Vol 6 Issue 10 July 2017

ISSN No : 2249-894X

*Monthly Multidisciplinary Research Journal*

# *Review Of Research Journal*

**Welcome to Review Of Research**

**RNI MAHMUL/2011/38595** **ISSN No.2249-894X**

Review Of Research Journal is a multidisciplinary research journal, published monthly in English, Hindi & Marathi Language. All research papers submitted to the journal will be double - blind peer reviewed referred by members of the editorial Board readers will include investigator in universities, research institutes government and industry with research interest in the general subjects.

## Regional Editor



## *Advisory Board*



More.....

Review Of Research

ISSN: 2249-894X
Impact Factor : 3.8014(UIF)
Volume - 6 | Issue - 10 | July - 2017

# BACKUP AND DISASTER RECOVERY SYSTEM FOR RESTORATION OF DATA SERVICES


Nitin D. Magdum[1] and Prof. Manik A. Raichurakar[2]
[1]JSPM's RajarshiShahu School of Engineering & Research , Narhe, Pune, Maharashtra, India.
[2]JSPM's RajarshiShahu School of Engineering & Research,Narhe, Pune, Maharashtra, India.

ABSTRACT :

*We propose a backup and Disaster recovery system for instant restoration of data services, focusing on a generalpurpose automatic backup and recovery approach to protect production database data. The system will automatically start on failover of production database on standby database. The recurrent need of excess storage space on production server gives birth to this idea. As the production servers deal with activities such as report generation etc., that turn increases the load on the server causing the performance to degrade. Also, there are more possibilities of threats like hardware failure, loss of machine etc., where data recovery can only be the option left, for the survival of the data. The functionality will be provided by standby server for backup of the running database and its immediate failover during failure. Furthermore, backup and recovery can be directly applied to any existing system and it database locks,functions,triggers and User management.*

KEYWORDS : *objectrelational database management system (ORDBMS) ,write ahead logs (WAL),Backup and recovery)*

## INTRODUCTION:

Nowadays , handling data in data centric operations has become a challenging task because of volume of data. There were many scenarios where a database system is used to control the collection of data items and plays vital role in contemporary applications, such as administration, social sites, searchengines, and banking systems. Somes of its properties include, abstractions, data consistency and concurrent databases, due to which it gains an immense attention in real world applications. Some of its features such as data integrity, consistency, storage and retrieval are noticeable and can be very useful in real time data operations. There is wide variety of relational databases available in the market to serve the purpose of the business needs from small to high end mission critical systems. The available databases can be classified into 2 types namely: commercial databases and open source databases. Commercial databases like Oracle, Microsoft SQL Server, DB2, Informix, MySQL Enterprise Edition (Oracle*) etc. and open source databases like PostgreSQL, MySQL, MongoDB etc.Commercial databases are not free and making modifications as per business need is difficult while open source means that it's possible for anyone to use and modify the software. The open source databases are the ones that changes the market drastically, due to two main features. Firstly, they are easily available for no cost. Secondly, they can be change according to business needs. The paper focuses on PostgreSQL database mainly. Postgresql Is An Objectrelational database management system (ORDBMS) based on POSTGRES, developed at the University of California at Berkeley Computer Science Department.Objectrelational database management system (ORDBMS)

is an another type of database. These are most powerful database model which stores the data in terms of rows and columns helping to select and manage data very easily with utmost reliability,consistency and performance by satisfying ACID properties.While talking all the good things about databases,there should always be a consideration of two key things, that may play a major role in the success of the same. These are performance and efficiency.Many times it has been seen that, the single database server is utilised for the purpose of serving the data storage and retrieval. When more of such requests come for processing, the server may be unable to complete the request and fail somewhere, to serve the purpose. Also, bottlenecks happens many times for the servers, that in place degrade the performance and efficiency.

