

CHINA: Big data analysis practice in retail and consumer goods industry

Limeng Wang, Senior Market Analyst
Yanxia Lu, Assistant Research Director
June 2022

IDC Doc# CHC45878420

Market changes shape digital transformation, while big data helps interpret consumer insights

- The retail industry is closely related to the factors that affect peoples' daily lives. In the past decade -- 2012 to 2022 – China's retail industry has expanded greatly with ecommerce, new retailing, and the COVID-19 pandemic. Amid tremendous changes in digital infrastructure, China's retail industry outpaces its upstream counterparts in digital transformation (such as supply chain), but still lags behind the leading "indicator-industries" such as the Internet industry.
- With the increasing focus on digital channels, consumers' preferences and behaviors Foot traffic to physical stores decreased and sales growth slowed dramatically through traditional channels. At the same time, new retailers are also meeting new challenges, such as 1) protecting consumers' data and privacy, tackling stagnant supply chain logistics, and dealing with abrupt disruptions in the region.
- The retail business has embraced multiple revolutions, such as vertical ecommerce, live-streaming ecommerce in social media, and the "forward warehouse" model of ecommerce for fresh produce, club warehouse, and community group buying. These changing cycles have made digital application the core competitiveness of the retail industry. For retailers, the challenges of the market are being met with increasing digital prowess.
- As we enter the age of epidemic prevention and control, the retail industry needs continuous improvement and upgraded responses to the changes in consumers' behavioral pattern. The industry is increasingly geared toward consumers, paying close attention to consumption scenarios, deep-seated consumption demands, and experience. Focus is shifting from "selling goods" to "designing experiences".
- To survive and grow, retailers must embrace a seamless integration of physical and digitalized products, services, and experiences. The customer journey can be analyzed through the retail ecosystem, while big data technology has already become the basis for refined operation in retail enterprises.



The key challenge in retail transformation hinges on value, cost, and talent within the organization

In 2021, many Chinese retailers are investing heavily in scaling digital capability. Retailers are faced with a large uptick in accumulated data -- consumer behavior data, supply chain data, etc. And, in turn, face challenges in exploring the value of the data and integrating data from multiple channels to generate comprehensive insights.

The impending digital / data transformations for the entirety of the retail industry leads to two main challenges:



Business value: Compared with industries such as finance and government, the retail industry is plagued by a low profit margin. As such, the cost to change is imposing. Only enterprises with high revenue --and a strong desire for business expansion-- are willing to make such investment.

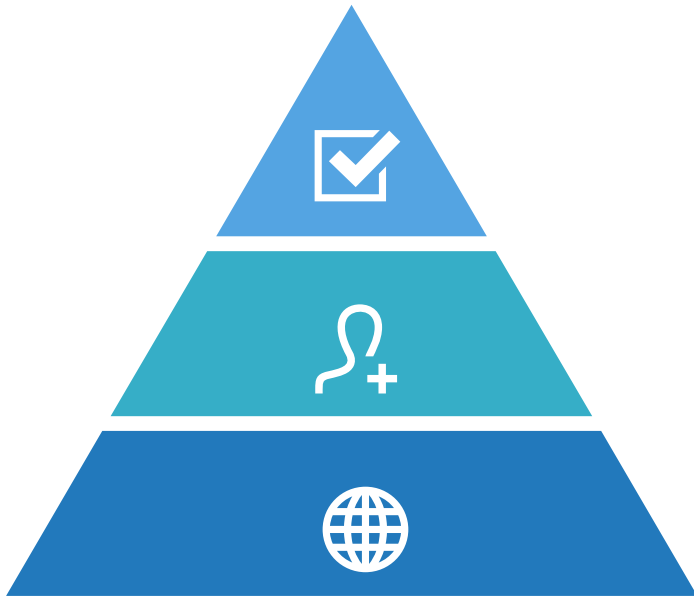


Costs: The cost of an in-house big data platform and other software systems is high. For example, few shopping malls and supermarkets can afford the investment needed to develop a new datacenter or recruit top-notch technology talent. Thus, traditional fast-moving consumer goods (FMCG) brand owners suffer from insufficient technical capabilities.



