

The Study of Backup Method based on RMAN and Streams

CHEN Bo^{1, a}, HU Ning-yu^{2, b}, Qiu Shao-ming³

¹ College of Information Engineering, Dalian University, Dalian, 116622, China

² College of Information Engineering, Dalian University, Dalian, 116622, China

³ College of Information Engineering, Dalian University, Dalian, 116622, China

^aemail: 1044046967@qq.com, ^bemail:524809261@qq.com

Keywords: Real-time backup; Database backup; physical backup

Abstract. For the problems of big backup files and long recovery time after traditional real-time backup method has been backup for a long time, A new method base on RMAN and Streams technology is presented in this paper. This method uses the Streams technology to make real-time backup for database. in the meantime, Timing detect the busy degree of the database system. When the database system is not busy and real-time backup file is larger than the preset threshold, using RMAN for a physical backup. In the recovery, this method implements physical recovery at first, and then, real-time backup file's SQL acts on database. This method avoids the disadvantages of real-time backup file growing over time, recovering from the beginning, and effectively solves the problems of big real-time backup files and long recovery time.

Introduction

With the extensive application of the database, the disaster management of the database has become an important part of the construction of information system. When the database crashed, this will result in data loss, system instability or even paralysis. Therefore, it has become a critical issue of the disaster management system that how to achieve the backup and recovery of the database effectively. The literature [1-5] presented a real-time backup method based on dual hot standby. This method can transmitting the changed information to the remote database system to operate replay by monitoring the change of local database in time, which can keep synchronous between the remote standby database and the main database. When a disaster occurs, this method uses the standby database to take control the business of local applications quickly to ensure the continuity of the business. Although this backup scheme can maintain the continuity of the business, the cost is high and it needs to equip with one or more same database system for each target database in the distal. At the same time, the network performance will affect the consistency of the standby database and the local database. The literature [6] presented the method of backup and recovery based on the redo log. This method will storage the log files on remote server, it uses the redo log files to restore when a fault occurs in the database. This method exists the weakness that backup files will take up a lot of storage media; Literature [7] presented another real-time backup method that based on redo log, The method can capture the change of the database from redo log, and then translate the change to SQL records and store them in local backup file in a specific format. When the database crashes, the recover processes read SQL records from the backup file directly while acting on the database. In general, the operations of the database will be too many to count from the establishment of the database to the crash of the database. It will consume a large amount of recovery time and storage medium if execute the first SQL record to the last SQL record sequentially. The literature [8-9] presented the method of database timing backup , which has a fast backup and restore function, but the database can only be restored to the last backup time when the database collapse.

This paper proposed a backup method of RMAN combine with Streams based on above issues. The method is based on the traditional real-time backup method, and executing physical backup method according to stetted strategy, avoiding long-term cumulating of backup file, so it solves the problems that file size and the recovery speed of traditional real-time backup method. Then it will

introduce the working principle of the backup method and the implementation process in detail. And then set up the simulation system to verify and analysis the method.

The Principle of Streams and RMAN

(1) The principle of Oracle Streams and RMAN

Oracle Streams is a Stream replication technology offered by Oracle since 9iR2. Its realized principle is very simple: It copies the data which changed to the target database through the three processes of capture, propagation and apply. Oracle Streams replication not only can be a single point of replication, but also can be multi point of replication. The internal principle is that the capture process find DML and DDL from redo log. Then generates a unique LCR (Logical Change Record) and preserves it in the queue, propagation process will transmits the LCR records to the queue of the target database. Apply process resolves the LCR records from LCR queue, while applying DML or DDL operation to the target database to complete the copy of data change. Diagram is shown in Figure 1.

The copying of oracle Streams only need to scan the redo log from the source database to dig out useful information and conversion to LCR to store in the advanced queue, other works can be done by other applications, that greatly reduces the influence on the performance of the source database and has the advantage of small system overhead. Streams can dig out the online redo log that can reduce delay and have the advantage of strong. When copying, this method did not need additional hardware support, so that it has the advantage of low cost.

RMAN is a backup tool which owns to Oracle, which is also a medium tool and has strong function of compress. RMAN will read the files that need to backup to the buffer by the block way, only writes data blocks in the backup file when the data block is not a blank data block. When the backup is completed, the SCN of the backup documental head will be updated. At the same time, the backup file information such as the file name, the size of files, the path of files, the time that operate backup, the system change number and so on, will be updated to the recovery catalog database, In order to restore the source database.

