

PGCluster Round Table

PostgreSQL Anniversary Summit

July 2006 At.Mitani

Agenda

- Requirement of the enterprise usage
- Does replication help of those requirement?
- About the PGCluster
 - Feature
 - Procedure
 - Replication feature
 - HA feature
- Yet another requirement for PGCluster
- Concept of the PGCluster-II
 - Feature
 - How to share IPC during distributed servers
 - Structure
 - Expected effect and problem
- Discussion

Requirement of Enterprise usage

- A huge amount of data operation
 - Use the huge storage with SAN / NAS
- A quick response with a lot of request.
 - Search (ex. Web application)
 - Update (ex. ERP system)
- High availability (HA)
 - 24-hour & 7-day a week operation
 - Disaster recovery
- Easy backup and restore
 - Differential back up / restore

Does replication help of those requirement ? (1/4)

- Huge data operation
 - Not good
 - It does not reduce the amount of operation data
 - Other solution
 - PostgresForest / pgpool-II

Does replication help of those requirement ? (2/4)

- Load distribution
 - Good
 - It reduce the search load
 - Not good
 - Update load becomes more heavy!
 - Other solution
 - Cluster / grid
 - I don't know the concrete tools

Does replication help of those requirement ? (3/4)

- High availability
 - Good
 - In master-slave composition, there is a service stop time during take over.
 - In multi-master composition, there is no service stop time during take over.
 - Other solution
 - There are some commercial tools for server level HA.

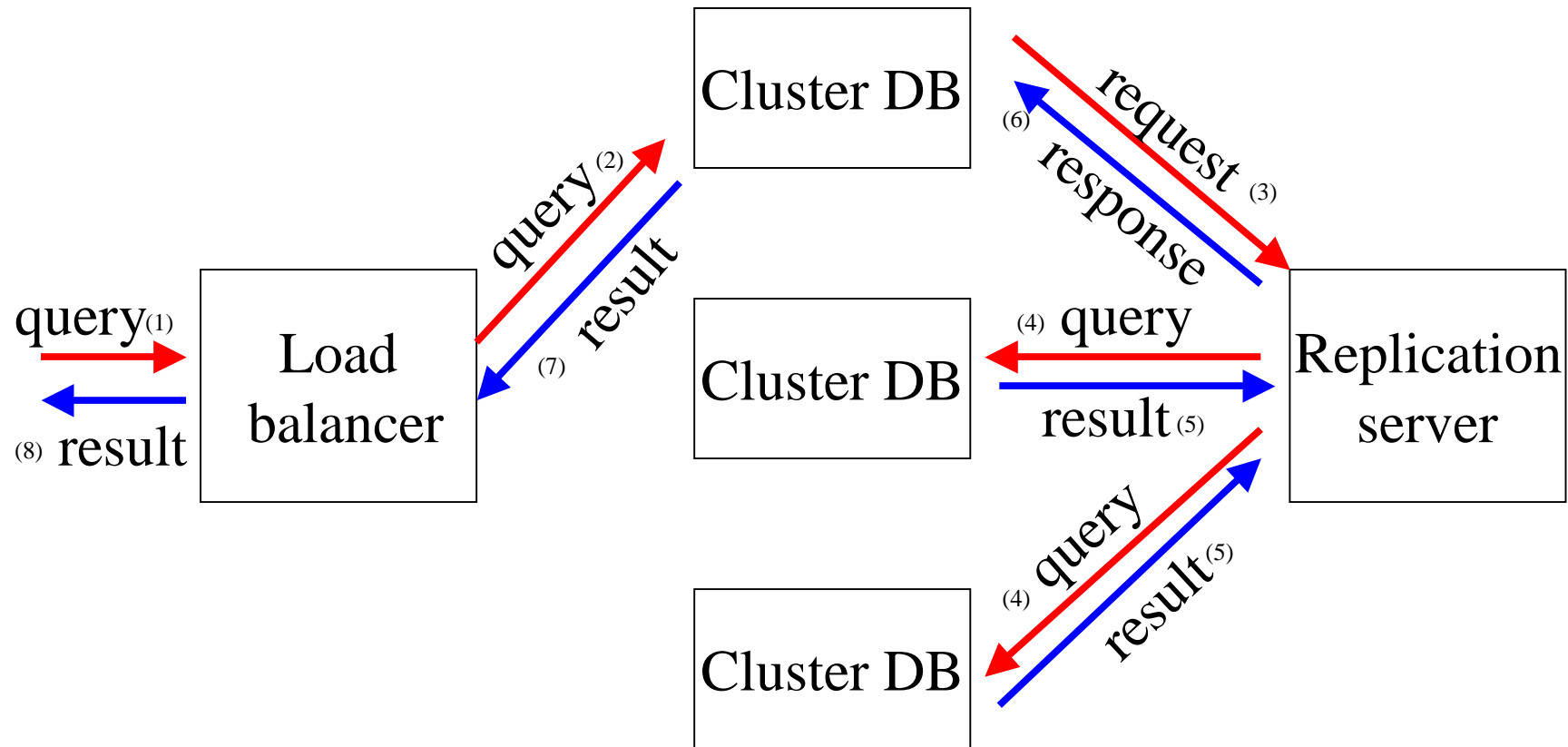
Does replication help of those requirement ? (4/4)

