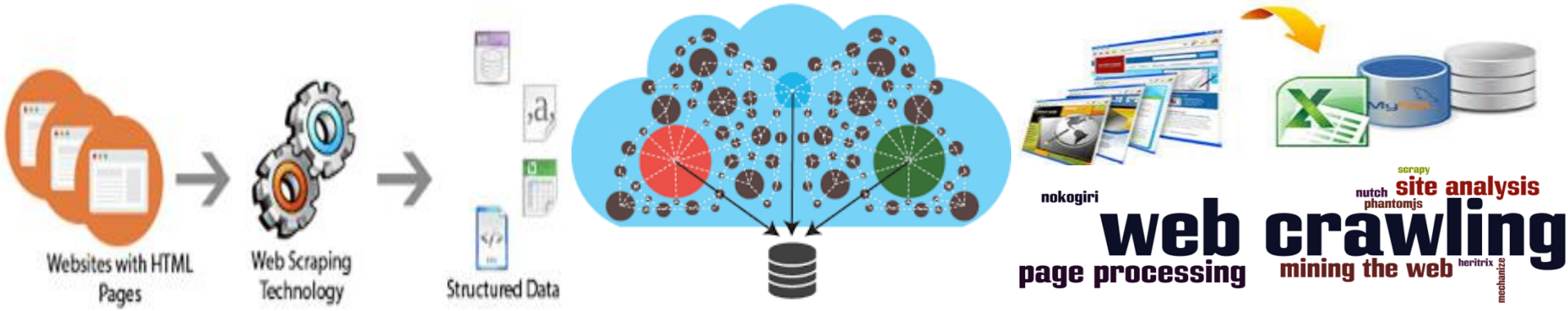
Snapshot Crawling to Produce and Disseminate Time Series Data of E-Commerce: A Case Study in Marketplace Site in Indonesia



Viktor Suwiyanto and Gusnisa Siswayu
BPS – Statistics Indonesia

Introduction

Constraints

- **E-commerce activity is one emerging issue in Indonesia (Perpres No.74 of Electronic national Trading System)**
- **Important role of e-commerce in SDG's (5.b,8.3,9.3,17.11)**
- **Until now, BPS-Statistics Indonesia as the NSO doesn't have e-commerce data.**
- **Conventional method of survey requires high cost, longer time and needs more resources.**

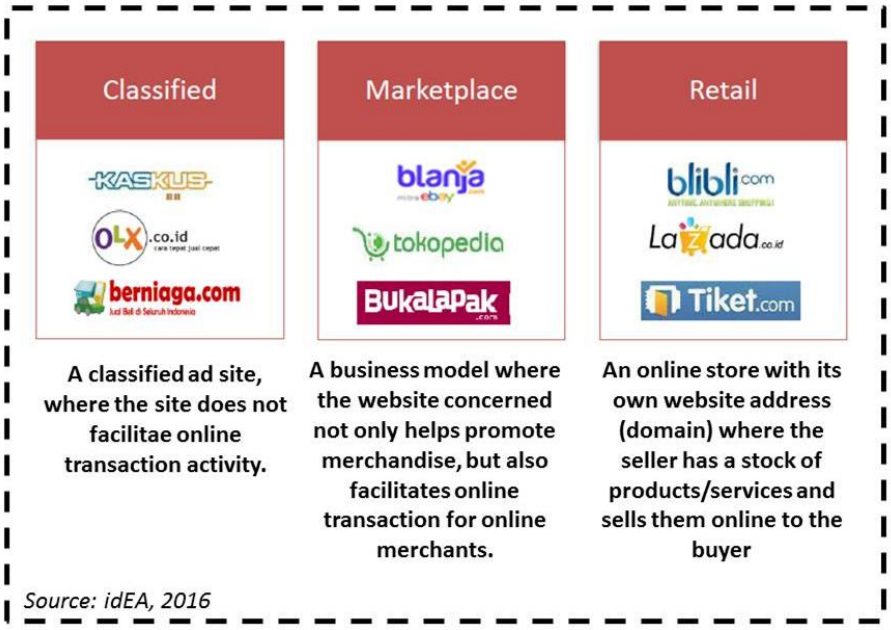


A new method for producing time series data of e-commerce using big data from several commercial sites in Indonesia

