GoldenGate 10.4.x

Zbigniew Baranowski

Outline

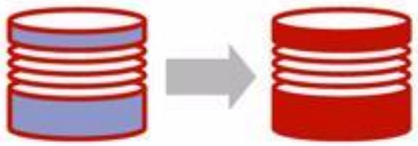
- What is GoldenGate?
- Architecture
- Monitoring
- Performance
- GoldenGate vs. Streams
- Summary

What is GoldenGate?

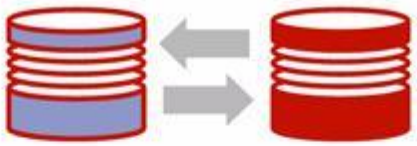
- Independent project started in mid 90s
- Aim: real-time data integration solutions
- Continuous data synchronization across heterogeneous environments
  - Oracle, DB2, SQL Server, MySQL and more
- purchased by Oracle in late 2009

Data integration using GG

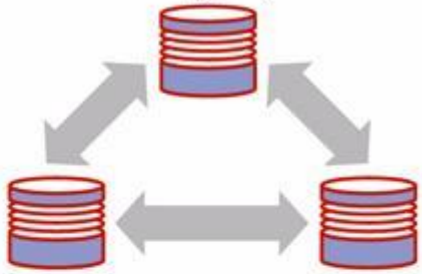
UNIDIRECTIONAL
Reporting Instance



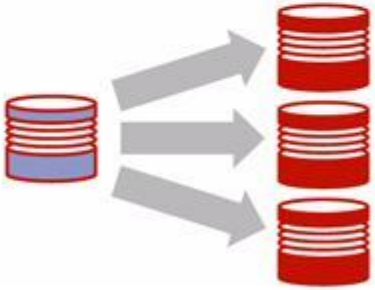
BI-DIRECTIONAL
Instant Failover "Active-Active"



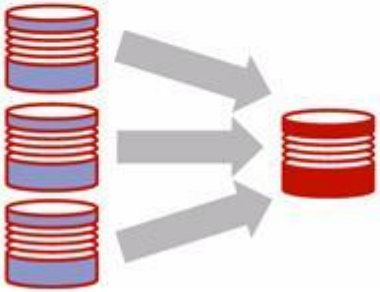
PEER-TO-PEER
Load Balancing, High Availability