Organizational strategy and talents: If retail enterprises do not truly reorganize their original business models and systems as a whole, the challenge is steep – legacy physical systems, lack of technology talent, etc. Without a focused overhaul and investment, many shopping malls, supermarkets, and brand owners will find it increasingly difficult to survive.

Advantages of using big data analysis in the retail industry: extensive data, basic IT infrastructure, and maturity scenarios



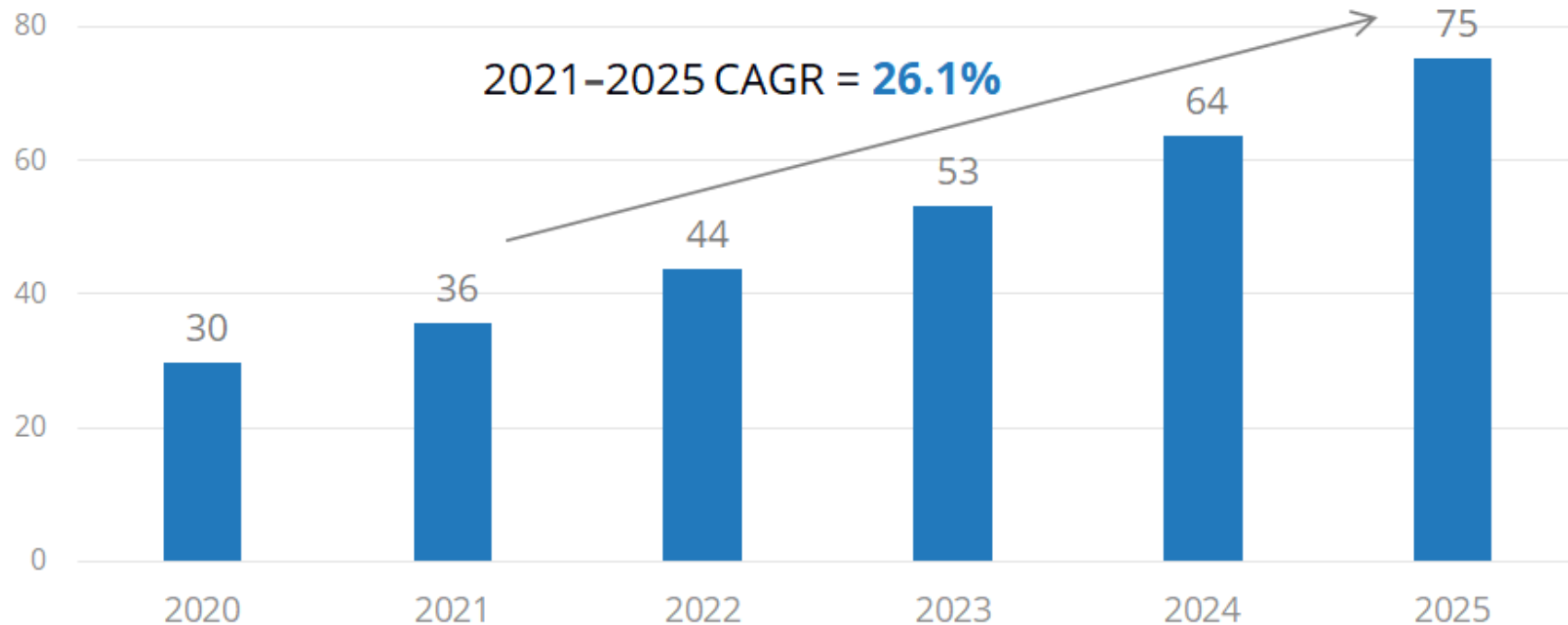
- **Retail should look to other industries to as a success template:** Leading industries, such as Financial Services and Internet “digital native” companies, can offer both functional points and practical experience that retailers can learn from. With a focus on: 1) transforming data into business value and 2) increasing the technical prowess of their in-house talent base, retailers can accelerate their path to change.
- For example, cloud service interfaces are convenient and easy to use, and adding technical talent in managing data warehouse and operations can be introduced.
- The industry has certain levels of **IT application and data analysis experience in its business system**, targeting the large-scale endpoint consumers. The retail industry is one of the mainstream industries for traditional data analysis with deep consumer data, a high frequency of transactions, and abundant available data.
- For example, systems such as ERP, CRM, and BI reporting have been in operation for many years, with high frequency of use and standardized product parameters and business processes. Business personnel are increasingly reliant on data year over year with improved ability to analyze and interpret data.
- Industry as a whole has been further digitized with extensive **data sources**. As the cost of data collection and labeling has steadily decreased over time. With the concept of data governance is gaining ground, and the scale of available data expanded, it becomes easier to obtain rich data products and services.
- Case in point: Data points are easier than ever to acquire through various sources: 1) pricing, 2) categorization, 3) purchase-sales-inventory (PSI) of competitive products, 4) consumer behavior data and 5) external data that can be used for analysis, such as weather, real estate, and social media content.