When the database recovery, the process check whether the control file of database has damaged , If damaged, the process will recover the control file first, And then look for a full backup file that is the nearest from the control file SCN, and use it for media recovery, if it exists incremental backup files between the SCN of full backup file and the SCN of recovered control file, The process will use them for media recovery, after media recovering, the process will do recovery by redo file as much as possible.

(2) The execution process of Streams combine with RMAN.

This paper puts forward the backup method of RMAN combine with Streams which has the advantages of strong real-time performance, fast recovery, and file size stable. It effectively solve the problems that traditional real-time backup has, such as bigger backup file and longtime recovery.

This method perform the physical backup according to the size of backup file and the busy degree of database system, Physical backup is performed only when the database system is not busy and real-time backup file is larger than 1/3 the estimated size of physical backup file. When the database crashes, firstly, using RMAN to do a quickly physical recovery, and then use SQL statements of the real-time backup file to recover the database, making the database quickly return to the collapse state.

The steps of perform a database backup as follows:

- ① The administrator set the storing address of the physical backup through task configuration module and evaluate the time of database system;
- ② The system keep performing real-time database backup all the time, and store the real-time backup file to the specified location;
- ③ Estimating the size of physical backup file and the busy degree of database system regular, if the database system is not busy and 1/3 physical backup file size is smaller than real-time backup

file, doing a physical backup of database;

④ Deleting the records from the real-time backup file, when SCN is smaller than the SCN of physical backup file.

The flow chart of system backup is divided into real-time backup flow chart and physical backup flow chart, as shown in figure 2:

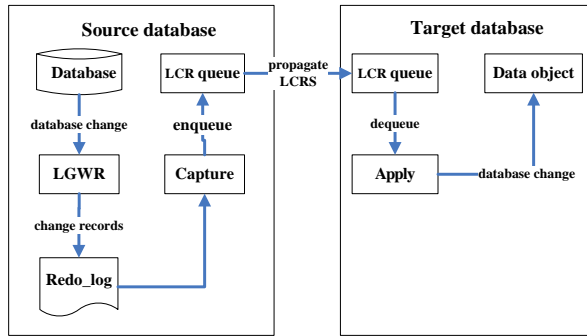


Fig.1. Oracle Streams replication Diagram

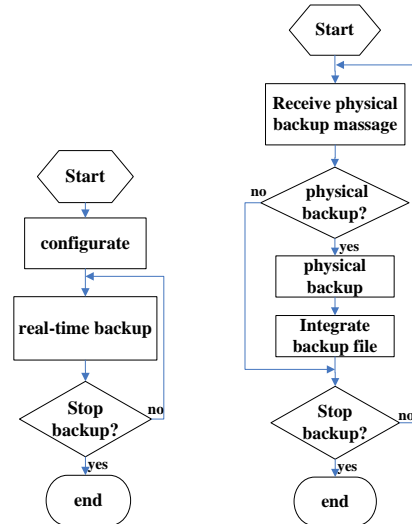


Fig.2. Flow chart of backup

Physical backup is decided by the judgment message that physical backup device return, when the timer event trigger, testing the busy degree of database system and estimating the physical backup files size of source database. It will do a physical backup of database just when the database system is not busy and the real-time backup file is bigger than 1/3 physical backup file size. Otherwise, continuing to do real-time backup. The flow chart of physical backup judgment device is shown in figure 3.

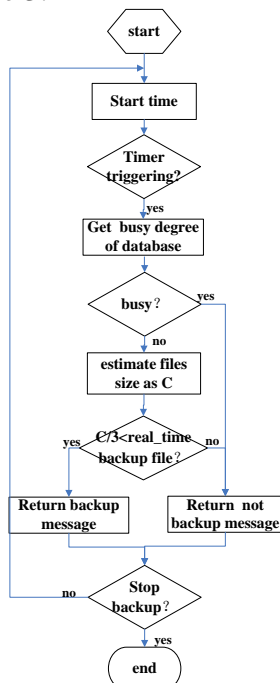


Fig.3. The flow chart of physical backup judgment device

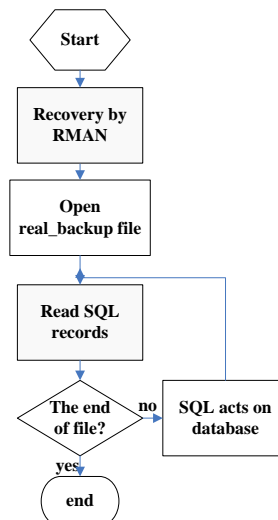


Fig.4. The flow chart of recovery

The steps of perform database recovery as follows:

- ① First, this method will do physical recovery with RMAN on the database;
- ② After physical recovery, using the records from the real-time backup file to recover the database until the last SQL statement.

The detail of recovery process as shown in figure 4:

The Realization of Combinational Backup Method

(1) The realization of the physical backup judgment device

The physical backup is controlled by decision device, which is composed of the timer, the busy degree of database system, the file size of real-time backup and the file size of physical backup. The following shows how to obtain these parameters:

Timer: this parameter is inputted through the task configuration module by user;

The busy degree of database system: Find non_idle wait events by querying the v\$session_wait table to determine the busy degree of system.

Real-time backup file size: Get from attribute in real-time backup file;

Forecast the physical backup file size: Physical backup file is mainly composed of data files and control files, data files size can be calculated by the dba_extents table, Control files size can obtain from its properties.

(2) The realization of the physical backup

When the process received backup message from judgment device, it immediately switch the backup method and do physical backup with RMAN.

The steps of doing RMAN backup:

① Configure the RMAN running environment, such as the establishment of the recovery catalog database, change the target database operation mode etc.;

② After receive a backup message, create the physical backup script according to the parameters from task configuration module and create the physical backup process;

③ Start physical backup process;

④ After complete the physical backup, integrate physical backup files and real-time backup files according to the SCN.

(3) The realization of the real-time backup

In the real-time backup only need to store the modification of information to the file, and don't need to act on another database immediately. Therefore, the source database and target database are the same for the real-time backup system. So it doesn't need the process of propagate. The principle of real-time backup is shown in figure 5.

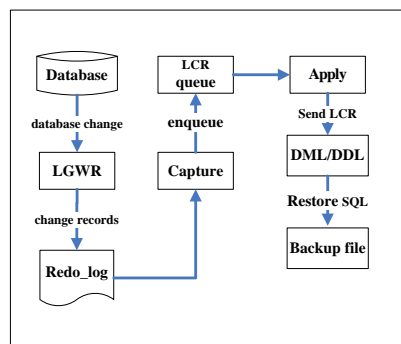


Fig.5. The principle of real time backup

The capture process catches database change information from the database redo log and store them in the LCR queue with the form of LCR , The application processes remove LCR from the queue and sent to DML Handler or DDL Handler. DML Handler and DDL Handler respectively parse LCR into DML and DDL statements, and then save all statements to a file. Once a new message into the LCR queue, process will notify the user program to read messages from LCR queue 、 handle it 、 store it to the real-time backup file and delete the LCR messages from the queue.

Analyze changes of the database from the database redo log by using Streams technology, it need to configure the Streams replication environment. The steps:

① Configure the environment of the system: change the work mode of the database, related variables of database system, create a Streams user who will be responsible for streams management etc.

② Configure the Streams user environment: Configure queue that user will manage and the processes that will be related to the queue, including the creation of agent, Queuing subscription, and Queue type that transformed mapping process.

③ The configuration of capture process and apply process: determine the capture object, the backup method need to capture all changes of the database; the way that application process handle message, the method used is a user-defined process treatment.

④ Application processing: include acquiring LCR, translate LCR to SQL and write to a backup file.

Test results

(1) The experiments and simulation of combinational backup method and Traditional real-time backup method.

The database uses the traditional real-time backup method and combinational backup method for database backup, after each backup, make the related files of the database lose completely by artificially, and then recover database. Repeat to do the operation; take notes about the backup file size and the time of recovery and draw the diagram of file size and recovery time. As shown in figure 6 and figure 7.

