ORACLE

MySQL

MySQL InnoDB ClusterSet

전양백

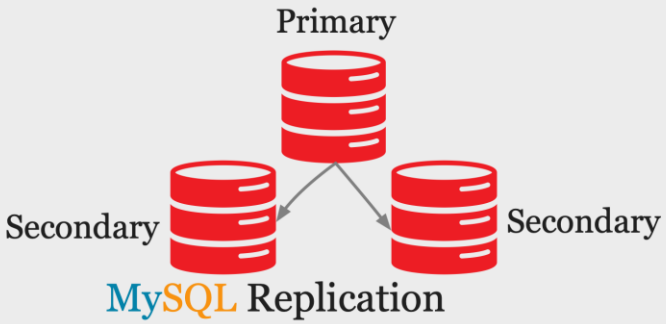
Principal Technical Support Engineer

Safe Harbor Statement

The following is intended to outline our general product direction. It is intended for information purpose only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied up in making purchasing decisions. The development, release and timing of any features or functionality described for Oracle´s product remains at the sole discretion of Oracle.

Past, Present & Future

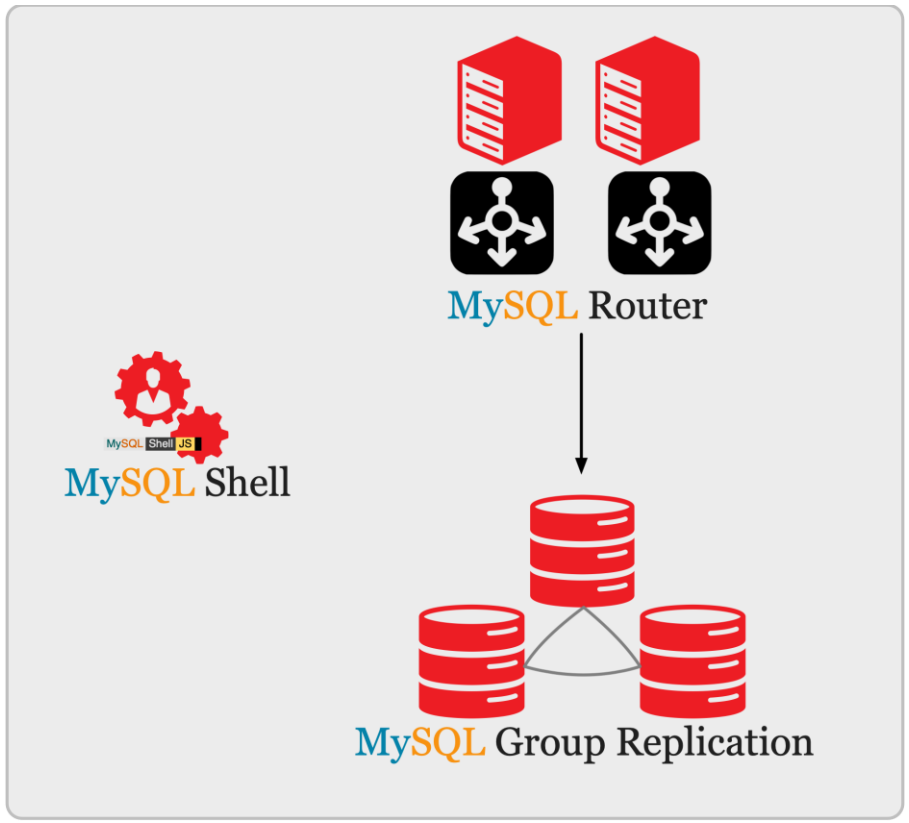
'Past' - Manual



MySQL Replication

- 기존의 복제 방식은 여러가지의 수동적인 스텝 필요
  - 복제 계정 생성, 소스로부터 백업 및 restore 절차 등
- 복제 방식
  - Async Replication
  - Semi-sync Replication
- Auto-Failover 미지원
  - Auto-failover를 위해서는 별도의 솔루션 필요
  - MMM, MHA, Orchestrator, OS layer HA solution..

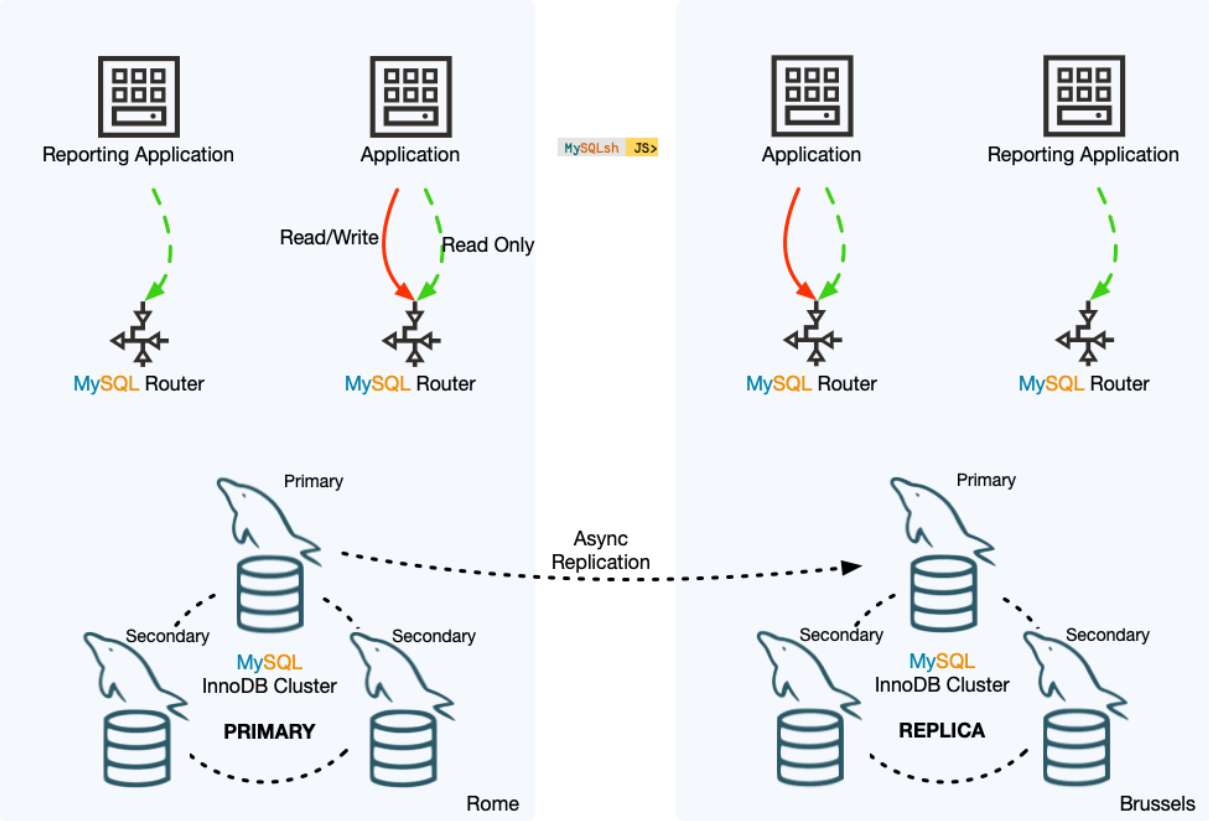
Present - Solutions!



MySQL InnoDB Cluster

- Group Replication
  - 멤버쉽 관리, 데이터 동기화, 네트워크 파티션 관리
- MySQL Shell
  - 파워풀한 인터페이스 제공
  - Java /python / sql 인터페이스
  - AdminAPI 제공 , 효율적인 DB 관리 툴
- MySQL Router
  - Read/Write 분산
  - Primary node의 auto detect

Present - Solutions!