- Easy backup and restore
 - Good
 - Replicated DB is used as real-time backup.
 - Other solution
 - Differential back up / restore

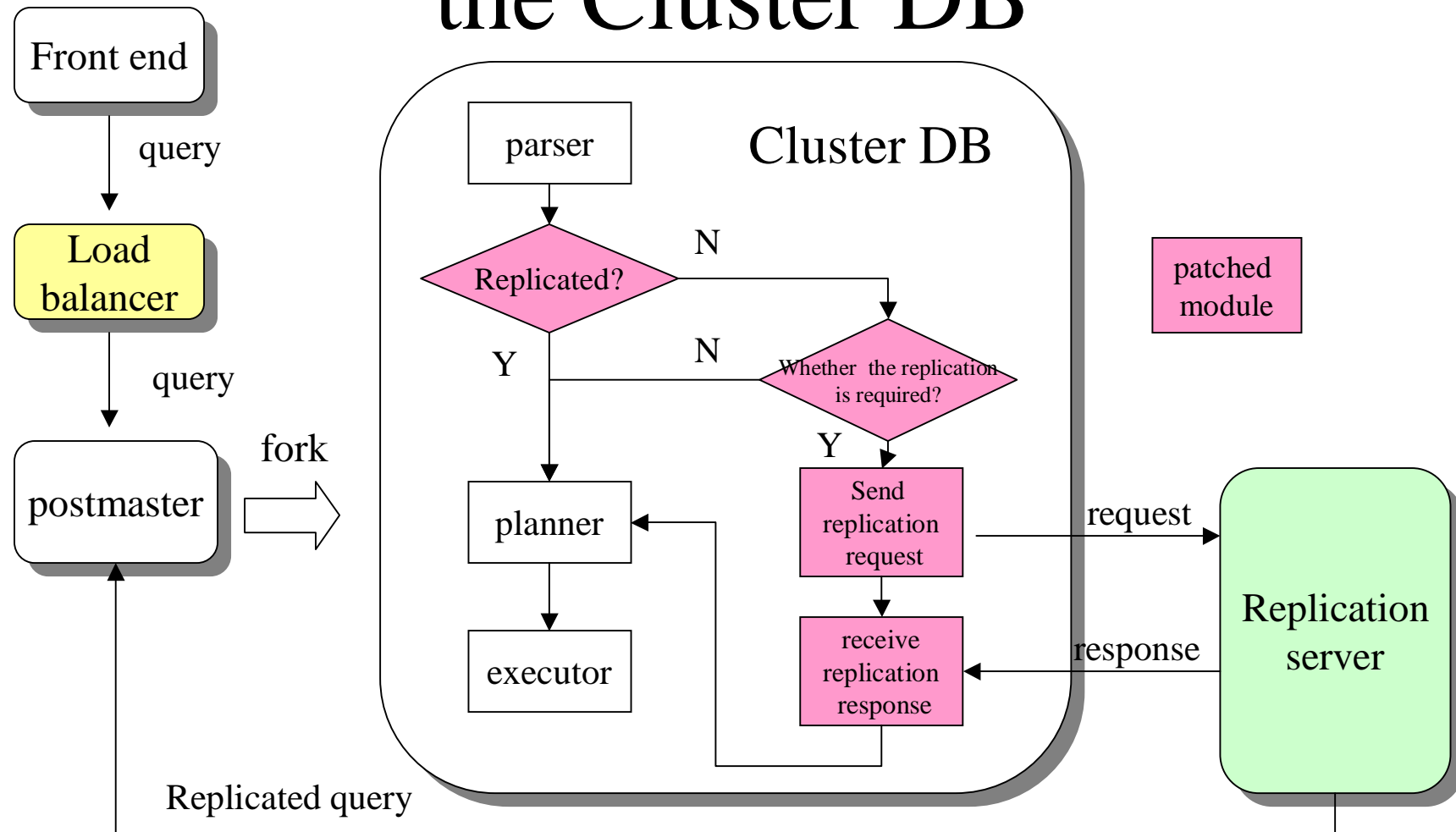
Feature of PGCluster

- Multi-master & synchronous replication system
 - Suitable for
 - Heavy read load system
 - HA system
 - Real time backup
 - Not suitable for
 - Heavy write load system
 - Huge size of database
 - Mobile database

Procedure of replication



Replication patch of the Cluster DB



What can be replicated

- Transaction
- Stored procedure
- Return value of internal function
 - now() ,random()
- Large object
- Prepared query
- Copy (from)
- Create / Drop
 - DB , table, domain, function, group ...

High Availability support

- Automatic take over
 - Cluster DB
 - Replication server
