Banking Comprehensive Risk Management System based on Big Data Architecture of Hybrid Processing Engines and Databases

Shenglan Ma[1], Fangkai Xie[1], Hao Wang[2*], Hong-Ning Dai[3], Ran Tao[1], Ruihua Yi[1], Tongsen Wang[1]
1. Division of Science and Technology, Fujian Rural Credit Union, Fujian, China
2. Department of ICT and Natural Sciences, Norwegian University of Sci. & Tech., Aalesund, Norway
3. Faculty of Information Tech., Macau University of Sci. and Tech., Macau, China
* Corresponding Author: Hao Wang, Norwegian University of Sci. & Tech., hawa@ntnu.no

***Abstract*—Banks are shifting from a simple credit risk management model to the comprehensive risk management model. Banking risks come from many channels and systems. Big data technology provides an innovative and effective solution for data management, and thus is suitable to be applied in the risk management scenarios that require high-quality data and complex data analysis. This paper firstly proposes big data architecture of hybrid processing engines and databases. This architecture uses Hadoop ecosystem with ETL and Spark processing engines, and using massive parallel processing databases (MPP), transactional databases, and HDFS. Then a banking comprehensive risk management system prototype based on the proposed big data architecture is implemented. Comparisons and evaluations clearly demonstrate that the proposed system has better performance.**

Keywords—comprehensive risk management, big data, hybrid architecture

INTRODUCTION

The bank is a special enterprise that operates risks [1]. The basic responsibility of the bank is to stick to the bottom line of not having systematic and regional financial risks. Risk management capabilities are not only become an important component of a bank’s overall competitiveness, but also a key factor in its long-term sustainability [2]. The signing of the New Basel Capital Accord [3] and its supplementary agreements have set clear standards for the supervision of the banking industry, prompts banks to shift from a simple credit risk management model to credit risk, market risk and operational risk [4].

The basis of risk management lies in the effective identification and measurement of risks and emphases on technological innovation to develop and use comprehensive risk management tools. The common-used risk management systems are individual systems built for specific risks or embedded in banking business systems. And the traditional risk management systems are focus on the formulation of response measures instead of prevention and management so that they cannot deal sudden and unforeseen risks. Moreover, there are many barriers in the traditional management systems, and then the data cannot be shared and handled timely.

Current researches have paid close attention on big data technology in the area of risk control. PayPal uses image mining technology in online transactions, historical data for each fraud transaction record [5]. ZestFinance provides credit assessment services for 15% US customers using more than a dozen assessment models to analyze thousands of raw data, including data from third-parties (such as phone bills and rental history) and borrowers themselves [6]. The development of big data technology will change the channels and mechanisms for information acquisition, analysis and application, and create technical conditions for risk management. On the one hand, effective data cleansing and data mining techniques can identify key information in the credit risk management from the massive amounts of data generated by customer transactions. On the other hand, the carrier of banking services is increasingly integrated with social media and e-commerce. Big data technology can integrate the structured and unstructured information generated by customers' online and offline behaviors [7], breaking the data boundaries and enabling banks to form three-dimensional tracking and evaluation of behaviors [8]. The application of big data technology will also effectively improve the bank's own liquidity risk management automation level, and vigorously enhance the decision support of measurement analysis for liquidity risk management [9]. Using the big data technology platform, it can realize the integrated operation of compliance, internal control and operational risk management systems [10].

Thus this paper proposes a comprehensive risk management system based on big data architecture. The architecture with Spark [13] and ETL engines uses hybrid databases including Massive Parallel Processing databases (MPP [11]) , transactional databases, and HDFS [12]). The proposed system supports comprehensive risk management services to achieve collateral management, internal assessment, risk early warning, risk combination and International Financial Reporting Standard (IFRS [14]). This paper is divided into five sections. Section II briefly introduces the comprehensive risk management business. Section III proposes a banking big data with hybrid processing engines and databases based on the analysis of risk data source. Section IV proposes a comprehensive risk management system prototype based on big data architecture and proposes some application cases. Finally, we draw our conclusion in Section V.

Comprehensive Risk Management

In the comprehensive risk management of banks, it mainly deals with credit risk, market risk, operational risk, liquidity and compliance risk [15].

Comprehensive risk management is a long-term and complex system engineering guided by advanced risk management concepts, involving development strategies, corporate governance, organizational structure, management processes, information systems and corporate culture. The so-called comprehensiveness is mainly reflected in various risks such as credit risk, market risk and operational risk. Fig 1 shows the bank's overall risk business from the perspective of data usage.

