFD4EA564D70
8	IC152 --> S93F88 --> S94ESAB8EAB02D
9	WE8405 --> B03E4FA92 --> E83639
10	NB5A61D80 --> P4E57 --> C766C134272F0180C
11	WD08829DA --> M25CCC --> PC7D350AF
12	WD08829DA --> M25CCC --> SBC0A 78A9018AS1088197

Each line represent ownership of other company



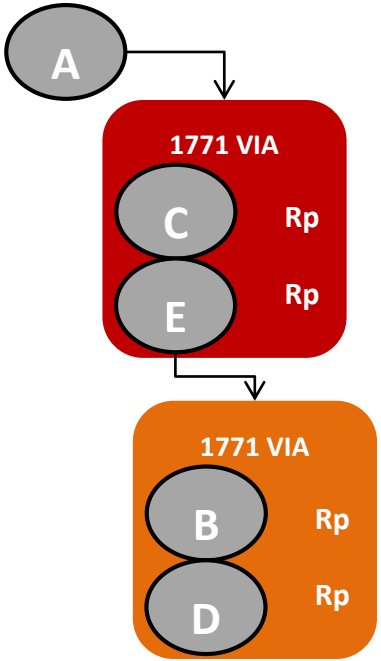
Potential Information– Company ownership structure

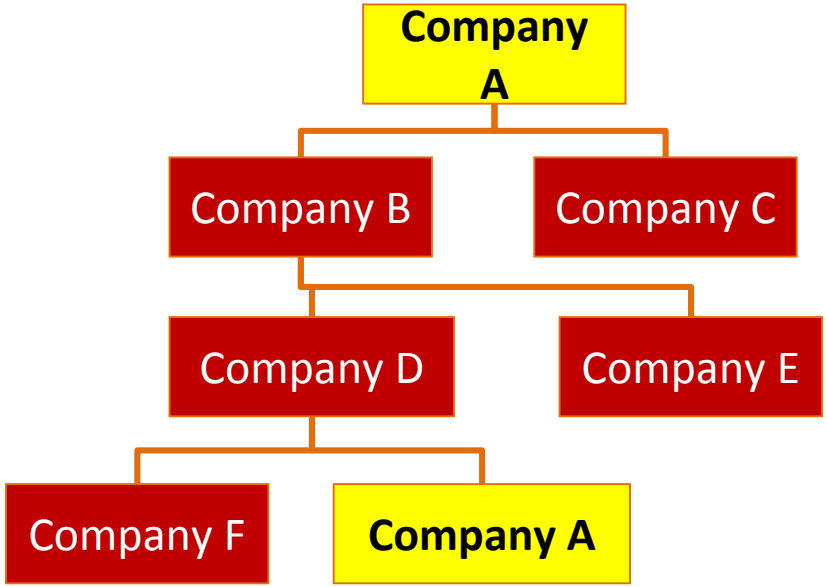
Whether company own or owned by a suspected fraud company?

Fraud anticipation on company who ownership relation with the suspected fraud company

Potential Information – Company ownership structure

Through the Big Data analysis, we can trace the parent-subsidiary shareholding to look for correlations of association arising therein to detect transfer pricing and fraudulant modus





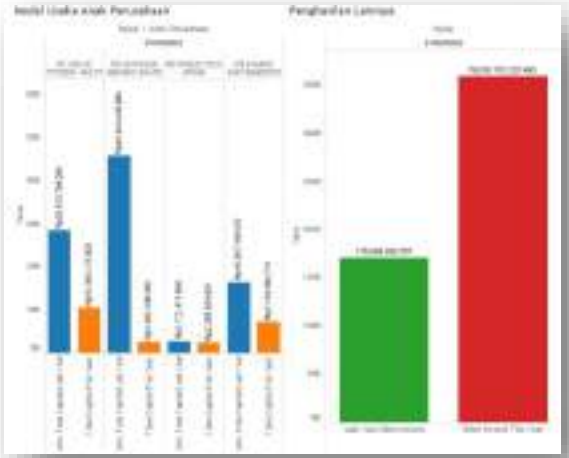
Potential information – Company sold their asset but didn't Record an additional income

According to regulation / *Undang undang pajak penghasilan (UU No.36/2008)- Pasal 4 (d)(1)*, one of Tax object : ***Income from assets transfer to other companies as share enclosing***

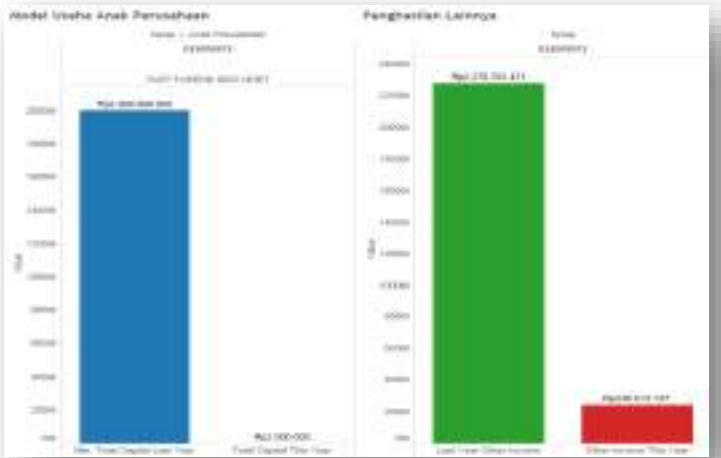
Which company sold their asset / share ownerships but **didn't experience significant increase on their income?**