Methodology

Step1: Coverage and Preparation of Data Extraction

E-Commerce Business Model in Indonesia

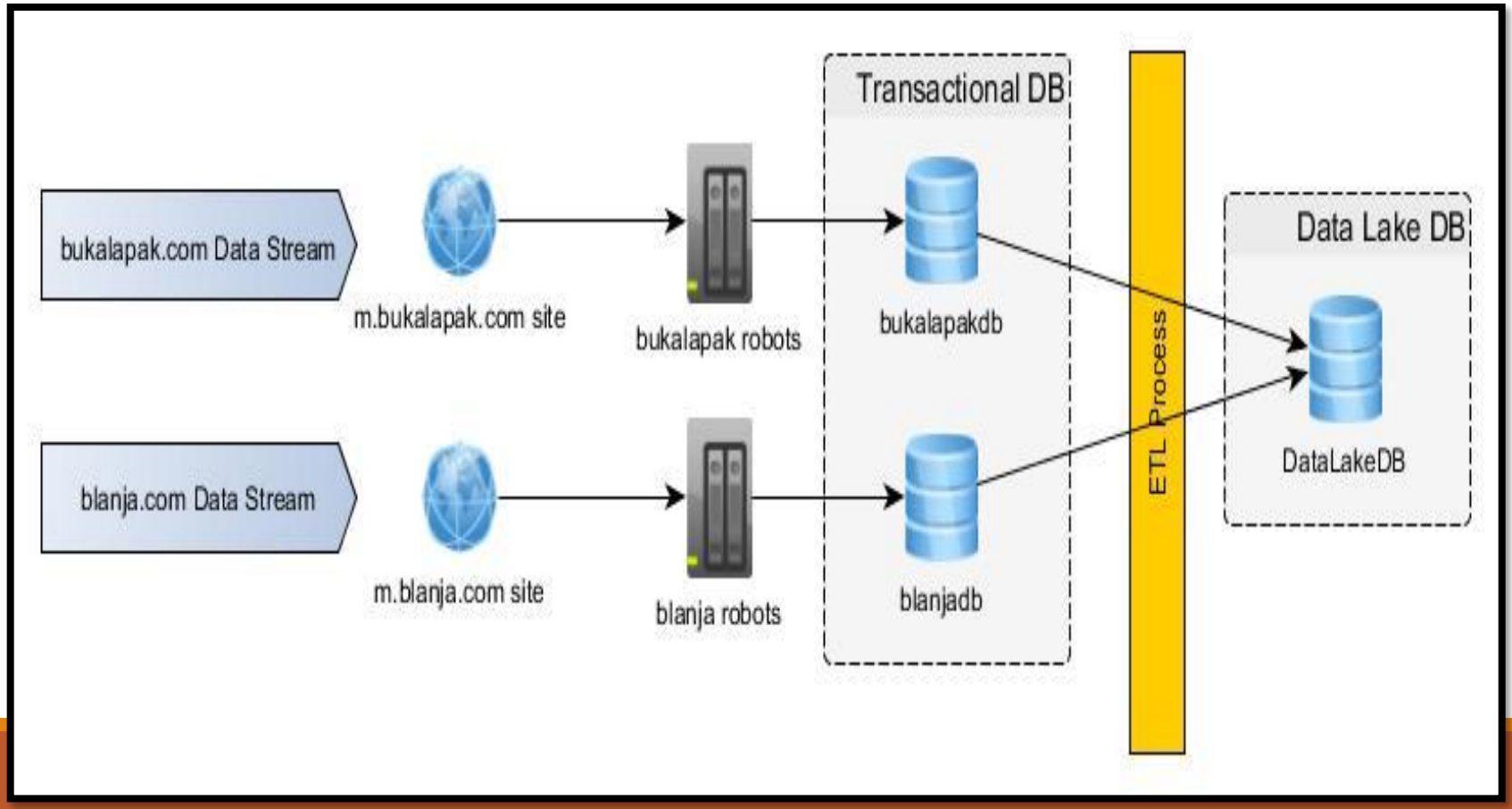


- **The coverage of collection data in this study is marketplace business model.**
- **Blanja and BukaLapak are choosen as the e-commerce company for data crawling subjects.**
- **Categorization of products from Blanja and BukaLapak into 12 Categories (Eurostat, 2016).**