- Dynamic server addition
 - Cluster DB
 - Replication server
- There is no single point of failure
 - All servers can set up multiple

Yet another requirement for PGCluster

- Write load dispersion
- Large size of database
- Short time recovery

Write load dispersion & Large size of database

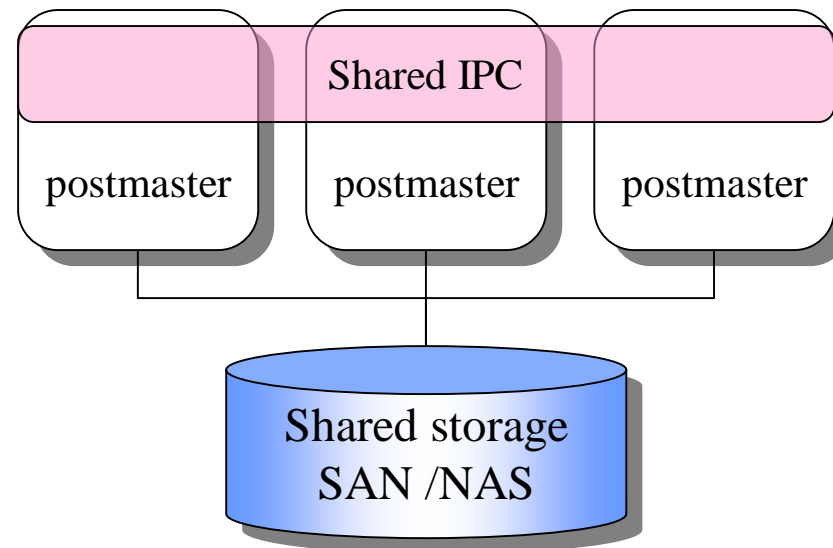
- Replication can not help for write load distribution
 - It is outside of scope for current version.
- Required distributed database
 - Parallel operation
 - PostgresForest
 - Pgpool-II
 - Shared IPC during distributed servers

Short time recovery

- Required by mobile users
 - The files under \$PGDATA are copied using ‘rsync’ command
 - Only the file with a difference is copied
 - Even if a difference is small, a file is all copied.
- Can an archive log be used?

