Institute and Faculty of Actuaries

众安保险 ZhongAn Insurance

Common Data Model Platform: Driving Insurance Digital Transformation to the Next Stage

ZhongAn Insurance – DIC 2019

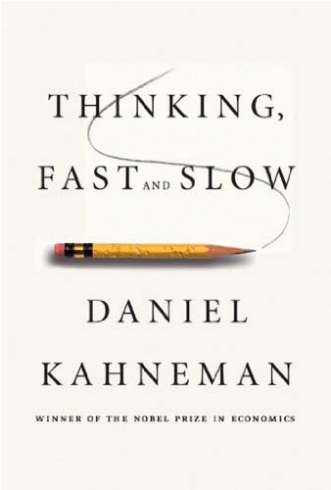
Institute and Faculty of Actuaries

众安保险 ZhongAn Insurance

Outline

- Trend in data intelligence application
- What is CDM platform?
- How to build CDM platform?
  - Architecture, path and methodology
- Applications and outlook

Thinking, Fast and Slow



Source: 2018 Alibaba Yunxi

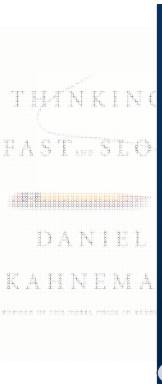
Thinking, Fast and Slow

System I (Fast)

Unconscious, fast, based on impression and subjective feelings
Autonomous execution

Can be done in seconds
Identifying person, driving...

Deep Learning



System II (Slow)

Concentrated
Considerations and decisions

Professional knowledge, longer thinking time
Operation decisions, stock management, scheduling...

Operations Research

Why less applications about slow thinking?
It requires higher data availability!

Source: 2018 Alibaba Yunxi

Challenges in data management

Complex data sources	Fragmented application	Long chain of dependency
• External data sources • Internal business systems • Inconsistent data format	• Spread sheets • SQL on EDW • Reporting database • Customize applications(mobile, data visualization etc.)	• Analyst, data scientist depends on data engineer • Data engineering depends on ETL process • ETL process depends on business system DB

Is Data Lake a perfect solution?

Since 2013, the concept of Data Lake become popular. Many big organizations consider it as the data management solution for big data era. It is optimized to ingest raw data quickly as received from both new and traditional sources.

Data Lake holds not only structured data but also unstructured data(text, image, audio etc.), this feature makes it popular again when AI become the new trend.

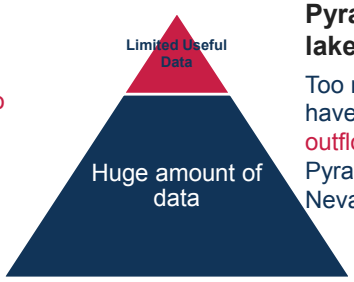
However, many applications of Data Lake focus too much on the “low cost” and “rapid” parts, but ignored a principle of “governing the data when needed”. This kind of application pattern lead to some problems...

Pyramid like data lake

Pyramid Lake in Nevada

Pyramid Lake is fed by the Truckee River, but has no outflow.

What most real data lakes look like



Pyramid like data lake

Too many data lakes have inflow, but no outflow, just like the Pyramid Lake in Nevada.

Institute and Faculty of Actuaries

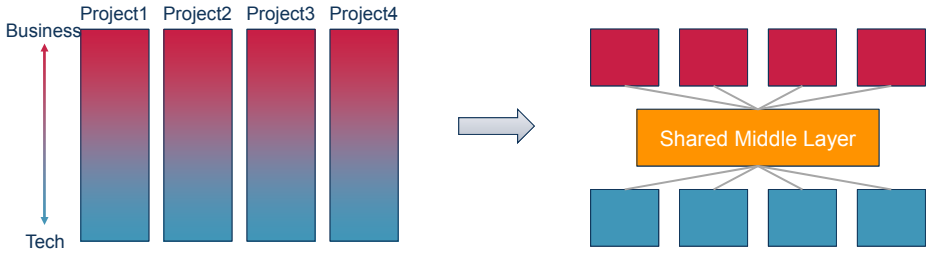
What is the new trend?

- What is “Big” for “Big Data”
  - 4 V-feature (Volume, Varity, Velocity, Value)
- Data vs. data asset
  - Data is worthless without data map
- CDM Platform vs big data system
  - CDM Platform is a concept proposed by Alibaba, firstly introduced as a middle layer of data processing
  - CDM Platform uses big data system as technical solution for handling the lifecycle of CDM

Institute and Faculty of Actuaries

Why do we always need a middle layer?