Rapid increase of spending on big data and analysis in the retail industry

The retail industry is getting further digitalized. New technologies such as big data technology fueled the upgrade of this industry, with new retail scenarios created and consumer experience reshaped.

IDC PRC Retail Big Data Market Spending Forecast, 2020–2025

Unit: RMB 100 Million



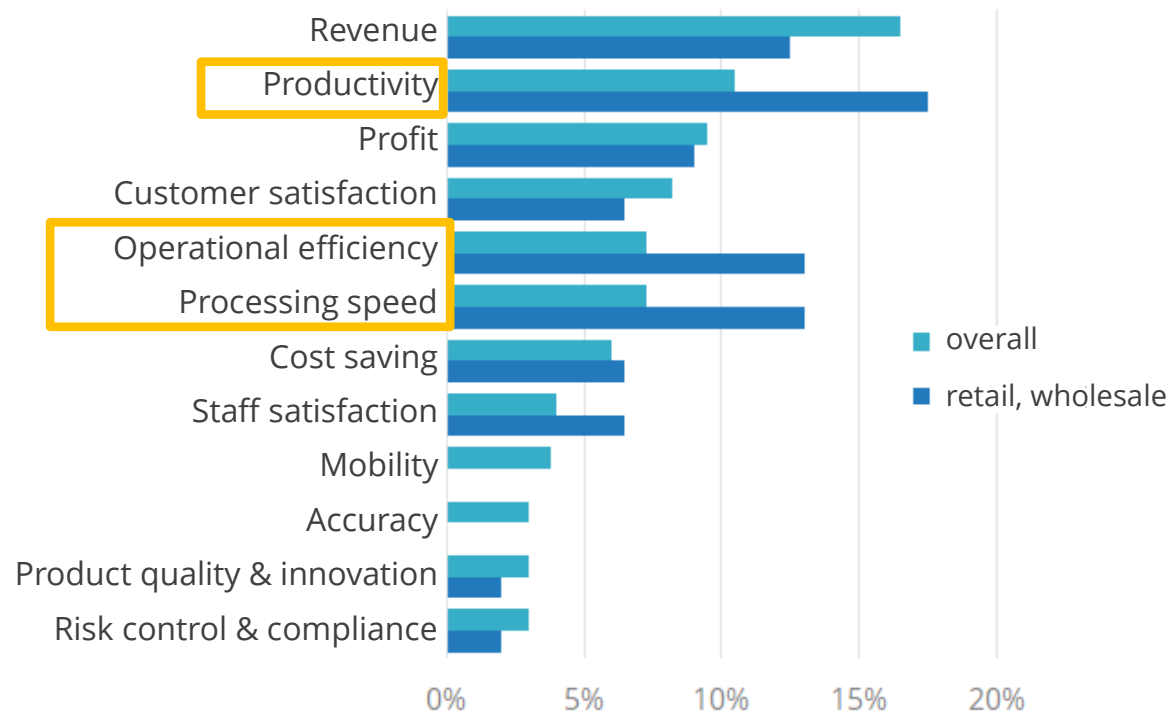
Note: The scope of the industry includes retail and wholesale, excluding consumer product manufacturing, which is consistent with the scope of the survey on the next slide.

