Ideas for evolution of replication technology @ CERN

Distributed Databases Operations Workshop
November 16th, 2010

Zbigniew Baranowski, IT-DB

Database
SERVICES

- Replication use cases at CERN

- Motivation for evolution of replication

- Oracle replication technologies

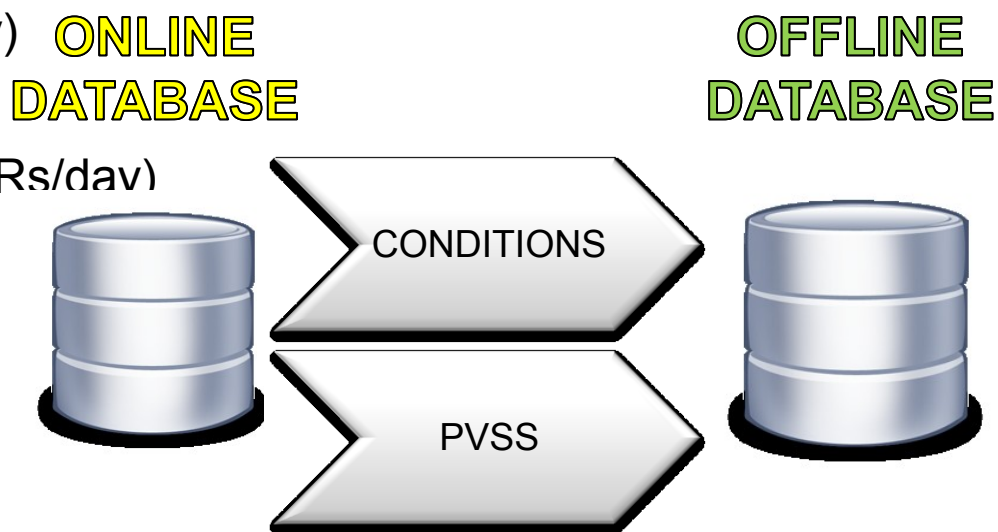- Possible future replication solutions for LCG

- Summary

- **ATLAS**
  - CONDITIONS (4M LCRs/day)
  - PVSS (60M LCRs/day)
- **CMS**
  - CONDITIONS (6M LCRs/day)
  - PVSS (20M LCRs/day)
- **LHCb**
  - CONDITIONS (6K LCRs/day)

- **ALICE**
  - PVSS (4M LCRs/day)
- **COMPASS**
  - PVSS (4M LCRs/day)

ONLINE DATABASE

OFFLINE DATABASE

CONDITIONS

PVSS

- **LHCb** (in addition to ONLINE-OFFLINE)
  - CONDITIONS (8K LCRs/day)

OFFLINE DATABASE

ONLINE DATABASE

- **ATLAS**
  - CONDITIONS (4M LCRs/day)

T1 DATABASES

**OFFLINE
DATABASE**

- **LHCb**
  - LFC (235K LCRs/day)
  - CONDITIONS (15K LCRs/day)

- **ATLAS**
 - AMI (800K LCRs/day)
  - Muon (700K LCRs/day)

T1/T2
DATABASE

OFFLINE
DATABASE

- Need of stable and reliable replication service

- Streams 10g require frequent interventions (at least once per week)
  - Consistency problems
  - Blocking sessions
  - Memory pools shortage
  - Logminer crashes
  - Users unsupported changes

- Streams administration is time consuming and requires expert knowledge

- Migration to 11gR2 in 2012

- Is there a solution which can simplify maintenance of replication?
  - Satisfies physics data workload
  - Requires minimum maintenance effort
  - Is resilient to user's unsupported operations
  - Ensures replicated data consistency
  - Utilizes minimum amount of resources