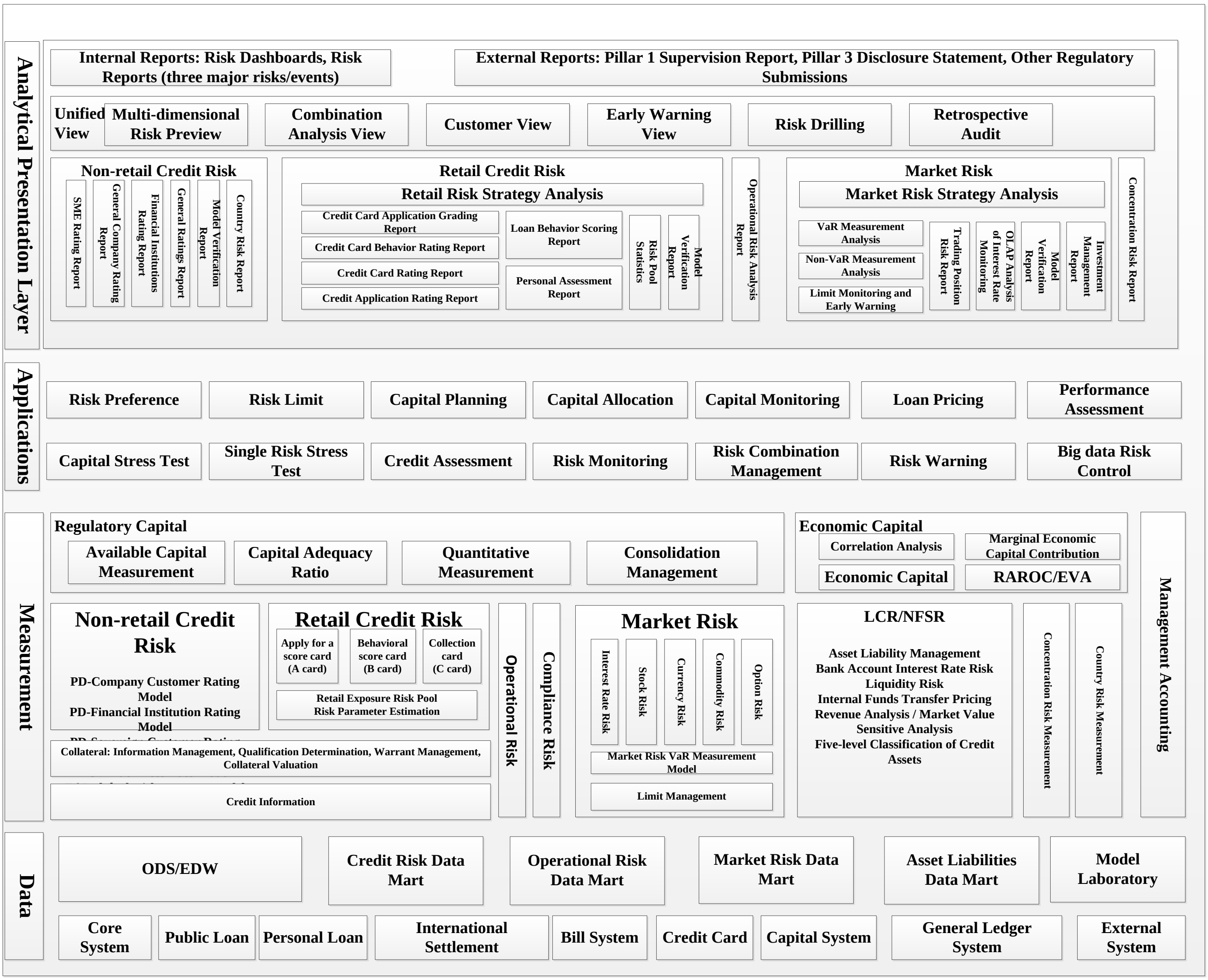

Fig. 1. Bank's Overall Risk Business Figure

In the comprehensive risk management, banks need to develop risk strategies when formulating development strategies, business plans and management performance assessments. Under the framework of risk preference, banks need to establish risk management processes that meet the development needs and regulatory requirements of each business as the basis for various business management activities. Banks need to improve their technical support capabilities and improve the data governance framework for risk and capital management.

1. Collateral Management. The collateral management implements the whole process monitoring and management of bank repossessed assets, including collateral management, registration and warrants management, valuation management, third-party organization management, risk early warning, report management and system management.

2. Internal Assessment. The credit risk management system is divided into risk identification, risk measurement and internal assessment applications. Risk identification is based on bank account classification, including company, bank, sovereignty, equity, retail and others. Risk measurement

includes customer rating and debt rating. Internal assessment applications are divided into non-retail internal assessment and retail internal assessment. Non-retail internal assessments require multiple adjustments and calculations with complex division of exposures, multiple models and complicated logic. Retail internal assessments have large variation in product caliber with multiple fraud methods and frequently updating control policies. The new form of Internet makes the new retail credit risk control models need to quickly adapt to the requirements of the new economy. Therefore, the internal assessment applications are required to support for diversified data collection and processing, various internal assessment application processes, and various flexible internal assessment reports and customer risk views.