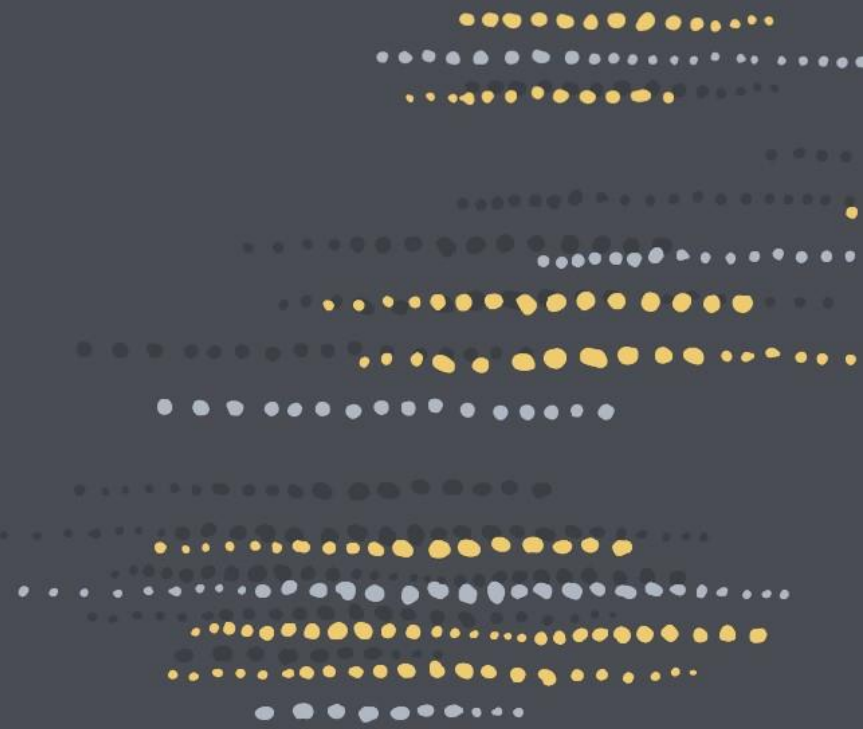
MySQL InnoDB ClusterSet (8.0.27 이후)

- For Multi-Region (c.g. 서울 DC ➡ 춘천 DC)

- Local Region : MySQL Router + InnoDB Cluster

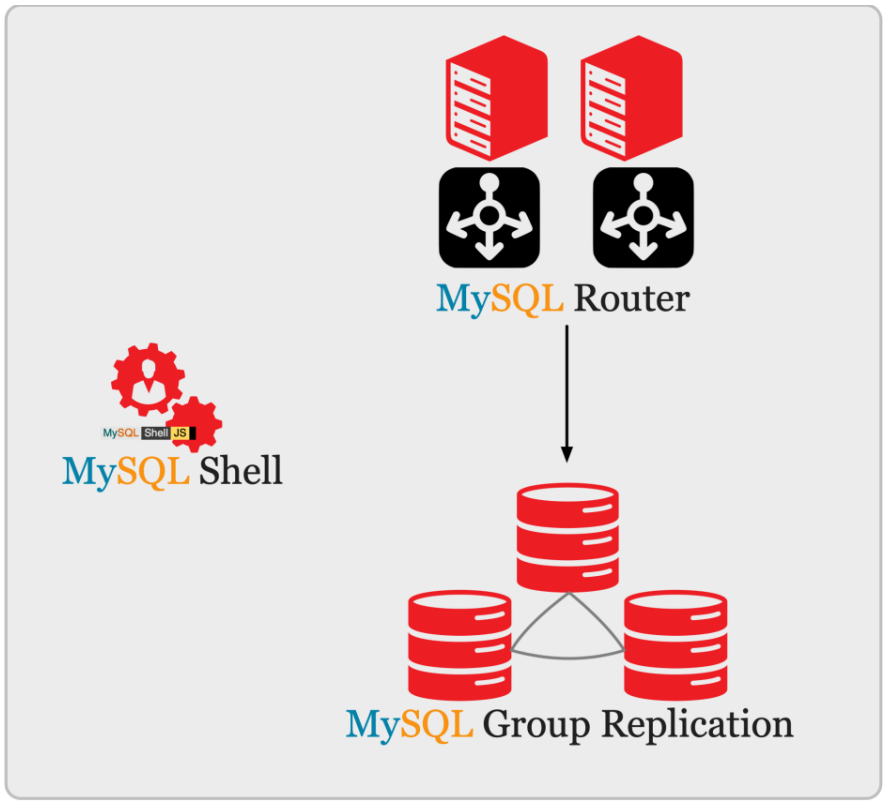
- WAN 구간 : 비동기식 방식으로 서울 데이터 센터 -> 춘천 데이터센터 간의 데이터 복제

MySQL InnoDB Cluster



MySQL InnoDB Cluster

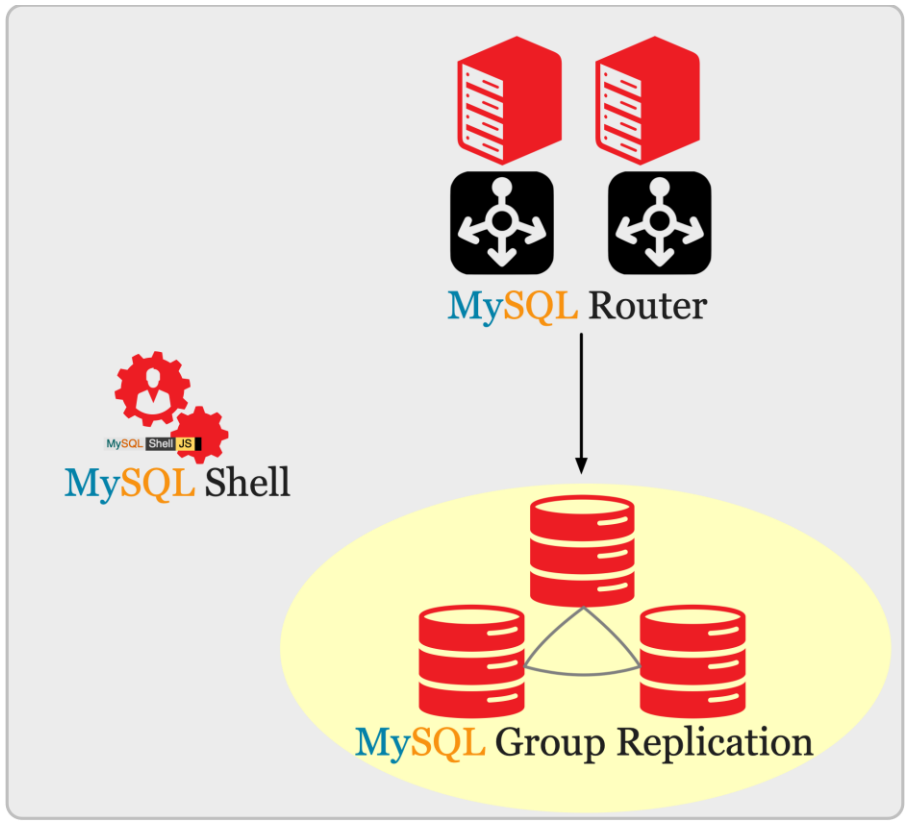
MySQL InnoDB Cluster는 MySQL DB에 대한 완벽한 HA 기능을 제공합니다.



InnoDB Cluster의 주요 컴포넌트

- MySQL Server
  - Group Replication
  - Clone Plugin
- MySQL Shell
- MySQL Router

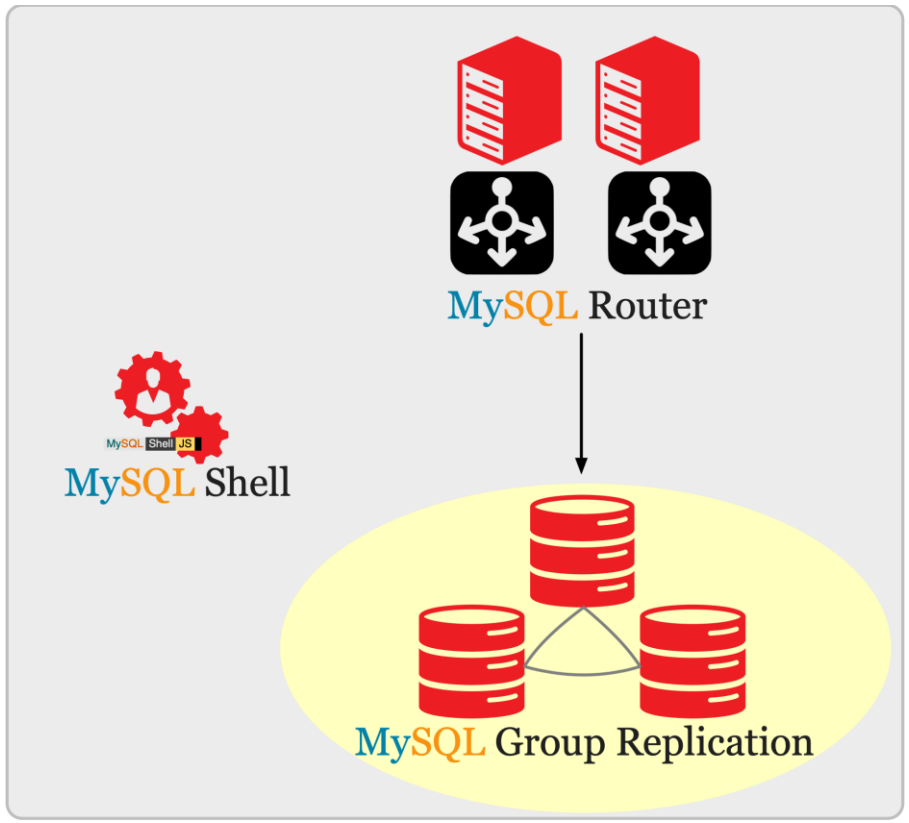
MySQL Group Replication



High Available Distributed MySQL DB

- Fault tolerance
- Confliction 감지 후 자동 Failover
- Single Primary (recommend), Multi Primary 선택
- Membership 관리의 자동화
- 온라인 노드 추가/삭제
- Split-brain 관리
- Transaction의 certification 을 통한 노드 별 데이터 일관성 유지
- 데이터 유실 예방