The scalability might be one of the problem, where the system fails sometimes or is dependent on, certain backup and recovery systems, in case of failures. Database replication can be the way to handle such cases.Data replication is the process of copying the data from one database to the other, which can be in the geographically separated area or in the same area as well. The replication facilitates alternate data accessing options, that inturn, will provide high data availability, performance and fault tolerance of applications. The DBA is responsible for monitoring the database server on a regular basis to identify bottlenecks (parts of the system that are slowing down processing) and remedy them. For data replication, in order to maintain and monitor a standby server, the cost of operation and human intervention required increases. This is an disadvantage of such systems. In this paper, we propose a novel approach which requires minimum human intervention, as a result reduces the overall cost the system. There are systems available for performing data replication and instant data restoration in case of failovers e.g. A baremetal recovery system for instant restoration of data services (BIRDS). These systems generally take snapshot of database which means data from one server is simply copied to another server.

Our contributions: As mentioned earlier, we design and implement a data recovery system for an open source database namely PostgreSQL. Recovery of data is performed based on write ahead logs (WAL) rather than actual data. This system can be implemented on production server or standby server.The system allows automatic monitoring of locks and database backup security. The system will provide an efficient mechanism to replicate the data and maintain backups that can be used when a disaster occurs. Other facilities like monitoring, debugging,logging or security management services can prove to be useful for certain users. Remaining paper is organised as: Section 2 discusses the available literature about this topic. Section 3 explain existing BIRDS system and our system in details. Section 4 presents results and discussions about the same. Final section deals with concluding remarks of the topic and future work.

## LITERATURE SURVEY

A lot of work is carried out to protect the privacy of individuals and showed that the utilization of the purpose should be aimed on the basis for backup and recovery. An alternative backup strategy is to directly copy the files that PostgreSQL uses to store the data in the database.

For example, the method perform to take file system backups is: tar cf backup.tar/usr/local/pgsql/data There are mainly two disadvantages of above method for backup.

1) The database server must be shut down in order to get a usable backup when it is running.

2) If you dug into the details of the file system layout of the database, you might be tempted to try to take backup or restore only certain individual tables or databases from their respective files or directories. This will not work because the information contained in these files are not usable without the commit log files, pg clog/*, which contain the commit status of all transactions. A table file is only usable with this information. Of course, it is also impossible to restore only a table and the associated pg clog data because that would render all other tables in the database cluster are useless. So, file system backups only work for complete backup and restoration of an entire database cluster.

Tarandeep Singh et.al (2013) deals with the method of replication as a backup technique for protecting data from any threat of failure. It highlights the significance of increasing the storage space on the production server so as to protect the database. Matthias Wiesmann and Andre Schiper (2005) paper does a comparative

study of various database replication techniques based on total order broadcast alongside its evaluation using a finer network model. It found that cost of synchronization between replicas is minimum and thus usage of total order broadcast based techniques is bright. Salman Abdul Moiz and Sailaja explains the features of replication and disaster recovery. It further deals with various strategies of database replication especially in the context of disaster recovery plan. It deals and compares various commercial tools and open sources that are available for the same and discusses the challenges involved.

Zuhoor A. AlKhanjari emphasizes on the tester which ensures that enterprise transactions retrieve correct data. The tester that's developed, DBSoft is a composite of five components; a parser, input generator, output generator, output validator and report generator. It is found that his tester avoids transaction runtime errors.

## METHODOLOGY

In this section, we discuss about existing system which is used for data recovery and backup. Next part of the section explains details about our system and algorithms for database recovery.