Source: IDC's *Worldwide Big Data and Analytics Spending Guide: Release V1*, March 2022

The retail industry pays more attention to productivity, operational efficiency, and processing speed in intelligent applications

Question: Please describe the key performance indicators (KPIs) you use to measure effectiveness

Top 12 KPI categories



n = 333

Note: IDC divides respondents' answers into different "KPI categories," which are calculated as a percentage of respondents. Only the top 12 KPI categories from global respondents are shown.

Source: IDC's *Analytics and AI Services Survey*, June 2021

China: Seven best practices in the industry

Best practices of big data analysis conducted by different types of enterprises in the China market.

As is shown in Figure 1, seven typical cases are organized into four quadrants based on two dimensions — how long has the company been established, and what are their main business categories.

- I. **CaptainBI** and **Bianlifeng** are start-ups in the retail industry with each having five years of history. Between the two, Bianlifeng is an early pioneer in digital application, helping itself with self-developed technologies, while CaptainBI serves Amazon ecommerce sellers.
- II. **Bestore** and **Bliss Cake** are relatively traditional retail enterprises with each having around 15 years of history. Having explored the datacenter, now they are pragmatic players focusing more on business flexibility.
- III. Other traditional retail enterprises include two consumer brands, **Nature Home** and **Shinho**, which have been established for around 30 years. These two legacy/long-standing domestic brands are making new attempts as the representatives for consumer durables and FMCG.
- IV. Similar to CaptainBI, **Yunlizhi** is also a big data technology supplier serving terminal enterprises in the industry. It has been incubated under a traditional FMCG Group in recent years.