3. Risk Early Warning. Risk early warning considers the company's financial situation, account behavior, credit information and other qualitative information, and then analyzes the bank's real-time solvency, operating conditions, profitability and growth ability to ensure the capacity of debt repayment is consistent with or improved with the pre-lending access setting time. For enterprises whose repayment ability is in a downward trend, corresponding measures shall be taken according to the level of warning indicators to reduce and eliminate potential risks of customers. Therefore, it is necessary to collect the basic data of customers and businesses to form an indicator system with the use of rules for real-time monitoring [16].

4. Risk Combination. The risk combination mainly includes economic capital measurement, risk and income measurement, concentration risk measurement, risk dashboard, risk combination report, value combination analysis, limited planning and risk-benefit optimization allocation, concentration monitoring, risk warning and stress testing. Comprehensive stress testing covers all major risks and businesses areas inside and outside of the table. It fully considers the interaction and feedback effects of various businesses and the possible non-linear relationship between risk factors and pressure indicators. The stress test reflects the overall risk profile of banks and banking groups.

5. International Financial Reporting Standard (IFRS [14]). IFRS is promulgated by the International Accounting Standards Board (IASB). It addresses the accounting for financial instruments, which contains three main topics: classification and measurement of financial instruments, impairment of financial assets and hedge accounting. On the measurement layer, IFRS includes classification decision rules, impairment engine, valuation engine and accounting. The classification decision rules access upstream transaction data, screening unclassified financial products for classification [17]. The accounting module first sets the relevant elements in the system, configures corresponding rules of accounting, and centrally processes the business events that occur in the valuation engine and the devaluation engine, and generates accounting documents.

BANKING BIG DATA ARCHITECTURE

This section will propose a banking big data architecture with hybrid processing engines and databases based on the risk data analysis.

Risk Data Analysis

1. Collection Content

The risk data includes both structured and unstructured data (S and uS) [18]. Structured data is stored in the common business systems and can be directly loaded in to the data warehouse. The unstructured data such as business credentials should be specially deal to mine the knowledge. These two types' data has financial information (F), such as basic information, financial information, account flow, money laundering classifications, etc., as well as non-financial information (nF), such as hobbies, behavior information, location records, etc.

2. Collection Source

The collection source comes from internal and external source (I and E). For internal data, information required for risk management from various banking business systems. For external data, risk data comes from Internet platforms (such as banking WeChat customer services), government agencies or third-party agencies (such as communications operators, credit bureaus, etc.).

3. Collection Method

For internal data, data are downloaded from business systems, and loaded into risk data mart after Extraction–Transformation–Loading (ETL [19]). For external data, structured data can be directly introduced by Spark or Flume based on unstructured data processing engines (uDPE), while unstructured data needs to undergo text or image analysis for data processing.

Main risk data is shown in Table I.

TABLE I. RISK DATA

Data	Source	Content	Method
Transaction	I or E	S+F,uS+F	uDPE,ETL
Money	I	S+F	ETL
Contact	I	S+nF	ETL
Recognition	I	S+nF	ETL
Contract	I	S+F or S+nF or uS+F	ETL
Basis	I	S+nF	ETL
Resource	I or E	S+F or uS+F	ETL
Regulatory	I	S+nF	ETL
Relation	I or E	S+nF	uDPE,ETL
Event	I or E	S+F or S+nF or uS+F or uS+nF	uDPE,ETL
Statics	I	S+nF	ETL

Therefore, based on the complicated data form, the big data technologies are proposed to process the comprehensive risk management [20-21].

Banking Big Data Architecture

a) Big Data Engines

The current common big data processing engines include batch processing framework, stream processing framework, and hybrid framework, as shown in table II.

TABLE II. COMMON BIG DATA ENGINES

Framework	Engines	Contents
Batch Processing	Apache Hadoop	Hadoop's processing capabilities come from the MapReduce engine. MapReduce processing technology meets the map, shuffle, and reduce algorithm requirements using key-value pairs. It is deal for dealing with very large-scale datasets that are not time-critical.
	ETL	ETL is the process of extracting, cleaning, and transforming the business system data and then loading it into the data warehouse. The purpose is to integrate the scattered, disorderly, and inconsistent data in the enterprise and provide analysis basis for the decision-making of the enterprise.
Stream Processing	Apache Storm	Storm is the best solution for near real-time processing. This technology can process data with very low latency and can be used to obtain the lowest latency workload.
	Apache Samza	Apache Samza is a stream processing framework tightly tied to the Apache Kafka messaging system. The technology can provide fault tolerance, buffering, and state storage through Kafka.
Hybrid	Apache Spark	Spark batch capabilities provide unparalleled speed advantages at the expense of higher memory footprint. Another consequence of this design of Spark memory computing is that if deployed in a shared cluster, you may encounter resource shortages. Spark consumes more resources than Hadoop MapReduce and may affect other tasks that need to use the cluster at the same time.
	Apache Flink	Flink provides low latency stream processing while supporting traditional batch processing tasks. Flink is suitable for organisations with extremely high flow processing requirements and few batch processing tasks. This technology is compatible with native Storm and Hadoop programs and can run on YARN-managed clusters.