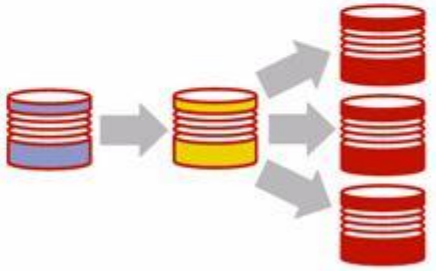
BROADCAST
Data Distribution



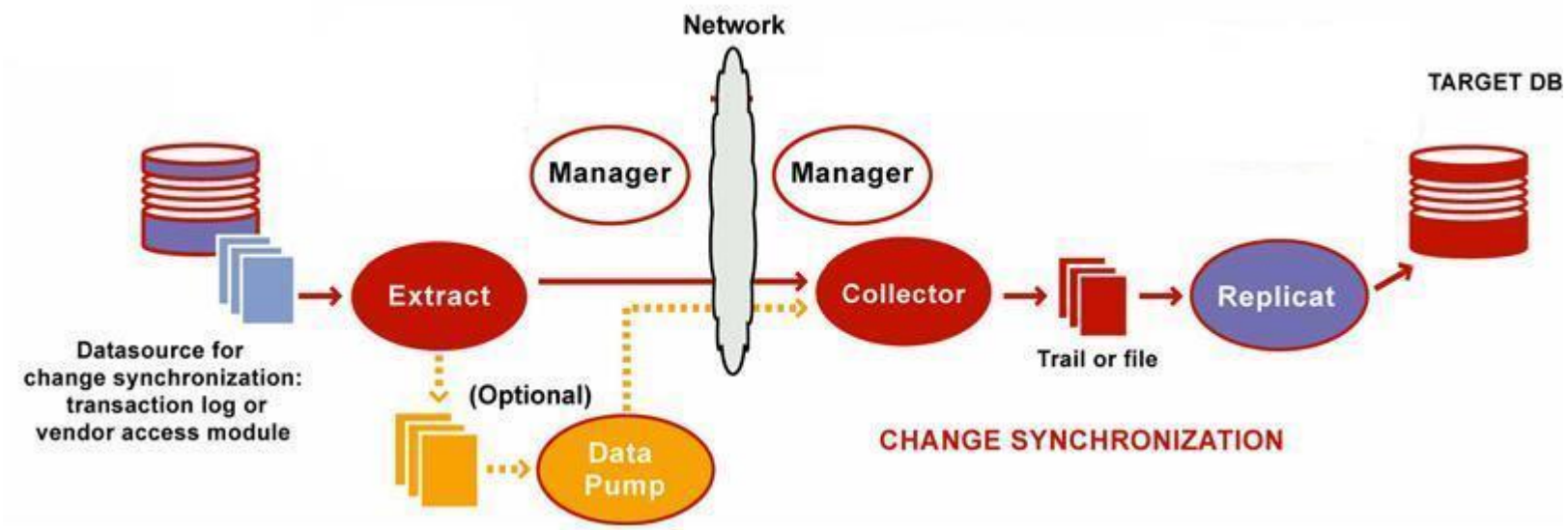
CONSOLIDATION
Data Warehouse/Mart/Store



CASCADING
Scalability, Database Tiering

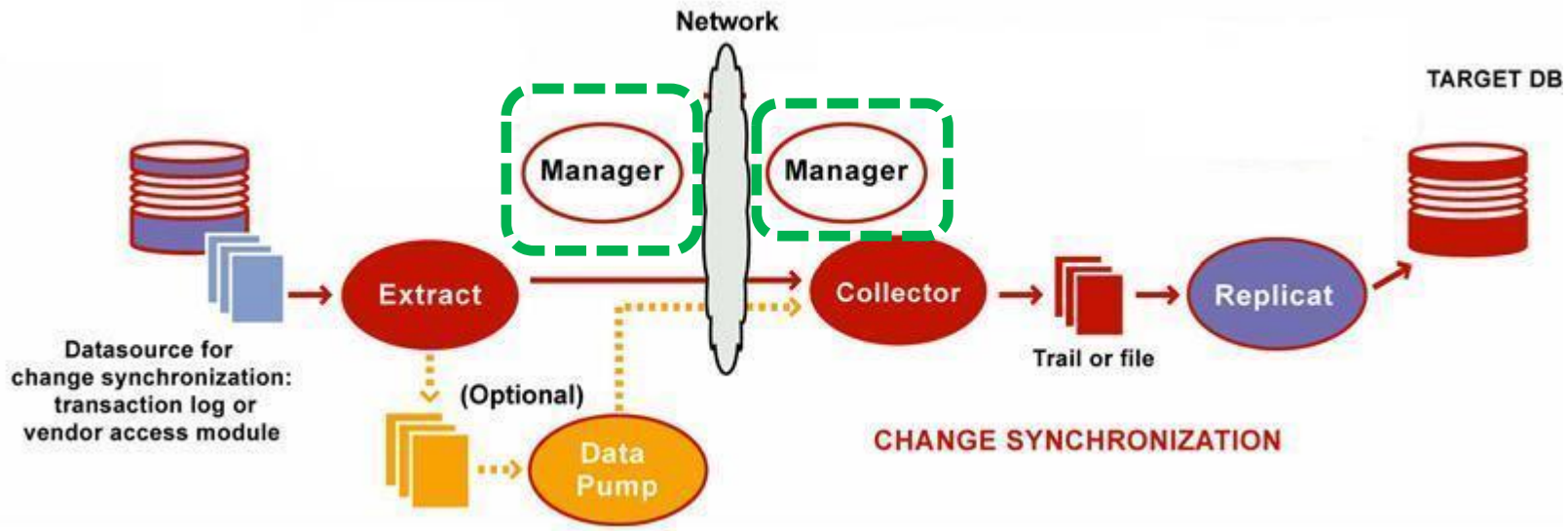