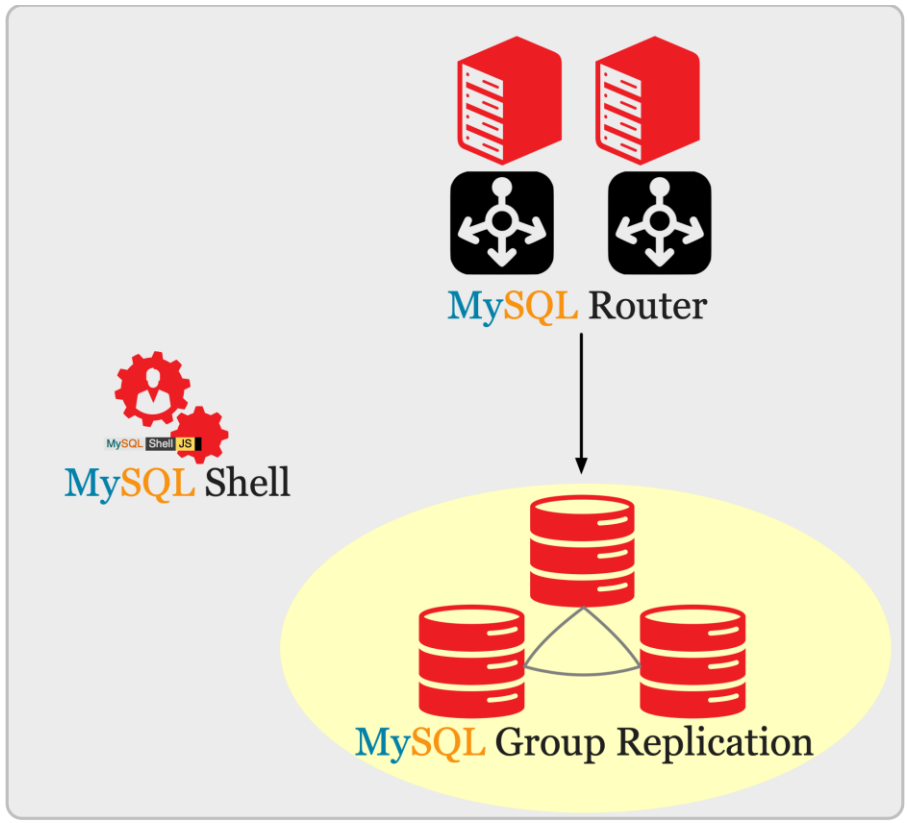
MySQL Shell



Console 기반의 개발 및 관리 기능을 제공하는 MySQL 클라이언트 툴

- JavaScript, Python 및 SQL 지원
- Document Store 및 관계형 모델 지원 (mysqlX protocol)
- AdminAPI를 통한 InnoDB Cluster 생성 및 관리
- MySQL 운영에 필요한 여러가지 utility 제공 (util.dump, util.load)
- 명령어, 배치 스크립트 지원

MySQL Router



어플리케이션과 GR을 구성한 MySQL 노드사이에서

투명한 클라이언트 라우팅을 담당하는

경량의 미들웨어 프로그램

- 애플리케이션 연결에 대한 Failover
- Read/Write에 대한 LoadBalancing
- InnoDB Cluster에 대한 metadata information 제공
- Stateless 디자인

MySQL InnoDB Cluster – Requirements / Limitations

Requirements

- InnoDB Storage Engines Only
- Primary Key
- Network Performance
- GTID / binlog_format=ROW
- Parallel Replication applier 사용

Limitations

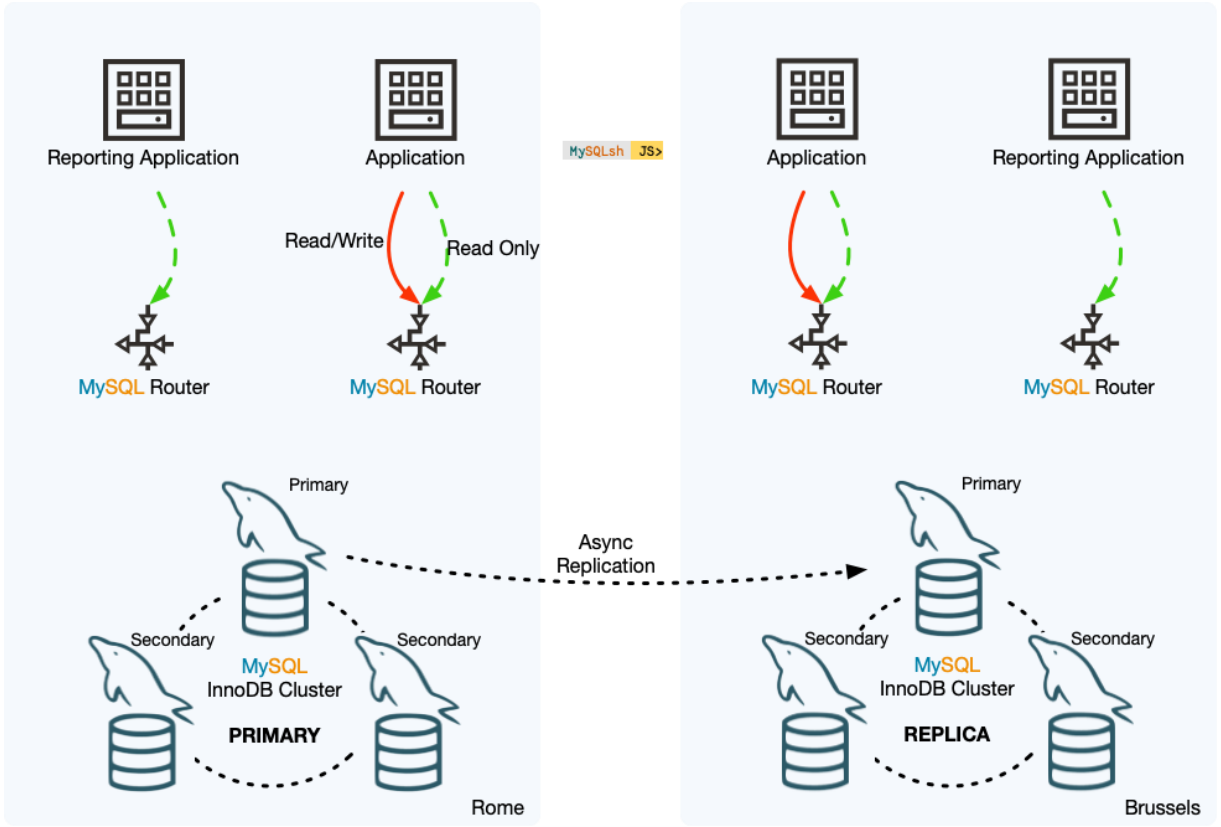
- **최대 노드 개수는 9**
- **대용량 트랜잭션**
- **READ-COMMITTED 권장 (Gap lock 지원불가)**
- **Multi-Primary node 제약**
  - Concurrent DDL/DML 지원 불가
  - FK 지원 불가
  - SELECT.. FOR UPDATE => Deadlock 발생

✓ https://dev.mysql.com/doc/mysql-shell/8.0/en/mysql-innodb-cluster-requirements.html
✓ https://dev.mysql.com/doc/mysql-shell/8.0/en/mysql-innodb-cluster-limitations.html

MySQL InnoDB ClusterSet

MySQL InnoDB ClusterSet

각 데이터 센터에 각각 InnoDB Cluster로 구성하며,
Primary Cluster와 Replica Cluster는 Async 방식으로 데이터 복제