Among them, Apache Hadoop can be seen as a processing framework that uses MapReduce as the default processing engine. Engines and frameworks can often be used interchangeably or simultaneously. For example, another framework, Apache Spark, can incorporate Hadoop and replace MapReduce. This interoperability between components is one of the reasons why the flexibility of big data systems is so high. Therefore, banking big data architecture always is based on the Hadoop ecosystem and uses the Spark engine to process the real-time tasks and batch computing, and the ETL engines to the data warehouse.

b) Databases

With the exponential growth of data volume in banks, the speed of data analysis and processing is significantly increasing, and the traditional transaction databases (such as Oracle, which is commonly used for banks with large data volumes and high security capabilities) can no longer meet the demand and hinders the development of the business.

Massive Paralleling Processing Database adopts SHARE-NOTHING distributed parallel processing flat architecture. Each data nodes in the database clusters enjoys its own hardware resources to achieve the purpose of sharing tasks in parallel. It can provide a cost-effective universal data computing platform for ultra-large-scale data management of banks with high-performance, high-available, and high-scalable ability.

MPP database can achieve the following advantages from hardware deployment and software development:

1. Low hardware cost and high scalability. The PC SERVER using the X86 architecture does not require expensive UNIX servers and disk arrays. It has simple operation and maintenance, and support online expansion.

2. Columnar storage and parallel computing. The parallel distributed processing technology based on columnar storage is applied in the big data platform which can avoid single-point performance bottlenecks and single points of failure. Each node can support 100TB of raw data. Nodes in the cluster are shared-free, with peer-to-peer computing capabilities, and can support the storage and calculation of up to 10PB data.

3. Efficient compression storage. Use hash or random distribution strategy for data distributed storage to reduce the space by 1 to 20 times, and improve I/O performance accordingly. It supports instance level, library level, and table level compression.

4. Smart Indexing. With coarse-grained intelligent indexing technology, the index expansion rate is no more than 1%. Compared with traditional Oracle indexes, the index space is greatly saved. The smart indexing contains column-based statistics, which can be directly used during data retrieval and positioning, while effectively filtering data, significantly reducing database disk I/O.

5. Redundancy mechanism. Redundancy mechanism is used to ensure high availability of the cluster. Automatic synchronization can be implemented among the mutually-supplied fragmented data. Through the replica, the MPP provides redundant protection, automatic fault detection and management, and automatic synchronization of metadata and service data.

Therefore, banking big data architecture chooses MPP for the massive data storage and transactional databases for the operational applications.

c) Architecture

This paper presents a Hadoop ecosystem based big data architecture, which consists of transactional databases, MPP databases and HDFS [23]. The transactional databases are used to perform On-Line Transaction Processing (OLTP [22]) to process online business data and the MPP databases handle high-value density structured data.

Bank big data architecture is divided into data access layer, data exchange layer, data service layer and data application layer, as shown in Fig 2. The details of the processing engines and databases are shown in Table III.