GoldenGate architecture



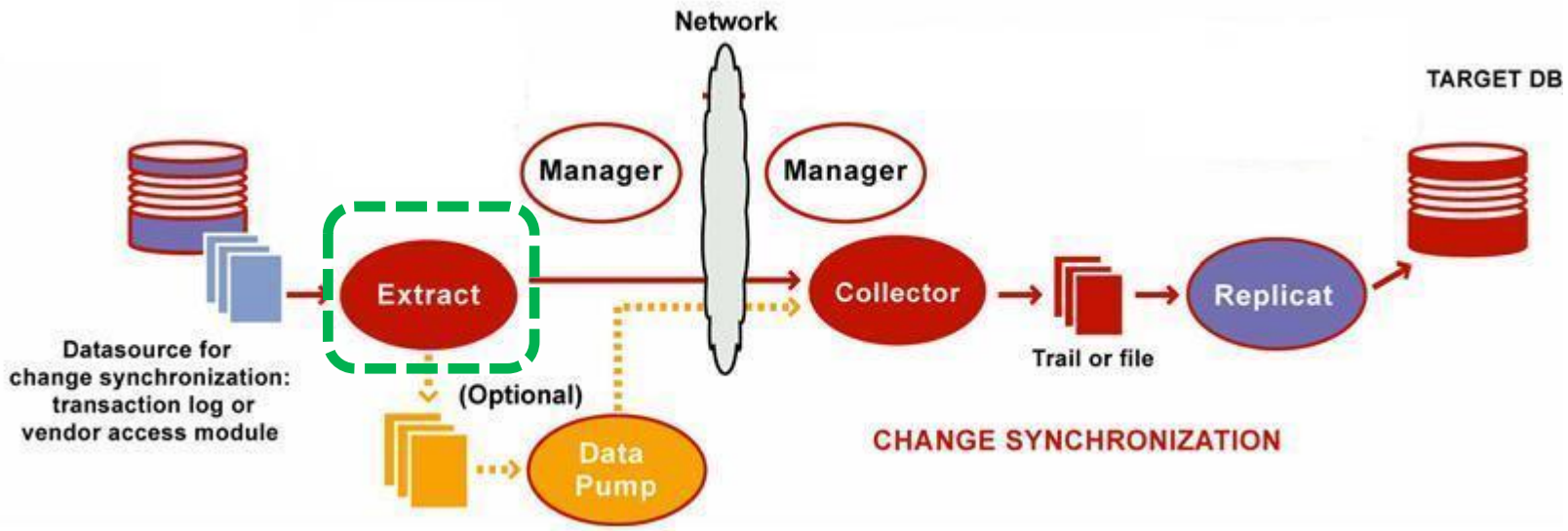
GoldenGate components

- Manager
  - Runs on both databases
  - Starts and monitors GG processes
  - Manages trail files
  - Reporting



