Strategy and Practice of Data Middle Platform Construction of NBS

Zhang Chunzhen
National Bureau of Statistics of China

June 2020

Contents

1. Data Island problem

Three aspects of data islandization

- ## Data organization

  Data belongs to different departments and operates independently.

- ## Data storage

  Data is stored independently, maintained independently, and isolated from each other.

- ## Data logic

  Data is loosely connected or unconnected, and cannot be integrated.

1. Data Island problem

Two directions to solve Data Island

- ## Top-down (business-driven)

  It starts from the upper level of the application system, considers the data dependence among the application systems, and develops interface services on demand, so as to achieve the mutual use and integration of the data among the respective systems.

  E.g., the Enterprise Service Bus

- ## Bottom-up (data-driven)

1. Data Island problem

Two directions to solve Data Island

- Top-down (business-driven)

- Bottom-up (data-driven)

It gives the priority to considering global data characteristics, formulating data specifications, defining data standards, unifying the cognition of data among different departments, and thus realizing integration between upper-layer applications and wider external services of data with the integration of lower-layer data.

E.g., information-centric distributed structure, data sharing mechanism based

on blockchain, data middle platform, data interoperability

2. Methodology

Middle Platform

a support platform that provides agile response to front-end applications in the form of reuse capabilities. (Alibabagroup.com)

- Five characteristics of Middle Platform
  - ➢ with reuse as its core value;
  - ➢ born for the front platform;
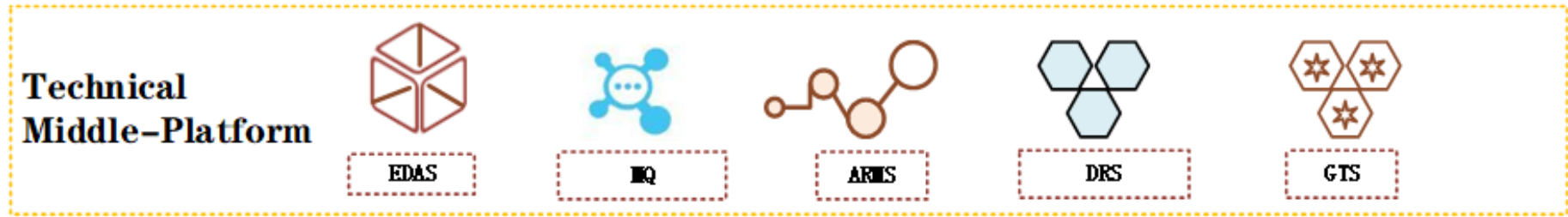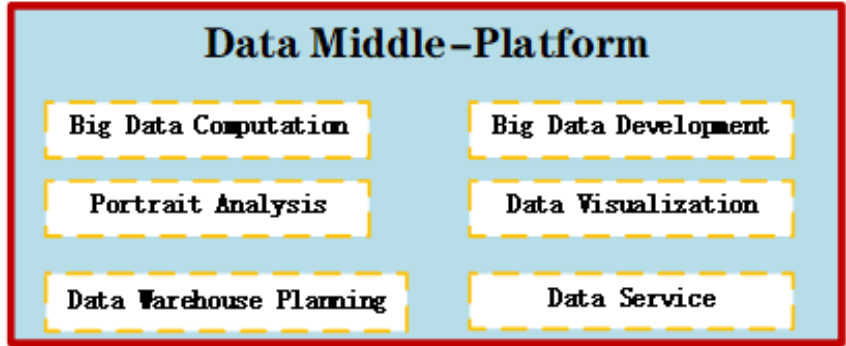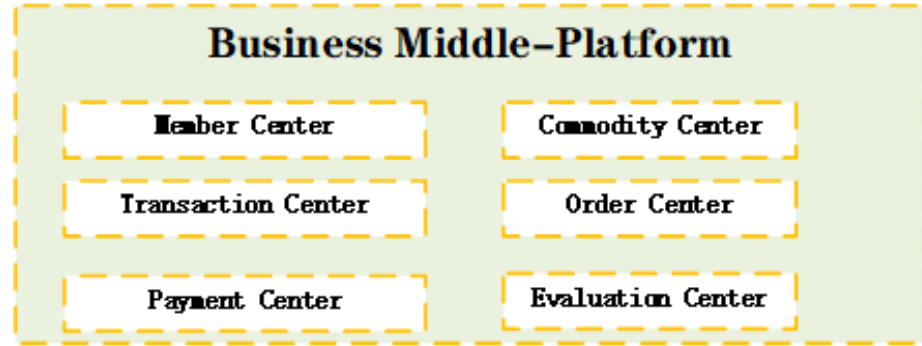 - ➢ decoupled natively;
  - ➢ relatively stable;
  - ➢ evolving.

