



Design and Implementation of A2 System Regional Center Data Synchronization Scheme

Yang Jiao^{1,2}(✉), Shan Xie^{1,2}, Hong-mei Deng^{1,2}, and Jian-hua Su^{1,2}

¹ Research Institute of Exploration and Development, Changqing Oilfield Company, Xi'an, Shaanxi, China

jyang_cq@petrochina.com.cn

² National Engineering Laboratory for Exploration and Development of Low Permeability Oil and Gas Fields, Xi'an, Shaanxi, China

Abstract. Based on the production data management system of oil, gas and water wells (A2), Changqing Oilfield has developed comprehensive query, big data analysis and other deepening application functions, which play an important supporting role in oil and gas exploration deployment, production dynamic analysis and other work. At present, Changqing A2 Regional Center has many problems in daily production data synchronization, such as long synchronization time, low timeliness, weak verification logic, and data timeliness and accuracy are facing challenges. In view of the above difficulties, this paper designed a new data synchronization scheme and completed the feasibility test of key technical points by analyzing the publish and unlock process of A2 data and the current synchronization logic, and built a full process closed-loop synchronization system of “publish awareness, decentralized synchronization, unlock identification and local update”. Through actual operation, the average data synchronization time was shortened from 83 min to 2 min, and the unlock recognition rate was improved to 100%, which realized the efficient application of A2 production data and boosted the high-quality secondary accelerated development of Changqing Oilfield. At the same time, the scheme has good guidance and reference significance for other oilfield companies that need to build A2 regional data center.

Keywords: A2 System Regional Center · Data synchronization · Unlock identification · Full process closed loop

1 Introduction

The production data management system for oil, gas and water wells (A2) is a professional information system built by China National Petroleum Corporation (CNPC), aiming at the organization and management of the core data of oil and gas field production and development. Changqing Oilfield has developed a series of comprehensive query, big data analysis and other in-depth application function modules based on A2, which play an important supporting role in oil and gas exploration deployment, development plan adjustment, production performance analysis and other work [1]. In 2018, the

A2 system of Changqing Oilfield was migrated from Xi'an to Beijing Changqing Data Center. After the migration, due to network bandwidth, link stability and other reasons, the A2 related oilfield self-built system cannot connect to the Beijing A2 database (Beijing database), resulting in the normal operation of each system. Therefore, Changqing Oilfield has redeployed a set of databases (Xi'an database) that are isomorphic to Beijing Database in Xi'an. At first, GoldenGate software was used for real-time data synchronization, but due to frequent failures, it brought huge workload of operation. Therefore, a programmed automatic data synchronization mode was adopted every night to support the data application of the oilfield self-built system. This synchronization mode does not judge A2 data modification, which makes it difficult to ensure data accuracy. In addition, the data published today can be used only the second day, which is difficult to meet the new needs of the secondary accelerated development of the oilfield. It is urgent to build a new data synchronization system [2].

2 Research on the Mechanism of A2 Data Synchronization and Unlocking

Currently, A2 data synchronization consists of three phases: company data publishing, basic entity synchronization and daily data synchronization. Company data publishing needs to go through three levels of auditing, including operation area, oil/gas production plant and oil field company. When the second level auditing is completed by the oil/gas production plant, A2 system will automatically summarize and check the data of every plants under the jurisdiction of Changqing Oilfield Company, and complete the first level auditing and the official publish of company data, usually before 15:00 a day. Basic entities, including well basic information, well bore information, well status history information, etc., start up at 20:00 a day and take an average of 3 min. Daily data synchronization involves common data such as production day data, mining machine data and test sample data of 36 plants. The process starts at 22:00 and takes an average of 83 min to complete deletion and increase of data in the last three days.

A2 data unlocking refers to the publish of company-level data after the completion of the publish, for the existing abnormal data, each plant can apply to unlock all the data this month from the date of the abnormal, the unlocked data can be modified, and need to publish and synchronize again. However, the current synchronization logic only complete deletion and increase of data in the last three days. It cannot accurately identify the unlocking plants and date range, and cannot process data for more than three days. It requires manual problem detection and manual day-by-day processing.