- # Logical (SQL based) replication
  - ## Streams11gR2
  - ## GoldenGate

- # Physical (block-level) replication
  - ## Active DataGuard11gR2

- # Combinations of physical and logical replication
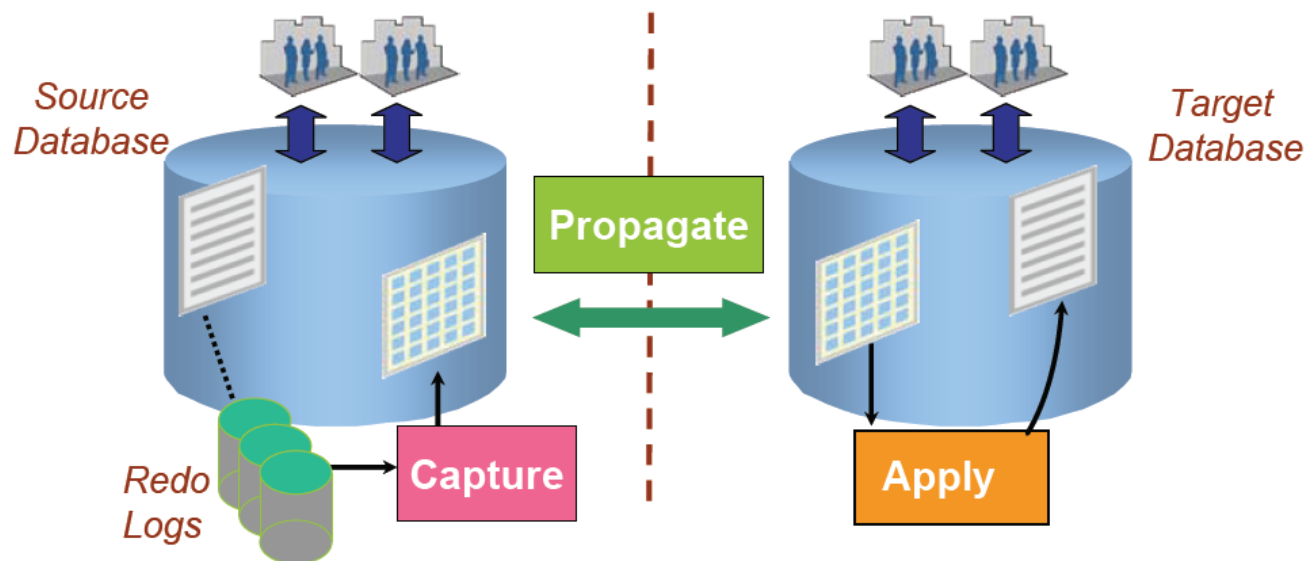
- Technology features
  - ☹ Considerable maintenance effort
    - but in 11g should be less than in 10g
  - ☺ No additional license required
  - ☺ Many improvements
    - stability, management, monitoring, verification of data consistency
  - ☺ Very good performance (30K-40K LCRs/s)
  - ☺ Best practices identified – a lot of experience
  - ☺ Source and destination database fully accessible for reads and writes

- ## As ONLINE – OFFLINE replication

  - ☹ Users and data content can abort the replication

  - ☹ streams processes may affect performance of online database

  - ☺ no extra hardware needed

  - ☺ bi-directional replication

ONLINE DATABASE → SQLs → **OFFLINE DATABASE** ← SQLs ←
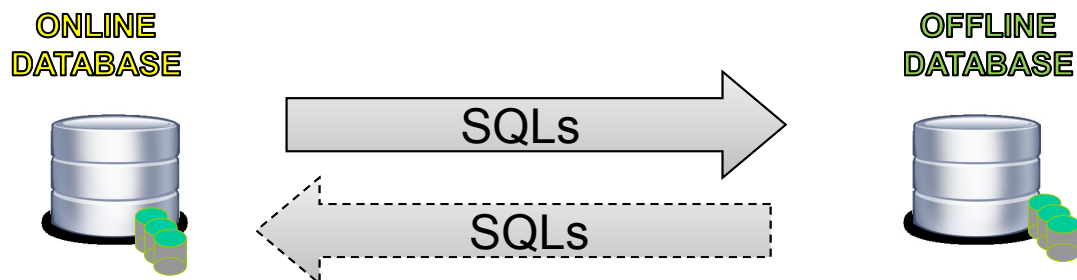
- ## As OFFLINE – T1s

  - Recovery of replica requires
    - ☹ coordination between T1 and other T1, T0
    - ☹ expert knowledge of procedures

  - Downstream capture
    - ☹ additional hardware required
    - ☺ complete isolation from OFFLINE database
    - ☺ standby database can be source of replication

  - ☺ T1s databases is read/write accessible

  - ☺ Good monitoring for distributed streams deployment (strmmon, EM)

