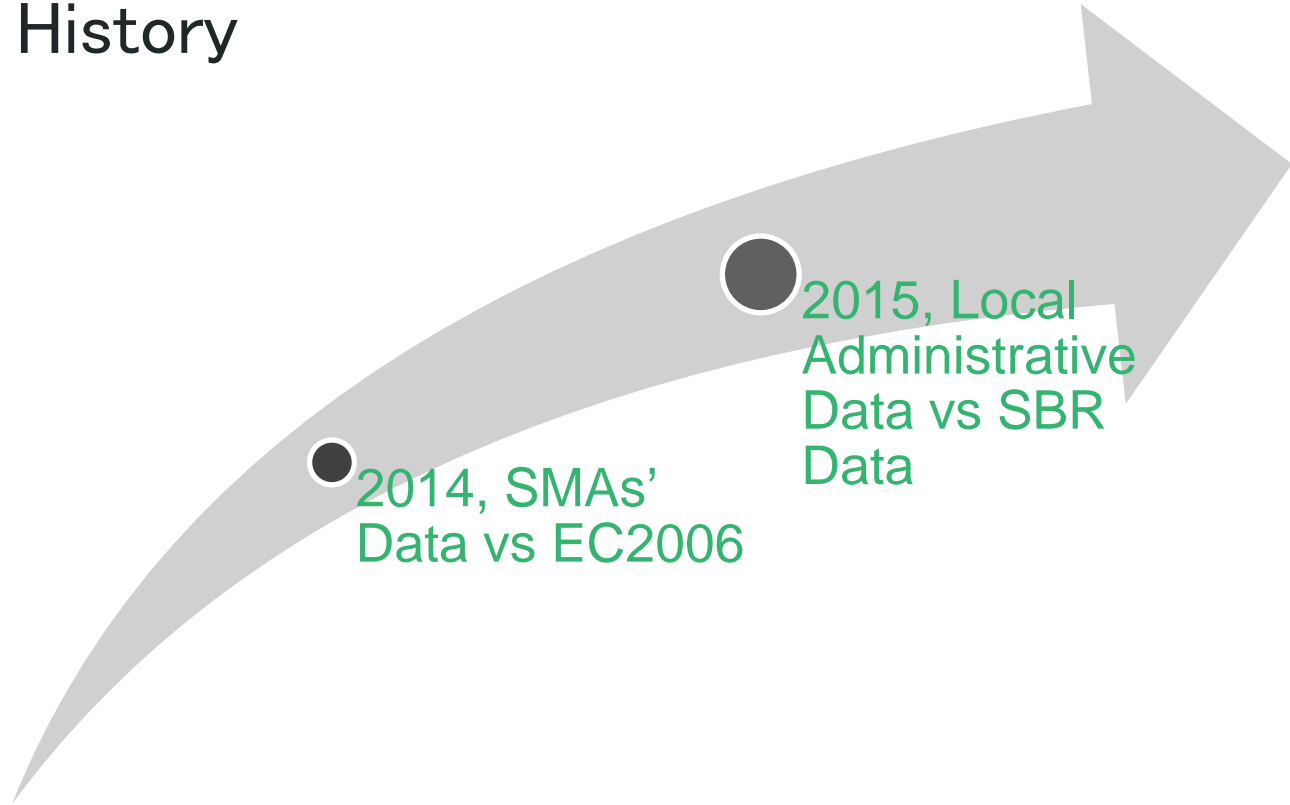# Tuning SBR Matching Feature

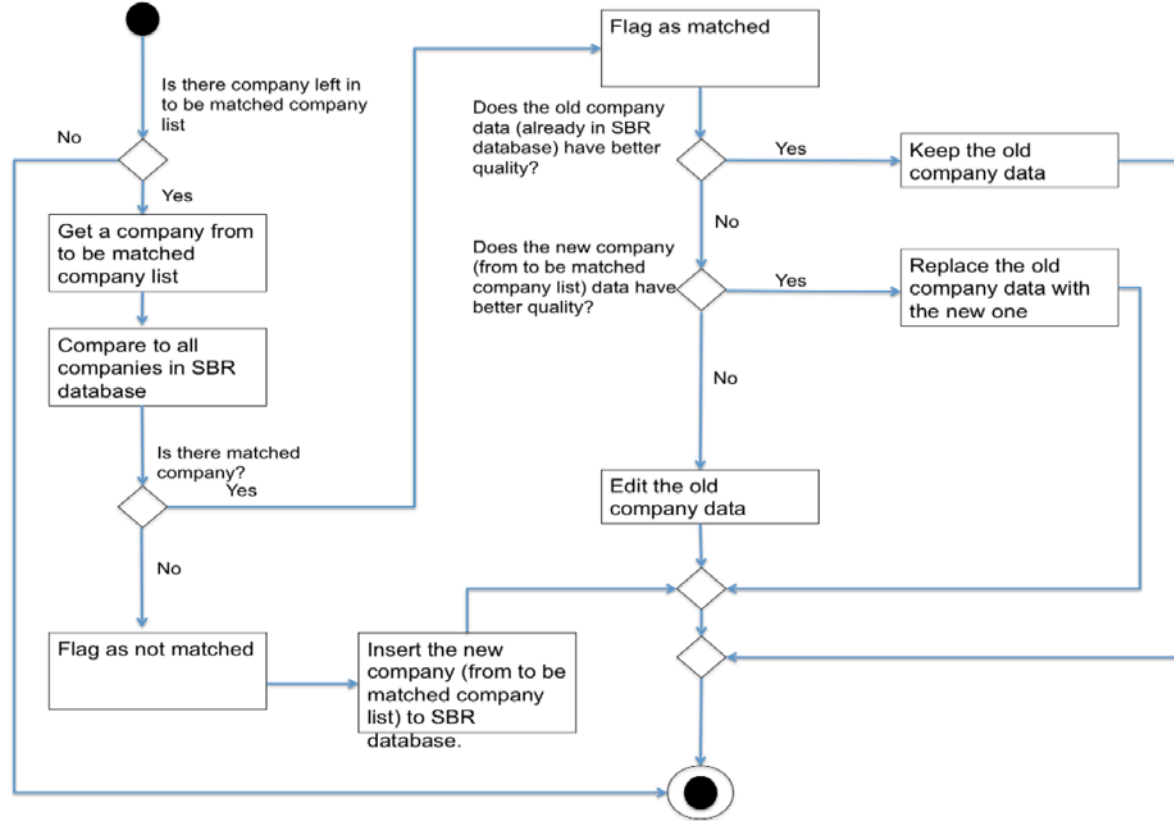BPS-Statistics Indonesia

Rr. Nefriana, Said Mirza Pahlevi, Irien Kamaratih

Matching History

2014, SMAs' Data vs EC2006

2015, Local Administrative Data vs SBR Data

The Algorithm

Why did we do this research?