In summary, the existing A2 data synchronization mechanism has simple logic, less code, and artificial operation, which achieves the data "published today, applied the next day". However, with the increasing speed of scientific research and production, more and more oil/gas production plants put forward the urgent need for data "published today, applied today". It is urgent to build an agile synchronization system with high efficiency, accurate judgment and reliable data [3].

3 Data Synchronization Scheme Design

3.1 Overall Scheme Design

Based on the analysis of existing synchronization logic, the overall implementation scheme of A2 data synchronization designed in this paper includes basic entity synchronization, daily data synchronization and unlocked plant data resynchronization. The scheme can synchronize the data in time after the initial release of the data of each plant, and accurately identify and resynchronize the data unlock of the current month in the subsequent synchronization process [4, 5]. The scheme changes the data synchronization mode from synchronizing the data of 36 plants for 3 days at a time to real-time synchronizing the data published by each plant (See Fig. 1).

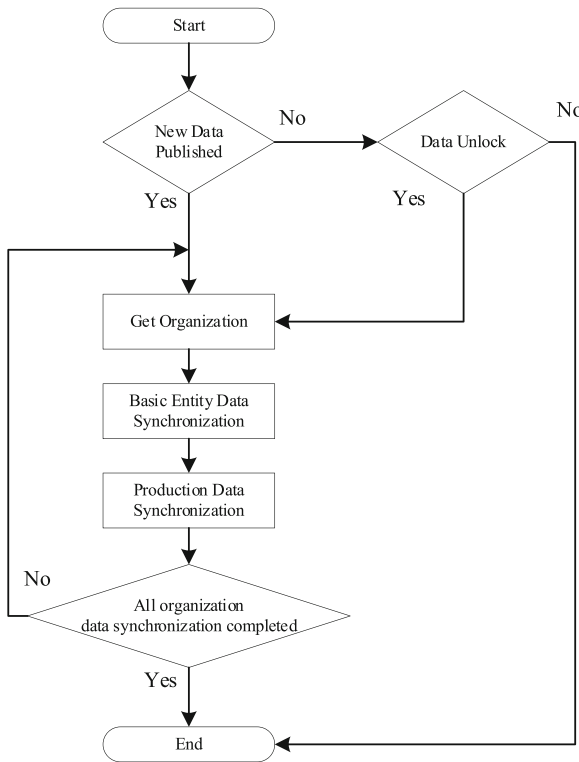


Fig. 1. Overall scheme of data synchronization

3.2 Release Status Judgment of Oil/Gas Production Plant

The data audit at the oil/gas production plant level is a secondary audit. The audit shows that the production data to be synchronized on that day has been released at the plant level. Since the plant-level release status is a sufficient and unnecessary condition for

each data table to complete the data entry on the current day, the data table containing the approval status field within the synchronization scope can be used as the starting data synchronization condition.

3.3 Unlock Status Judgment

The essence of unlocking is the re-publishing of data, which is completely opposite to the process of data publishing. If the changed data is in the operation area, it needs to be unlocked by the company, then by the plant level, and finally by the operation area, and the data will be reprocessed. For data unlocking over a period of time, it is necessary to unlock data day by day from the current data release date until the start date of the data change. Since data needs to be republished after unlocking, data unlocking can change the data release time. By comparing the data release time, it can be determined whether there is data unlocking or multi-day unlocking in the oil/gas production plant [6].

4 Construction of Whole-Process Synchronization System

4.1 Data Table Structure Analysis

In terms of daily data synchronization, the old delete all add all data synchronization method does not need to determine which company the data in the data table belongs to, but the new synchronization method needs to specify how the company relates to the data in different data tables, so that the data of a company in the table can be accurately and completely synchronized after the data of a company is published. Daily data synchronization involves more than 20 data tables, including well completion layer table, production layer status history table, daily data of injection well status, and daily data of production well status. The primary key is divided into two types: well ID and entity ID. Well ID is directly associated with the organization ID. Entity ID includes geological unit, combination unit and station library, and each type is associated with the organization ID. Through the above association relationship, the data to be synchronized in the data table can be accurately identified after the oil/gas production plant data is released [7].