Early detection to indicate unreported income

Company who recorded an additional income in Other income column



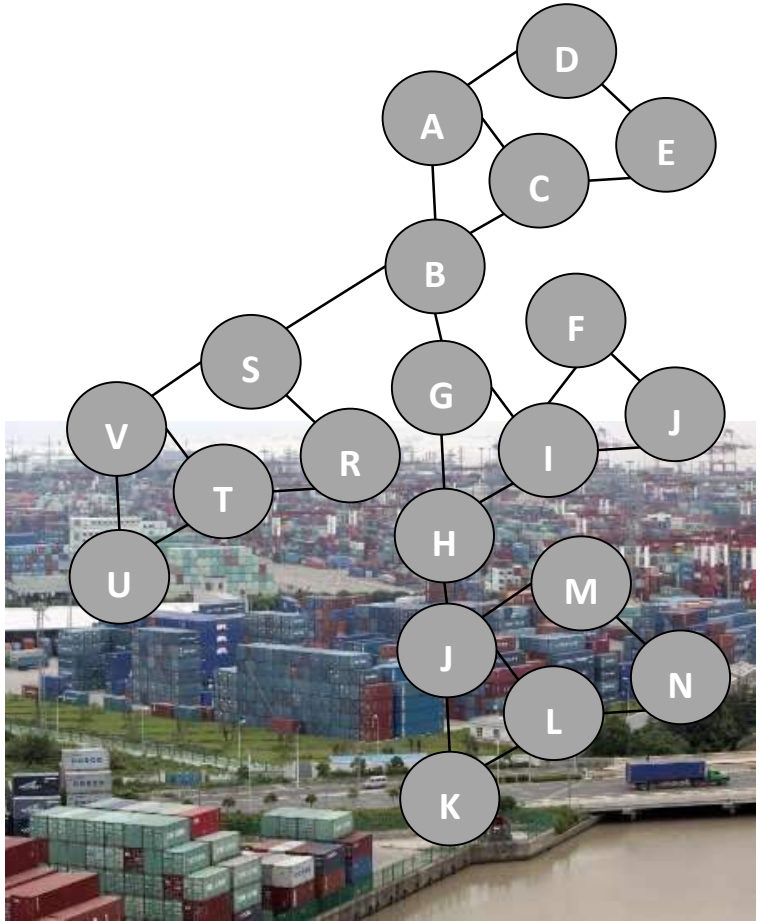
Company who didn't record an additional income in Other income column



What would DGT do if knew
Transaction relation between company?

Identification – Transaction relation between company

Company can be connected to others company trough transaction

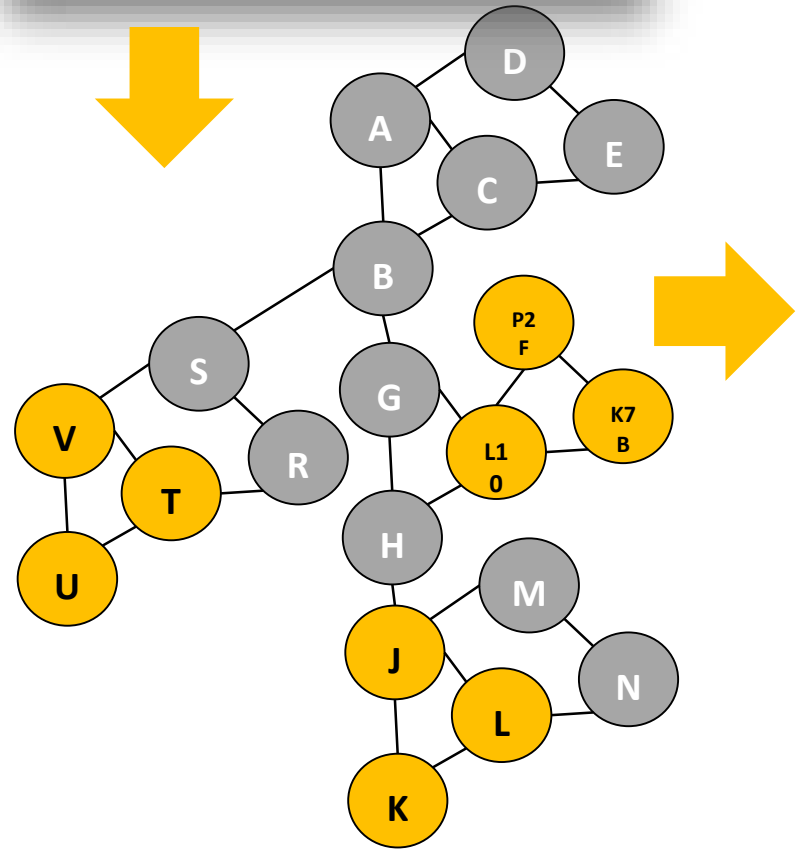


From the sample data, we found **103 million transaction** from **23,128 companies**