High Availability (Failure within a Region)

- RPO=0
- RTO=seconds (automatic failover)

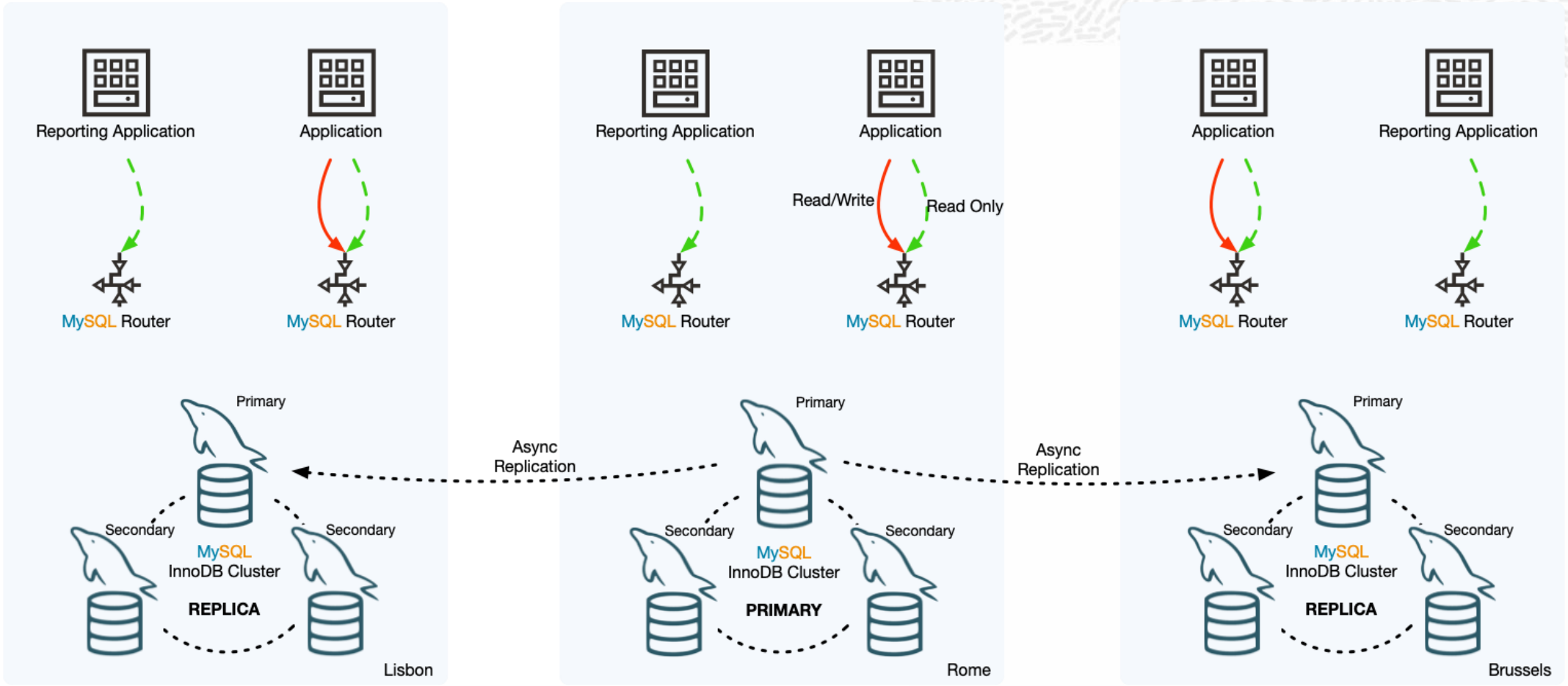
Disaster Recovery (Region Failure)

- RPO !=0
- RTO = mintues or more (manual failover)
- No write performance impact

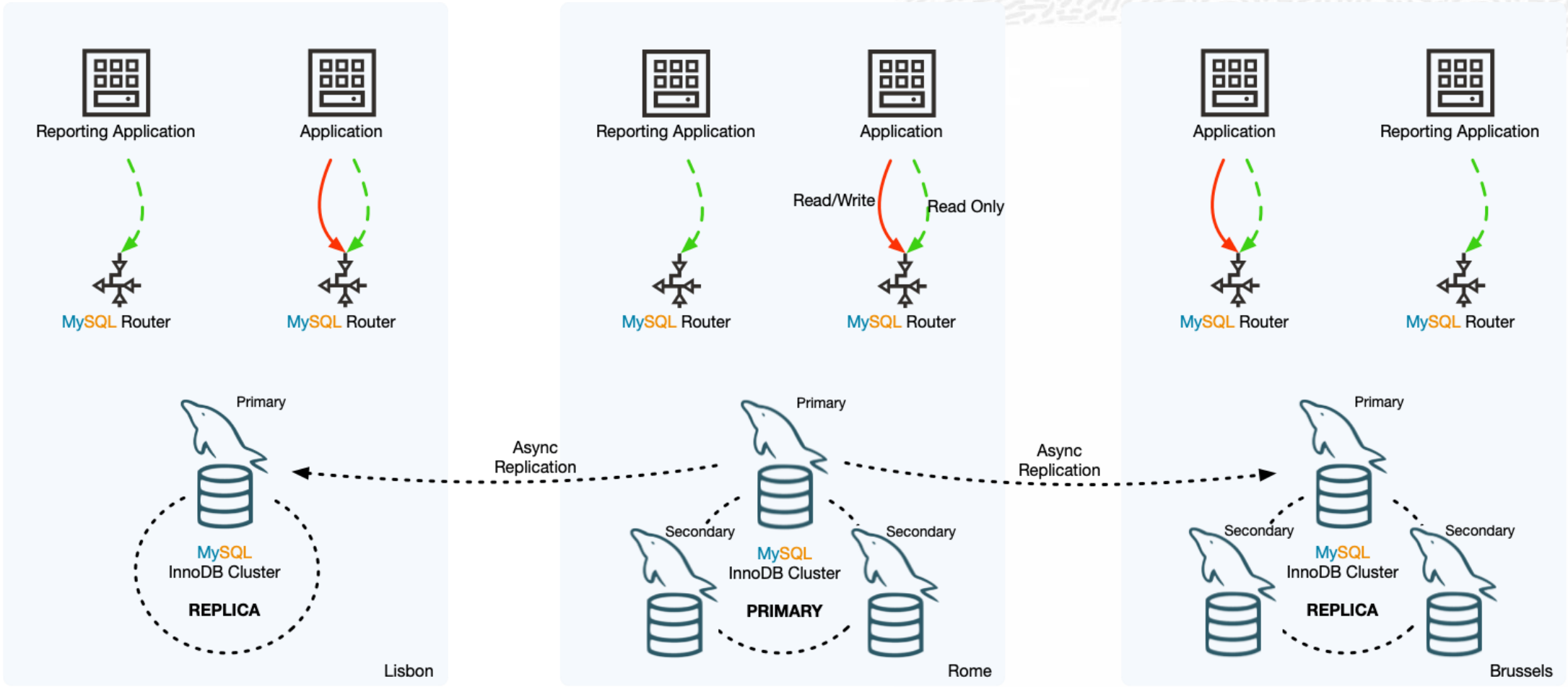
Features

- 사용 편리성
- 친숙한 환경 (mysqlsh, CLONE,…)
- Online 노드 추가/삭제
- Application 변경없이 Replica Cluster로 변경

MySQL InnoDB ClusterSet – 3 Data Centers



MySQL InnoDB ClusterSet – Not every Cluster has to be 3 nodes



Replication Enhancements

Features in replication that made ClusterSet possible:

- 8.0.22: Automatic connection failover for Async Replication Channels
- 8.0.23: Automatic connection failover for Async Replication Channels using Group Replication
- 8.0.24: Make skip-replica-start a global, persistable, read-only system variable.
- 8.0.26: Group Replication Member actions (configurable **`super_read_only`** on **`PRIMARY`** member)
- 8.0.26: Specify the UUID used to log **`View_change_log_event`**
- 8.0.27: Asynchronous Replication Channel conguration automatically follows the **`PRIMARY`** member

MySQL InnoDB ClusterSet – Requirements / Limitations

Requirements

- 모든 컴포넌트는 8.0.27 이상 버전 요구
- InnoDB Cluster의 메타버전이 2.1.0 이상
- Single-Primary 구성만 적용
- In-bound replication이 존재할 경우 구성 불가