Methodology

Step2: Scraping Robot and Schedulling Development

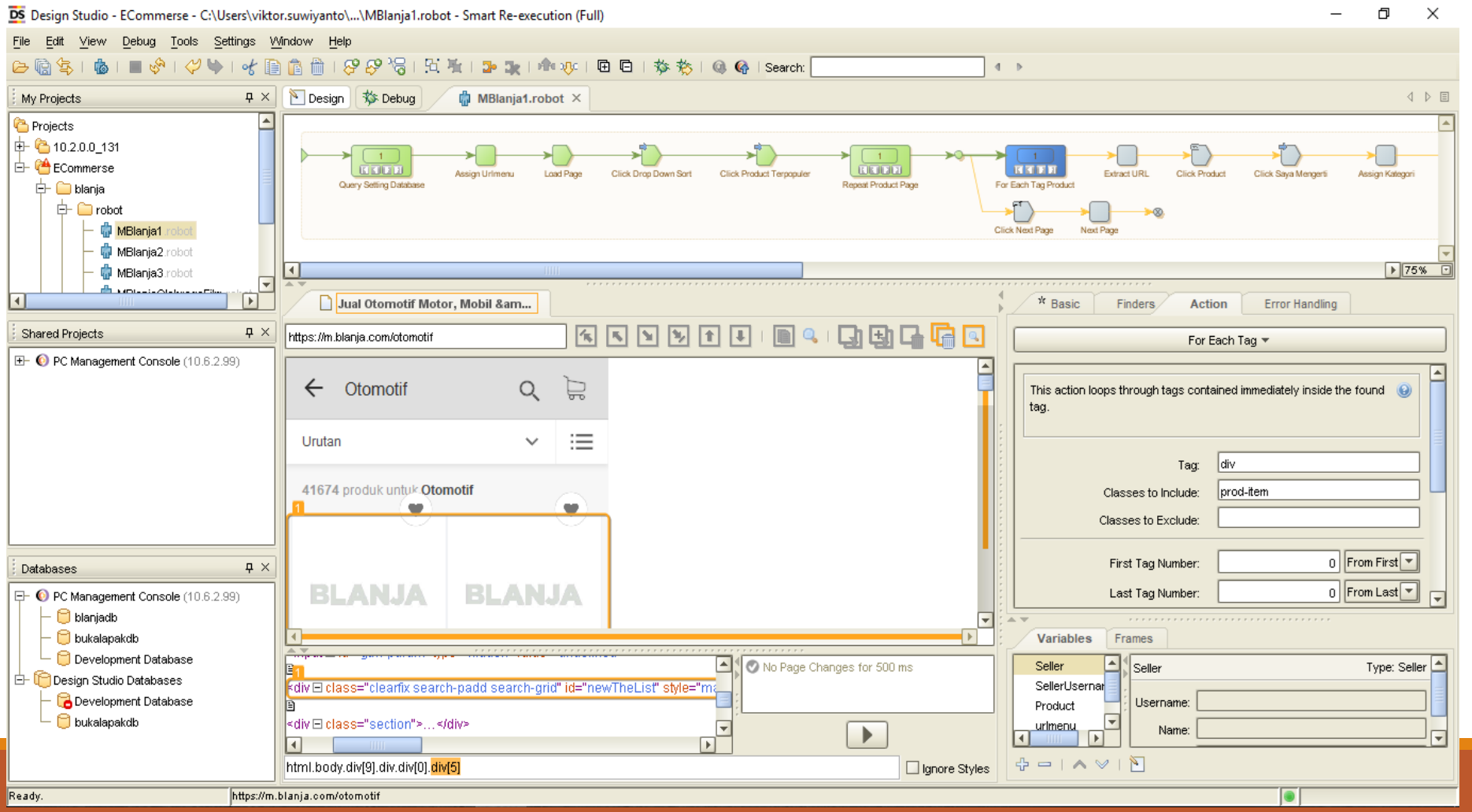
E-commerce Snapshot System Architecture Design