- It's a simplification of organization
- It's a simplification of technique architecture
- The core of a middle layer is "share" and "standardize"

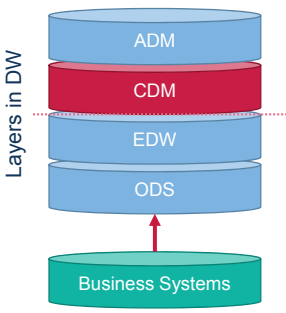


A shared middle layer can reduce redundancy invest in technique and manpower.

Institute and Faculty of Actuaries

Why do we call this middle layer CDM Platform?

- **CDM(Common Data Model)** is a conceptual layer in data warehouse, by using across the whole organization, consistency and efficiency could be guaranteed.



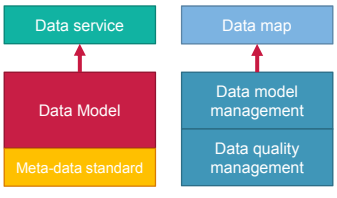
- Organized by analytic purposes(themes)
- Keep consistency in format and logic

Good concept but poor execution in real projects...

- Organized by system functions
- Contain the complexity in business systems

CDM Platform

A system and corresponding standard handling the whole lifecycle of common data model, including building, managing and servicing.



Data and meta data

Systems and Tools

Institute and Faculty of Actuaries

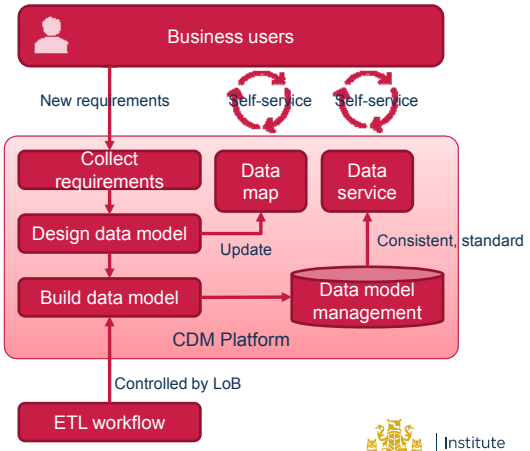
Workflow with a CDM Platform: self-service best practice

Traditional Workflow

- Business users might not know how many data property were produced, thus new requirements might be fragmented and redundant
- These requirements make optimization on computation difficult, so that users have to use spread sheets to access transaction detail
- Business users depends on data engineers too much produced a very long change cycles

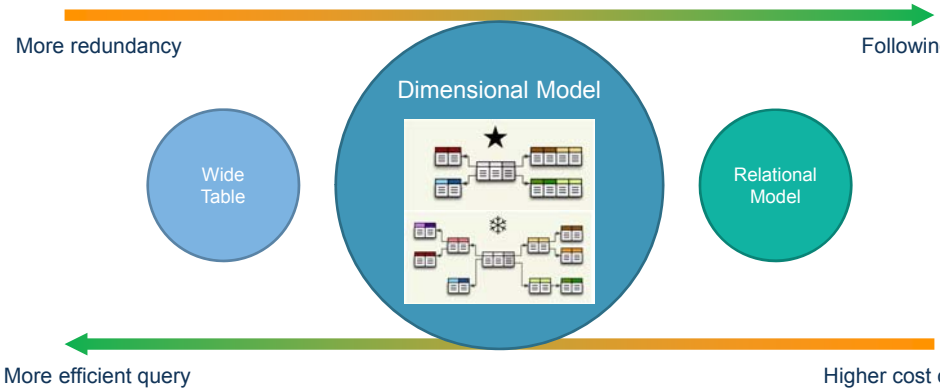
Workflow with a CDM Platform

- Requirements are collected and combined, everyone can access data asset catalog
- Data model can leverage new technologies, so that business users can access data in the finest granularity
- Data model isolated the complexity of business systems, business users can access consistent, standard data with self-services.
- More importantly, this consistency and standardization is promised by system, not human



Institute and Faculty of Actuaries

Design principle: data model

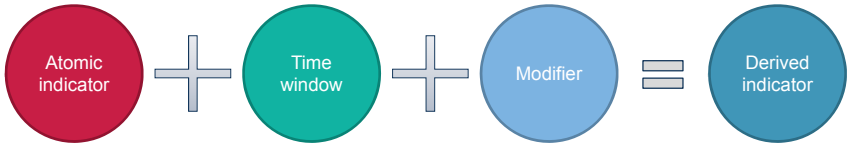


Institute and Faculty of Actuaries

Design principle: indicator decomposition

Decomposition pattern: CDM only hold atomic indicator, modifier can be decomposed by application