Concept of PGCluster-II

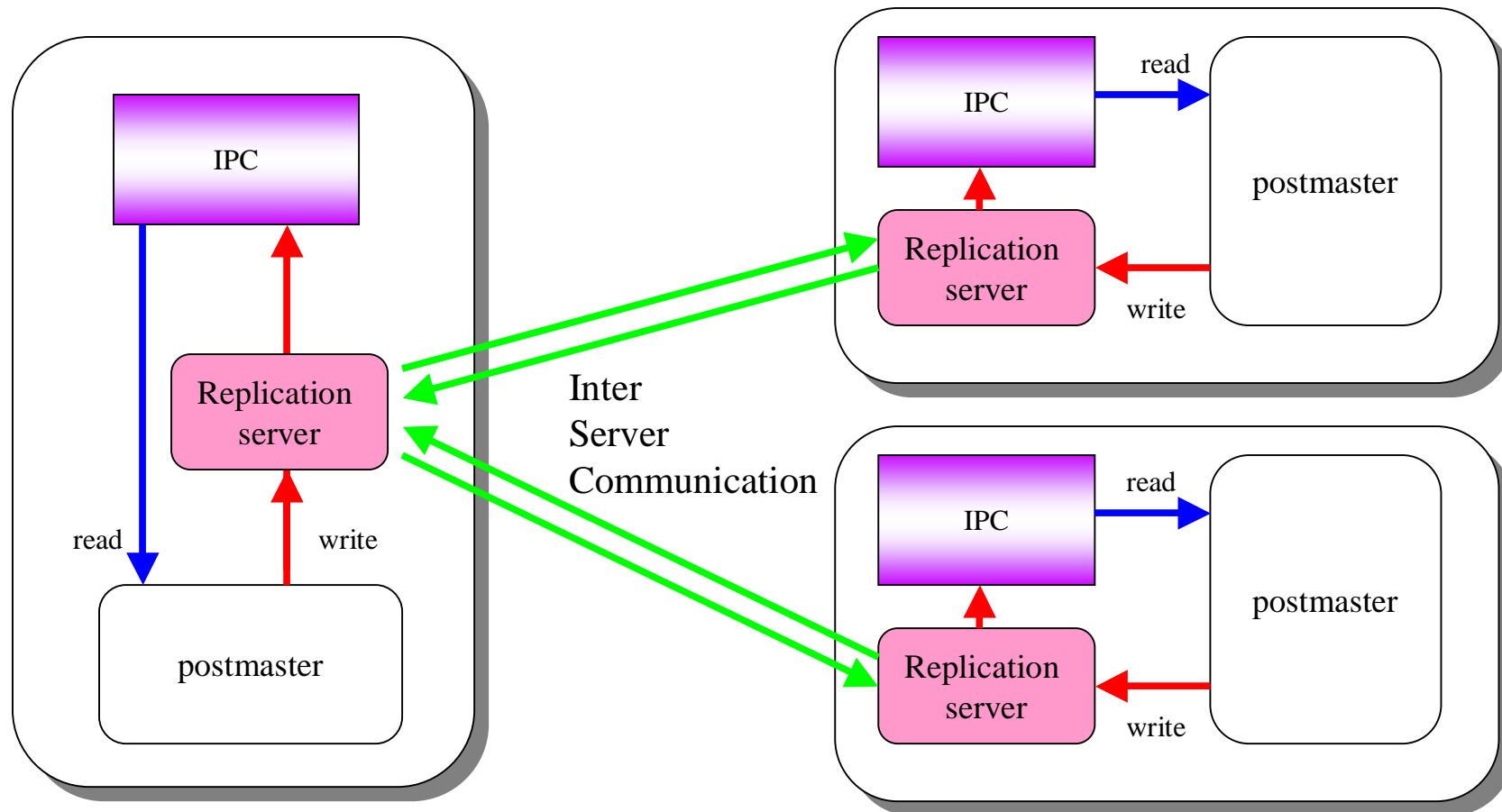
- Shared DB cluster
 - Shared storage (using SAN / NAS)
- Distributed postmaster
 - Shared IPC during distributed servers



How to share IPC during distributed servers

- Basically, the IPC is the fastest communication method during process
 - Therefore, any sharing method becomes an overhead
- It should use local IPC as much as possible
 - Read locally
 - Write synchronously (replication)
 - Required the multi-master & synchronous replication

Server composition of PGCluster-II



Expected effect and problem

- Effect
 - Performance improvement of write load
- Problem
 - Over head of the shared IPC
 - Increase in hardware cost
 - Storage
 - Network

Discussion

- How to improve the performance
 - With pgpool-II ?
- How can backup the huge data in a short time

Thank you

- When you have a question || issue || requirement for PGCluster, would you please send email.
 - mitani@sraw.co.jp
 - pgcluster-general@pgfoundry.org
- You can download from following site,
 - <http://pgfoundry.org/projects/pgcluster/>