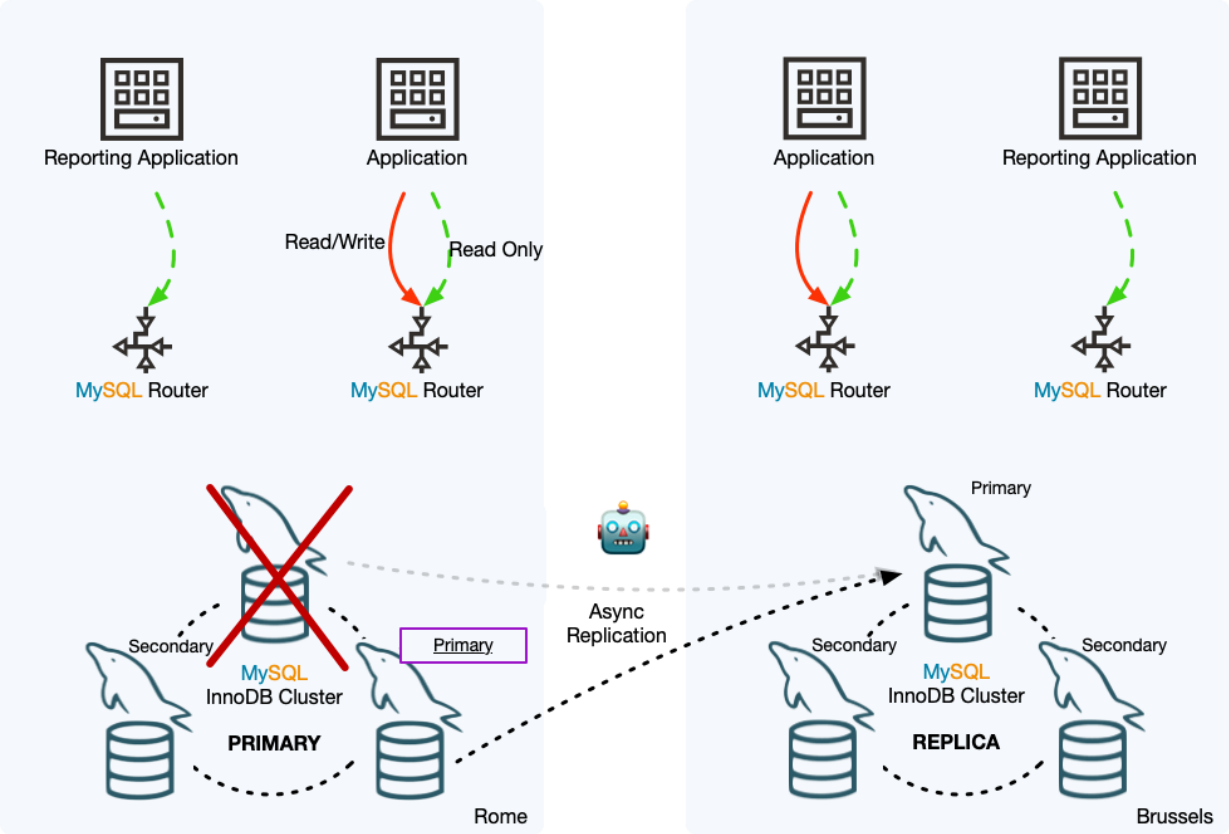
Limitations

- **일관성보다 가용성을 우선시하는 구성**
- **Failover 시에는 수동으로 실행**
- **Semi-sync 지원불가**

✓ https://dev.mysql.com/doc/mysql-shell/8.0/en/innodb-clusterset-requirements.html
✓ https://dev.mysql.com/doc/mysql-shell/8.0/en/innodb-clusterset-limitations.html

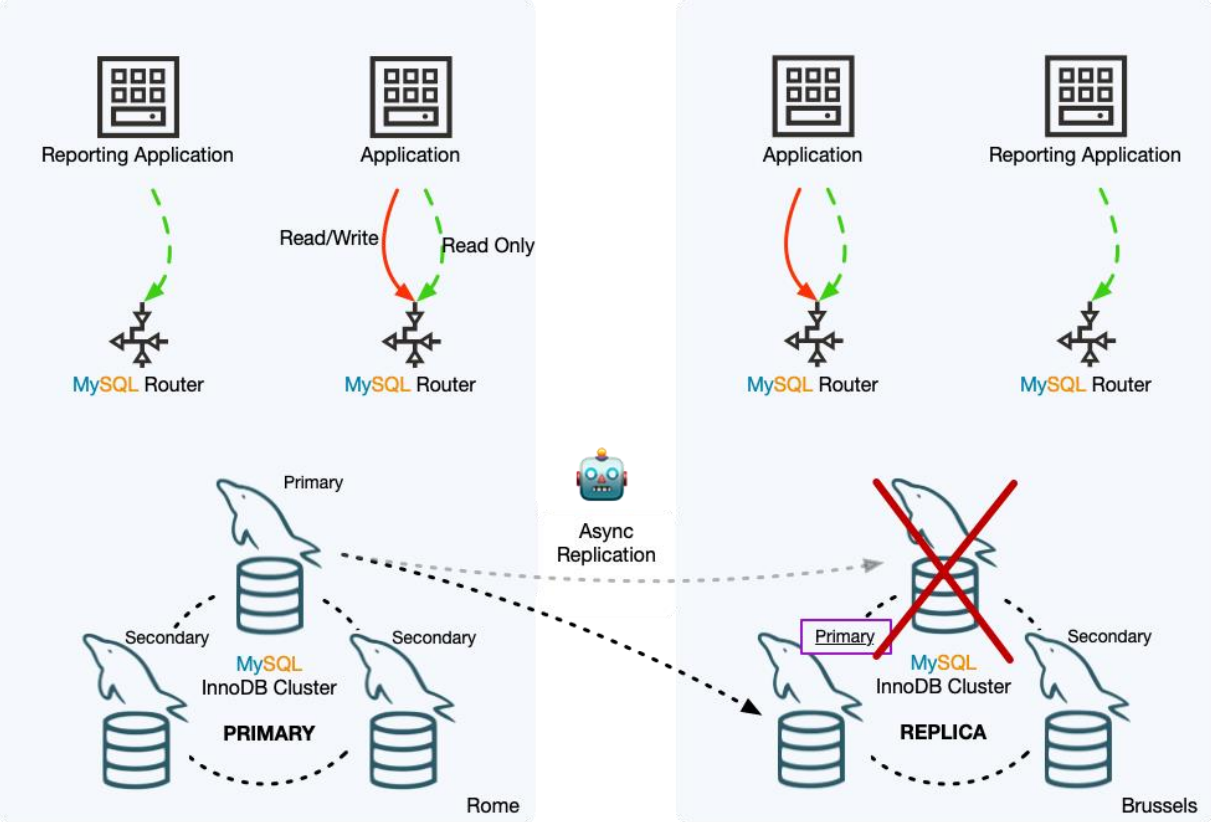
InnoDB ClusterSet 시나리오

PRIMARY Cluster PRIMARY member Crash/Partition - Automatic



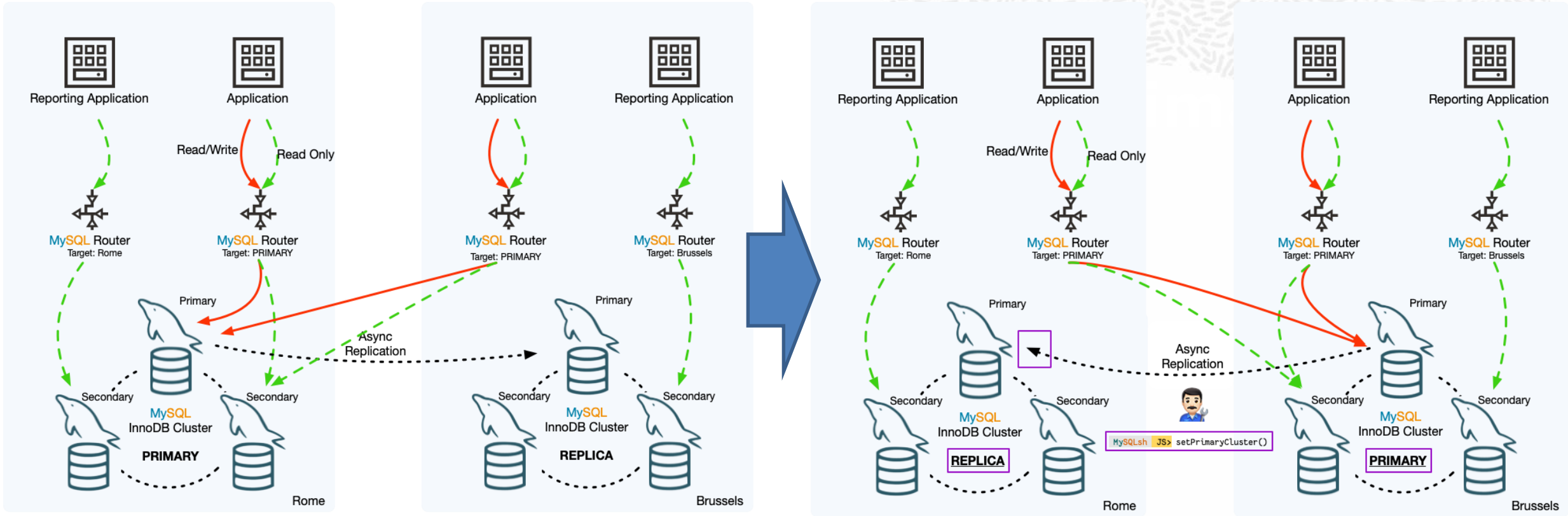
- Primary node 장애 시 Secondary node 중에서 To-Be Primary 노드를 자동 선출
- Replica Cluster에서는 Primary Cluster 에서 변경된 Primary node를 자동으로 인지하여 복제 연결