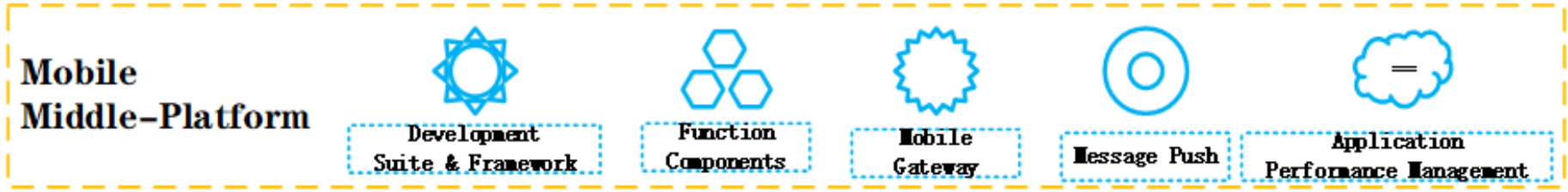# 2. Methodology

- Classification of Middle Platform
  - ➤ Technical middle platform
  - ➤ Business middle platform
  - ➤ **Data middle platform**
  - ➤ ......

Business of the Alibaba Group

| Tmall | Taobao | AliPay | Youku | Amap | Cainiao Express&Logistics | UC | ... |

Mobile Middle-Platform

- Development Suite & Framework
- Function Components
- Mobile Gateway
- Message Push
- Application Performance Management

Business Middle-Platform

Member Center	Commodity Center
Transaction Center	Order Center
Payment Center	Evaluation Center

Data Middle-Platform

Big Data Computation	Big Data Development
Portrait Analysis	Data Visualization
Data Warehouse Planning	Data Service

Technical Middle-Platform

| EDAS | MQ | ARMS | DRS | GTS |

IaaS

| ECS | SLB | OSS | RDS | CDN |

2. Methodology

Data Middle Platform (DMP)

- an enterprise-level comprehensive data capability platform
- focusing on the reuse of data-related capabilities
- breaking through the technical barriers of data production, storage, analysis, service, and circulation
- a global bottom-up solution for data island problem

2. Methodology

Three cognitions of DMP construction

- ## Cognition 1

  The basic premise for an enterprise to build a data middle platform is that it has a certain scale of informatizable business, the stock or the expected incremental business is diversified, and the business is coupled.

- ## Cognition 2

  Data middle platform is a strategic choice for enterprise development.

- ## Cognition 3

  Data middle platform is a necessary condition for enterprise innovation.

3. DMP practice of NBS

DMP construction basis for NBS

- To Cognition 1

  ➢ statistical application softwares are all based on unified metadata, regime design, report design, user and authority management, etc.

  ➢ use the same source roster, survey objects, and the same norms and standards for sampling

  ➢ the processes of data collection, examination, summary, and acceptance in data processing are mostly similar

  ➢ data processing and publishing, archiving management, and further in-depth analysis also follow consistent management standards

3. DMP practice of NBS

DMP construction basis for NBS

- ## To Cognition 2

  It is based on the development of the Statistical Cloud, which is positioned as a strategic project of national statistics and provides a good opportunity for overall planning and governance of statistical data resources.

- ## To Cognition 3

  With the "three onto-cloud" goals, the Statistical Cloud aims to create a "cloud statistics" to deeply change the statistical production method and innovate the way statistical data serves government decision-making and social governance.
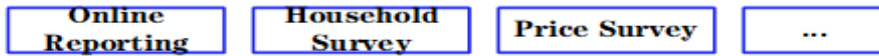
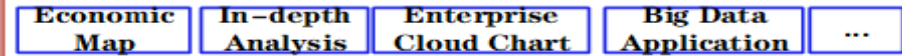# 3. DMP practice of NBS

DMP architecture of NBS

- large middle platform, small front end

- cloud-based

- horizontal layered structure

- service-oriented
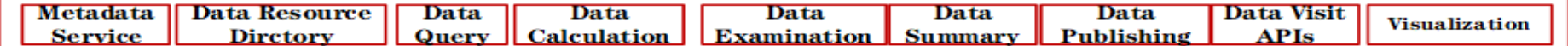
Statistical Data Production Systems

| Online Reporting | Household Survey | Price Survey | ... |

Statistical Data Development and Application Systems

| Economic Map | In-depth Analysis | Enterprise Cloud Chart | Big Data Application | ... |

