Oracle STREAMS
for database replication in LCG

Openlab Technical Review 07 June @ CERN

Eva Dafonte Pérez

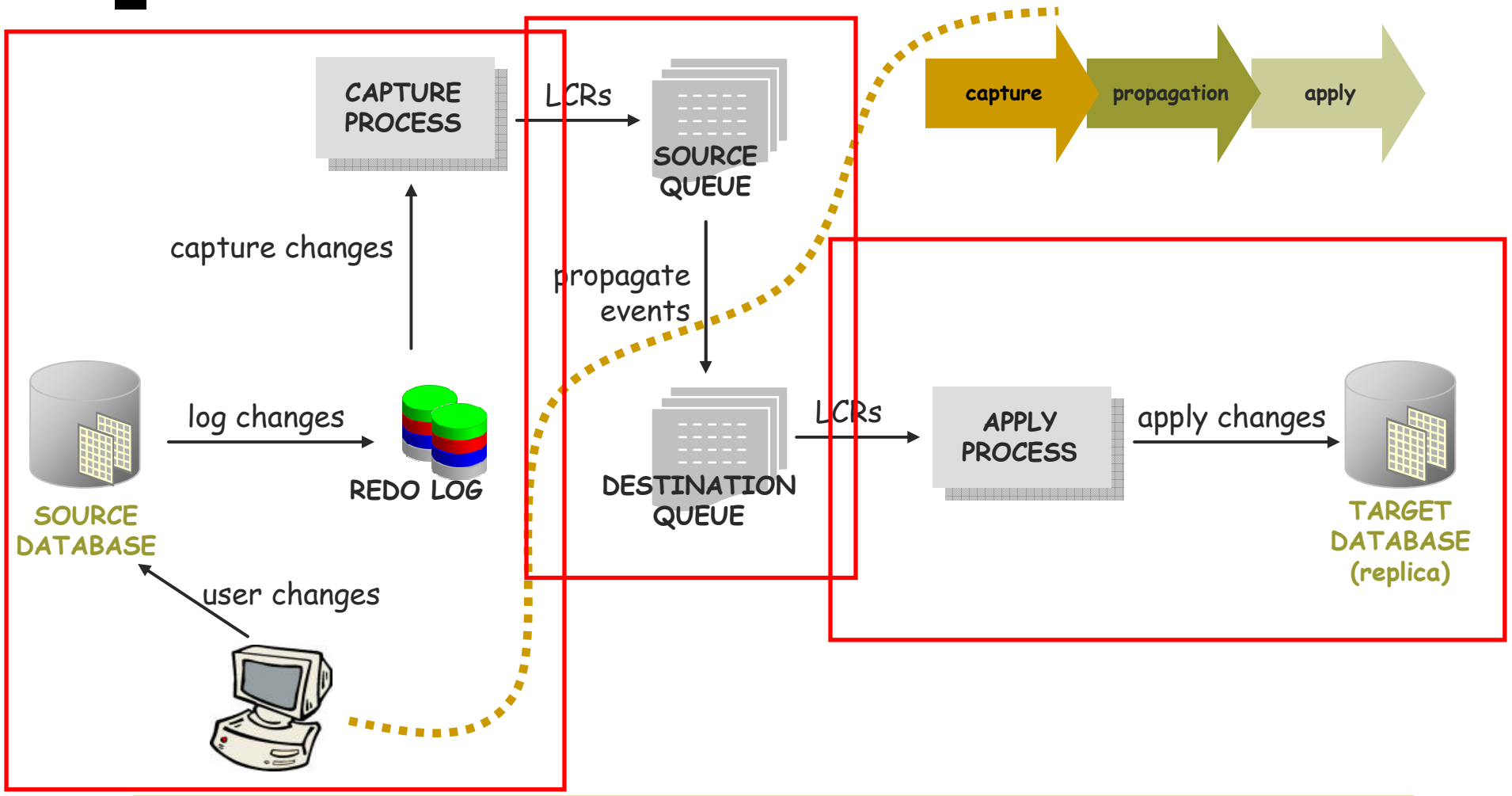
Agenda

- STREAMS overview
- STREAMS on the TESTBED
- Configuration PROBLEMS
- STREAMS log mining configuration survey
  - Downstream capture
  - Split & Merge solution
- STREAMS monitoring
- Throughput TESTS

