Research on the construction and analysis of a green finance market research platform based on the background of big data research

ZhaoXuan Lv*

International education institute , Shandong-Agriculture-University, Taian 271000，Shandong, China

E-mail: lyu.zhaoxuan.578892@outlook.com

Abstract. In recent years, the term "big data" has become a topic that people have to talk about, but the term is so abstract that people can't seem to tell who came up with it first. Some say it's the Google team, some say it's McKinsey, a globally renowned consulting firm, and others say that futurists and data scientists came up with the concept of big data in the 1980s. There is a part of the population that looks at data macro-processing and considers big data as any form of data that cannot be processed well in a limited time with traditional software, a process that includes acquisition, storage, sharing, transformation, analysis, visualisation, etc. Some people think of big data as being primarily about data management, data analysis, and data visualisation. Others believe that big data is just data. This paper is based on the research background of big data combined with the development of the platform of green finance for research and analysis, the new financial field of research collection of popular big data technology for research and analysis.

Keywords: Big Data Context; Green finance; Market Research; Financial platforms;

1 Introduction

China's Internet-integrated financial platform has roughly experienced three stages of development: the first stage is the explosive growth stage. Marked by 27 companies obtaining third-party payment licences from the Central Bank in 2011, several Internet giants (BATJ, etc.) stepped up their layout of the financial sector, successively establishing fintech platforms such as Baidu Finance, Ant Financial, Tencent FIT, and Jingdong Finance, and rapidly expanding their business to financial areas other than payments [1]. At this stage, big tech companies are groping for financial business, and the scale is growing at a high rate [2].The third stage is the period of comprehensive rectification. Since November 2020, the People's Bank of China, the CBIRC and other four departments have conducted three regulatory interviews with Ant Financial Services, unveiling a new round of strong regulation of Internet finance. In April 2021, four

Y. Jiao et al. (eds.), *Proceedings of the 3rd International Conference on Internet Finance and Digital Economy (ICIFDE 2023)*, Atlantis Highlights in Economics, Business and Management 1,
https://doi.org/10.2991/978-94-6463-270-5_44

departments jointly interviewed 13 online platform companies engaged in financial business, including Tencent and Du Xiaoman Financial, further clarifying the determination to rectify the big tech financial platforms [3]. The platform economy is the foundation of the digital economy, but also a key link in the integration and development of the digital economy and the real economy, and is the organizer and coordinator of all kinds of digital economic activities. Accompanied by the rapid development of the digital economy over the years, Internet platforms have become more and more widely involved in resource allocation.

According to the China Academy of Information and Communication Research, the total value of China's super-billion-dollar Internet platforms grew from $770.2 billion to $350.43 billion in 2015-2020, with a compound annual growth rate of 35.4%. However, the rapid development of Internet platforms has also brought about many problems, such as platform "two choices", big data "familiarization", inter-enterprise shielding, data leakage, and pinch-point acquisitions, etc., which have affected the healthy development of the digital economy.

Based on the reality of the development of China's digital economy, the Classification and Rating Guidelines draw on international regulatory thinking to scientifically define the categories of platforms, rationally classify platforms, and enhance the pertinence and effectiveness of platform regulation.

All in all, the appearance of these two guides means that "the rules of platform governance in China have taken a big step forward." The biggest difficulty in the previous Anti-Monopoly Law was the determination of the dominant position of the platform, but now, after the platform classification, there is a quantifiable standard for the platform level, so that the responsibility of the platform is clearly delineated, to a certain extent, the platform supervision can be more targeted, which is also more conducive to the healthy development of the industry as a whole.

2 Characteristics and major risks of the development of Internet-integrated financial platforms

There is currently no academic consensus on the classification of Internet finance, mainly because of the very different ways and development paths of technology companies entering the financial industry in different countries, and the complexity of Internet financial business innovation itself [4][5]. In this paper, the financial services currently operated by China's Internet-integrated financial platforms are classified into three major categories of payment, financing [6], and wealth management according to the nature of their business, as shown in Table 1.

Table 1. Business classification of Internet financial holding platforms

form	Specific operations	typical example
payment category	non-bank payment industry	Alipay, Paypal

financing category	consumer finance	Suning Consumer Finance
	microcredit	short term loan
	private bank	Qianhai Microbank
	creditworthiness	Sesame Credit, Tencent Credit
Wealth Management	Insurance brokers, internet insurance	Cathay Property and Casualty Insurance, Micromin Insurance Agency
	Public fund management, fund sales	Tianhong Fund, Teng An Fund Sales

2.1 Risk characteristics of Internet-integrated financial platforms

Internet financial platforms can achieve a surge in the number of users and business scale in a short period of time by virtue of their unique ecosystem and traffic advantages [7][8]. For example, the size of the third-party payments market grew rapidly from $1.2 trillion in 2013 to nearly $250 trillion in just seven years (see Figure 1), and its growth rate remained consistently high at over 100 per cent until 2018.