T1S DATABASES

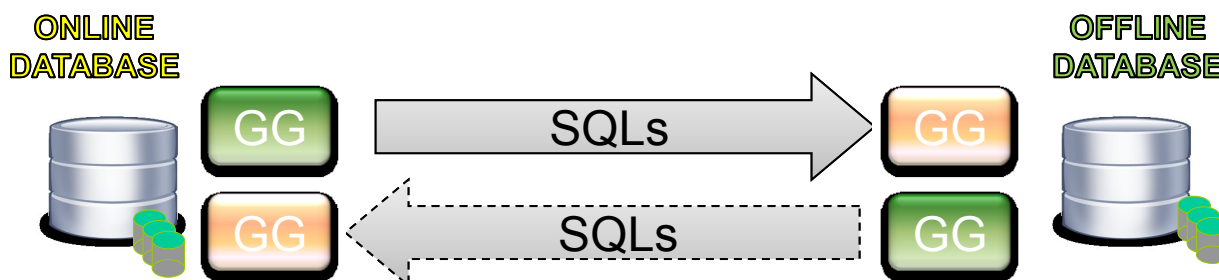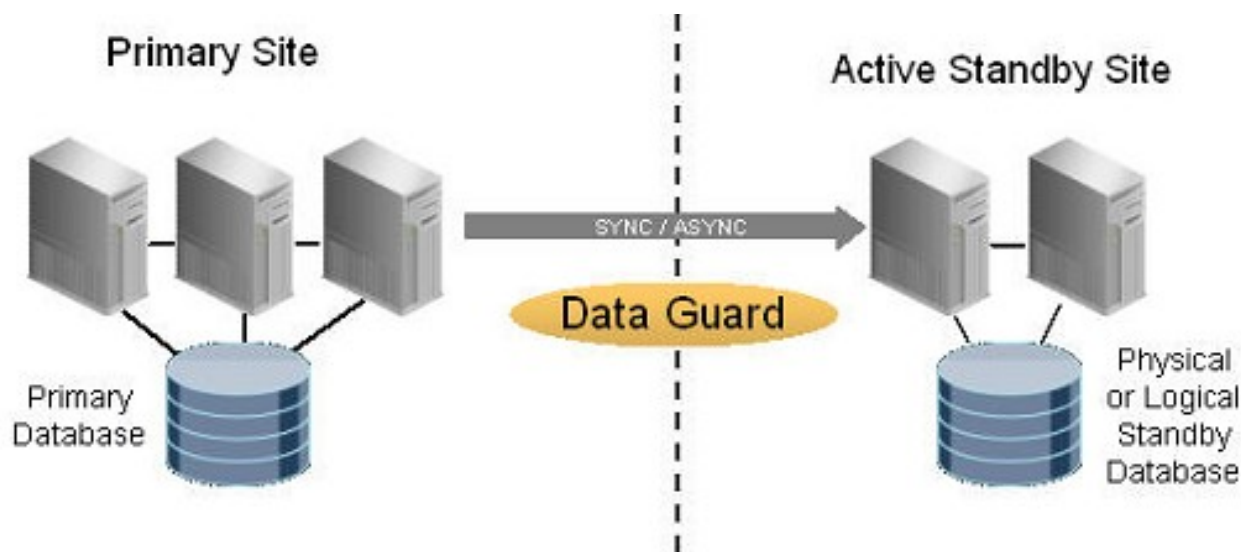OFFLINE DATABASE