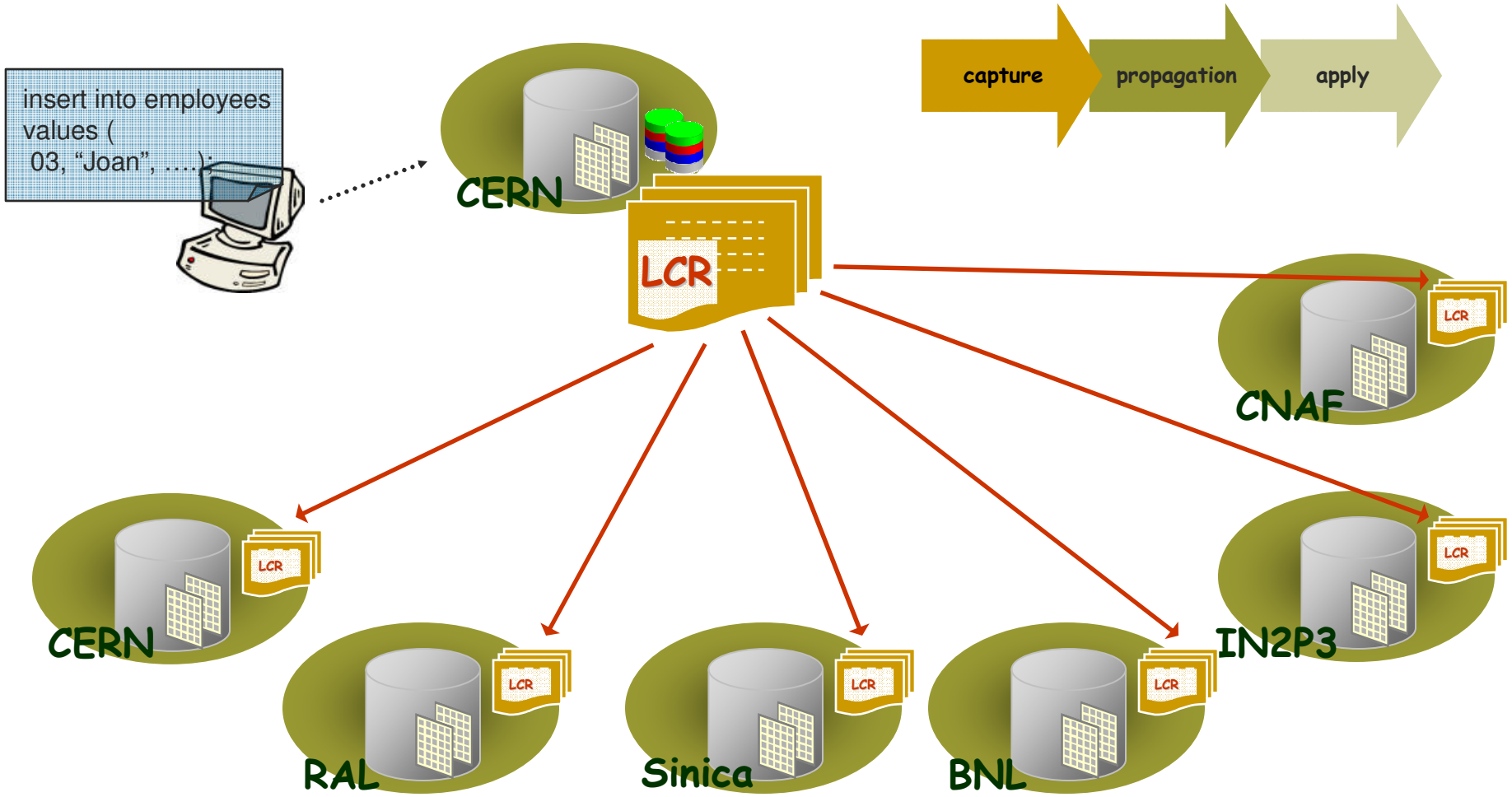
STREAMS Overview

- Flexible feature for information sharing
- Basic elements:
  - Capture
  - Staging
  - Consumption
- Replicate data from one database to one or more databases
- Databases can be non identical copies