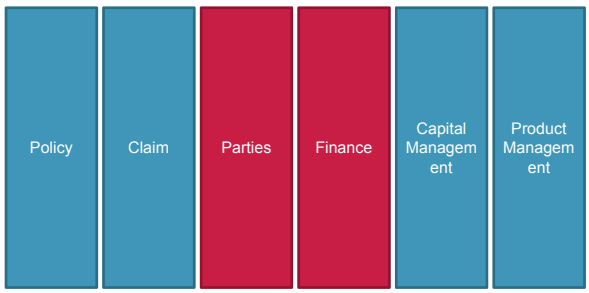
Pushing down atomic indicator into CDM layer can leverage new technologies like MOLAP



Example: Premium income + Last 7 days + from channel X = Channel X's last 7-day income

Application path: start from a vertical theme

- Shared CDM Platform is applicable to the whole organization
- But cost of building and transition is considerable
- Design for a bigger picture but start from a vertical theme

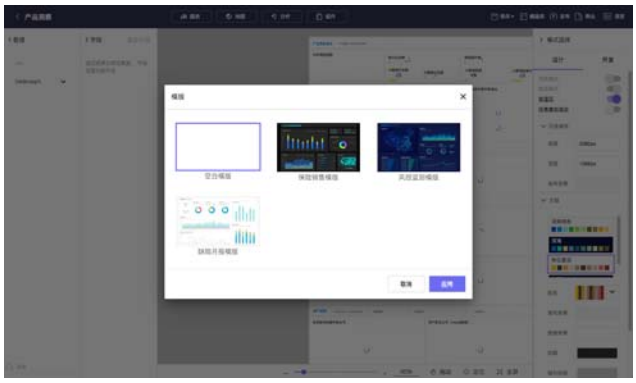


Themes in insurance data warehouse

Data asset management: data map

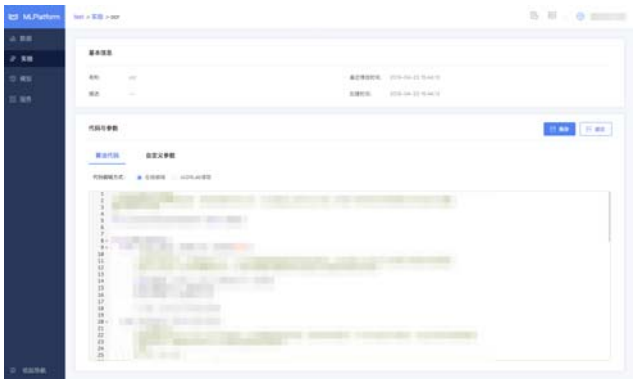
- Asset catalog, report, search engine
- Metadata, changelog
- Kin relationship among tables
- Consistency and standardization check

Data service: self-service data analysis platform



- Integrated with OA approval process
- Building, maintaining, connecting data model by self-service
- Defining derived indicators in real-time
- Customized data visualization without coding
- Realtime multi-dimension analysis
- Support integrating with business systems

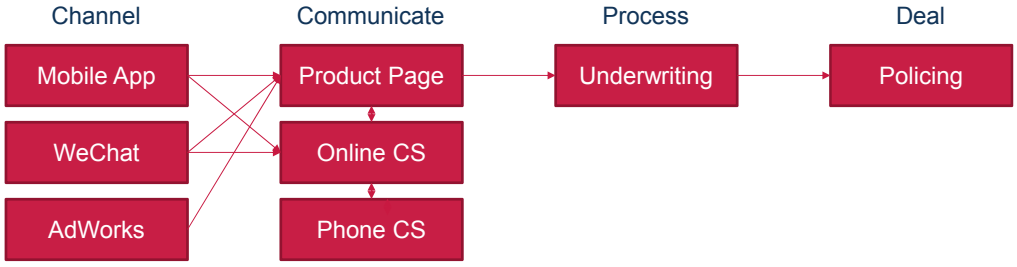
Data service: machine learning platform



- Integrated with OA approval process
- Provide models, algorithms
- Customizable algorithm component
- Lifecycle management of models
- Support back testing