DOWNSTEAM DATABASE

Redo Transport
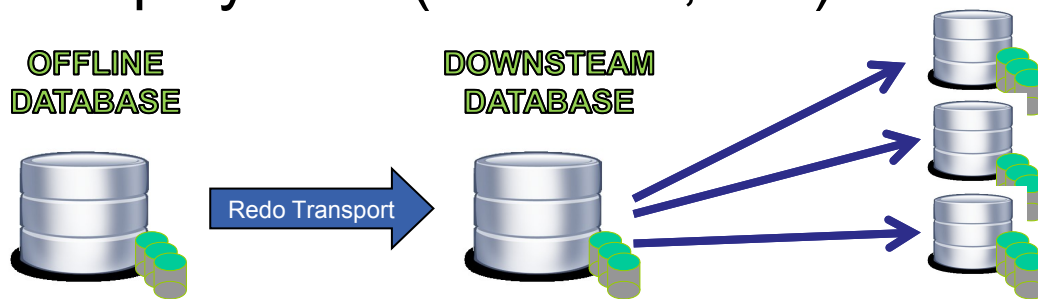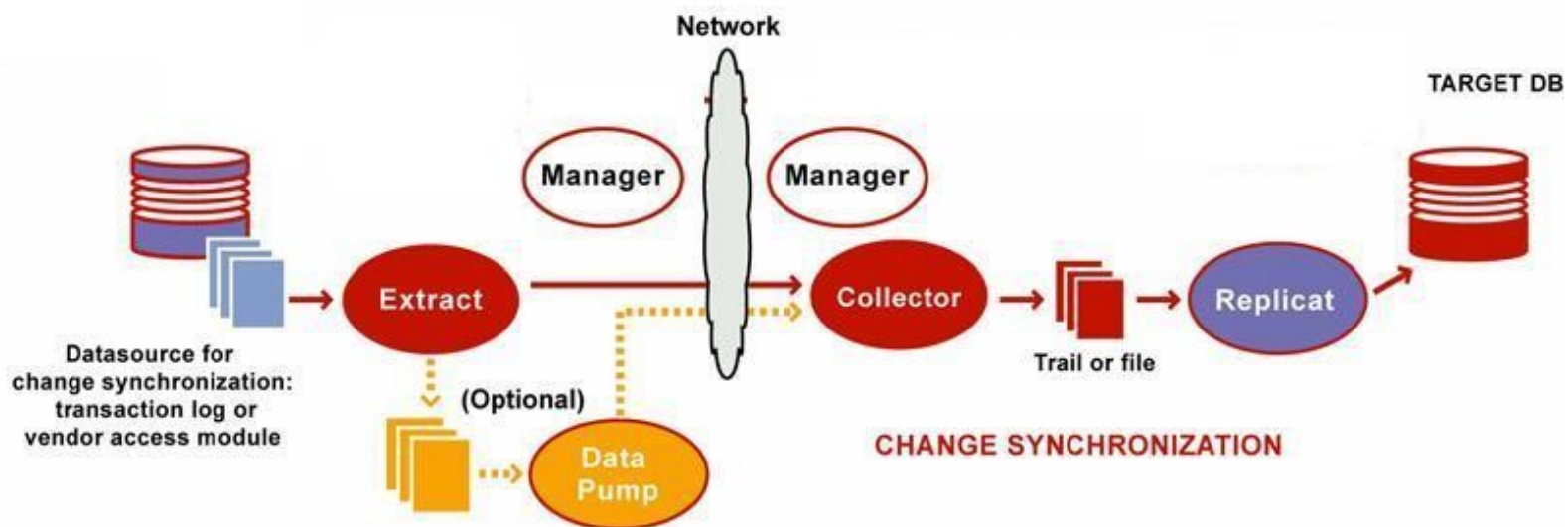
Source: Oracle.com

- Technology features

  - ☺ source and destination database fully accessible for reads and writes

  - ☺ good quality of software (very stable, free of locks, almost transparent for databases)

  - ☺ good performance 7 – 12K LCRs /s

  - ☹ additional license required

  - ☹ standby database cannot be used as source

  - ☹ no in-house experience

  - ☹ additional dedicated disk space required for trail files

  - ☹ additional software to be installed and maintained on database's machines

GoldenGate solution

- ## As ONLINE-OFFLINE replication
  - ☺ no extra hardware needed
  - ☺ possible loops back in replication
  - ☺ minor impact on source database
  - ☹ users and data content can abort the replication

ONLINE DATABASE   GG   SQLs   GG   **OFFLINE DATABASE**

GG   SQLs   GG

- ## As OFFLINE – T1s

  - ☺ easier maintenance
    - No side effects on source when target is down
    - No split of replication required
    - Trail files can be used for T1 recovery
  - ☹ no remote administration - access to nodes required
  - ☹ no monitoring for distributed environment
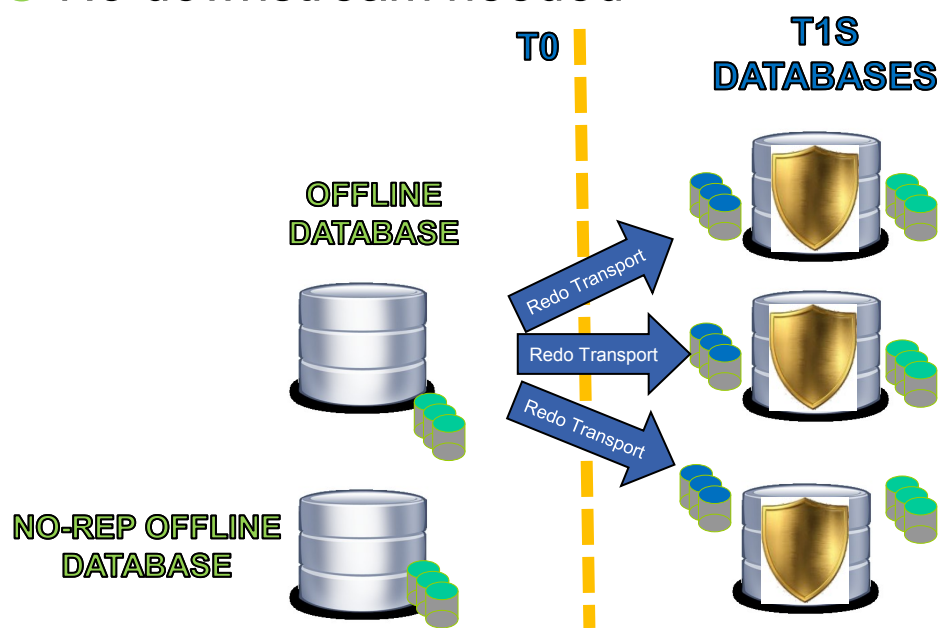 - ☹ cannot use standby database (i.e. Active Dataguard) as a source of replication