- Degradation in precision and performance (loading time) as the number of businesses increased
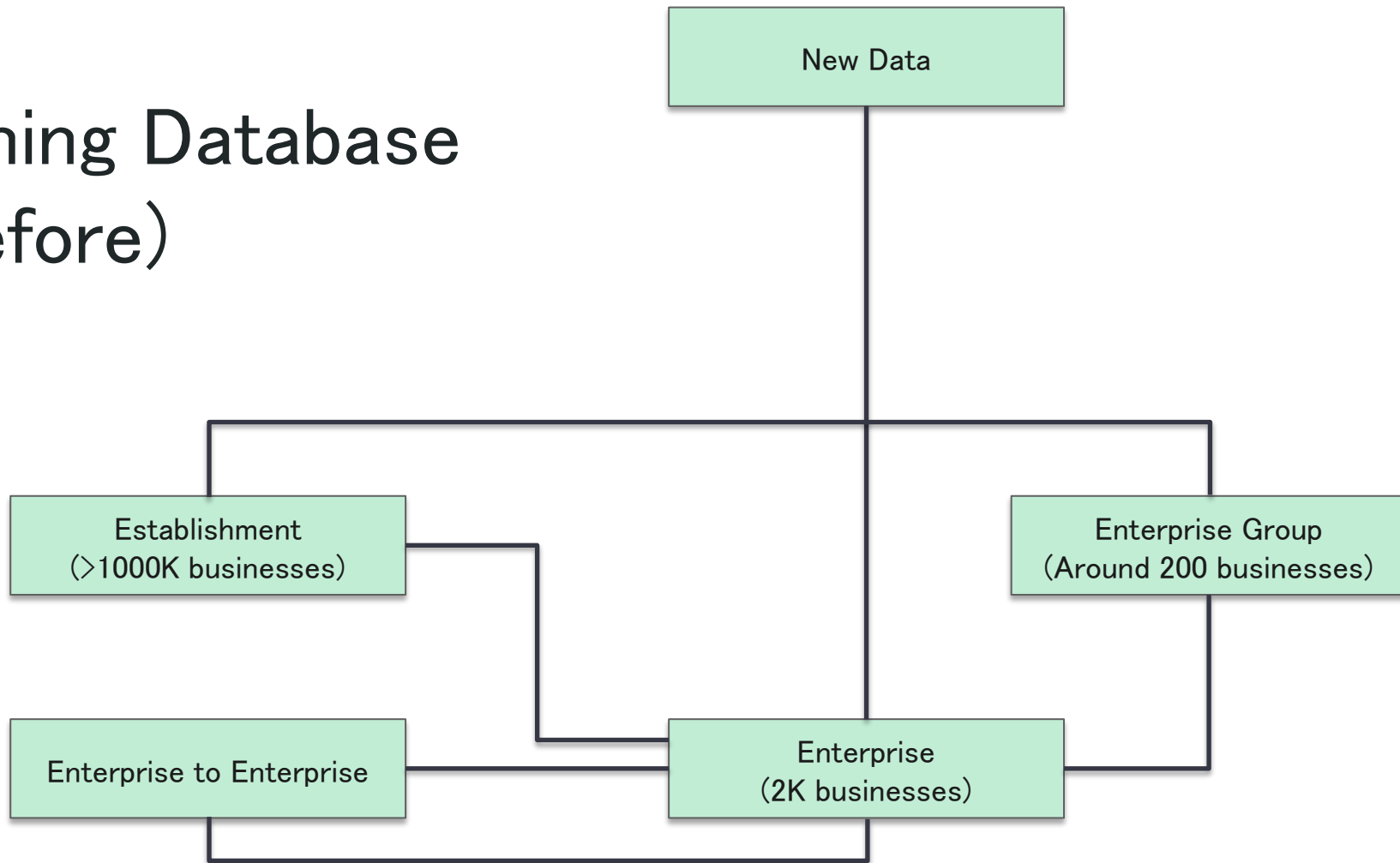
- Complains from operators