REPLICA Cluster PRIMARY member Crash/Partition - Automatic



- Replica Cluster 의 Primary node 장애 시 Secondary node 중에서 To-Be Primary 노드를 자동 선출
- Primary Cluster의 Primary node는 Replica Cluster의 변경된 Primary node를 자동으로 인지하여 복제 연결

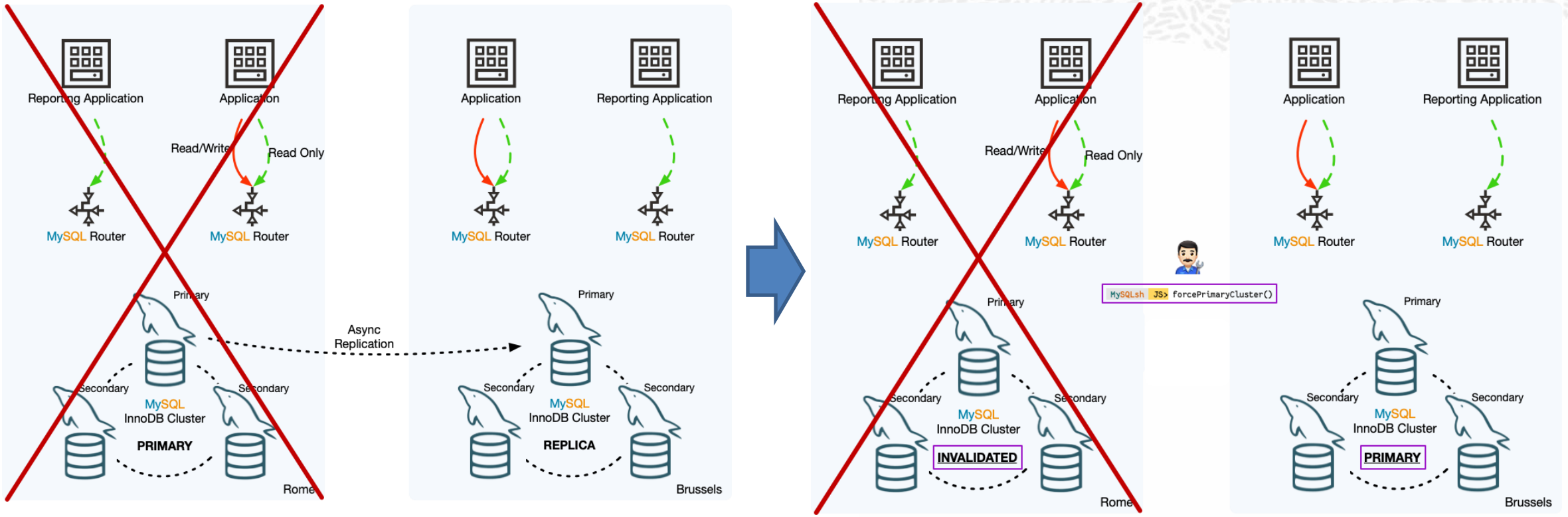
Changing Primary - Change Primary Cluster on Healthy System



Switchover

- **setPrimaryCluster()** : 이 명령어로 Replica Cluster를 Primary Cluster로 변경
- Async replicate channel은 자동 변경되며 데이터의 일관성을 보장
- 모든 mysql router의 라우팅 경로도 즉시 변경

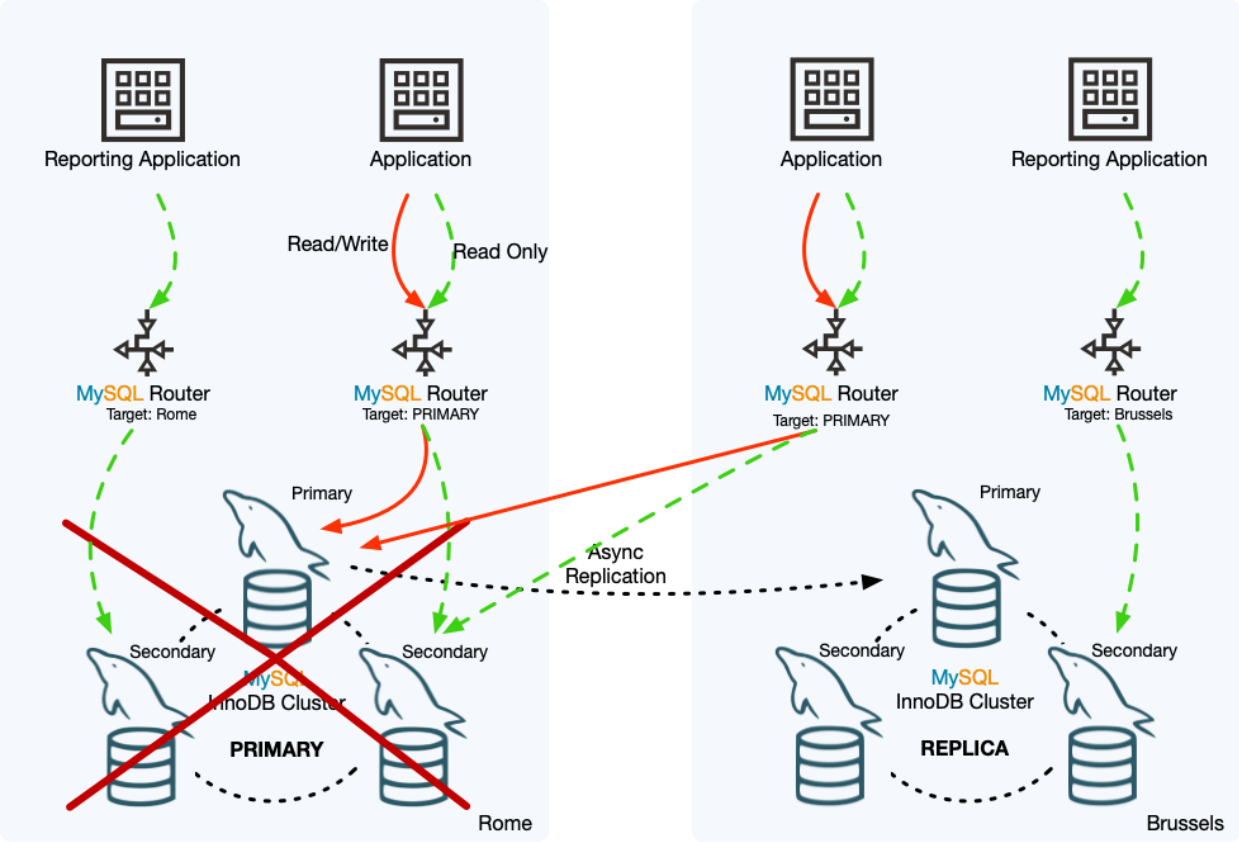
Datacenter Crash/Partition



Failover to another Cluster

- **forcePrimaryCluster()**
- Primary Cluster 그룹에 장애가 발생할 경우, 강제로 Primary Cluster를 Promote 하기 위한 명령어
- **Split-Brain 발생 가능성**

Group Replication Crash/Partition



Router Integration

When GR is offline:

- network partition
- no quorum
- full cluster lost (e.g. power outage)