4.2 Effective Field Analysis

A2 system is developed based on EPDM model, and there are a certain number of invalid fields in actual application. After testing, reducing the number of synchronized fields can improve the efficiency of data synchronization. Therefore, invalid fields need to be judged and filtered. First of all, the fields with the number of duplicate values less than 10 should be initially filtered, and then the total data volume of the above fields containing duplicate values, the data volume of non-null values, and the specific data of non-null values and other auxiliary information should be obtained. Finally, whether the field is invalid can be determined through manual judgment of the auxiliary information. For example, Data table X contains field Y, in which the number of non-repeating values is 1, the total data amount is 34452797, the total effective data amount is 128629, and the total effective field value is 0. It can be determined from the above information that this field is invalid.

4.3 Synchronous Log Construction

By comparing the data release time of Beijing Database with the data synchronization start time of Xi'an Database, the logical trend of the synchronization scheme can be obtained (see Fig. 2). Therefore, date and time are the key data that must be recorded. Before building the data synchronization logic, the data synchronization log table needs to be established first. The log table mainly includes six aspects: name of the oil/gas production plant, data release time, data synchronization start time, data synchronization end time, synchronization consumption time, and remarks. The data synchronization log table not only determines the direction of synchronization logic, but also records the synchronization and unlocking status of data in a timely manner to assist in the analysis of synchronization efficiency and troubleshooting of synchronization errors [8].

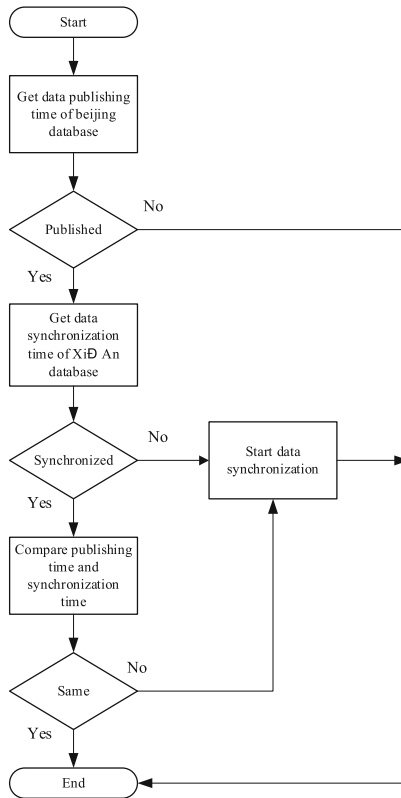


Fig. 2. Data synchronization and unlocking judgment logic

4.4 Synchronization Logic Construction

The data synchronization logic of A2 regional center is mainly based on the data release time of Beijing database and the data synchronization time of Xi'an database. According to this scheme, data synchronization is started at 9:00 a.m. every day. First, the data

synchronization oil/gas production plants are obtained and written to the data synchronization log table. Then, the oil/gas production plants in the data table are periodically traversed, and the data release and data synchronization time of each plants are queried circularly. According to the data synchronization and unlocking judgment logic in Fig. 2, the initial synchronization or unlocking resynchronization of data is started. After the synchronization of all oil/gas production plants are completed, continue to judge whether the data of the current day is unlocked until 11:00 p.m. At 10:45 p.m., a special judgment will be made. If the number of units completing data synchronization on the current day is less than the total number in the synchronization log data table, a full data synchronization of the current day will be started, and the operation will be recorded in the synchronization log data table. The system administrator will be notified the next day to check the data integrity. At 10:45 p.m., a special judgment will be made. If the number of oil/gas production plant completing data synchronization on the current day is less than the total number in the data synchronization log table, a full synchronization of the current day's data will be started, and the operation will be recorded in the data synchronization log table, and the system administrator will be notified to check the data integrity [9].