The Challenge

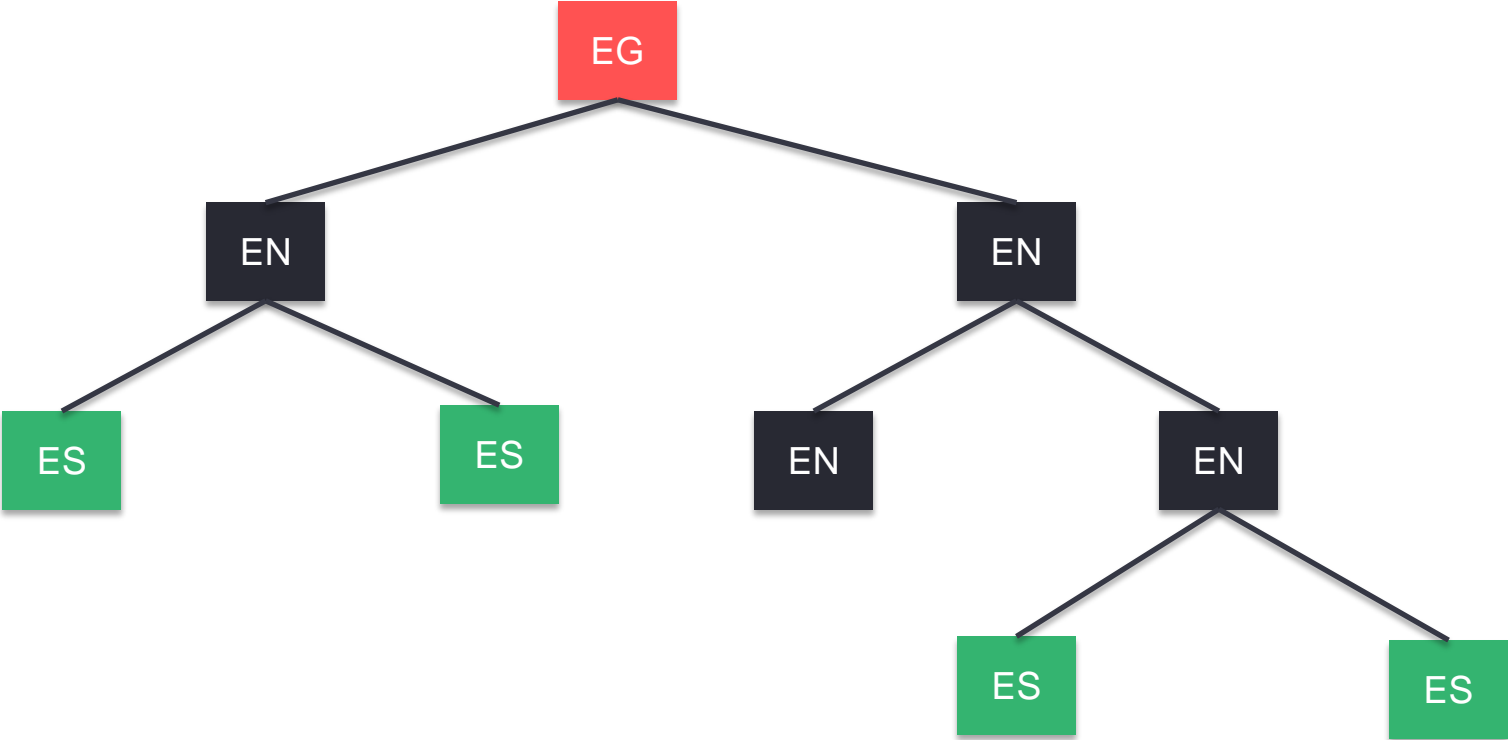How to improve the precision and performance of the matching feature?