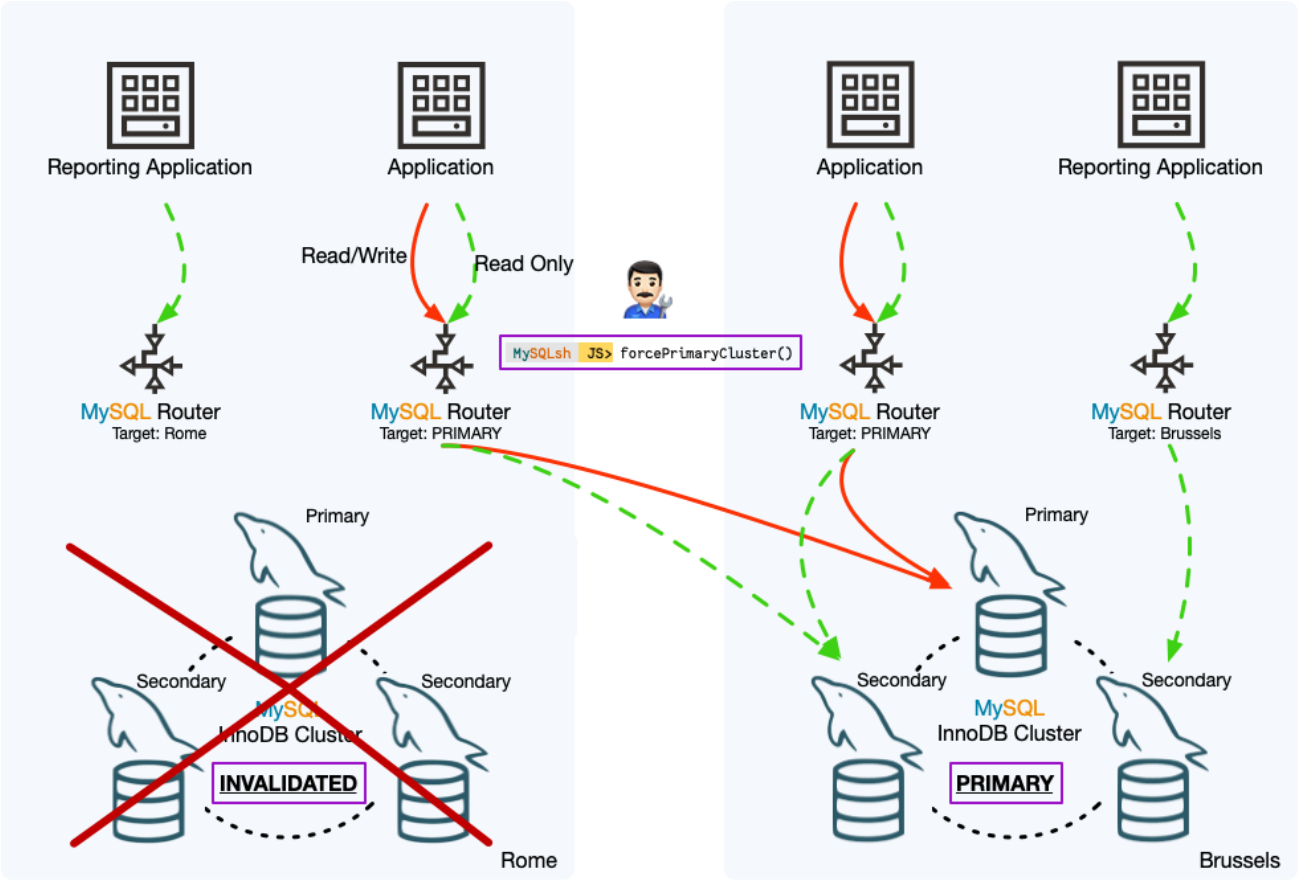
Failover to another Cluster

- **forcePrimaryCluster()**

  - 새로운 Primary Cluster로 강제 promotion

- Router 인스턴스들은 새로운 Primary Cluster로 트래픽 변경

Group Replication Crash/Partition - forcePrimaryCluster() & Router



Router Integration

When GR is offline:

- network partition
- no quorum
- full cluster lost (e.g. power outage)