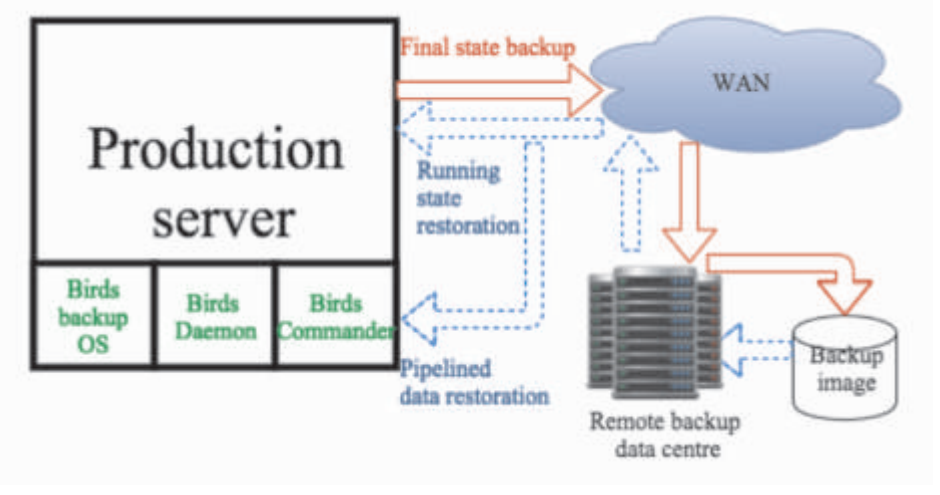
### 3.1 Existing System

BIRDS stands for A baremetal recovery system for instant restoration of data services. BIRDS uses novel recovery strategy and parallel recovery. BIRDS takes continuous backup of operating system (OS) using virtualization. Architecture of BIRDS is as shown in Figure 1. In the figure,solid line indicate data flow and dash line indicate recovery data flow A BIRDS Commander run in the application layer of the backup OS, issuing commands for backup or recovery when needed.

BIRDS stands for A baremetal recovery system for instant restoration of data services. Mainly two contributions in BIRD First is automates the backupandrecovery process by isolating the operating system and second is BIRDS uses novel recovery policy named as parallel recovery of operating system(OS). BIRDS takes continuous backup of operating system (OS) using virtualization.Baremetal recovery(BIRD) process involves five main steps as below

1 boot from external drive and start the recovery process.
2 restore data from the backup data center to the production site.
3 reboot the production site.
4 resume data services on the production site.
5 restart the backup agent for the next failure.

Architecture of BIRDS is as shown in Figure 1. In the figure, solid line indicate data flow and dash line indicate recovery data flow.



A BIRDS Commander run in the application layer of the backup OS, issuing commands for backup or recovery when needed. Commands are handled by a BIRDS Daemon which resides in the kernel of the backup OS. The BIRDS Daemon monitors the running of the production OS and backups or recovers the checkpoints of the production OSs running states together with the persistent storage states to and from the backup center upon the requests of the BIRDS Commander. For initialization, a full copy of the persistent storage and the initial running states are replicated to the remote backup center. After that, any storage data changes are caught by the BIRDS Daemon and replicated incrementally. For backup, checkpoints of the protected production OS are stored in a file, together with incremental snapshots of the containers persistent storage stored in the form of logical volumes (LV).

For recovery, after new hardware is prepared and ready to use, BIRDS can immediately start the recovery process in two stages. First, the backup OS is booted using the hardware and the prerecorded running states are transferred back from the remote backup center.Second, the BIRDS Daemon monitors the running of the resumed service which requests to access data in the local storage device. Data misses will happen soon because no data is really restored yet.

## Proposed System:

Proposed System: The proposed system which overcomes shortcoming of birds tool. It is designed for database level architecture. We implemented it for PostgreSQL database.

## Mainly following goals of proposed system

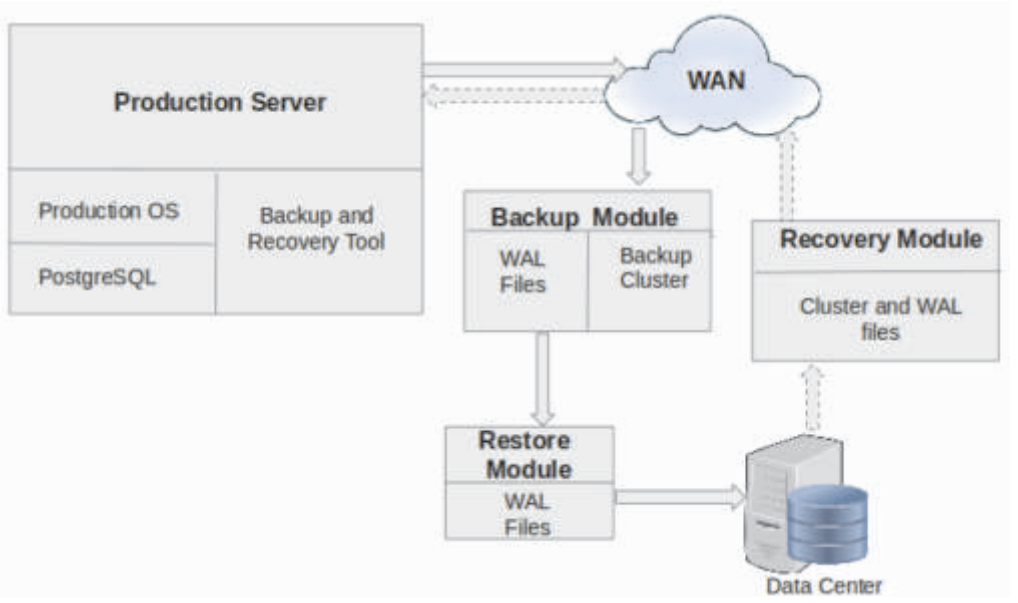
1. Disaster recovery is to prevent data loss and to restore data available after corruptions.
2. Smaller the Recovery Time Object(RTO)
3. Backupandrecoveryoperations should only introduce low overhead.