Will **tuning database** structure and **removing stop words** improve its precision?
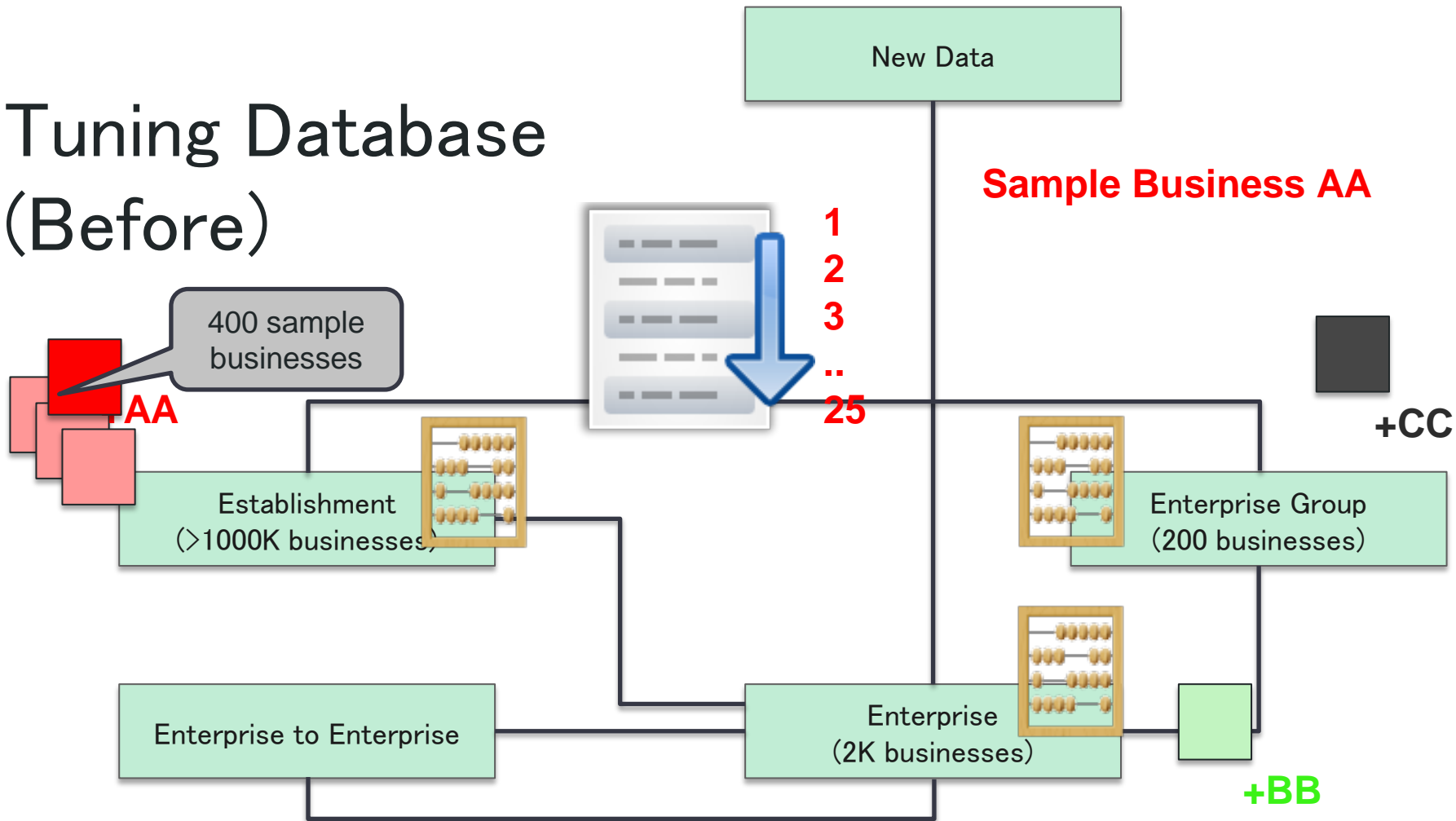
Tuning Database (Before)

```
                                    ┌─────────────────┐
                                    │    New Data     │
                                    └────────┬────────┘
                                             │
     ┌───────────────────────────────────────┼──────────────────────────┐
     │                                        │                          │
┌────┴──────────────┐                         │              ┌───────────┴──────────────┐
│ Establishment     │                         │              │ Enterprise Group         │