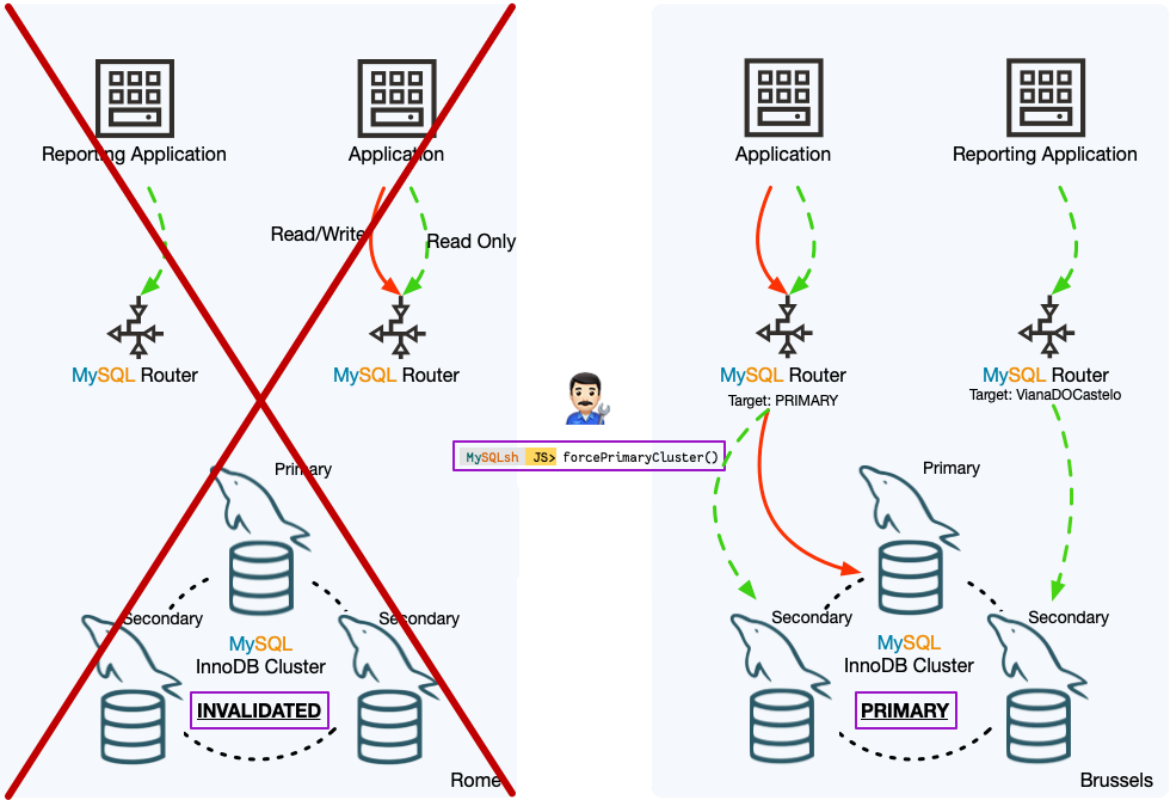
Failover to another Cluster

- **forcePrimaryCluster()**

  - 새로운 Primary Cluster로 강제 promotion

- Router 인스턴스들은 새로운 Primary Cluster로 트래픽 변경

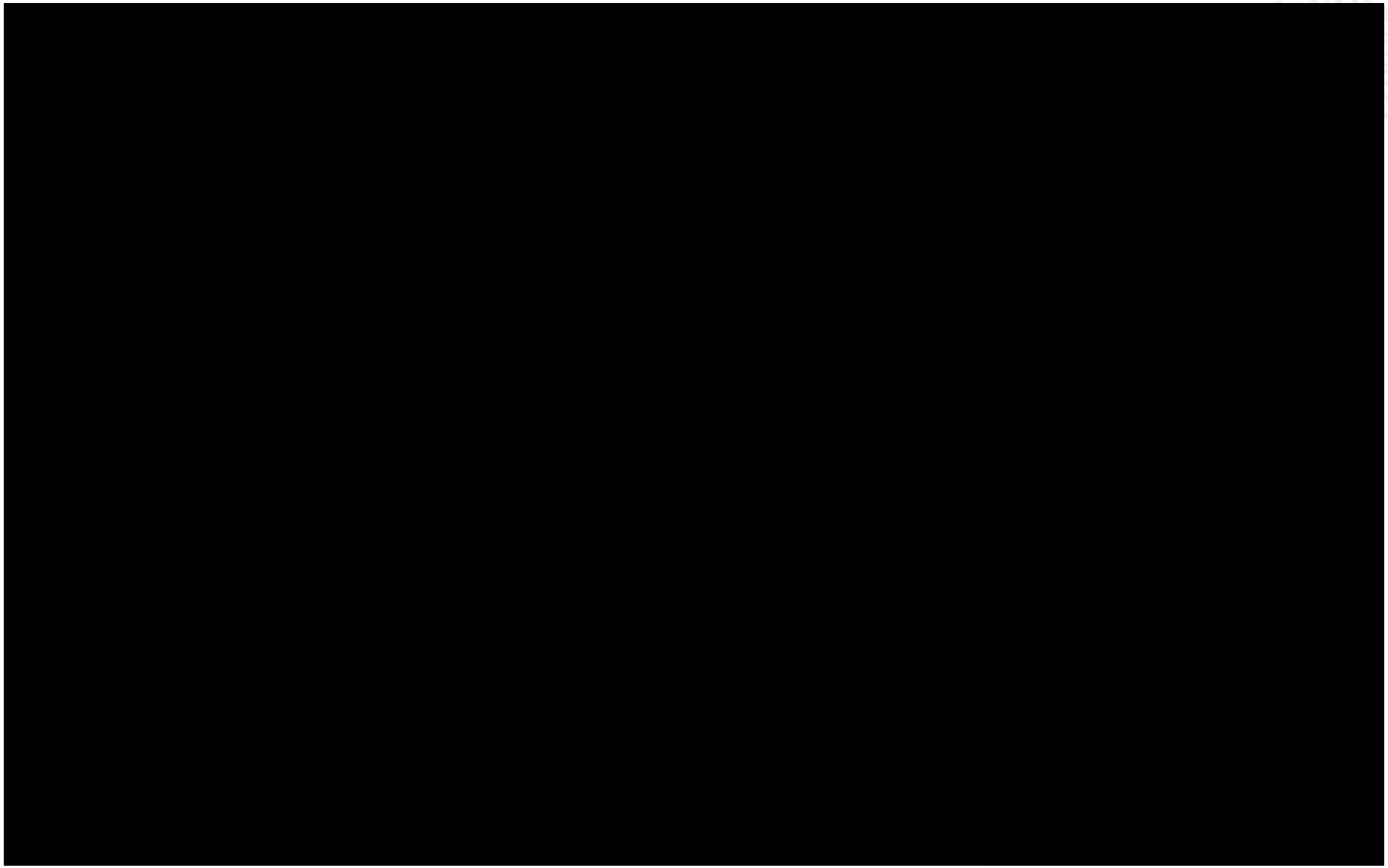
Datacenter Crash/Partition – Router Integration



Router Integration

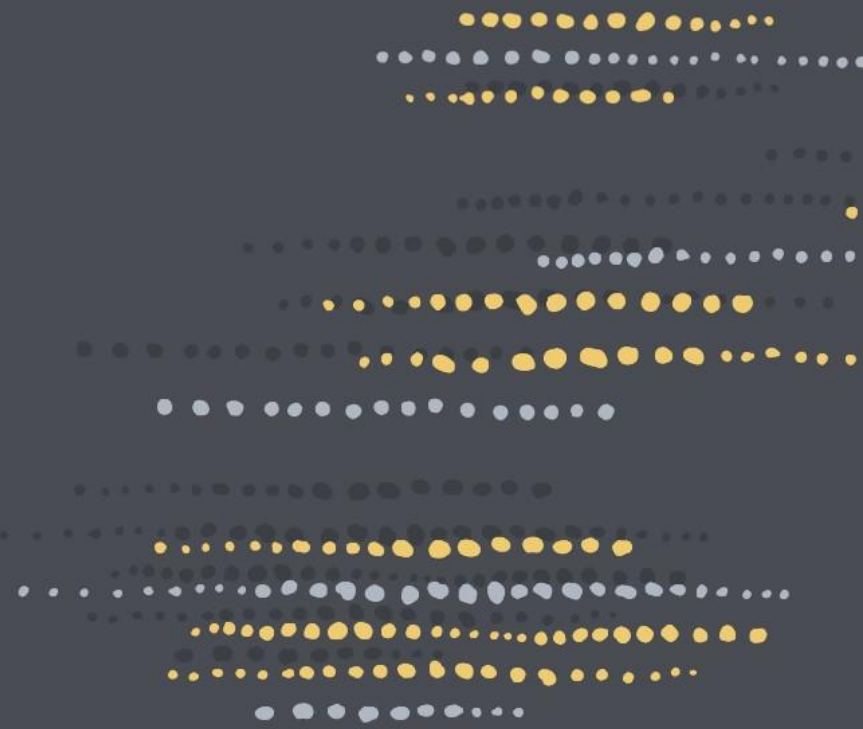
- MySQL Router는 새로운 토폴로지에 대해서 인지해야 하며, 트래픽을 redirect 해야 한다.
- Old Primary cluster를 라우팅했던 MySQL Router들은 old primary cluster에 대한 정보를 포기해야 하며 새로운 topology를 배워야 한다.

MySQL InnoDB ClusterSet Demo



MySQL Database Architectures

Summary



Business Requirements

RTO & RPO ?

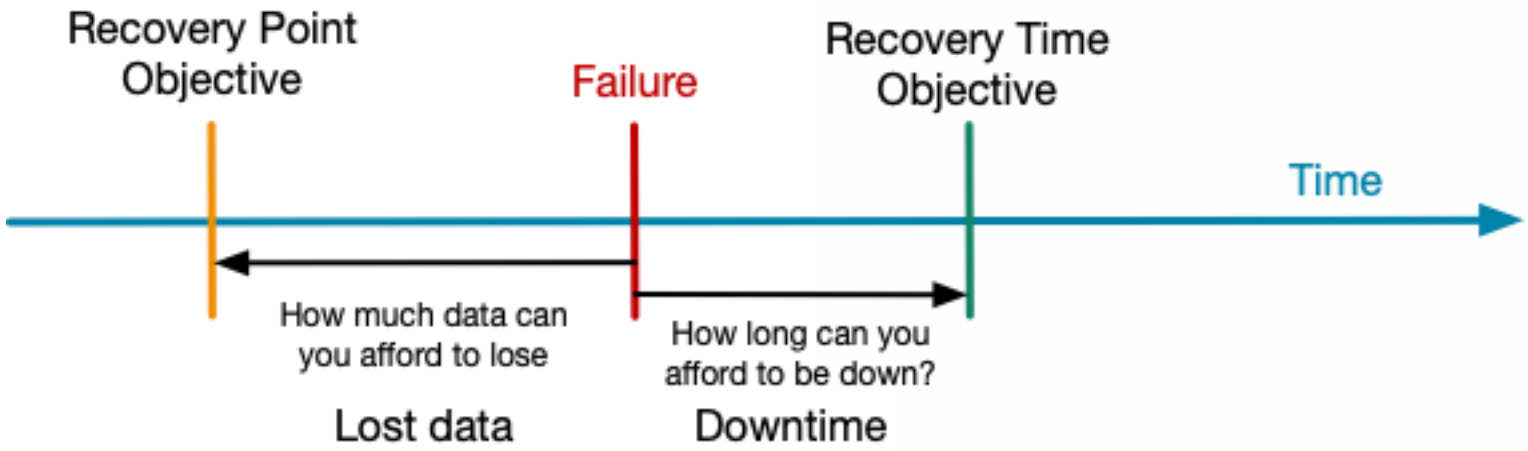
- RTO: Recovery Time Objective (= 복구 시간)
  How long does it take to recover from a single failure
- RPO: Recovery Point Objective (= 장애 시 데이터 유실)
  How much data can be lost when a failure occurs

Types of Failure

High Availability: Single Server Failure, Network Partition

Disaster Recovery: Full Region/Network Failure

Human Error: Little Bobby Tables



MySQL Database Architectures

Single Region

Requirement	Solution
RTO = hours, RPO = minutes	MySQL Server w. Backups & Binary Log Sync
RTO = hours, RPO = less than a second	MySQL Server w. Backups & Binary Log Stream
RTO = minutes, RPO = less than a second	MySQL InnoDB ReplicaSet
RTO = seconds, RPO = 0	MySQL InnoDB Cluster

Multi Region

Requirement	Solution
RTO = minutes, RPO = seconds	MySQL InnoDB Cluster w. asynchronous replica
RTO = seconds and/or RPO = 0	MySQL InnoDB ClusterSet

Q&A