Backup and Recovery tool protect databases against the data loss, it also ensures the database recovery where it enables the database maintenance and performance.Database maintenance is the key factor while data size is increasing. Every database needs regular maintenance to increase the database performance. There are various maintenance tasks that need to perform such as database level user management, monitoring locks, functions, triggers and space monitoring.

Backup & recovery tool compress the backup and encrypt it, and also provides backup security for the overall database level security. Mathematical Model of Proposed System:Backup and Restore tool provide multiple component such as

1) Backup Module
2) Recovery Module
3) Restore Module

Figure 2 shows the overall architecture of our system. Detail explanation is as follows:




Backup Module: As with everything that contains valuable data, PostgreSQL databases should be backup regularly.

PostgreSQL database support following types of backups:1.

1. Logical Backup
pg_dump & pg_restore
2. Physical Backup
Offline Backup (called as Cold Backup)
Online Backup (called as Hot Backup)

Backup module takes physical online backup of running database cluster and restore in same production server or data center. We implemented following algorithm to table write ahead log backup algorithm.

Backup Algorithm
// Input :Take a Input from Administrator
//output: Restoring Data on New Server
start
max_checkpint;
for (i=0 ; i<#checkpoint; i++)
move (;logfile!=NULL;)
end for
end

Restore Module:

PostgreSQL database maintains WAL (write ahead log) in data cluster pg_xlog/subdirectory directory. Using the WAL file contain transaction statement runs on productiondatabase. Using the write ahead continuous restoring production database on same server or standby server.

Restore Algorithm
// Input : Log file of Server1
//output: Log file of Server2

```
start
move;
while (log!=NULL; )
move (;logfile!=NULL;)
move –r log server2
end while
end
```

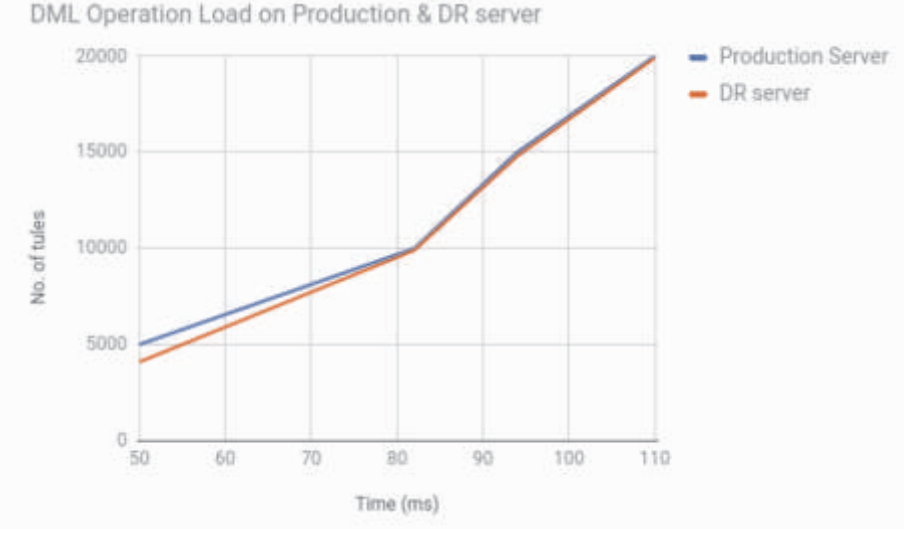
Recovery Module: To recover data we can combine a filesystemlevel backup with backup of the WAL files. When recovery is needed, we restore the file system backup and then replay from the backedup WAL files to bring the system to a current state. As the failure can cause data loss on any server on the productions Recovery Module recover all the data up to last commit transaction.