│ (>1000K businesses)├──────┐                  │              │ (Around 200 businesses)  │
└───────────────────┘      │                  │              └──────────────────────────┘
                           │                  │                          │
                           │                  │                          │
┌───────────────────┐      │     ┌────────────┴────────┐                 │
│ Enterprise to     ├──────┴─────┤ Enterprise          ├─────────────────┘
│ Enterprise        │            │ (2K businesses)     │
└────────┬──────────┘            └──────────┬──────────┘
         │                                  │
         └──────────────────────────────────┘
```
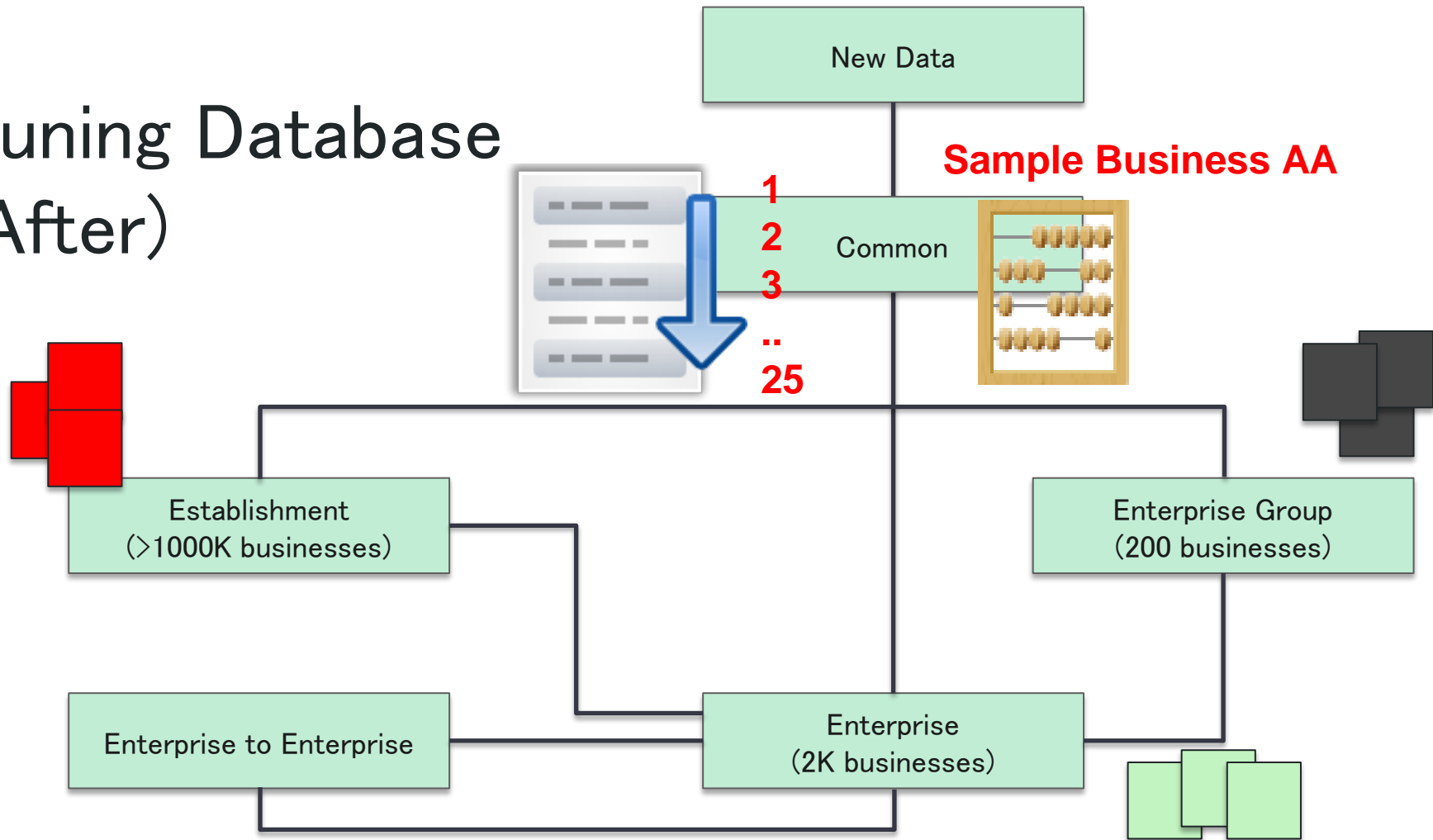
Business Relationship Model