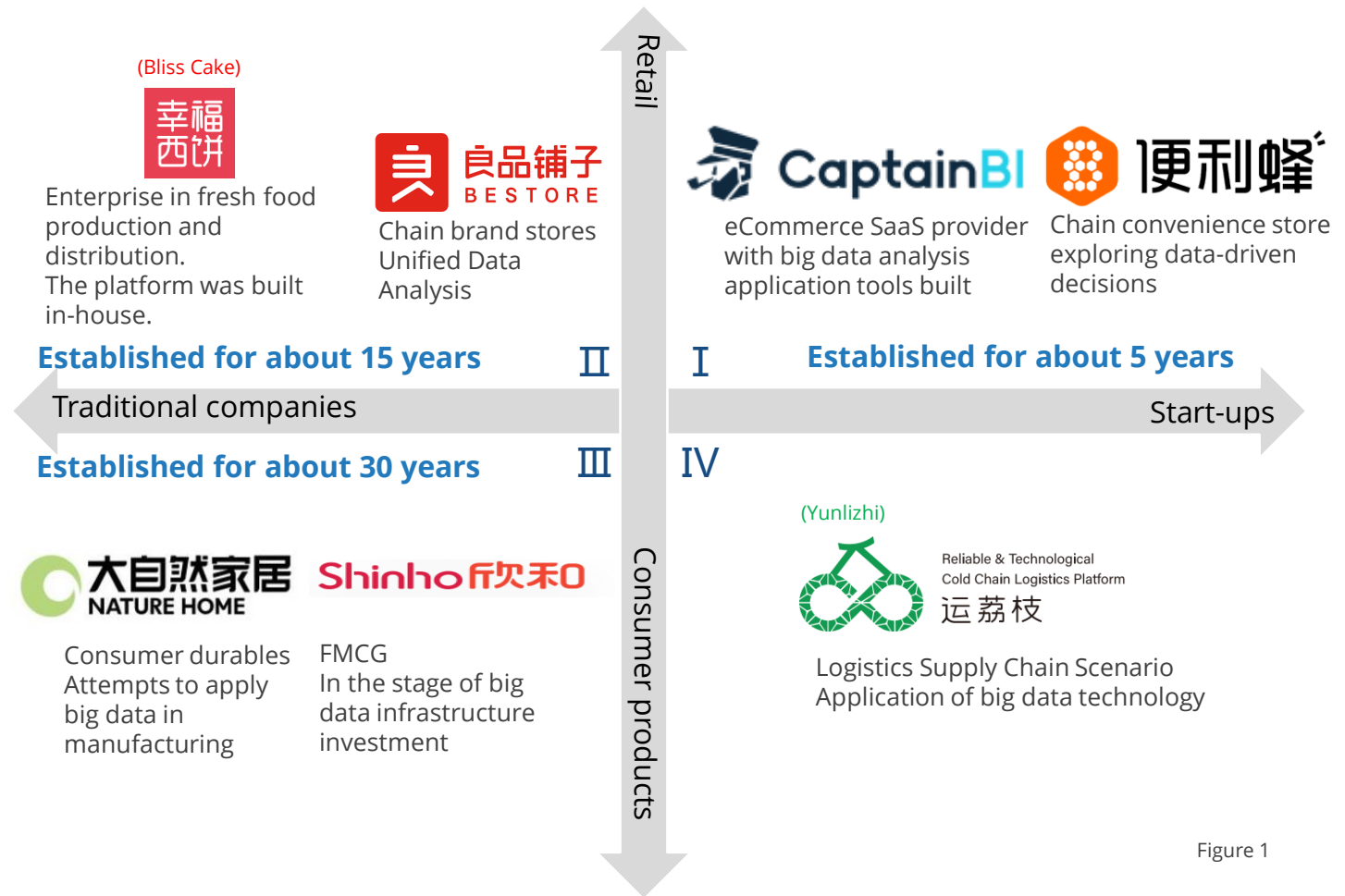


Figure 1



Case Study: FMCG in big data infrastructure investment: **Shincho**

Shincho 欣和

Established: 1992

Main business categories: FMCG — condiments and snacks

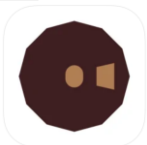
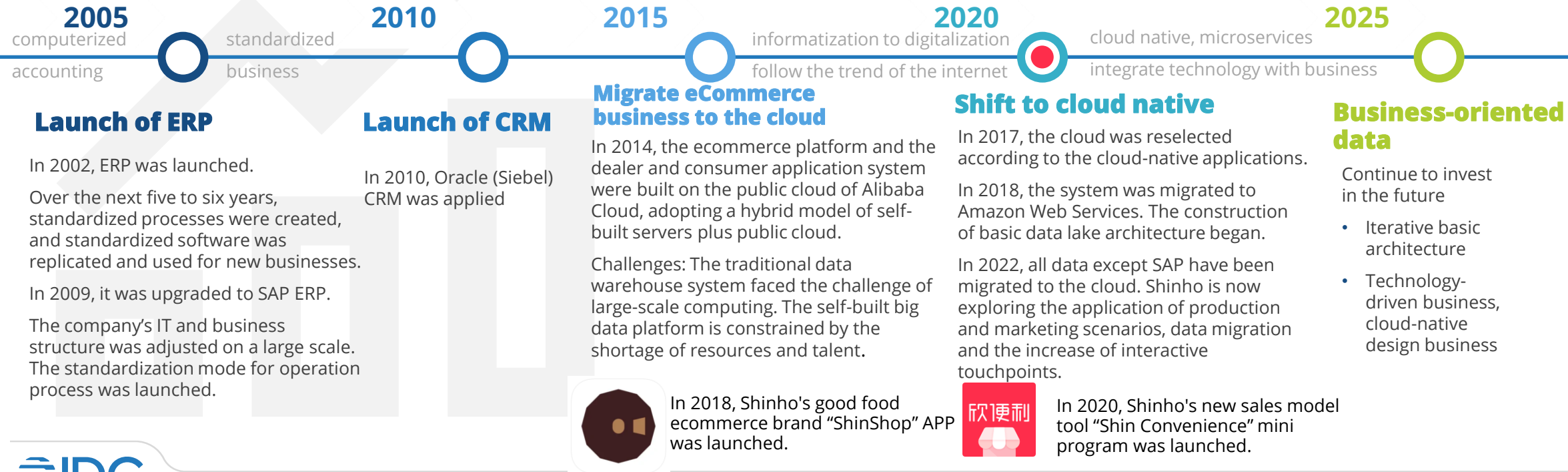
Shincho owns 10 plants, and has 11 product brands / more than a hundred products.