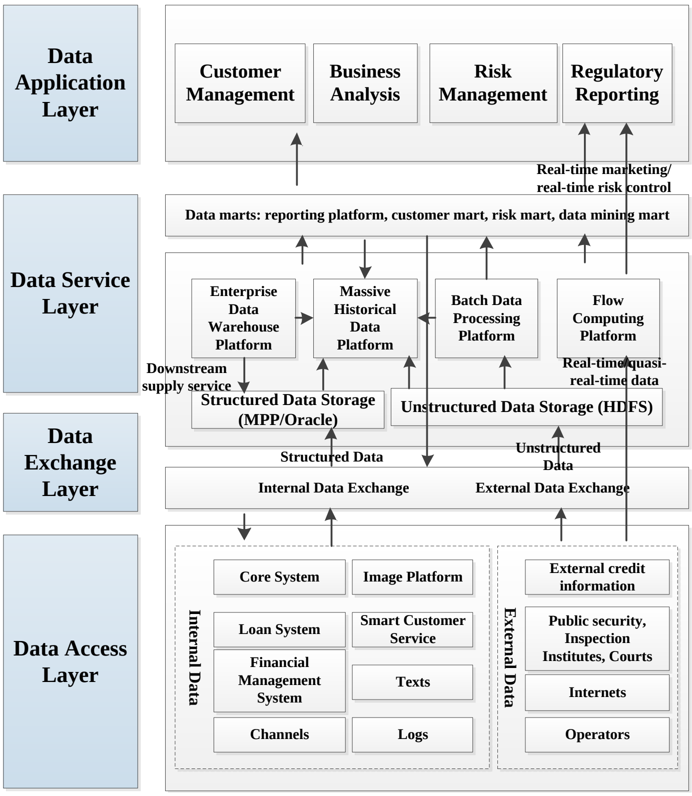

Fig. 2. Bank Big Data Architecture

TABLE III. TECHNOLOGIES OF DATA SERVICES LAYER

Platform	Engines	Databases	Contents
Enterprise Data Warehouse Platform (EDWP)	ETL	MPP	The MPP databases serve as a back-end database engine mainly for structured data processing. The cleaned data canbe loaded into the MPP directly for the processing of data marts.
		Oracle	The transactional database mainly deals with online business data and adapts to a large number of business scenes that need to add, delete and modify. The decision data from the data warehouse is saved to Oracle.
Massive Historial Data Platform (MHDP)	Apache Spark	MPP	Use Spark to batch dump the current data in the MPP to the history database of the MPP, and simultaneously dump the unstructured data in the HDFS to the MPP history database in rich text format.
Batch Data Processing Platform (BDPP)	Apache Spark	HDFS	Use Spark to process unstructured data in HDFS in parallel and generate data mart to to MHDP and EDWP.
Flow Computing Platform (FCP)	Apache Spark	-	Use Spark to extract real-time from data sourcs and comput the results to the clients.

1. Data Access Layer.

Data access layer accesses to bank internal and external data. Internal data includes major business system transaction data (such as core system, loan system), image data (such as business credentials), system logs, etc. External data includes governments, and regulatory authorities, etc. Currently, there are often some financial blockchain data accessed to the bank data architecture.

2. Data Exchange Layer.

Data exchange aims to two-way exchanges from business systems, bank outlets and external systems. The business systems include raw transaction data generated by business systems, and at the same time, data such as model data, indicators and market data refined by the basic platform is supplied to each business systems and supports business operations. The data exchange layer distributes data such as model data, indicators and marts extracted from the basic data platform to outlets to support individualized business analysis and also accesses personalized data of each bank outlet. The external systems are important supplements to in-row data by accessing data from various external industries. It also provides data realization and data services.

3. Data Service Layer

The data service layer includes three types of databases: MPP, transactional database and Hadoop Distributed File System (HDFS). Two of the key data storage areas are structured data areas and unstructured data areas. Structured data is the row data, and the value is stored in the database. Most of the internal data in banks are structured data. Bank structured data area consists of data origination area, data theme area and data aggregation area. The processing of the structured data area uses the MPP database, which mainly deals with the data in the near, medium and long term. Through the efficient processing of the MPP database, transactional data is periodically pushed to the data application layer to provide fast calculation and services. Unstructured data area processing and structured data area processing is different. Through the establishment of big data platform (Hadoop cluster and Spark cluster, that Hadoop is responsible for off-line analysis services and Spark is responsible for data pre-processing services), the real-time data collection and processing scenarios can be solved. The unstructured data area processes the data and pushes the computational data into the transactional database of the data application layer, such as the fund net value, large amount of funds changed information, and so on.

4. Data Application Layer

The data application layer uses the data service layer interface to build data analysis application systems. Through the complete indicator sets (as shown in the measurement of Fig 1) and common summarized data, the front-end portal improves the ability of self-service analysis by introducing BI tools, multi-dimensional analysis and data mining.

MPP-based EDWP

MPP-based data warehouse is the main processing link of structured and risk-based marts. It is a service level system based on big data architecture. Therefore the details of EDWP in the Fig 2 are shown in Fig 3.

PC Server infrastructures support horizontal expansion of the system, including MPP database and Hadoop platform. The components include metadata management, unified scheduling monitoring, ETL service, batch file exchange, real-time data exchange, data quality management platforms. Data access is used to control the security and the standardization of the data warehouse, and provides access using web service [24] or API.

The data exchange layer deals internal and external data, structured and unstructured data. The MPP database is used for the processing of recent and medium-term data, including the standard data model (SDM), foundation data model (FDM), analysis data model (ADM) and index data model (IDM). Unstructured data areas are processed differently from structured data areas by the Spark engine. Historical data archive store archived data for all areas and provides archival data query services by Spark engine.