If we see from reporting perspective, there are million of connection within thousand of companies

What insight that can be discovered further ?

```
SELECT * FROM Triangle_Finder (
       ON (SELECT 1)
       PARTITION BY 1
       database('beehive')
       userid('*****')
       password('*****')
       inputTable('pajak_sri.augmented_ppn1111b2_edge_small')
       outputTable('pajak_sri.ppn1111b2_triangles')
);
```



Discovery – Transaction relation between company

Big Data Analytics can *find relation between company who form triangle or rectangle by the connected transaction*

Result :

No	Node 1	Node 2	Node 3
1	P2F73310C7AC	K7BD2 D6302F9581A	L10DC 9356877E 7742F8ADA
2	P2F73310C7AC	K7BD2 D6302F9581A	L10DC 9356877E 7742F8ADA
3	P2F73310C7AC	K7BD2 D6302F9581A	L10DC 9356877E 7742F8ADA
4	T2D9DF A6	K7BD2 D6302F9581A	L10DC 9356877E 7742F8ADA
5	T2D9DF A6	K7BD2 D6302F9581A	L10DC 9356877E 7742F8ADA
6	T2D9DF A6	K7BD2 D6302F9581A	L10DC 9356877E 7742F8ADA

Each line represent triangle connection between company

Potential information– Transaction relation between company

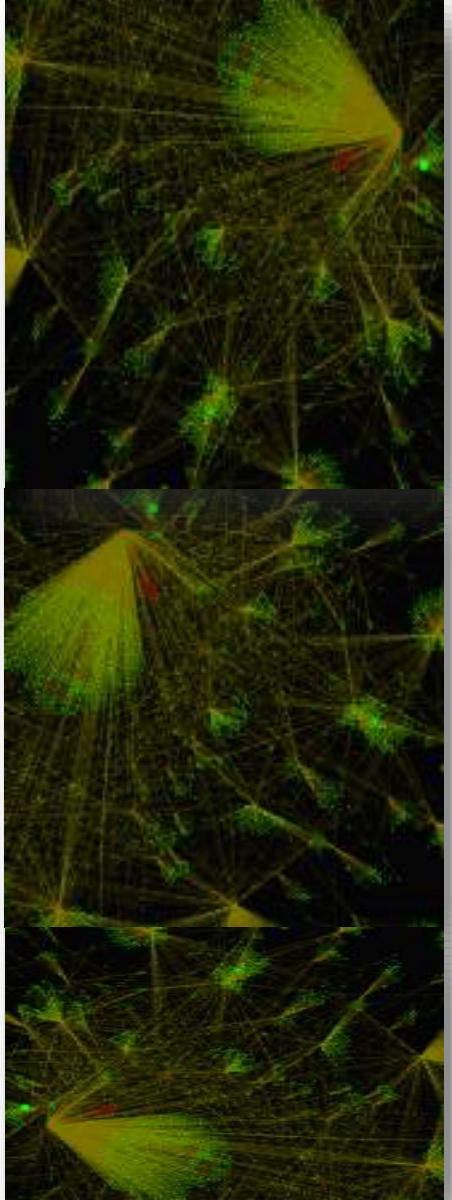
- Which company that have connection with fraud suspected company?

Extend the depth audit to the connected company

- Is there any company who have been connected with three other companies with different industry (business line) ?

Indication of fictitious transaction between the company

What can we find if we can **visualize transaction between company** ?



Identification – Transaction between company in food industry

Data are taken from transaction between company which are recorded in form 1111a2 (Output VAT) dan 1111b2 (Input VAT)

Business group of **food industry** (business group code 10) year **2011**

There are **1 million ++** of transaction from **400++** companies

Discovery – Transaction between company in food industry

Big Data Analytics can dig up some inter data correlation, especially for transfer pricing prevention. Each transaction pool could be mapped, clarifying the transfer pricing indication

What other **interesting** things can be done?

Early detection on **transfer pricing** with additional data **of overseas transaction** and detail of sold goods / item

Equalization of salary data between the reported salary in **Witholding Tax Return** and **Income Tax Return** **to identify the difference**

Indication of profit manipulation by identify uncommon **increase or decrease of profit**, from year to year

Detection the possibility of company to have a fraud based on their relation with fraud suspected company

Detection the possibility of company to have a fraud based on similar pattern with fraud suspected company

CONCLUSION

- Analysis process by the Big Data is effective to speed up the achievement of DGT's target
- Big data is relatively fast, accurate, and reliable when used in analysis process
- With its capability to recognize pattern, Big Data can reveal more about tax fraud / tax avoidence indication
- From the standpoint of MIS, Big Data is an example of relevant efforts in accelerating the fulfillment of key objectives

THANK YOU