Sales scope: More than 60 countries and regions around the world

Total investment in IT application: Nearly RMB 100 million per year, big data accounting for a half of the investment.

Achievements

- ✓ The champion in digital application in the condiment industry.
- ✓ Redesign the technical architecture of the entire system based on Amazon Web Services, decoupled applications, and deployment of cloud-native services, in order to pave the way for improving business agility.
- ✓ Investing in AWS, Shincho plans on saving on intangible costs such as investment resources and technical team building in the big data infrastructure. Shincho is expecting a better return on investment in the future.



Case Study: eCommerce SaaS suppliers building application tools: CaptainBI



Established: 2016



Main business categories: Cross-border ecommerce software as a service (SaaS) service-operation analysis data platform/big data analysis service

Application of big data analysis: Internal business analysis system and CaptainBI for external sales

User scale: Serving more than 400,000 registered sellers in total on the Amazon ecommerce platform

Construction investment: Tens of millions of RMB was spent in total on Amazon Web Services (AWS)

Achievements

- ✓ Scalability: With the same cost, it can support a user increase by four to five times in 2021 and an increase by three times in the scale of daily data analysis.
- ✓ Data processing efficiency was increased, with time shortening from 5–6 hours to 20 minutes. The time for data analysis was also shortened from more than 10 seconds to 2 seconds.

Changes in the market: Demand increase

Main factors: COVID-19 pandemic → cross-border ecommerce industry
The global pandemic has triggered an explosive growth in China's cross-border ecommerce since 2020. As competition heats up, accompanied by rising logistics and marketing cost for sellers, the company is faced with increasing demand for refined operations.

Goal
Make sure that customers are provided with safe, stable, and reliable data services.

Selection process (optional scheme)
Self-built big data platforms and agile data warehouse solution, such as Oracle ADW, Greenplum, ClickHouse, Alibaba Cloud AnalyticDB (ADB), etc.

Stable online operation

The big data architecture that integrates data lake with data warehouse can efficiently expand the future demands for data mining and machine learning.

Challenges in the business
Rapid growth of users and data; low query efficiency of traditional relational database.

Approach
Optimize the technical architecture (high-quality big data architecture technology stack provided by cloud service providers)

Current choice (online solution)
Adopt Amazon EMR
Data storage — Amazon S3
Data processing component — Amazon Hudi
Real-time query engine — Amazon Presto, Amazon Athena

Case Study: Applying big data technology in the logistics supply chain scenario: Yunlizhi Technology



Reliable & Technological
Cold Chain Logistics Platform

运荔枝

Establishment date: 2019

Company background: A subsidiary of Fresh Life Style Supply Chain Management under New Hope Group

Main business categories: Cold chain/supply chain SaaS service

Application of big data analysis: Wholistic food supply chain. It serves both internal clients such as New Hope Group and the Fresh Life Supply Chain Management system, and external clients such as restaurant chains, food industry and trade companies, as well as shopping malls and supermarket retailers.

User scale: By the end of 2021, client base has grown to 2,500 clients.

Achievements

- ✓ Yunlizhi has helped clients reduce costs, improve efficiency and ensure commodity quality in cold chain logistics
- ✓ In terms of overall supply chain optimization, Yunlizhi has coordinated with Yunxian to help a restaurant chain client achieve a **55% cost reduction** in procurement labor and a **30% increase** in inventory turnover.
- ✓ In terms of store supply guarantee, Yunlizhi has helped a food factory client reduce days of inventory from **an average of 45 days to 14 days**.

Big data system (self-developed)

- Data services (supporting data collection and governance solutions)
- Big data resource platform based on Hadoop ecology
- Big data management tools (data pipelines, metadata, unified interface platform, etc.)
- Data application products (Data3BI, profile system, panorama analytics platform, digital twin–simulation platform, etc.)