Methodology

Step2: Scraping Robot and Schedulling Development

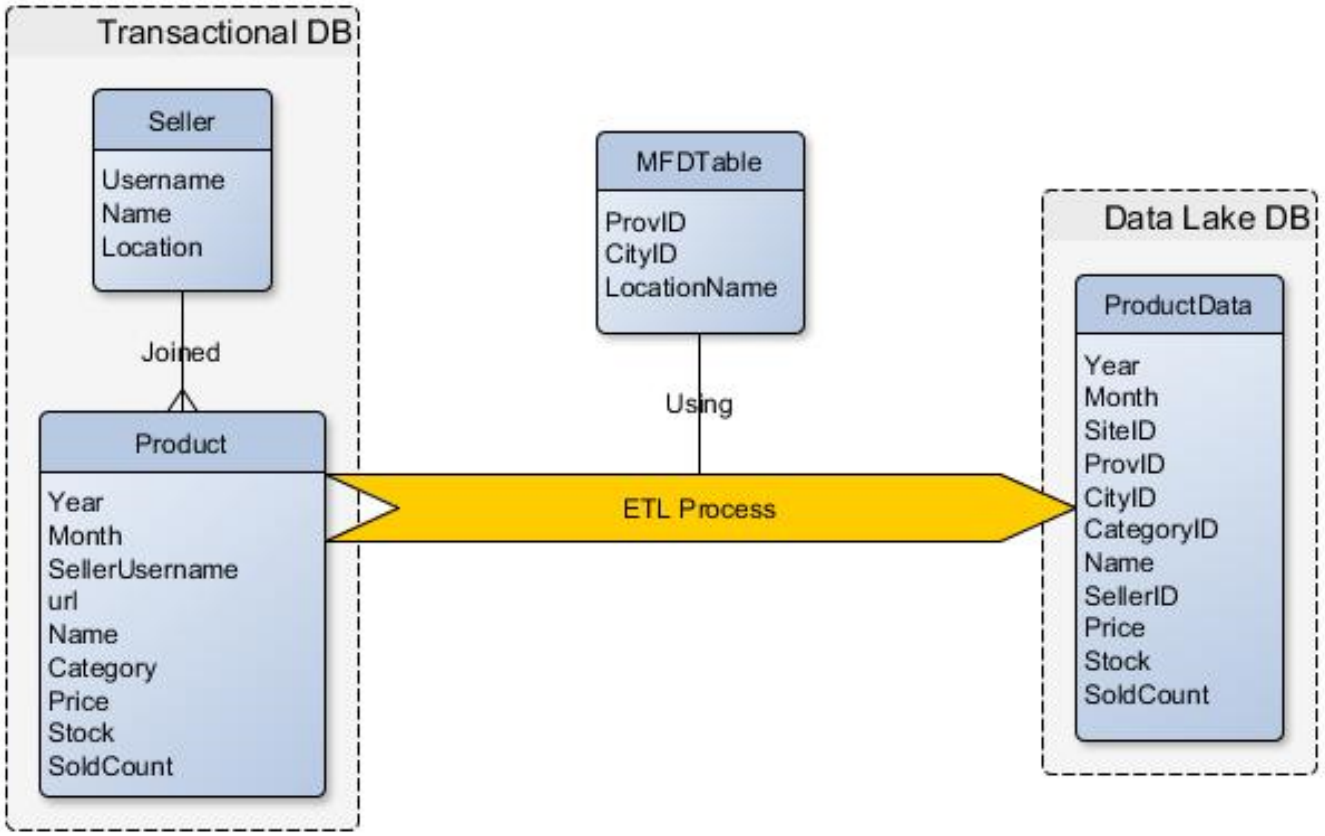
Design Studio of Kofax Kapow for Developing Scraping Robot



Methodology

Step2: Scraping Robot and Schedulling Development

ETL Process of Product Data from Transactional DB to Data Lake DB



Methodology

Step 3: Dashboard and Dynamic Table Development (Dissemination)