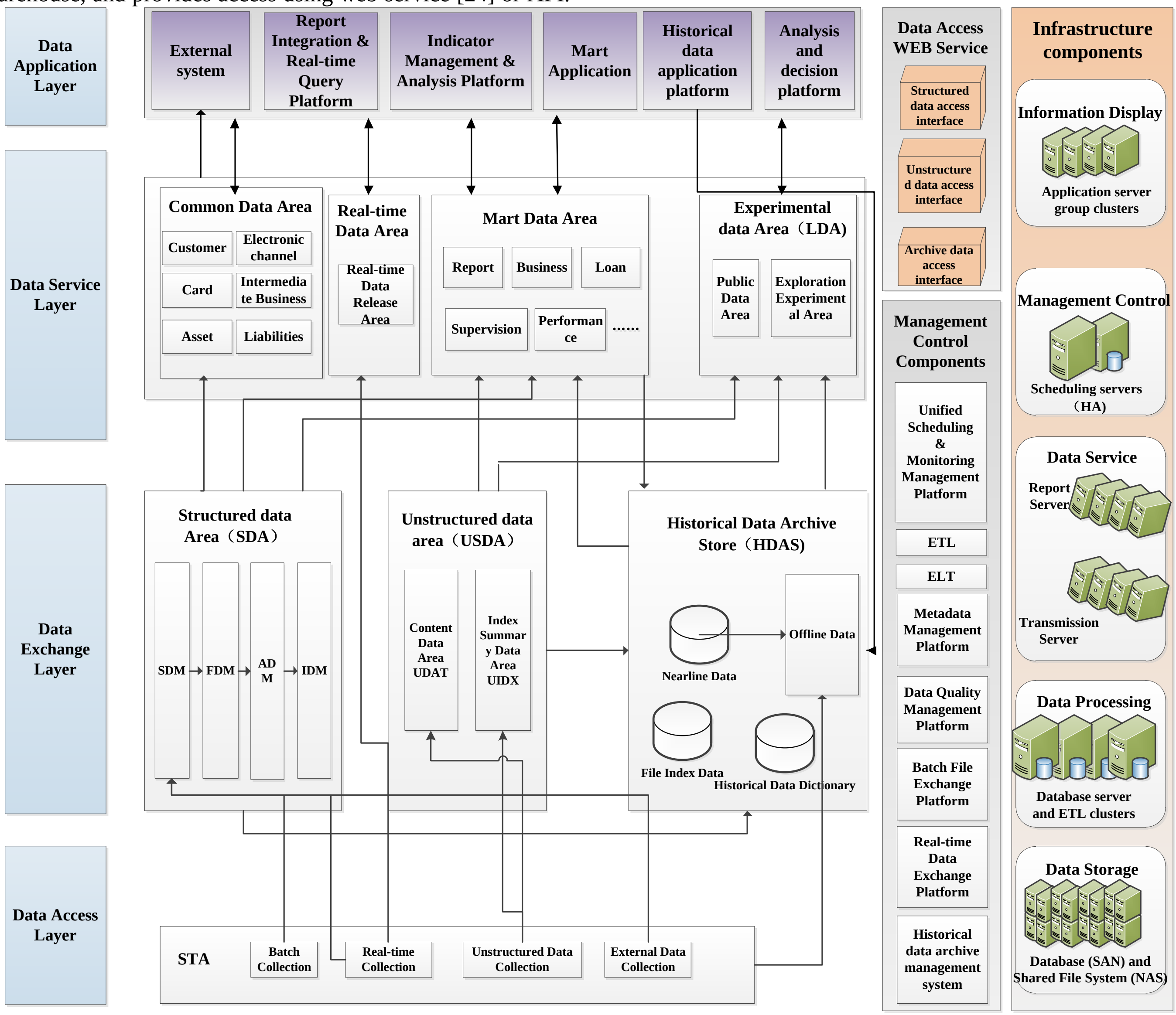

Fig. 3. MPP-based Data Warehouse

BANKING COMPREHENSIVE RISK MANAGEMENT SYSTEM BASED ON BIG DATA ARCHITECTURE

System Design

This section describes the realizations of comprehensive risk management systems based on big data architecture as shown in Fig 4. It is an interconnection of multiple business systems with the unified data. It focuses on the support of internal or external data application capabilities, real-time processing capabilities and machine learning model application capabilities. It mainly uses scenario-driven on-demand data storage.

In the data access layer and exchange layer, Spark and ETL engines are mainly used for data collection and exchange, including batch, real-time data collection and subsequent upgrades to a full-line data acquisition platform to achieve decoupling of upstream source data systems and big data platforms. The risk data from the business systems is retrieved by the \ETL to MPP databases. And then, the unstructured data is retrieved to HDFS by the Spark processing engine.

In the data service layer, some risk tasks such as data storage, value mining, and real-time risk detecting are processed. It is the core layer of big data platform, providing powerful computing and stream processing capabilities, and at the same time modeling storage of external data and on-demand data.

Data application layer is mainly used for interaction services between risk applications and big data platforms, and subsequently upgraded to a full-line data service platform to achieve decoupling between downstream applications and big data platforms.

In the red dotted line box, the real-time risk warning is computed in the FCP. The bottom two parts are data management, scheduling management, log management and other bank-wide unified technologies or platforms.