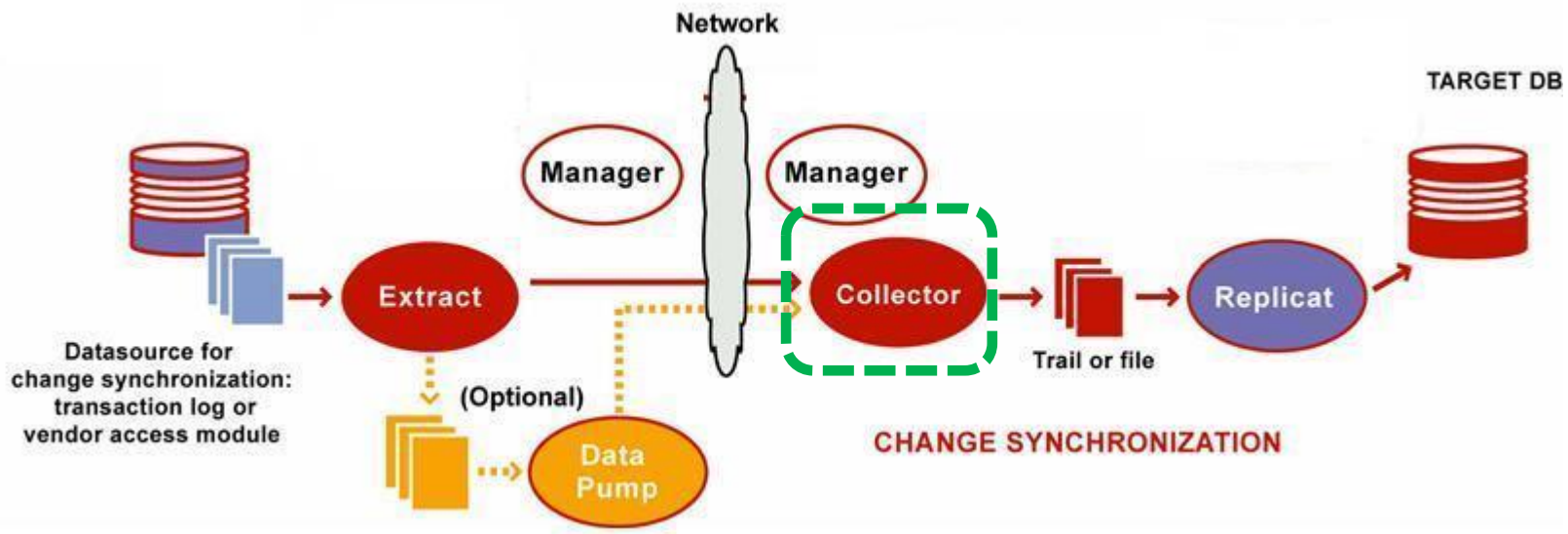
GoldenGate components

- Extract (capture process)
  - Runs on the source database
  - extracts data changes from redo logs
  - Writes transactional and DDL changes (in common GG format) to trail files



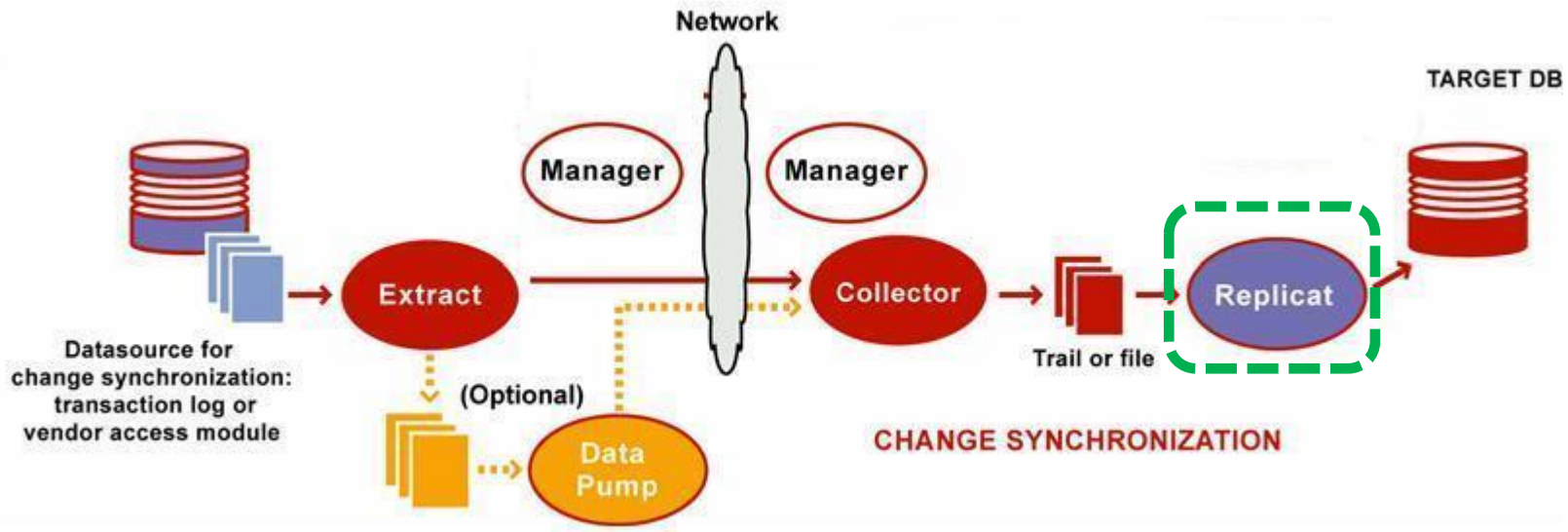
GoldenGate components

- Collector
  - Background process which runs on the target system
  - Two modes: static and dynamic (managed by manager)
  - Receives data changes from TCP/IP and writes to trail files