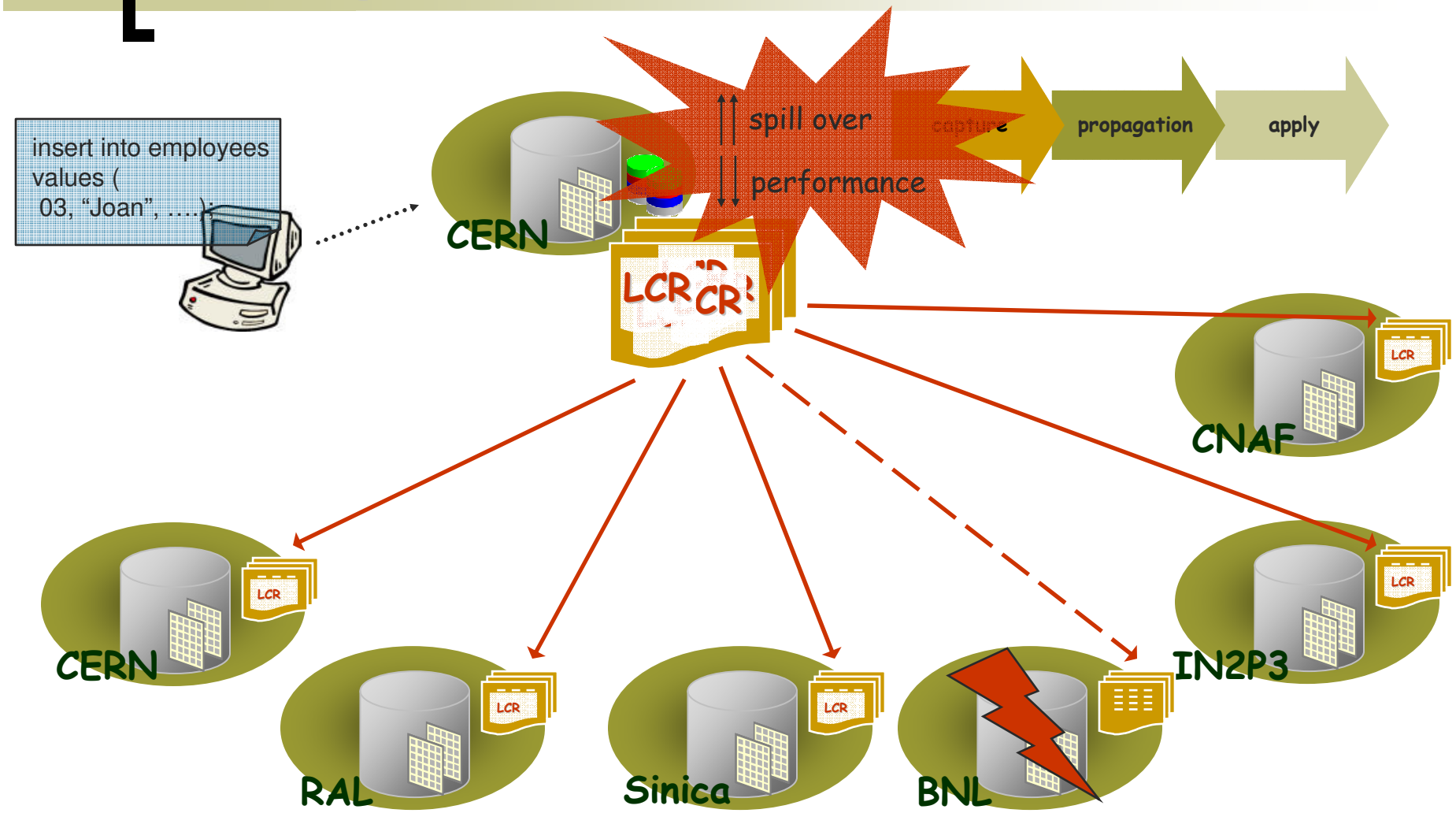
STREAMS Architecture



STREAMS on the TESTBED



Configuration PROBLEMS

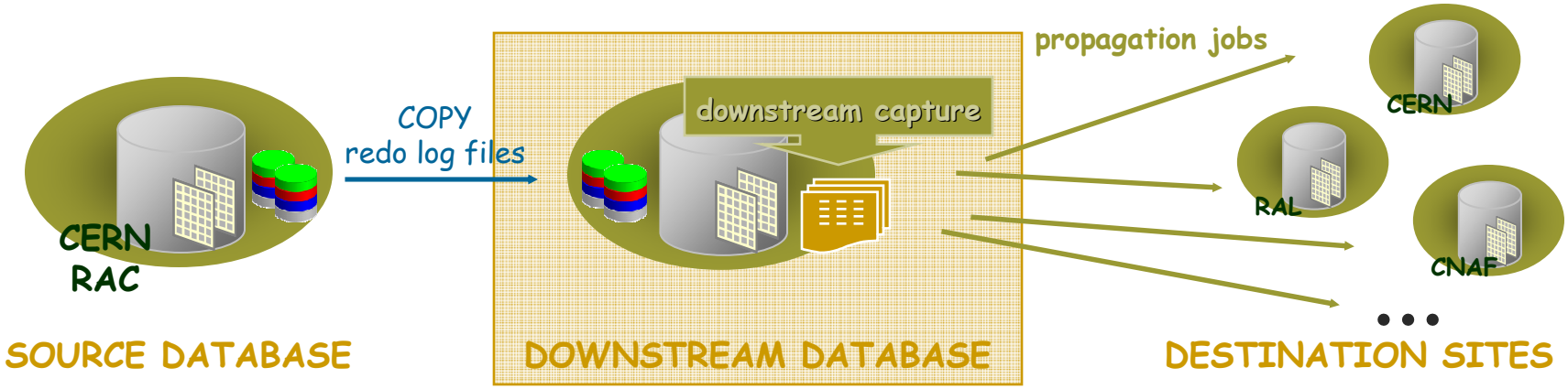


STREAMS log mining survey

- Objectives
  - Source database (CERN) isolation against network or database problems at the replicas
    - Downstream Capture
  - Replica sites isolation against each other
    - split & merge solution

DOWNSTREAM Capture

- Capture process runs on a different database
- Redo log files are copied from source to downstream
- Use of **fewer resources** and data loss protection
- Definite **latency** in the replication process