Active DataGuard 11gR2

Source: Oracle.com

Active DataGuard 11gR2

- Technology features
  - ☺ Physical replication
    - identical copy
  - ☺ Minimum maintenance effort
  - ☺ Outperforms other replication technologies
    - Oracle claims 200 MB/s of redo processing
  - ☺ Improved data reliability of primary database
    - failover
    - automatic recovery of corrupted blocks
  - ☺ Fast recovery with RMAN
  - ☹ Additional license required
  - ☹ Target/standby database is read only

- ## As ONLINE – OFFLINE replication

  - ☹ additional database installations needed for no replicated data (split of OFFLINE)

  - ☹ same version of software required (installation, upgrades)

  - ☺ online database is protected with another standby database

  - ☺ further replication to T1s is possible in sequential standbys configuration

ONLINE DATABASE → Redo Transport → **OFFLINE STANDBY DATABASE** → Redo Transport → **T1S DATABASES**

NO-REP OFFLINE DATABASE
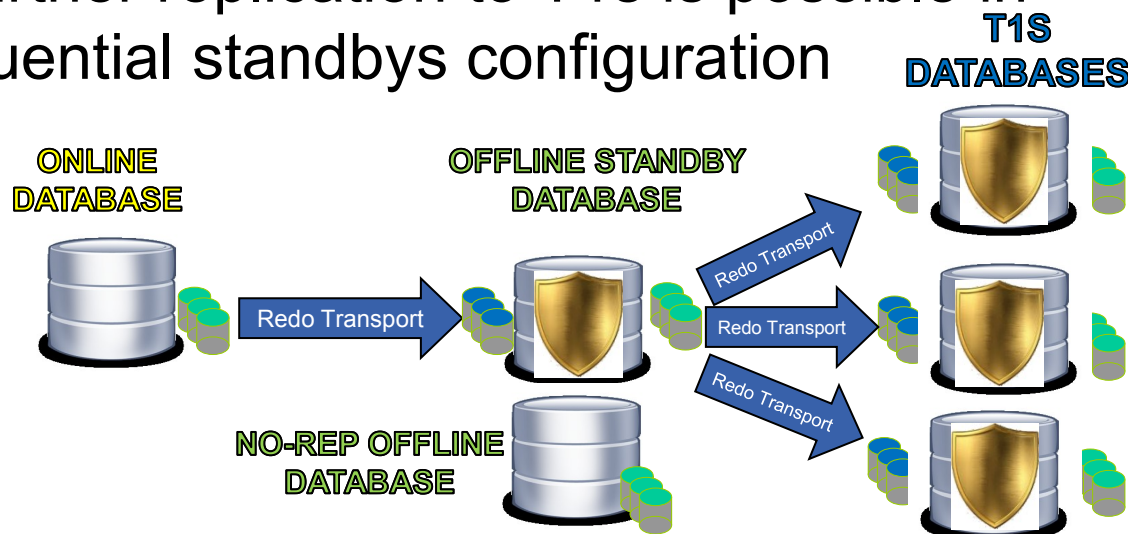
- ## As OFFLINE – T1s

  - ☹ same version required on all T1s DBs
    - Coordination of interventions becomes critical
  - ☹ T1 database is read only
  - ☹ additional database installations needed for no replicated data (split of OFFLINE)
  - ☺ Physical replication: lower maintenance effort
  - ☺ No downstream needed