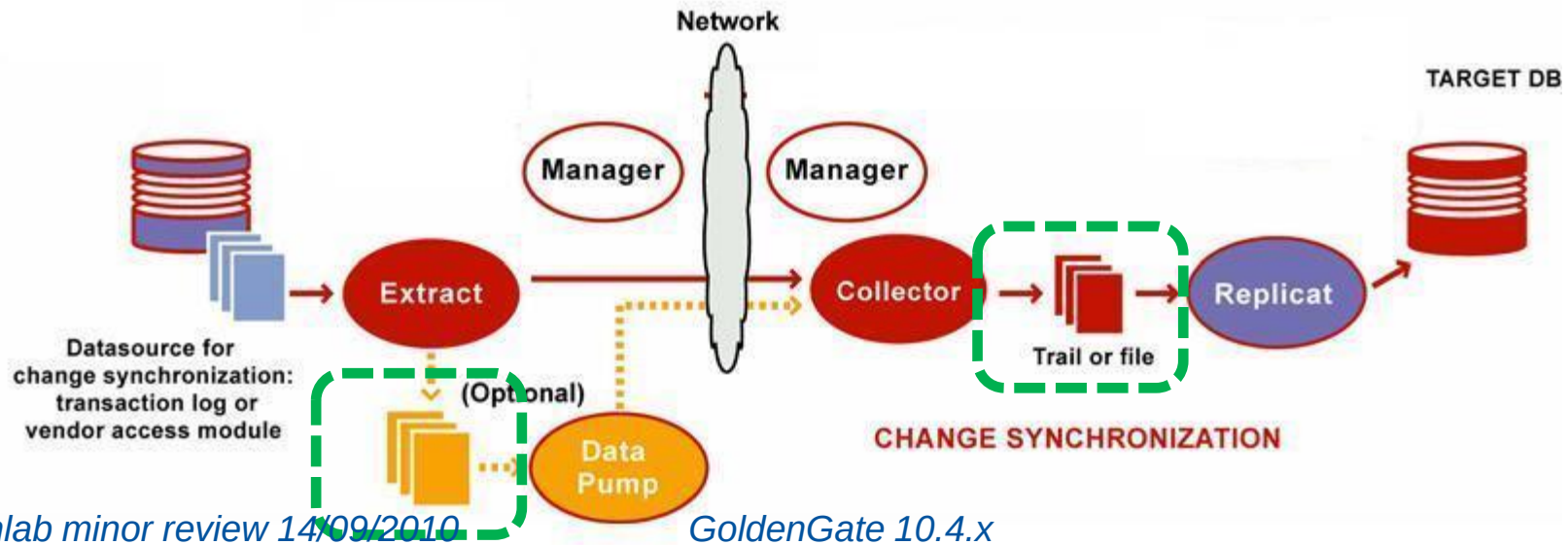
GoldenGate components

- Replicat (apply process)
  - Runs on the target system
  - Reads transactional data from trail files and applies them on target database
  - Can by parallelized