Recovery Algorithm

```
// Input : Read a transaction from log file
//output: Data available up to last checkpoint
start
recovery
for (i=0;i<=checkpoint ;i++)
{
read checkpoint
}
end for
end
```

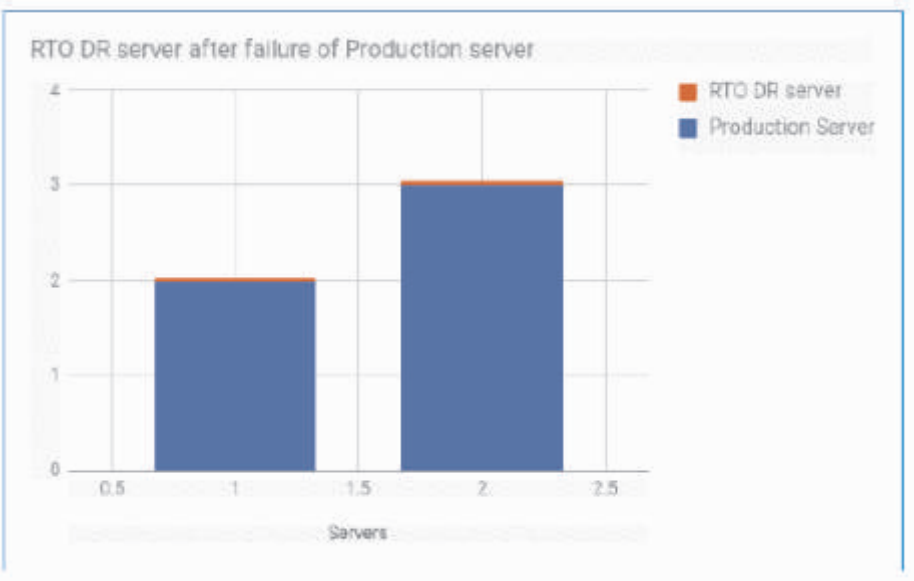
Disaster Recovery: Current DR approaches can be grouped into two categories according to their means of Failover and Failback.whole production site can be corrupted, all the data and system states are routinely replicated to a remote backup data center. After disasters corrupt the production site, run a production site on disaster server and start recovery program simultaneously started to fetch back data and states from the remote backup data center. In this system Disaster recovery time near to zero.

Below graph shows DML(D ata Manipulation Language ) operation performed on production server it will reflected on DR (Disaster server) with zero data loss and minimum time.




Graph . DML operation Load on production & DR server

Below graph show the time required for failure of production server running the application on DR(Disaster server) server.Its near to zero.




Server Page: Server page mainly used to monitor a database locks,functions & triggers also user can get detailed information of databases,schemas and tables. User can add multiple server in server page.

User Management : User management is mainly used to add user in database also manage the user permission and privileges in databases.

## C ONCLUSION

PostgreSQL WAL functionality, when configured with backup insistent recovery, enables you to run successful backups and restoration. This backup and instant restoration tool gives you the ability to recover data up to a last committed transaction or before specific actions, so that you can restore complete database systems as it was before a crash or corruption.

# Publish Research Article
# International Level Multidisciplinary Research Journal
# For All Subjects

Dear Sir/Mam,

We invite unpublished Research Paper,Summary of Research Project,Theses,Books and Books Review for publication,you will be pleased to know that our journals are

## Associated and Indexed,India

- Directory Of Research Journal Indexing
- International Scientific Journal Consortium Scientific
- OPEN J-GATE

## Associated and Indexed,USA

- DOAJ
- EBSCO
- Crossref DOI
- Index Copernicus
- Publication Index
- Academic Journal Database
- Contemporary Research Index
- Academic Paper Databse
- Digital Journals Database
- Current Index to Scholarly Journals
- Elite Scientific Journal Archive
- Directory Of Academic Resources
- Scholar Journal Index
- Recent Science Index
- Scientific Resources Database

Review Of Research Journal
258/34 Raviwar Peth Solapur-
413005,Maharashtra
Contact-9595359435
E-Mail-ayisrj@yahoo.in/ayisrj2011@gmail.com