5 Application Effect

5.1 Improve the Data Unlocking Recognition Rate

Based on the statistical analysis of data synchronization logs, the average daily data synchronization time of each unit is about 120 s (see Fig. 3). Compared with the original data synchronization time of 83 min, the efficiency is increased by more than 20 times.

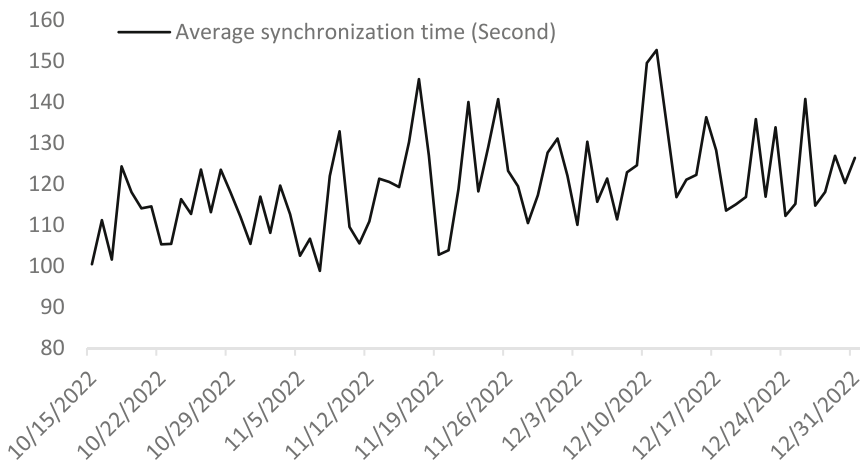


Fig. 3. Average synchronization time of new scheme

5.2 Improve the Accuracy of Data Unlocking

By comparing the data release time in reverse order between the Beijing database and the data synchronization log table, the accurate identification and resynchronization of cross-date data unlocking are realized, and the data accuracy and consistency are effectively improved.

6 Conclusion

The A2 system regional center data synchronization solution solves the problem that the oilfield self-built information system is difficult to directly apply A2 data. The core of the scheme is to identify the release status and unlock status of the data. At the same time, the efficiency of data synchronization has been improved again by simplifying the invalid fields, and finally the “published today, applied today” of A2 production data has been realized. It has good guidance and reference significance for other oilfield companies that need to build the regional center of A2 system, or the database cannot be directly connected to the application due to network, efficiency, synchronization software and other problems, the business scenario that needs to build a regional data center has good guidance and reference significance.

References

1. Tao, F.: Establishment and promotion of an integrated application system for oil, gas, and water well production data. *Commun. World* **06**, 272–273 (2017)
2. Tian, B.: Research on the application of logical data synchronization technology based on oracle database in practice. *Comput. Program. Skills Mainten.* **15**, 31–32 (2016)
3. Wang, H., Mu, L., et al.: Optimized management and fast query of decentralized storage of oil and gas production dynamic big data. *Pet. Explor. Dev.* **46**(05), 959–965 (2019)
4. Kou, Y., Wang, X., et al.: Research and application of database synchronization technology. *Comput. Knowl. Technol.* **15**(15), 3–5+34 (2019)
5. Chen, J., Yan, J., et al.: An optimized database synchronization method and device. *Guangxi Commun. Technol.* (02), 1–4+9 (2017)
6. Wang, Z., Xiu, W.: Database synchronous programming solution. *J. Liaoning Normal Univ. (Nat. Sci. Ed.)* **19**(01), 32–33+58 (2017)
7. Cui, J.: Application of master data management system in enterprise informatization construction projects. *Eng. Constr. Des.* **21**, 201–203 (2021)
8. Wang, G., Wang, Y., et al.: Research on interface technology for synchronized reporting of A2 and development database. *China Petrol. Chem. Stand. Qual.* **34**(06), 165–166 (2014)
9. Xia, B.: Research and application of database synchronization based on oracle advanced replication function. *Inf. Syst. Eng.* **03**, 87 (2015)