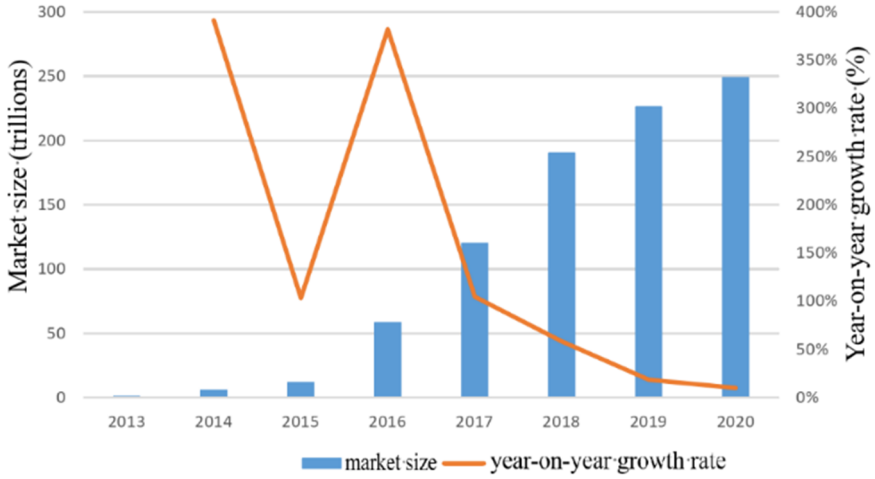

Fig. 1. China's Third-Party Mobile Payment Transaction Size and Growth Rate, 2013-2020

However, the rapid growth of financial business will pose a serious challenge to the risk-taking ability of Internet companies due to their limited level of financial services management [9].

3 Cluster task scheduling algorithm optimisation

3.1 Scheduling Algorithm Optimisation Principles

3.1.1 Metrics of data skew.

If there are a large number of computation tasks piled up by the system on a certain Worker node or several Worker nodes, it causes other computation nodes to wait for a large amount of time, which is called data skewing [10]. An example of Spark's join operation is shown in Figure 2.

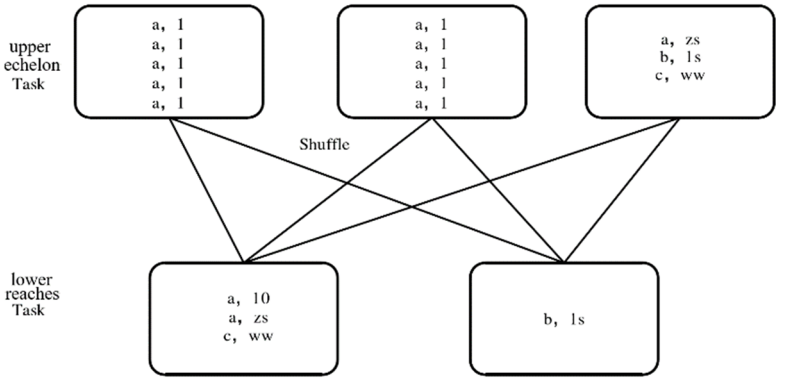

Fig. 2. Shuffle's join operation execution flow

(1) In the Stage phase: this paper uses the Coefficient of Variation (CV) to evaluate the skewness of the data tilt, as shown in the formula (1), and uses Cou_k to indicate the degree of skewness in the kth Stage. In Equation (2), Std_k is used to denote the standard deviation of the kth task, Data denotes the size of the data volume of the kth Stage, ω denotes the number of Tasks contained in that Stage, and l denotes the number of Tasks in each stage. Equation (3) represents the average of the number of Tasks contained in all Stages. According to the final result, the closer the result of Cou_k is to 1, the more skewed the Stage is; the closer the result of Cou_k is to 0, the less skewed the Stage is.

$$Cou_k = \frac{Stk_k}{Data_k} \tag{1}$$

$$Std_k = \sqrt{\frac{\sum_{l=1}^{\omega}(Data_{k.l} - \overline{Data_k})^2}{\omega_k}} \tag{2}$$

$$\overline{Data_k} \frac{\sum_{l=1}^{\omega k} Data_{k.l}}{\omega_k} \tag{3}$$

(2) In Task stage: this paper uses Factor of Skew (FS) to define the degree of skew of each Task in a certain Stage. In Eq. (4), $FS_{k,n}$is used in this paper to describe the

skewness of the nth Task of the kth Stage. This paper also assigns two critical thresholds to FS, which are upper and lower bounds. The upper bound is given a number greater than 0 and the lower bound is given a negative number.

$$FS_{k,n} = \frac{Data_{k.l} - \overline{Data_k}}{\overline{Data_k}} \tag{4}$$

3.1.2 Shuffle Optimisation Principle.

Shuffle is the bridge that distinguishes between wide and narrow dependencies in the DAG scheduling phase. operators that perform Shuffle operations are called wide dependencies, and those that don't are called narrow dependencies. the execution speed bottleneck of Spark's compute engine largely stems from the speed of the Shuffle phase. Distinguished in terms of broad categories, the entire Spark computation phase is divided into two important parts, ShuffleMapStage and ResultStage, as shown in Figure 3.