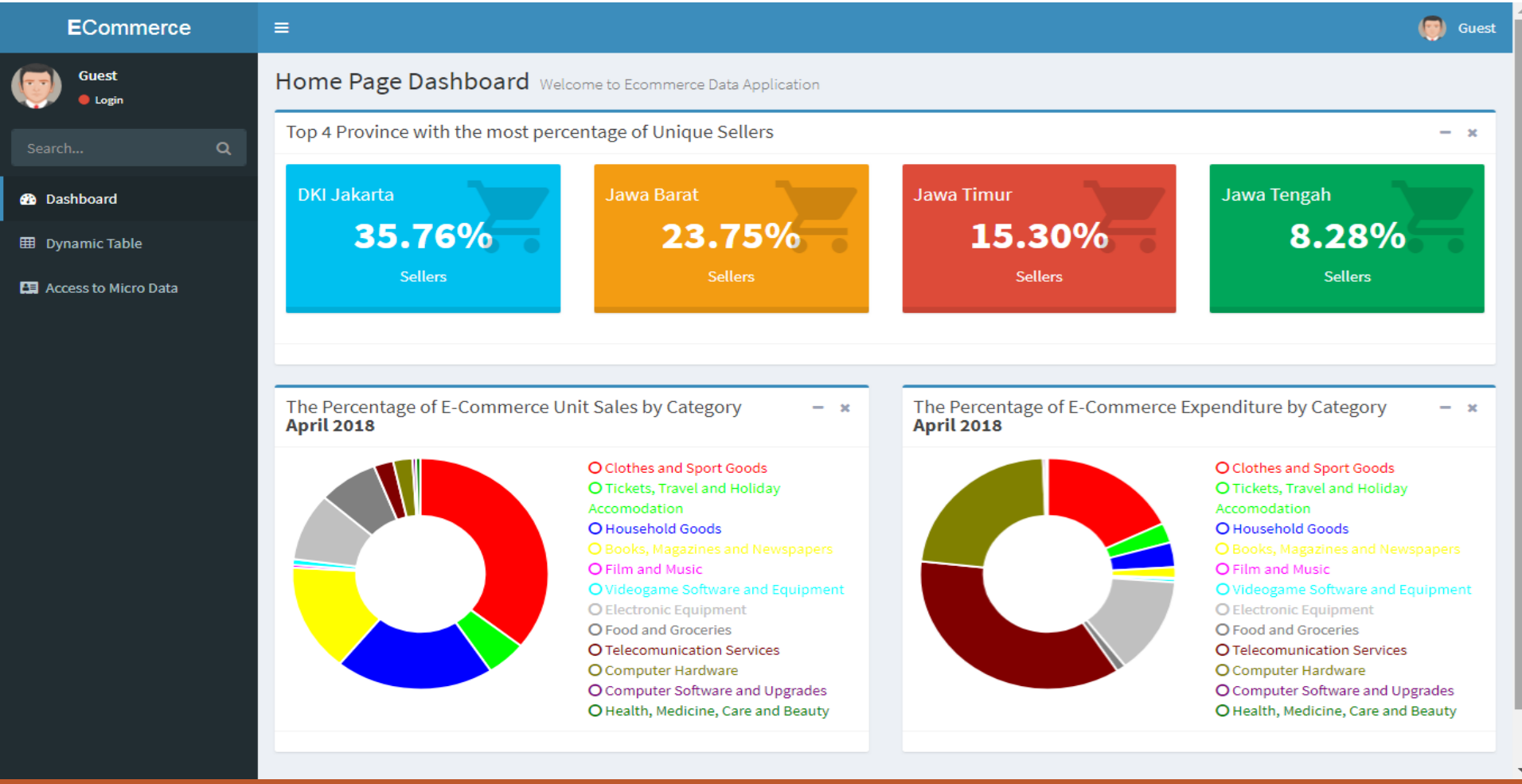
Develop a web application with features:

- **Dashboard**
- **Dynamic table**
- **Access to Microdata**

https://s.bps.go.id/ecommerce

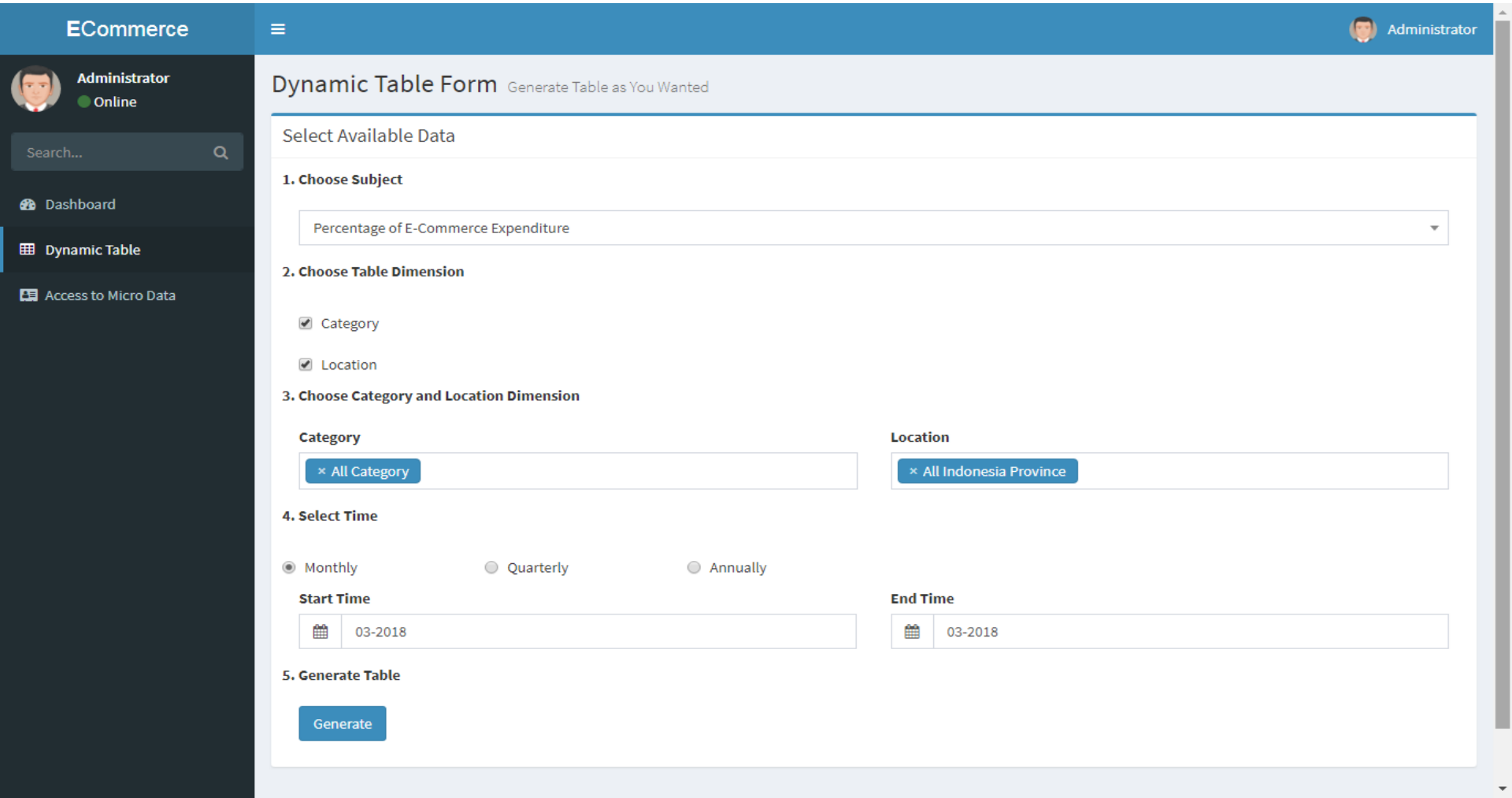
Dissemination

Home Page Dashboard



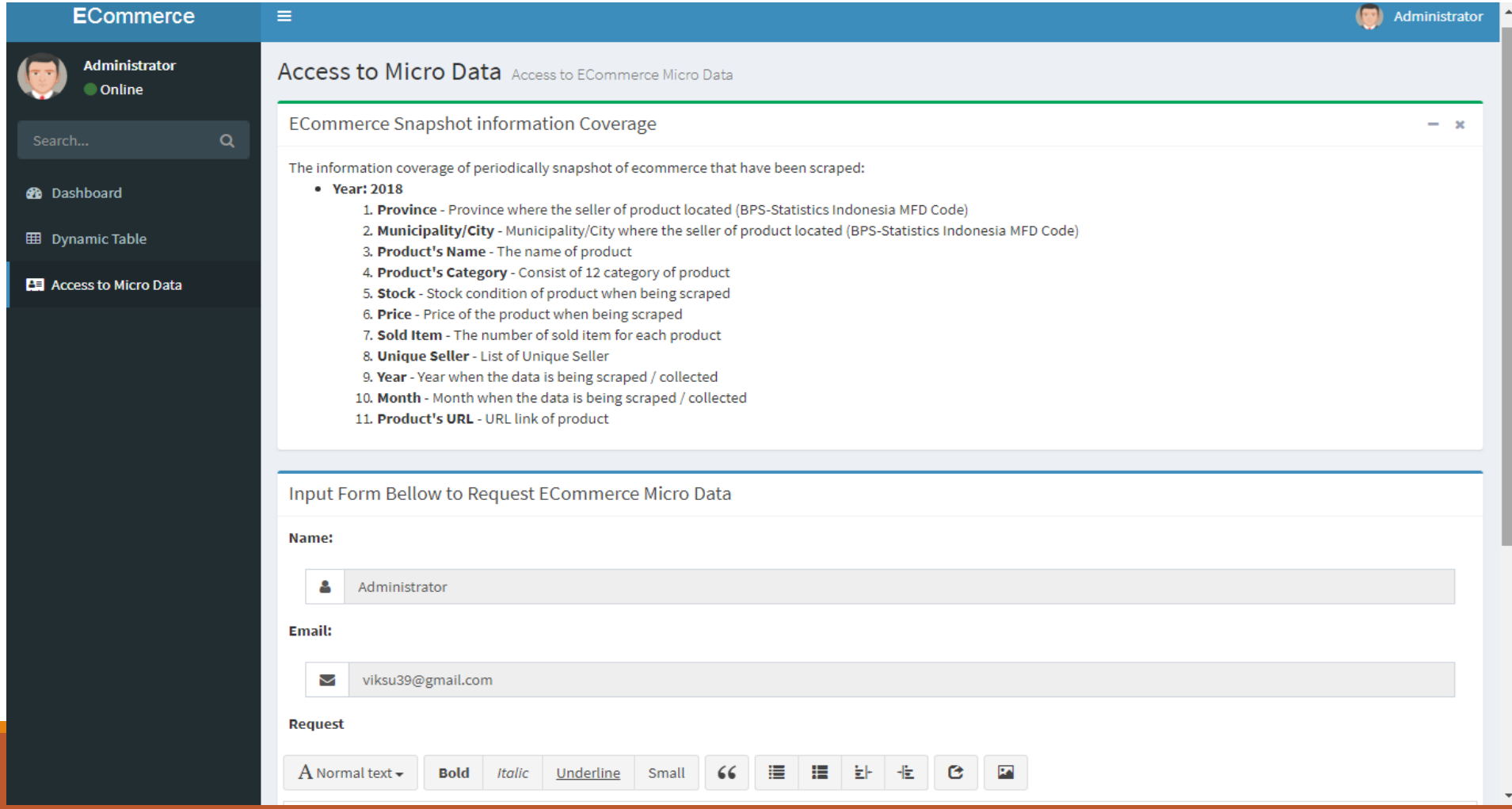
Dissemination

Dynamic Table Form



Dissemination

Access to Micro Data



Conclusion

- There are 9 variables successfully collected namely: 1) location of seller (province and municipality/city); 2) product's name; 3) product's category; 4) stock of the product; 5) price of the product; 6) number of sold item; 7) unique Seller; 8) time periods (year and month); and 9) product's URL.
- There are 6 statistics indicators of e-commerce: percentage of e-commerce expenditure, percentage of e-commerce unit sales, percentage of unique seller, growth of e-commerce expenditure, growth of e-commerce unit sales, and growth of unique seller.
- Data and indicators can be presented monthly, quarterly or annually in web application with URL address: https://s.bps.go.id/ecommerce not active

Discussion

- The Data was not authentic because there is no agreement from marketplace website
- If there are changes in the marketplace website the crawling robot cannot get the data because of errors
- We cannot publish this data because there is no legal framework for this data processing

E-Commerce Data Now

- In 2019, The Indonesian government provide Government Law (Peraturan Pemerintah/PP) No. 80 about E-Commerce Transaction or PMSE (Perdagangan Melalui Sistem Elektronik / Trade Using Electronic System) in bahasa. In This Law BPS-Statistics Indonesia gets the authority to collect e-commerce data from e-commerce actor.

- In 2023, BPS-Statistics Indonesia provide Institution Law No.4 about Submission and management of E-Commerce data and/or Information

Thank You

www.bps.go.id

viktor.suwiyanto@bps.go.id, siswayu@bps.go.id


Jl. Dr. Sutomo 6-8 Jakarta 10710


bpshq@bps.go.id


(021) 3841195, 3842508, 3810291


Badan Pusat Statistik (Page)


(021) 3857046


@bps_statistics