https://twiki.cern.ch/twiki/bin/view/PSSGroup/DownstreamDatabaseConfiguration

DOWNSTREAM Capture

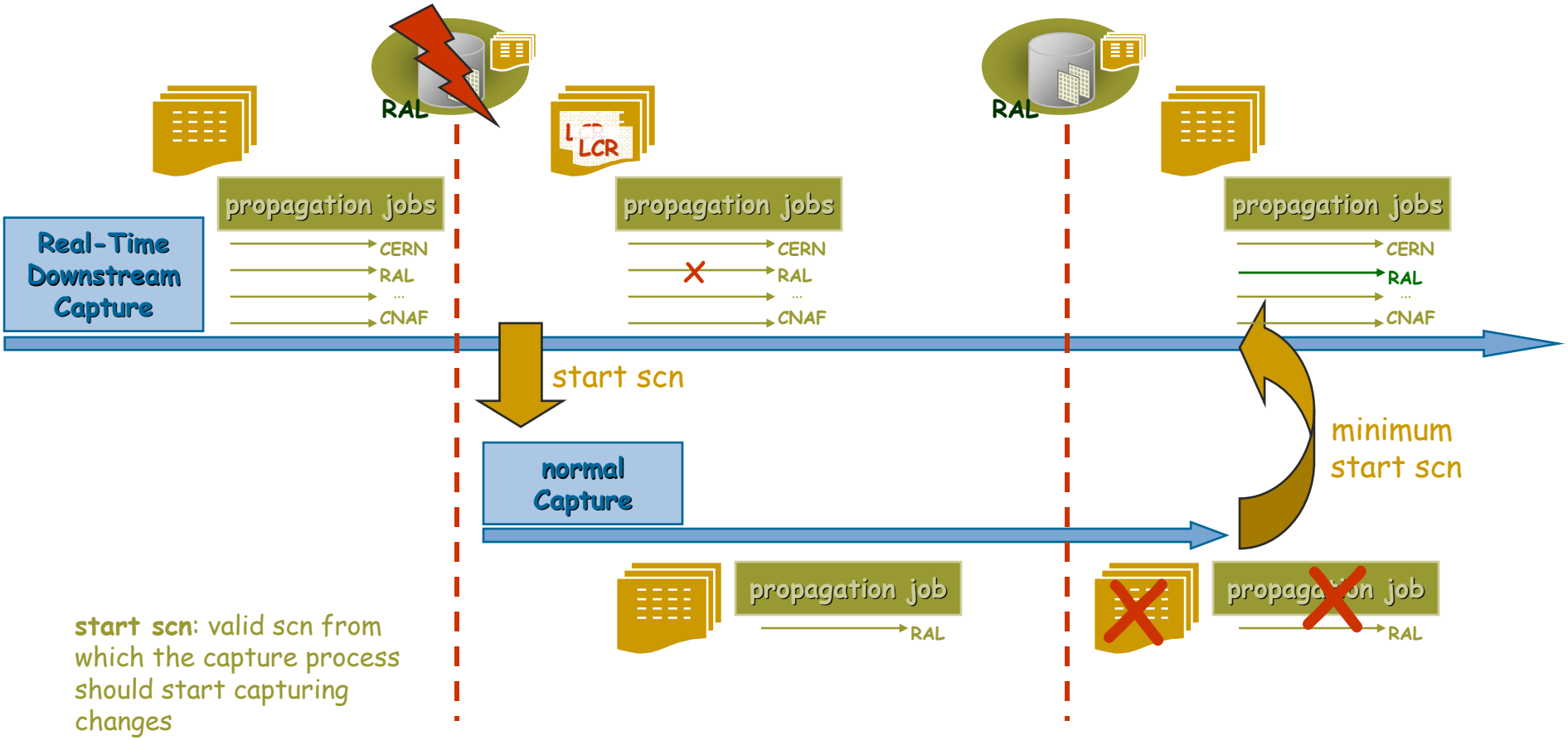
- Real-Time Downstream capture
  - Redo transport services use the LGWR
    - records data in the **online** redo log at the source db
  - The redo data is stored in the standby redo log at the destination db
  - Capture process captures changes from
    - **standby** redo log -> whenever possible
    - **archived** redo log files -> whenever necessary
  - Reduces the amount of time required to capture changes made at the source database

Split & Merge solution

- Split the capture process
  - (original) Real-Time capture for sites "in good shape"
  - (new) normal capture for site/s unstable/s
    - new capture queue and propagation job
    - original propagation job is dropped
      - → spilled LCRs are dropped from the original capture queue
- Merge the capture processes
  - Real-Time capture is used for all the sites
  - Resynchronization

suggested by Patricia McElroy (Principal Product Manager Distributed Systems/Replication)

Split & Merge solution



start scn: valid scn from which the capture process should start capturing changes

with the help of Volker Kuhr (Oracle consulting)

STREAMS monitoring

- "Home-made" scripts
  - capture, propagation and apply status
  - queues status
  - processes statistics
- STRMMON: Oracle Streams monitor tool
  - overview of the Streams activity
- Health Check report
  - information on the setup and operation of Streams

STREAMS monitoring

Display general information about each capture process

Capture Name	Serial Number	ID	Number State	Redo Entries Scanned	Total LCRs Enqueued
STRMADMIN_CAPTURE	C001	136	7 CAPTURING CHANGES	13394731	705854