The Evaluation Method
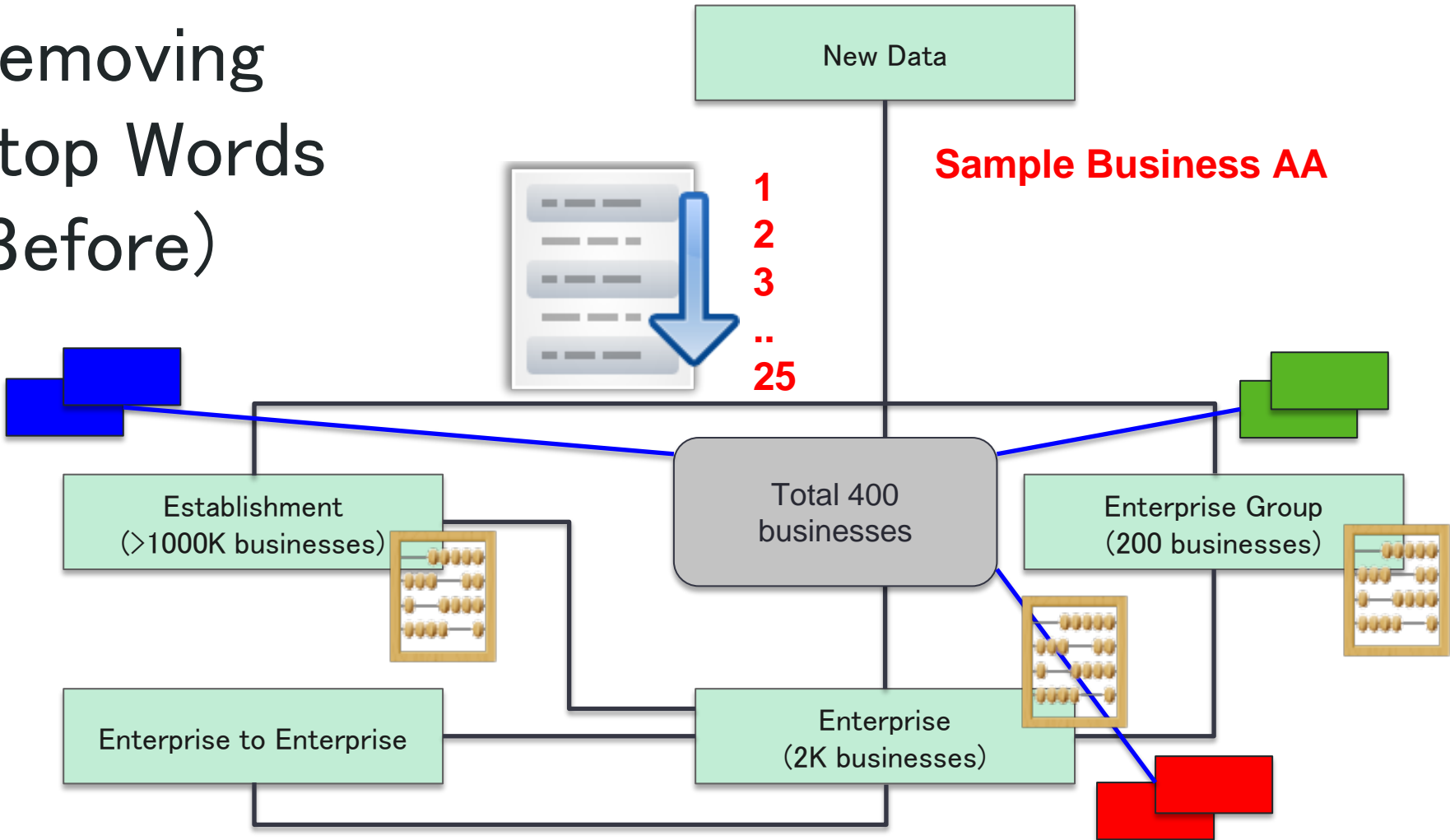