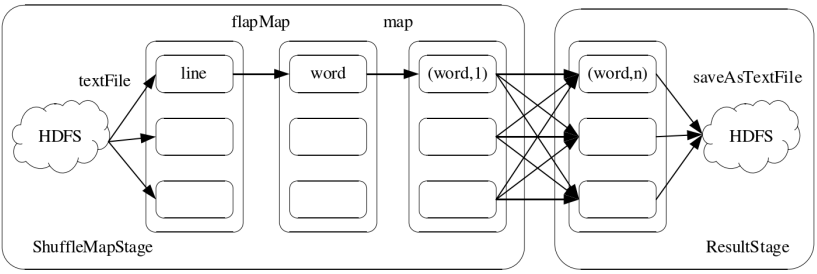

Fig. 3. Shuffle scheduling model

4 Conclusion

Combined with the current situation of platform development in China, and based on the platform's connecting objects and main functions, the Classification and Rating Guidelines categorize platforms into six major categories, namely, online sales platforms, life service platforms, social entertainment platforms, information and message platforms, financial service platforms, and computing application platforms.Internet platforms connect to different objects, provide different services, and have different impacts on the rights and interests of consumers, as well as on the operating order of the economy, society and politics.Through the classification of Internet platforms, the General Administration of Market Supervision can have a more in-depth and comprehensive understanding of the characteristics of different types of platforms, so as to accurately formulate the key points of governance of different types of platforms, and thus enhance the level of intelligence, globalization, personalization and refinement of China's platform governance, as well as enhance the level of equalization, universality and convenience of public services.This paper introduces the principle of the scheduling model of Spark big data engine and its performance optimisation scheme, and describes the two problems that still exist in the task scheduling model, generating a large

number of intermediate tem-porary files and the uneven allocation of tasks on the server to generate data skewing problems. Focus on the task execution process of Spark and the principle and opti-misation scheme of HashShuffleManager. In order to improve the efficiency, the TaskGroup concept is proposed for the scheduling model, which groups the Tasks of each Stage and performs caching and I/O operations on the data in groups. It also improves the default HashShuffleManager scheduling algorithm, which evenly allo-cates Tasks according to the CPU utilisation, memory utilisation, disk utilisation and queue length of each node as a reference unit, which solves the problem of multiple equals due to data skewing, and greatly enhances the execution efficiency of the system. The platform economy is a new type of economic form with the Internet platform as the main carrier, data as the key production factor, new-generation information technology as the core driving force and network information infrastructure as the important support. In recent years, China's platform economy has been developing rapidly, and its status and role in the overall economic and social development has become increasingly prominent. It

Reference

1. Ai Xinzhou, Bai Jiadan. Reflections on the high quality development of financial industry fuelled by financial technology[J]. Cooperative Economy and Technology,2023(18):48-49.DOI:10.13665/j.cnki.hzjjykj.2023.18.060.
2. Park Y. Artificial intelligence fuelling the development of green finance[J]. Finance and Economics,2023(18):12-14.DOI:10.19887/j.cnki.cn11-4098/f.2023.18.033.
3. Zhu Xinrong,Liu Xiaohan,Xu Jiayi. Collaborative digital technology and green financial development[J]. China Finance,2023(12):57-58.
4. Du Jinhuan. Analysis of green credit implementation effect of Huzhou Bank under the empowerment of financial technology[D]. Hebei Institute of Finance,2023. DOI:10.27837/d.cnki.ghbjr.2023.000115.
5. Ou Chen,Yao Shujun. Research on digital governance of green finance[J]. North Finance,2023(03):40-45.DOI:10.16459/j.cnki.15-1370/f.2023.03.001.
6. Chen Huimeng,Song Dandan. Practice and Suggestions of Digital Transformation Enabling the Development of Green Finance Business in Commercial Banks[J]. Financial Zonghengheng,2022(12):55-60.
7. Zhang Wuqiao,Guo Haiyan. Research on the mechanism of green industry development in Guizhou Guian New Area empowered by big data[J]. Southwest Academic,2022(00):167-181.
8. Zhang Yalei. Comparative Analysis of Green Financial Technology Application at Home and Abroad[J]. Financial Zonghengheng,2022(10):45-50.
9. Zheng Jianhui. Research on the digital transformation of green finance in China's commercial banks[J]. Southwest Finance,2022(09):69-80.
10. Huang Zhuo,Wang Pingping. Fintech Enabling Green Financial Development: Mechanisms, Challenges and Countermeasures[J]. Social Science Journal,2022(05):101-108.