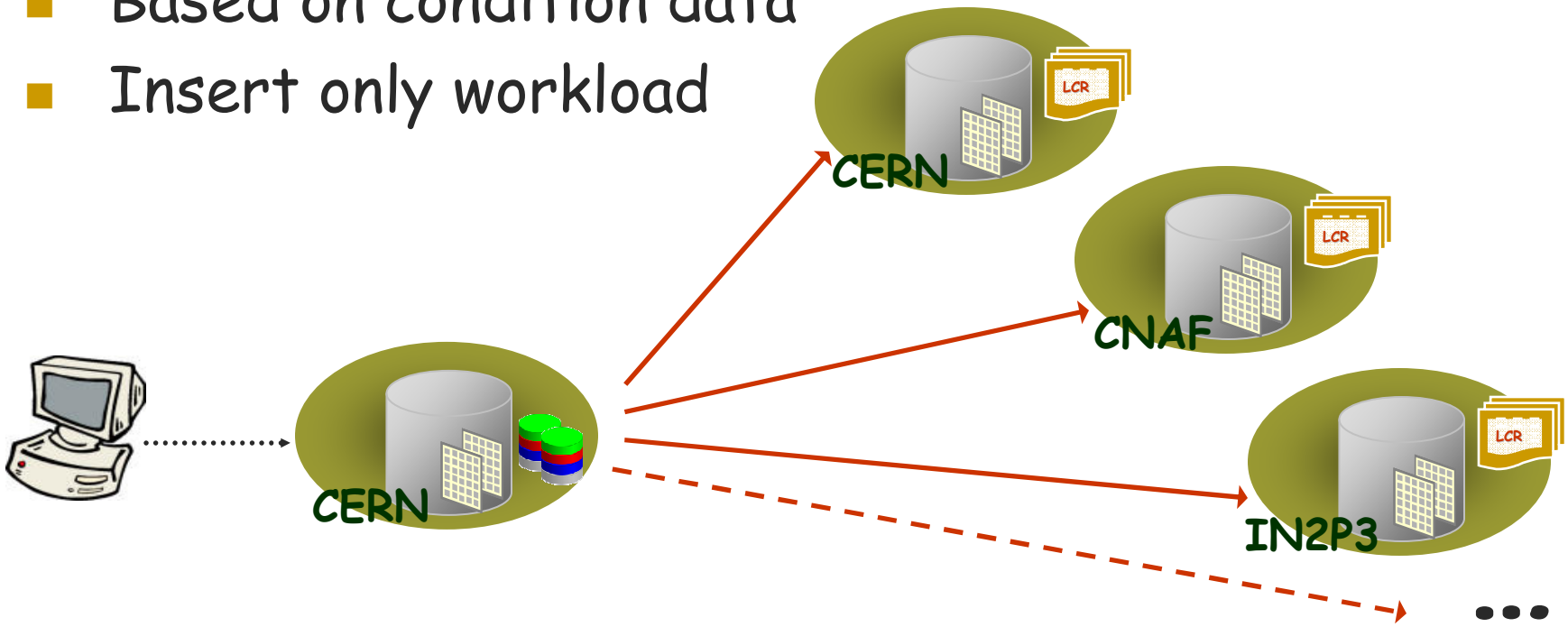
```
STREAMS Monitor, v 2.2  Copyright Oracle Corp. 2002, 2005.
Interval = 3, Count=1000
Logon= @ ORACLE 10.2.0.2.0
Streams Pool Size = 752M

LOG : <redo generated per sec>
NET: <client bytes per sec> <dblink bytes per sec>
Cxxx: <lcrs captured per sec> <lcrs enqueued per sec> <capture latency>
MEM : <percent of memory used> % <streams pool size>
PRxx: <messages received per sec>
Qx  : <msgs enqueued per sec> <msgs spilled per sec>
PSxx: <lcrs propagated per sec> <bytes propaged per sec>
Axxx: <lcrs applied per sec> <txns applied per sec> <dequeue latency>
<F>: flow control in effect
<B>: potential bottleneck
<x%I x%F x%xx>: <idle wait events percentage> <flow control wait events
percentage> <other wait event percentage and name>
xx->: database instance name

2006-06-6 16:25:26 || d3r1-> |      |      | MEM 6 % 752M
2006-06-6 16:25:26 || d3r1-> | LOG 512 | NET 6K 0 | <B> C001 0 0 3sec <0%I 0%F ->  | Q46190 0 0 | PS01 0 0 0 <89%I 0%F ->  | PS02 0 0 0 <0%I 0%F ->  MEM 6 % 752M
2006-06-6 16:25:29 || d3r1-> | LOG 0 | NET 6K 0 | <F> C001 0 0 3sec <0%I 0%F ->  | Q46190 0 0 | PS01 0 0 0 <100%I 0%F ->  | PS02 0 0 0 <0%I 0%F -> MEM 6 % 752M
```