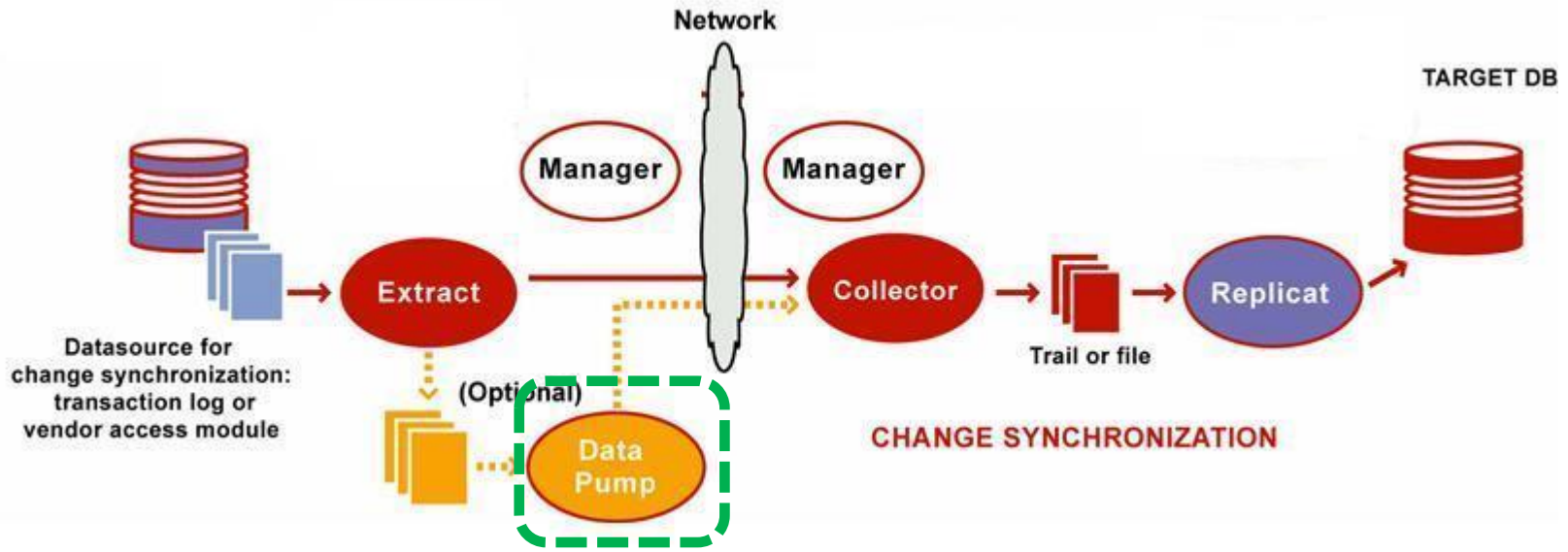
GoldenGate components

- Trail files
  - Contains data changes written in GG common format
  - Series of files that GoldenGate temporarily stores on disks
  - Two types
    - Extract Trail (located on source - optional)
    - Remote Trail (located on destination)