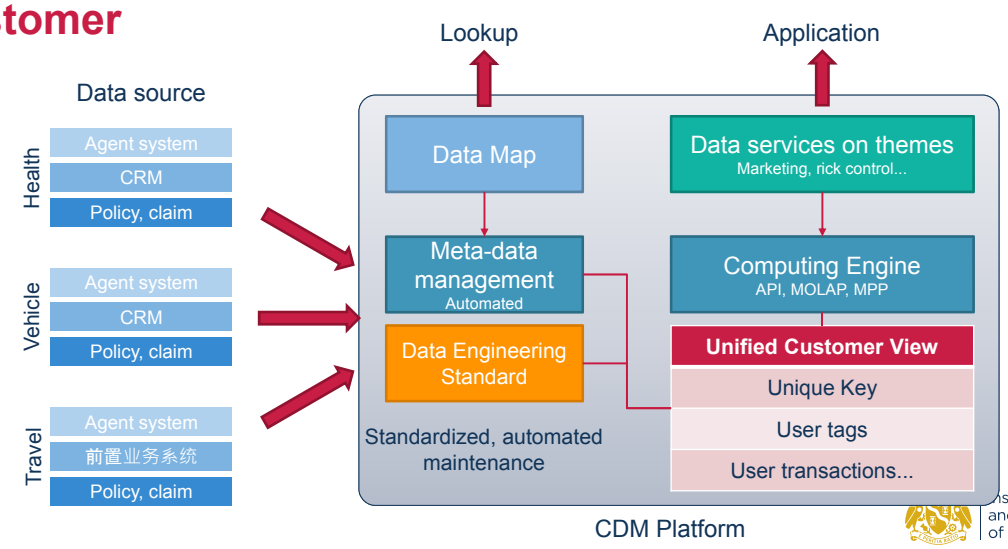
Application: customer conversion funnel

Target: By optimizing every stage of the funnel to improve the conversion rate, e.g. optimizing advertising strategy, pushing customized marketing information, simplifying underwriting process etc.

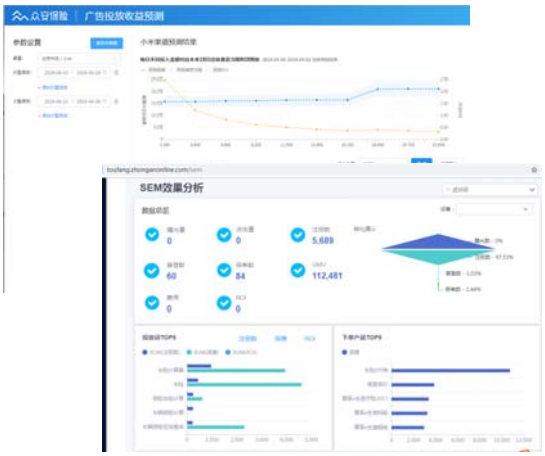


Difficulty: The data of this funnel flow through many business systems, and we cannot make sure every system has the same identifier of a customer, and the connection between each system is nested, so it's difficult to track the full context of a transaction.

Customer conversion data model: unified view of customer



Customer conversion: optimizing advertising strategy



- Joining data from different sources(ad-pages, product pages, app)
- Modeling user action event
- Predicting ROI of each channel on different group of people
- Optimize advertising strategy subject to constraints like budgets

Customer conversion: optimizing smart advisor chatbot

- Joining data from chatbot and policing system
- Intension understanding with NLP
- Tunnel analysis
- Predicting user sensitivity to certain corpus
- Optimizing process and corpus recommendation

Institute and Faculty of Actuaries

Outlook: data intelligence application pattern

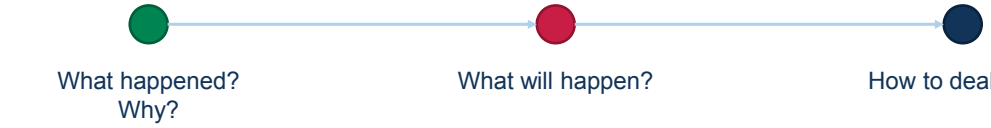
What happened? Why?

What will happen?

How to deal with?

Institute and Faculty of Actuaries

Outlook: data intelligence application pattern



What happened? Why?

What will happen?

How to deal with?

Get insight of data + **Prediction with ML** + **Optimize with OR**

Discovering underlying rules, connections and making inferences

Help making decisions satisfying objects and constraints

Institute and Faculty of Actuaries

Institute and Faculty of Actuaries

谢谢！

Thank you!

ありがとう！

감사합니다！

Merci!

Danke!