Data Services

Data Servitization Interface

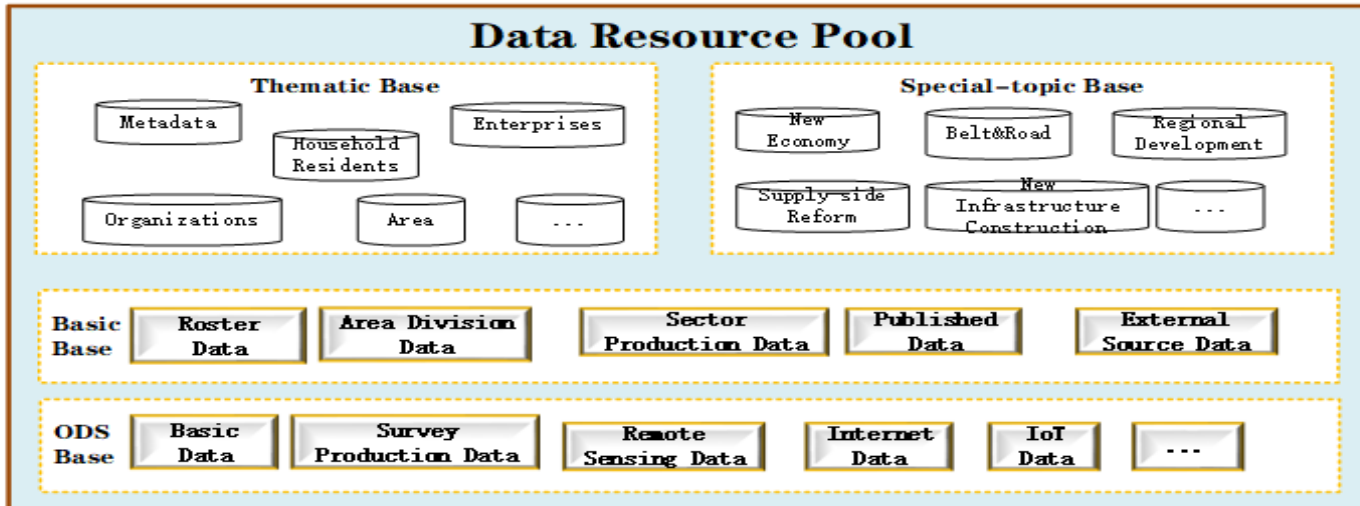| Metadata Service | Data Resource Dirctory | Data Query | Data Calculation | Data Examination | Data Summary | Data Publishing | Data Visit APIs | Visualization |

Modelling Analysis
- Model Design
- Interactive Analysis
- Flow Analysis
- Real-time Analysis
- Dynamic Inception
- Data Mining
- Machine Learning

Data Resource Pool

Thematic Base
Metadata | Household Residents | Enterprises | Organizations | Area | ...

Special-topic Base
New Economy | Belt&Road | Regional Development | Supply-side Reform | New Infrastructure Construction | ...

Basic Base: Roster Data | Area Division Data | Sector Production Data | Published Data | External Source Data

ODS Base: Basic Data | Survey Production Data | Remote Sensing Data | Internet Data | IoT Data | ...

Management Governance
- Standard Management
- Quality Management
- Asset Management
- Security Management
- Operation Management
- Metadata Management
- Service Management

Organization Processing
Theme Definition | Data Labling | Data Association | Dimension Definition | Storage Definition | Indicator Calculation

Aggregation Integration
Mode Definition | Design of Standard | Data Extraction | Data Transfer | Data Mapping | Data Integration | Data Validation

Collection Exchange
Data Synchronization | Data Import | Data Scraping | Breakpoint Resuming | Data Capturing | Task Management | Data Distribution

Data Sources
Foundation Support System | Statistical Data Production System | Government Departments | Internet | IoT | Remote Sensing System | ...

3. DMP practice of NBS

Six core capabilities

- Data collection and exchange
- Data aggregation and integration
- Data organization and processing
- Data management and governance
- Data modelling and analysis
- Data service and application

3. DMP practice of NBS

DMP operation carrier: Container + Microservices

- splits the monolithic application into multiple small services with high cohesion and low coupling according to the business field

- runs in an independent process (deployed in Container)

- developed and maintained by different teams using different protocol stacks, languages, and storage. .