- ## Streams11gR2 replication at all Tiers

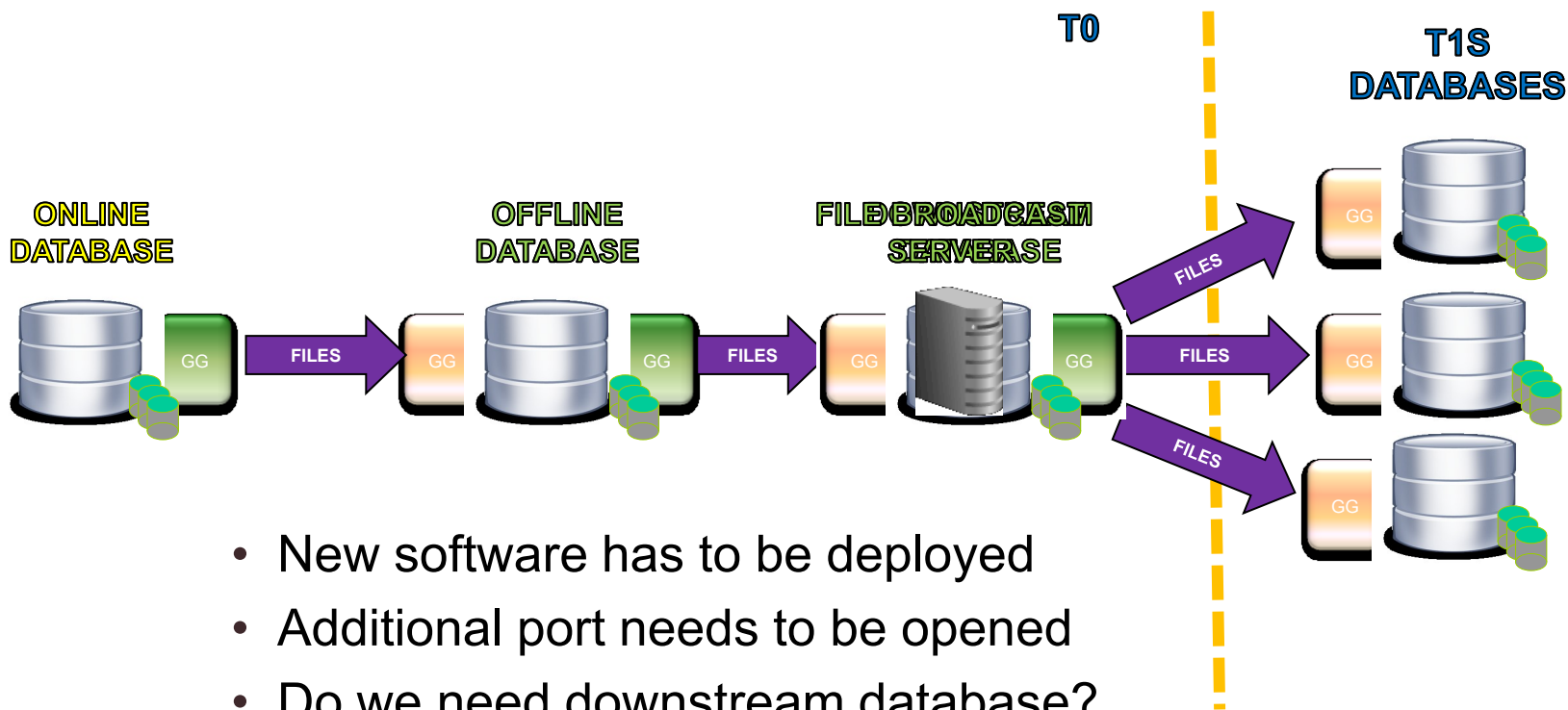ONLINE DATABASE → PROPAGATION → OFFLINE DATABASE → Redo Transport → DOWNSTREAM DATABASE → PROPAGATION → T0 / T1S DATABASES