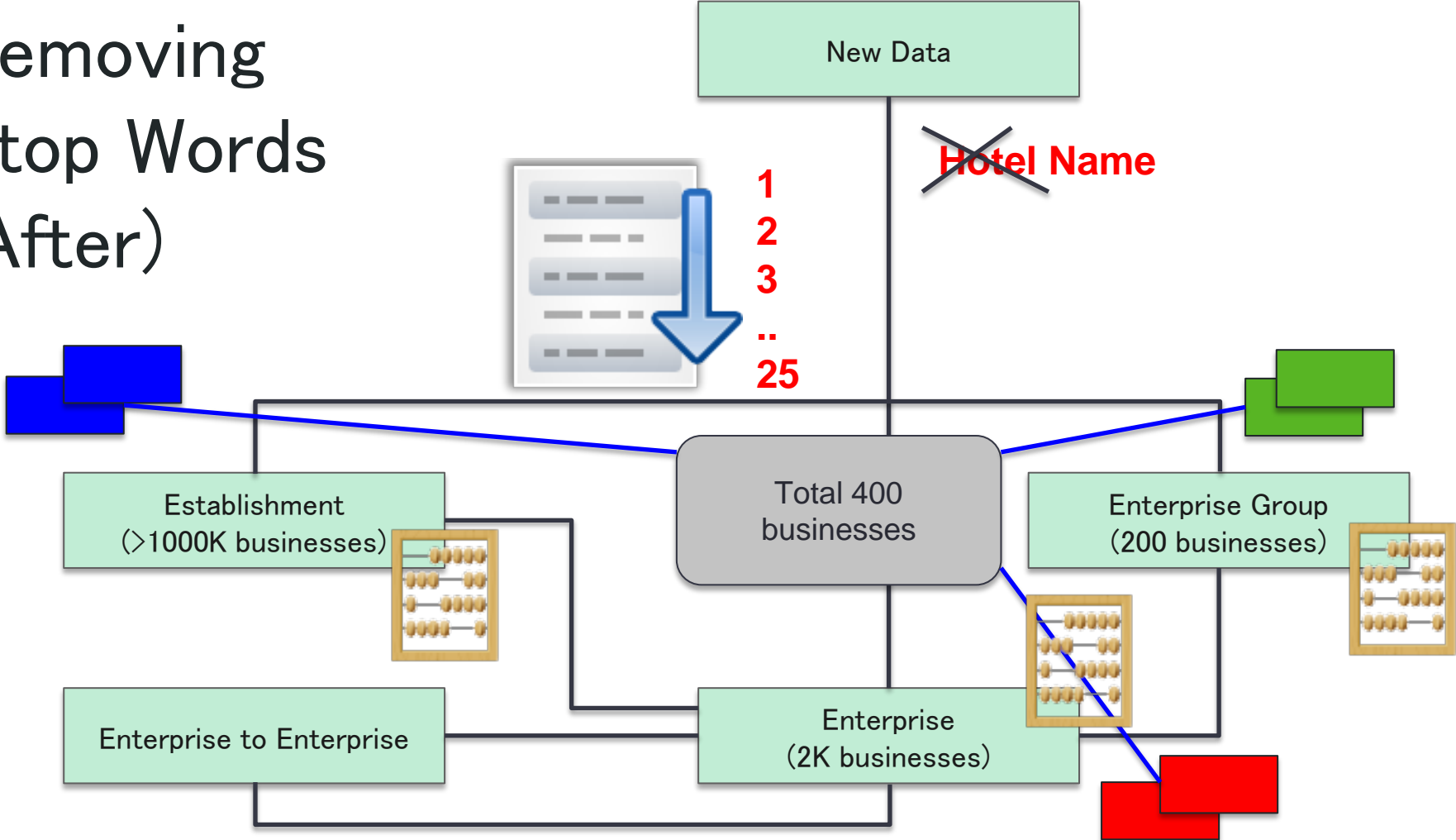
Removing Stop Words