Cloud services (Amazon Web Services)

- A large number of cloud native services and open-source components are used to reduce data mining costs.
- Learned from the best practices of leading clients in the industry and ensured data security and compliance.

Yunlizhi Technology Integrated Cold Chain Supply Chain Services

Customer demand	storage management scenario storage operation	urban distribution scenario Integration of warehouse and distribution Cars assigned for the region urban joint distribution	trunk route transportation scenario car-load shipment Less-than-carload shipment	Scenarios for digital upgrade and transformation transaction inventory dispatch performance monitoring payment					
solutions	food industry and trade	leisure goods	Chinese food chains	retail and e-commerce	Western food chains				
standard service	warehouse	trunk route	distribution	system	settlement				
digital/sm products	trunk network joint distribution and fitting	smart line arrangement cloud warehouse dictionary	fixed inventory management anti-channel conflict	automatic reconciliation warehouse digital twin	commission product performance cockpit	Lizhi Pay traceability system	warehouse and network optimization in-transit monitoring	recommended shipping capacity electronic receipt	business area profile green code for frozen products
Performance capacity	<p>A cold chain delivery network that covers the whole nation. The service covers 679 cities in 31 provinces across the country. The number of average daily order exceeds 50,000. There are more than 800,000 service stations across the country. More than 100,000 cold chain vehicles in cooperation. Cooperative cloud warehouse exceeds 3 million square meters.</p>				System empowerment OMS WMS TMS BMS FMS CRM APP Data monetization cargo owner driver vehicle goods station plan road network behavior Algorithm optimization GBST ARIMA k-meansS Dynamic programming Logistic regression trajectory data mining collaborative filtering				

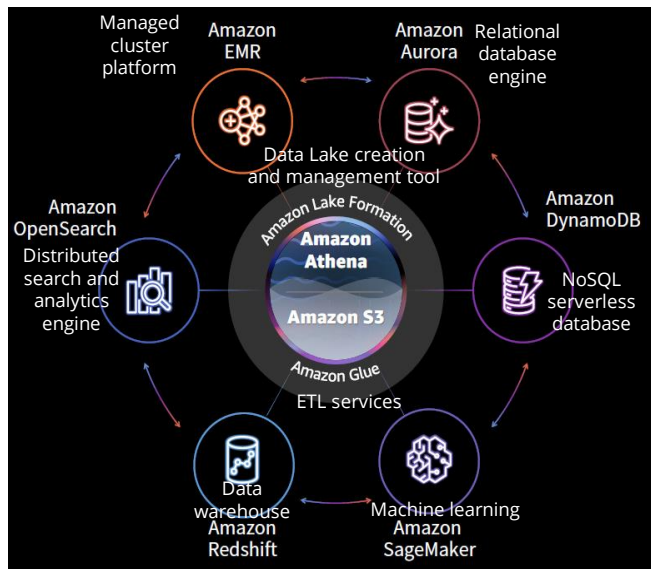
Amazon Web Services

Localized solutions are formed inline with clients' practical business needs, leveraging resources from the global product library, partners of Amazon Web Services across the world, and 20 years of experience in Amazon ecommerce platform to help domestic users go overseas and overseas users to land in the China market.

Solution Framework

Featuring the smart Lake House and personalized recommendations, it provides a set of mature basic service tools. Its rich product lineup are fully open in China.

Lake House Architecture on AWS



Advantages & Cases

- Based on 20 years of practical experience in Amazon ecommerce platform, results have been achieved in clients' businesses. It is internally driven to uphold "Day 1" spirit, by putting customers first and working backwards from their clients' customers' needs. $f(I) = (\text{mechanism} * \text{framework})^{\text{culture} * \text{organization}}$
- AWS cooperates with more than 1,000 retailers and ecological partners in China and abroad, selects the tools (AI/ML, IoT, microservices, data lake) needed according to their development directions, and conducts dedicated data analysis integrating multiple data sources for specific scenarios, so as to reduce data extraction and migration, improve performance and efficiency, and cut costs.