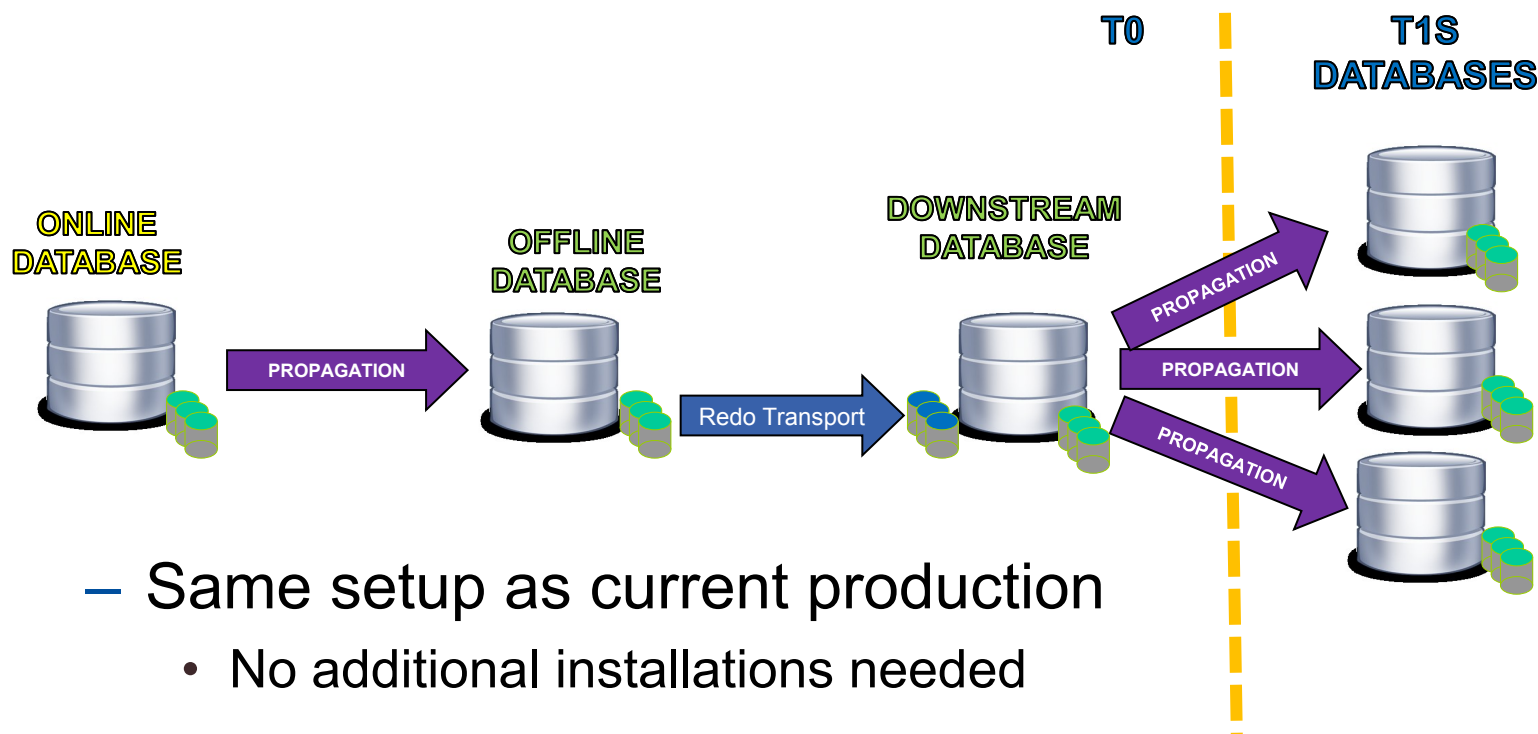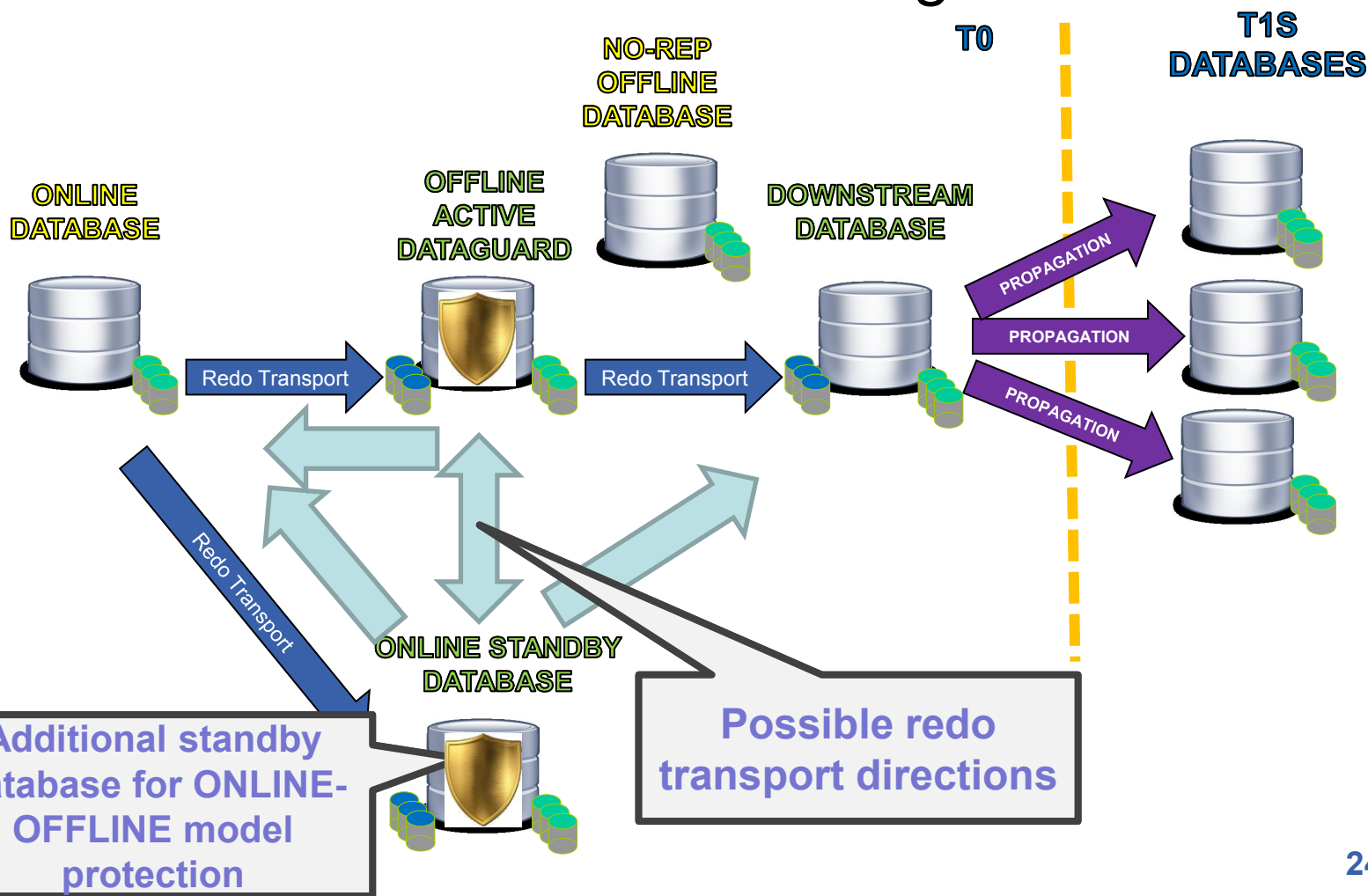
- – Same setup as current production
  - No additional installations needed