Throughput TESTS

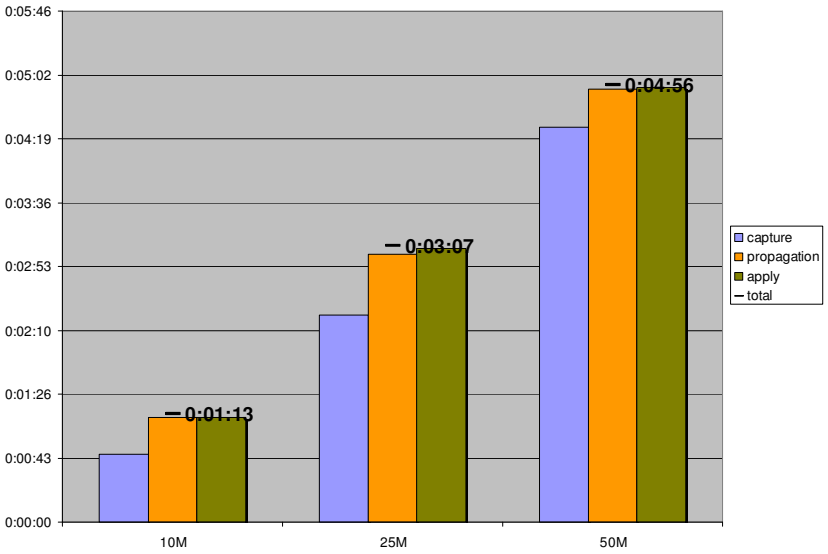
- Script written in python
- Based on condition data
- Insert only workload



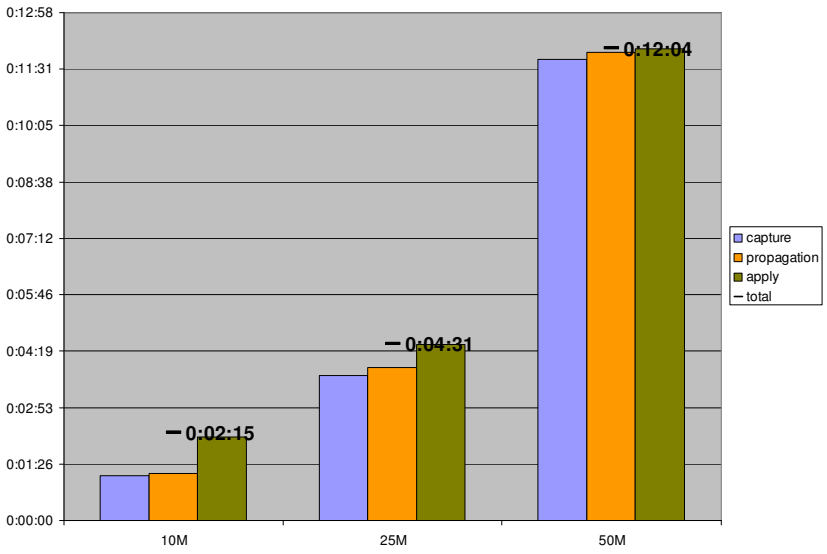
All databases are RAC systems

Throughput TESTS

- CERN to CERN replication



- CERN to CNAF replication



similar results CERN to IN2P3 replication

Throughput TESTS

- Preliminary numbers (~10MB/min)
- Observed
  - Apply process is the bottleneck
    - Apply parallelism does not help
  - Propagation job stops due to FLOW CONTROL
  - Queues @destinations are not sized appropriately
- Improvements together with ORACLE
  - Queue size
  - Processes optimization

in contact with Patricia McElroy (Principal Product Manager Distributed Systems/Replication)

Questions & Answers