- uses lightweight communication mechanisms (such as HTTP RESTful API, or RPC)
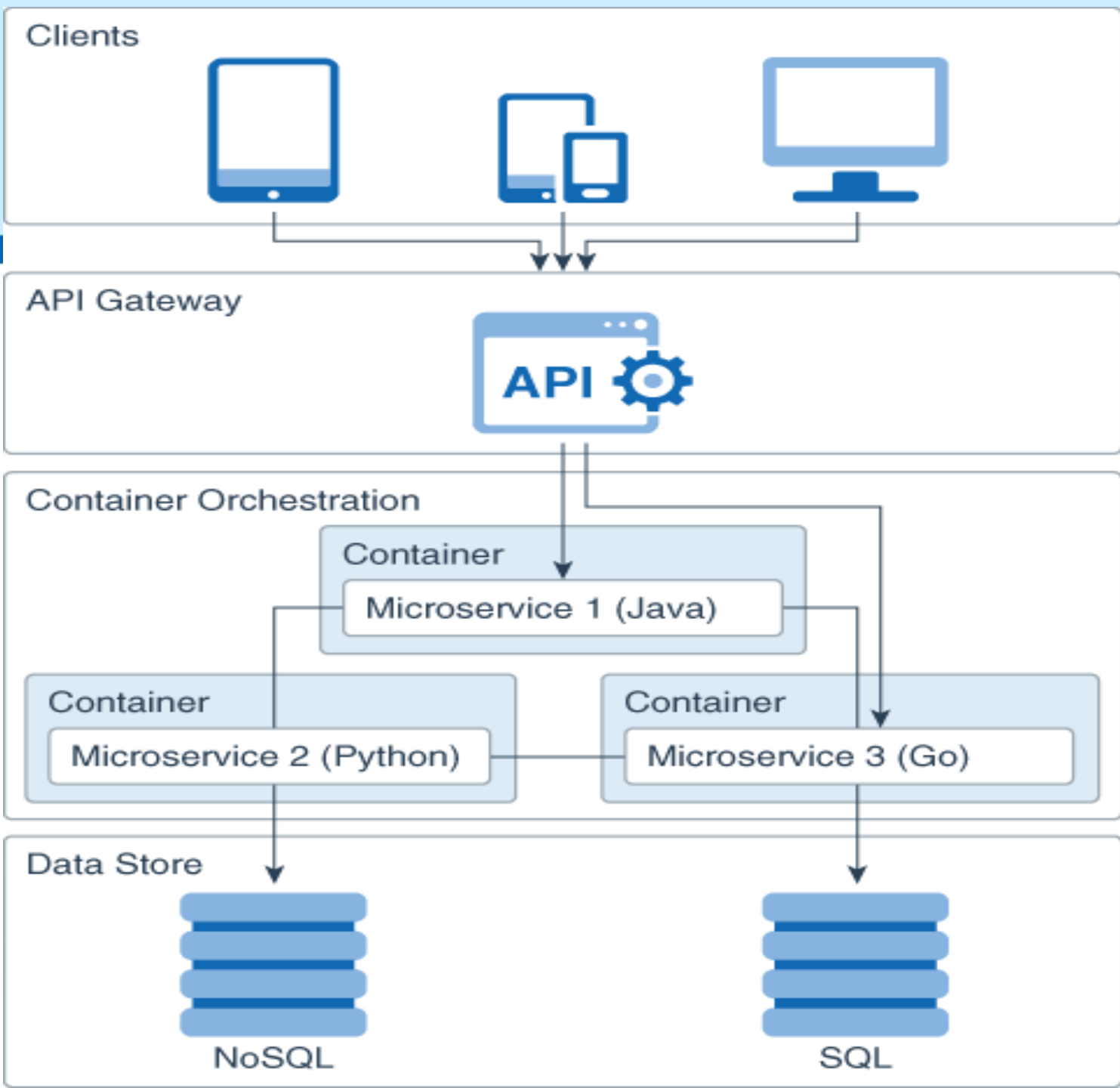
4. Discussion & Conclusion

Role of DMP

- Solving data islands
- Tolerating the entry of various data sources
- Building statistical big data platform
- Transforming official statistics
  - ➢ from statistical analysis to predictive analysis
  - ➢ from single-domain to cross-domain
  - ➢ from passive analysis to active analysis
  - ➢ from non-real-time to real-time analysis

4. Discussion & Conclusion

Future of DMP

- empowering data partners through data servitization
- cultivating healthier and more dynamic statistical data ecology
- integrating into the Web of Data

Thanks!