- ## GoldenGate replication at all Tiers

- New software has to be deployed
- Additional port needs to be opened
- Do we need downstream database?

- ## ONLINE –> OFFLINE: Active DataGuard
- ## OFFLINE –> T1s: Streams11g

NO-REP OFFLINE DATABASE

T0

T1S DATABASES

ONLINE DATABASE

OFFLINE ACTIVE DATAGUARD

DOWNSTREAM DATABASE

Redo Transport

Redo Transport

PROPAGATION

PROPAGATION

PROPAGATION

Redo Transport

ONLINE STANDBY DATABASE

Additional standby database for ONLINE-OFFLINE model protection

Possible redo transport directions

ONLINE-OFFLINE model is broken !!!

T0

T1S DATABASES

ONLINE DATABASE

OFFLINE ACTIVE DATAGUARD

DOWNSTREAM DATABASE

Redo Transport

Redo Transport

PROPAGATION

PROPAGATION

PROPAGATION

Redo Transport

Redo Transport

Recovery

ONLINE STANDBY DATABASE

ONLINE-OFFLINE model is broken !!!

T0

T1S
DATABASES

ONLINE
DATABASE

OFFLINE
ACTIVE
DATAGUARD

DOWNSTREAM
DATABASE

Redo Transport

PROPAGATION

PROPAGATION

PROPAGATION

Recovery

Reo Transport

Redo Transport

ONLINE STANDBY
DATABASE

Summary

- Migration to the new database versions (2012) gives an opportunity to re-design and improve the replication service
- Three candidate technologies are being investigated
    - Streams11gR 2
    - GoldenGate
    - Active DataGuard
    - Combined solutions
- Other proposals/ideas?
- Experiments requirements?

- ## Many thanks to all Physics DBAs, especially:
  - Luca
  - Jacek
  - Dawid

- ## Consultancy
  - Gancho
  - Stephen Balousek (Oracle)
  - Jagdev Dhillon (Oracle)

DB