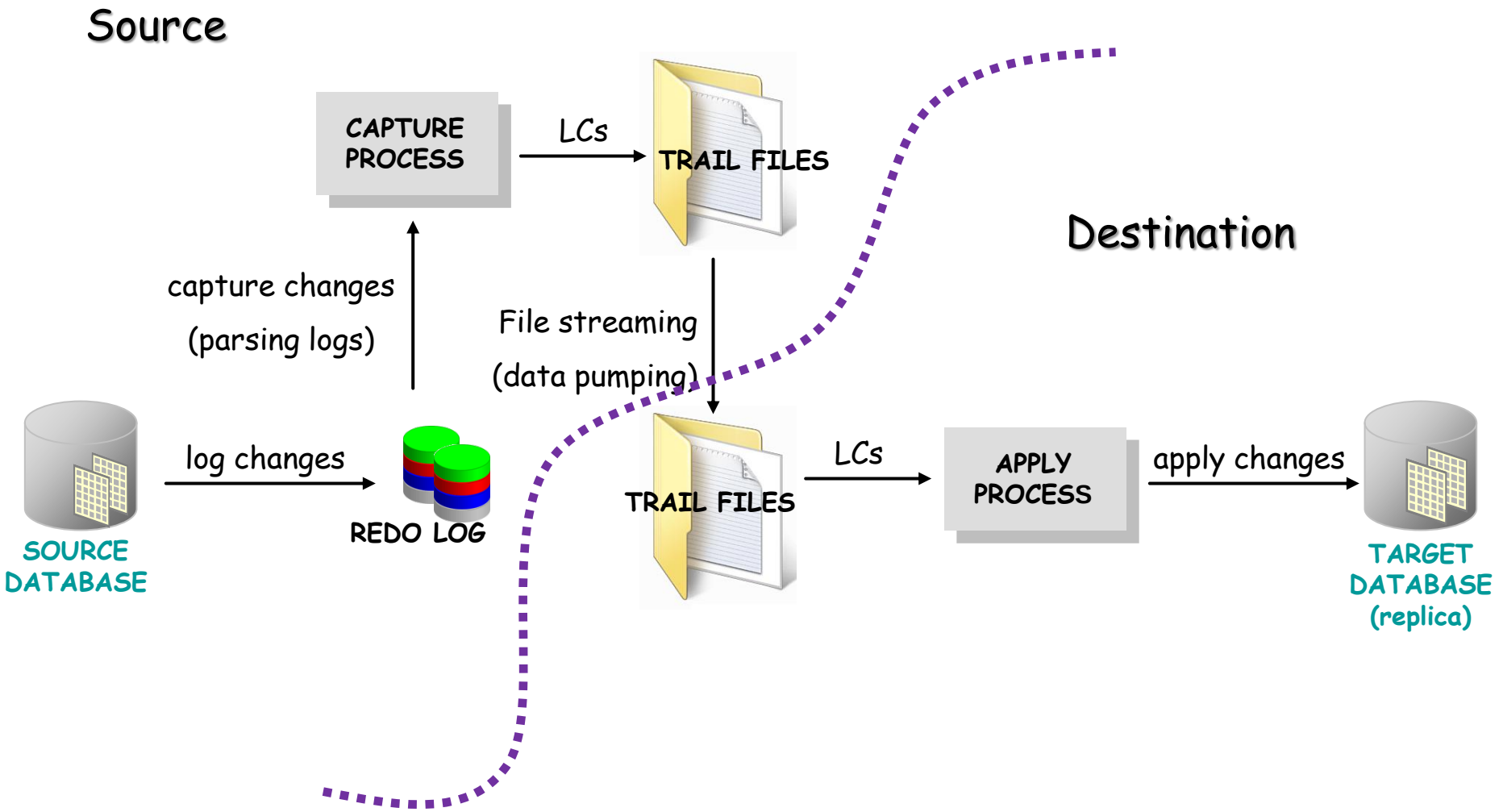


GoldenGate components

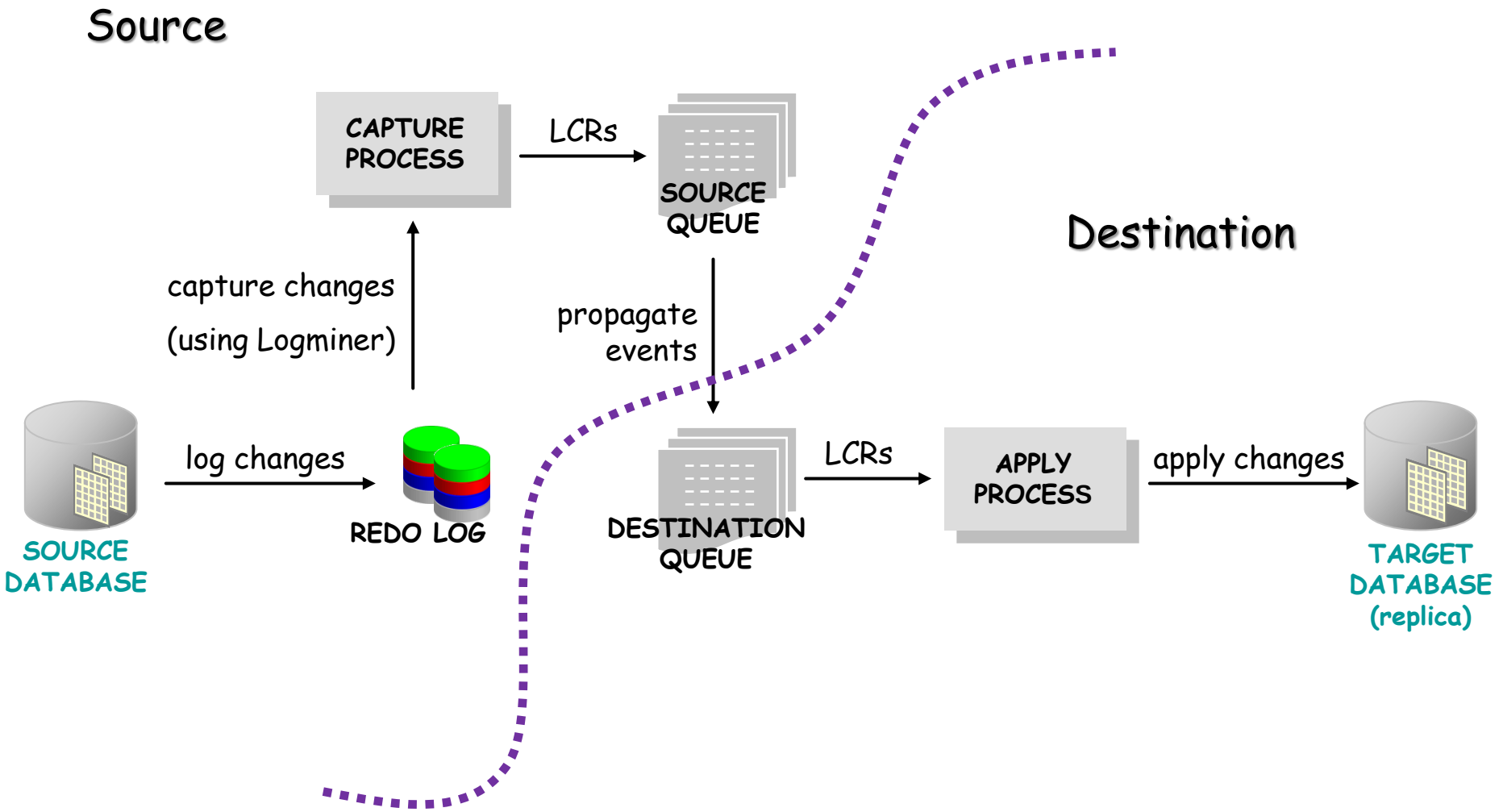
- Data pump
  - Sends data changes from extract trail via TCP/IP to remote trail on the target
  - Optional component
    - Extract process can sends changes directly to remote trail
  - Optional encryption and compression over TCP/IP