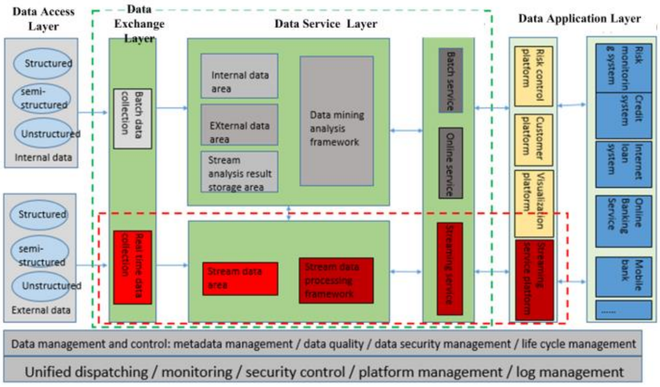

Fig. 4. Comprehensive Risk Management Platform

Application Cases

1. Loan Full Process Control

Loan full process control is carried out from three stages: before the loan, in the loan and after the loan.

In the stage that before the loan, the data mining module in the data service layer are carried out to find the high quality customers who meet the conditions of the loan access. Then, in the stage that in the loan, the customer's transaction information, basic information and external information in the EDWP can be analyzed. Then the system update the loan review information in real time, and timely alert the customer's abnormal information. Finally in the stage that after the loan, the system automatically generates monitoring reports, dynamically monitors customer risk, collateral risk and collateral valuation based on the BDPP and FCP. And the indicator sets are retrieved from the risk marts in the EDWP.

2. Enterprise Customer Risk Warning

In the traditional mode, the risk monitoring of the enterprise customers depends mostly on the customer visits, the public opinion of the enterprise, the lawsuit information, the analysis of the annual report of the enterprise and the risk information of the enterprise. The identification and early warning of risk are often lagging behind by low time limitation, high cost and low accuracy, and the risk identification cannot be present in the first time or even missing.

The way to carry out the enterprise customer risk early warning based on the proposed big data architecture. Firstly, to maximize the scope of all relevant information of the enterprise, the system exchange the whole lifecycle data such as the enterprise basic information, risk information, and public opinion information. Secondly, the system unifies the internal and external data, cleaning and standardization, providing data compatibility services through the integration of data by ETL. Then model the data into to the risk marts according to the characteristics of the enterprise, such as credit model, credit rating model, and so on. Further, using the reptile technology to continuously get the public opinion information and loaded into the BDPP by the Spark engine for the real-time monitor.

Evaluations

Table IV shows the comparisons among the comprehensive risk managemanet platform based on the proposed big data architecture with other conventional platforms. The conventional platform is always used as a single application, such as collateral management system, internal assessment system, risk early warning system, and etc.

TABLE IV. THE COMPARISIONS AMONG THE COMPREHENSIVE RISK MANAGEMANET PLATFORM BASED ON THE PROPOSED BIG DATA ARCHITECTURE WITH OTHER CONVENTIONAL PLATFORM

Comparisons	**Proposed Big Data Architecture**	**Commonly Used Single Systems**
Processing Efficiency	With the Spark processing engine and MPP database, the risk indicator can be quickly processed and pushed to the clients.	With the transactional database, the huge indicator may be runned slowly. See Table V.
Data Integrity	EDWP can store the whole risk data together and form the risk mart for analysing.	Each risk control application needs to process different data, and it will cover many application to complete one risk analysis task.
System Overhead	Using the big data architecture, the whole data analysis and process procedure canbe integrated into the server cluster.	The network loads by the system communication are costly.

Moreover, to prove that the MPP database outperforms the traditional Oracle one (Oracle is widely in the banks), some data processing types that have been used in the Oracle database, including “insert only”, “truncate & insert”, “delete & insert”, “insert & update” and “update only”. To omit the influence of the different operation capability, all the experiments are running on the virtual machine with the same CPU cores, memory capacity and hard drive storage. The experiments repeated for 50 times and the average time cost for different data processing type is shown in Table V. The data is credit transaction of customers, and is always used for the credit card fraud risk detection. The results clearly show that

MPP takes less time to finish the data processing task, especially for the “truncate & insert” and “update only” operation. And MPP costs less storage to save the same data volume, as shown in Table VI.

TABLE V. DATA VOLUME FOR THE EXPERIMENTS

Experiment Type	Table Name	Initial Data Volume (byte)	Changed Data Volume (byte)
Insert only	F_INB_VAS_DUKDTL	0	812318
Truncate & insert	F_AGR_DEP_DPSFWBXTJ	46359334	46497443
Delete & insert	A2_PRD_CRD_MAIN_M	2724012	350945
Insert & update	F_INB_LDG_BALCH	8927	598612
Update only	F_CIF_CIFE_CBI	469010	69844

TABLE VI. AVERAGE TIME COST AND DATA STORAGE OCCUPIED

Table Name	Oracle (time /s)	MPP (time /s)	Oracle (MB)	MPP (MB)
F_INB_VAS_DUKDTL	25	2	133	53.2
F_AGR_DEP_DPSFWBXTJ	3890	710	9819	3927.6
A2_PRD_CRD_MAIN_M	180	11	392	152.8
F_INB_LDG_BALCH	361	3	176	88
F_CIF_CIFE_CBI	1053	9	203	81.2

CONCLUSIONS

Banks face many challenges in conducting comprehensive risk management, including the lack of integration of existing data, distribution in different business systems, inconsistent data coding, and failure to record key business data into databases, etc. To solve the inadequacies of existing risk measurement systems, this paper proposes a comprehensive risk management system based on the big data architecture of hybrid processing engines and databases. The system supports the collateral management system, internal assessment system and risk early warning system with efficient data services. Qualitative and quantitative experimental comparisons are also shown that the proposed system and architecture has a better performance

ACKNOWLEDGEMENT

This work is partially funded by the Fujian Fumin Foundation and partially supported by the National Natural Science Foundation of China under Grant No. 61672170 and the Science and Technology Planning Project of Guangdong Province under Grant No. 2017A050501035.