Client Type	Business Demands	Solutions	Results
Retailers' digital transformation: a fashion brand and a food brand	The growth of product sales results in HANA performance bottlenecks in internet datacenters (iDCs). The agility of daily report systems should be improved.	Migrate daily report system data to the cloud; migrate clients' daily report system data in FineBI to Amazon Redshift.	<ul style="list-style-type: none"> Maximum delay time reduced from 36 seconds to 8 seconds. Concurrent user capacity increased from 200 to 1,000.
	The lack of machine learning (ML) capabilities results in the risk of new product information leakage from the image recognition model during the customization process.	Use Amazon SageMaker to build an ML platform to realize ML recognition of merchandise display of stores.	<ul style="list-style-type: none"> New product data is confidential. Recognition cost per picture is reduced by 10 times.
Overseas users landing in the Chinese market: a fashion brand	To build a private domain data platform.	Build a smart Lake House using Amazon S3, Amazon EMR, and Amazon Redshift.	<ul style="list-style-type: none"> The CDP serves as an important asset of the digital marketing department. The daily business value it generates is incorporated into the evaluation system.
Chinese brands going overseas: a clothing brand and an advertiser	A lack of relevant technical professionals who can build big data platforms and big data analysis systems.	Build a smart Lake House Architecture based on Amazon Redshift by using data analysis-related components and services and applying Amazon Personalize to create additional purchase recommendation scenarios.	<ul style="list-style-type: none"> Reduce the costs of data storage, management, operation, and maintenance. Overall revenue increased by more than 4%.
	During peak periods such as ecommerce activities and product promotions, requests for computing resources and data interfaces surges, yet the speed of data processing is slow.	Use Amazon Lambda and Amazon EKS to handle requests in peak periods. Conduct asynchronous operation for data processing and divide data capture and data analysis into two microservices.	<ul style="list-style-type: none"> Performance expanded and stability improved in the high concurrency scenario. Data capture and processing speed has more than doubled.

Summary: Advice for Technology Buyers

1. According to the enterprise's development stage, strategic direction, available resources and current level of IT application, (and referring to best practices in other industries), tech buyers can set staged goals and scope for the enterprise's overall application in big data analysis. In addition to IT departments, it is also necessary to coordinate within the organization and its businesses so that big data analysis can be integrated into business processes and daily decision making.
2. Priority should be given to preliminary research, sorting out businesses and data while paying attention to the demand feedback from senior executives. On the basis of traditional data analysis, buyers can sum up the experience accumulated and problems set as well as the skills gap. Consider adding adequate talent and skill sets, and/or working with an external consulting team to plan and execute.
3. **In terms of business**, evaluate the problems to be solved, the conditions of application scenarios and expected results. Are their mature solutions in the industry that you can use / learn from, what are the possible obstacles in project implementation, will further upgrades be needed after the project is launched, and what support can be gained from technology suppliers and industry solution ecosystems?
4. **In terms of data**, evaluate the consistency of the enterprise's internal system data. The scope of business demands should be defined. Governance should be followed by analysis. Use web crawlers cautiously to extract external data so as to avoid the leakage of key business data. Understand the types of external data services available, such as data collection, labeling, governance, etc.
5. **In terms of big data technology**, big data is moving fast, with many mainstream big data components which need to be selected in line with business development demands. For instance:
 - *Factors of technical architecture selection:* 1) If the business develops and changes rapidly, flexible architecture is needed; 2) The enterprise needs to manage data type, scale, and growth; 3) The speed and real-time performance of data processing required in the business; 4) The range and frequency of historical data accessed; 5) Whether the architecture supports ML expansion, and whether AI capabilities are needed to support intelligent applications such as decision making in the next few years.
 - *Application value:* Focus on the maturity and usability of the technology and on whether the technology provider has sufficient service teams to solve issues in follow-up operation and maintenance such as performance optimization.
 - *Data security and compliance:* Evaluate the level of security required for different types of business and data. Consider working with several technology providers to disperse risks. Explore the balance between customer experience and data privacy in marketing, with an eye on risks to your brands' reputation when it comes to consumer privacy.
 - *Overseas business demands:* Get support from cloud service providers on overseas big data services.