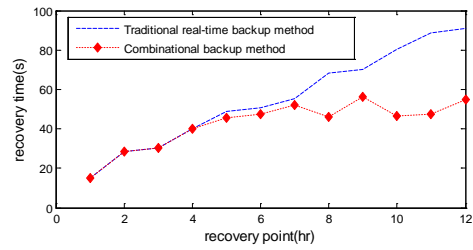
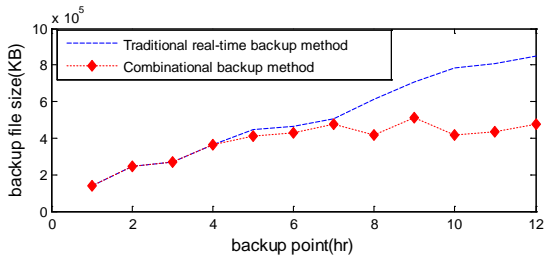


Fig.6. The comparison chart of backup file size Fig.7. The comparison chart of recovery time

From figure 6, 7 can be seen ,with the increase of time, the backup file size of traditional real-time backup method is growing and recovery time is increasing; The backup file size of combinational backup method only wandering between the 1~4/3 times of physical backup file size. Similarly, recovery time is similar to the trend of backup file size. The reason: the real time backup file stores the change of the database, as time go by, there are more and more changed information in the database, so the real time backup file size is bigger and bigger with the time; while the combinational backup method will perform physical backup when the real-time backup file is larger than the preset threshold (1/3 physical backup file size), the new backup file will replace all the old backup files. So the backup file size of combinational backup method only wandering between the 1~4/3 times of physical backup file size.

(2) The experiments of combinational backup method and physical timing backup

The database also uses the physical timing backup method and combinational backup method for database backup, after the completion of each backup, damaging database artificially, and then recover database. Repeat to do the operation, figure 1 shows the comparison chart of recovery situation.

SCN (System Change Number) is used to identify a version of database, the database is assigned a unique identifier SCN when a transaction is committed and the SCN also will be writing to the log file. Each database has a global SCN generator.

Table 1. Comparison table of recovery

	1	2	3	4
The backup scn	13618922	13621443	13645140	13659520
The failure scn	13620142	13635463	13656125	13665223
Damage	Completely	Completely	The log files ok	The log files ok
The SCN after combinational recover	13620142	13635463	13656125	13665223
The SCN after physical recover	13618922	13621443	13656125	13665223

As can be seen from table 1, only when the log files are intact, physical timing backup method can recover the database to the state of collapse, or it can only be restored to the state of last backup time; combinational backup method can recover the database to the state of collapse although the files are all damaged.

Conclusion

Aiming at the disadvantage of traditional real-time backup method, propose a new backup method of RMAN combine with Streams. The method can effectively solve problems that the big capacity of real-time backup files and long recovery time. The backup method set a threshold as 1/3 physical backup file size, it cannot guarantee the smallest of the backup file, Therefore, how to set the default threshold is the direction of further study.

References

- [1] Xianghui Lin. Design and implementation of a database disaster recovery technology[D]. University of electronic science and technology, 2011.
- [2] Xiongling Zhang. Design and implementation of a remote database backup system[D]. Hunan university, 2013.
- [3] Yibing Deng, Shanguang Chen, Wei Hu, etc. A Novel Block-level Continuous Data Protection System. New Trends in Information Science and Service Science (NISS), 2010 242~247.
- [4] Ruijun Fu. Real-time backup method for database based on packet capture[J].Computer engineering and design, 2010 (31) 5355-5358.
- [5] Yuwen WANG, Xiaojie LIU, Tao LI, etc. Multi-point and multi-level real-time backup method for database table[J].Computer engineering and design, 2009 (29) 2665-2667.
- [6] Ying Liu, Xiaowu WANG, Bing Zeng. Design on New System Architecture for Data Backup and Recovery[J].Communication technology, 2011 (44) 66-68.
- [7] Cheng Ya. Research real-time backup technologies for database based on Oracle Streams[D].Nanjing university of aeronautics and astronautics, 2012.
- [8] Yunfan Zhang. Backup and Recovery Strategy of Oracle Database[J].Computer engineering, 2009 (35) 85-87.
- [9] Lina Wang, Yuntao Yue, etc. Design and implementation a backup software for Oracle Database[J].Journal of Wuhan university, 2008 (54) 060-064.