REFERENCES

[1] Anderson, Ronald C., and Donald R. Fraser. "Corporate control, bank risk taking, and the health of the banking industry." Journal of Banking & Finance 24.8 (2000): 1383-1398.

[2] Cordella, Tito, Giovanni Dell'Ariccia, and Robert Marquez. "Government guarantees, transparency, and bank risk-taking." (2017).

[3] Hofbauer, Günter, Monika Klimontowicz, and Aleksandra Nocoń. "The changes in capital structure of selected banking markets as a result of new regulations." European Financial Systems 2017 (2017): 222.

[4] Bessis, Joel. Risk management in banking. John Wiley & Sons, 2015.

[5] Gupta, Rajendra, and Piyush Kumar Shukla. "Performance Analysis of Anti-Phishing Tools and Study of Classification Data Mining Algorithms for a Novel Anti-Phishing System." International Journal of Computer Network and Information Security 7.12 (2015): 70.

[6] Su, Fei, et al. "A Survey on Big Data Analytics Technologies." International Conference on 5G for Future Wireless Networks. Springer, Cham, 2017.

[7] Lam, James. Enterprise risk management: from incentives to controls. John Wiley & Sons, 2014.

[8] Spira, Laura F., and Michael Page. "Risk management: The reinvention of internal control and the changing role of internal audit." Accounting, Auditing & Accountability Journal 16.4 (2003): 640-661.

[9] Waemustafa, Waeibrorheem, and Suriani Sukri. "Systematic and unsystematic risk determinants of liquidity risk between Islamic and conventional banks." (2016).

[10] Kiron, David. "Lessons from becoming a data-driven organization." MIT Sloan Management Review 58.2 (2017).

[11] Java, Mayur, et al. "Collateral management: A changing landscape." Journal of Securities Operations & Custody 9.1 (2017): 25-29.

[12] Zikopoulos, Paul, and Chris Eaton. Understanding big data: Analytics for enterprise class hadoop and streaming data. McGraw-Hill Osborne Media, 2011.

[13] Zaharia, Matei, et al. "Apache spark: a unified engine for big data processing." Communications of the ACM 59.11 (2016): 56-65.

[14] Biddle, Gary C., et al. "Do Adoptions of International Financial Reporting Standards Enhance Capital Investment Efficiency?." (2016).

[15] Yeh, Quey-Jen. "The application of data envelopment analysis in conjunction with financial ratios for bank performance evaluation." Journal of the Operational Research Society 47.8 (1996): 980-988.

[16] Liu, Qing. "An Optimization Approach for Market Potential Indicator System Evaluation of Cultural Tourism Resources based on Cloud Platform." Revista de la Facultad de Ingenierá 32.5 (2017).

[17] Chawla, Gaurav, et al. "Point-in-time loss-given default rates and exposures at default models for IFRS 9/CECL and stress testing." Journal of Risk Management in Financial Institutions 9.3 (2016): 249-263.

[18] Blumberg, Robert, and Shaku Atre. "The problem with unstructured data." Dm Review 13.42-49 (2003): 62.

[19] El-Sappagh, Shaker H. Ali, Abdeltawab M. Ahmed Hendawi, and Ali Hamed El Bastawissy. "A proposed model for data warehouse ETL processes." Journal of King Saud University-Computer and Information Sciences 23.2 (2011): 91-104.

[20] Gandomi, Amir, and Murtaza Haider. "Beyond the hype: Big data concepts, methods, and analytics." International Journal of Information Management 35.2 (2015): 137-144.

[21] Hipperson, Tim. "The changing face of data insight–And its relationship to brand marketing." Journal of Database Marketing & Customer Strategy Management 17.3-4 (2010): 262-266.

[22] Plattner, Hasso. "A common database approach for OLTP and OLAP using an in-memory column database." Proceedings of the 2009 ACM SIGMOD International Conference on Management of data. ACM, 2009.

[23] Shvachko, Konstantin, et al. "The hadoop distributed file system." Mass storage systems and technologies (MSST), 2010 IEEE 26th symposium on. Ieee, 2010.

[24] Erl, Thomas, et al. Web service contract design and versioning for SOA. Prentice Hall, 2017.