- Stop words in businesses' name: LOSMEN, PENGINAPAN, WISMA, HOTEL, BUNGALOW, VILLA, COTTAGE, RESTORAN, RUMAH MAKAN, PT, CV, plus symbols
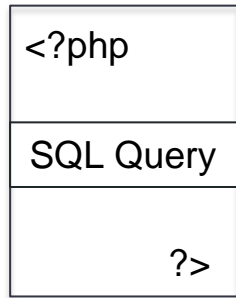
- Stop words in businesses' address: JLN, JL, plus symbols

Removing Stop Words （Before）

Sample Business AA

New Data

1
2
3
..
25

Total 400 businesses

Establishment （＞1000K businesses）

Enterprise Group （200 businesses）

Enterprise to Enterprise

Enterprise （2K businesses）

Removing Stop Words (After)

New Data

~~Hotel Name~~

1
2
3
..
25

Total 400 businesses

Establishment (>1000K businesses)

Enterprise Group (200 businesses)

Enterprise to Enterprise

Enterprise (2K businesses)

Tuning Procedure

<?php	
SQL Query	
?>	

Nonstored Query Procedure

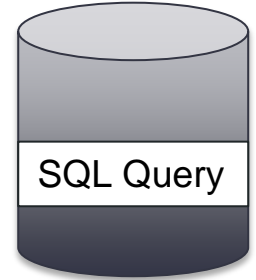→ Stored Query Procedure

SQL Query

(Before)

(After)

Experiment Data

	Precision (Distant to 1st Rank)		Performance (Loading Time per Second)	
	Before	**After**	**Before**	**After**
Tuning Database	0.93	0.00	7.18	5.78
Removing Stop Words	1.22	1.12	11.25	9.59
Tuning Procedure	1.01	1.01	7.37	3.42

Experiment Results

What will we do next?

Removing stop words has been implemented in our system. Next, we will implement tuning database architecture and tuning query procedure solutions.

Future Research

- What about other free information retrieval tools beside SQL Server's FTS?