GoldenGate architecture (overview)



Streams architecture (overview)



GoldenGate vs Streams (setting up)

	Streams	GoldenGate
Installation	•Embedded in Oracle database	•Unzipping binaries
Management	•Using SQL or EM	•Mainly with GGSCI command tool •Director application
Replication setup	•TNS + database links config •Executions of few procedures	•Editing parameter files + executions of few command •Additional port required
RAC environment	•No additional steps required	•Shared storage, CRS configuration, additional parameters
Schemas selection	•Setting rules for process – SQL procedure	•Editing mapping files
Supervisor account	•STRMADMIN – owner of processes, jobs, queues and links •Executes transactions on target	•GGADMIN – keeps metadata about replicated schema •Executes transactions on target
Process watch	•None	•Manager process

GoldenGate vs Streams (performance)

	Streams	GoldenGate
Bottleneck	•Capture process – log mining	•Delivery process – limitation of single session resources
Parallelism of apply / delivery	•By setting process's parameter •Apply coordinator takes care about serialization	•By manual addition of processes and specification of filters •No coordinator process – no serialization guaranteed
Big parallel transactions	•Serialized capture	•Parallel capture
Replication processes impact on system	•Quite significant	•Minimal
Potential improvements	•None	•Parallelism of schemas + BatchSQL + compression
Stability	•Not stable. A lot of aborts due to database memory issues •Hangs due to session blocking	•Very stable

GoldenGate vs Streams (functionality and maintenance)

	Streams	GoldenGate
Initial load	•Data pump	•Initial load configuration
Monitoring	•SQL + EM + Custom tools	•Director
Administration	•All task can be done through SQLPlus •No direct access to machines required	•Using Director (most of tasks) •Using GGSCI requires direct access to node
Error handling	•Errors well documented •Errors expressed in database language •All error handling procedures fully understood	•Errors not well documented and not fully understandable •Handling procedures not recognized yet
Data replication (DDL+DML)	•Replication of data definitions and modifications	•Special handling of data definitions. •Only delivery without parallelism can guarantee smooth replication
Specific data handling	•Apply handlers	•None

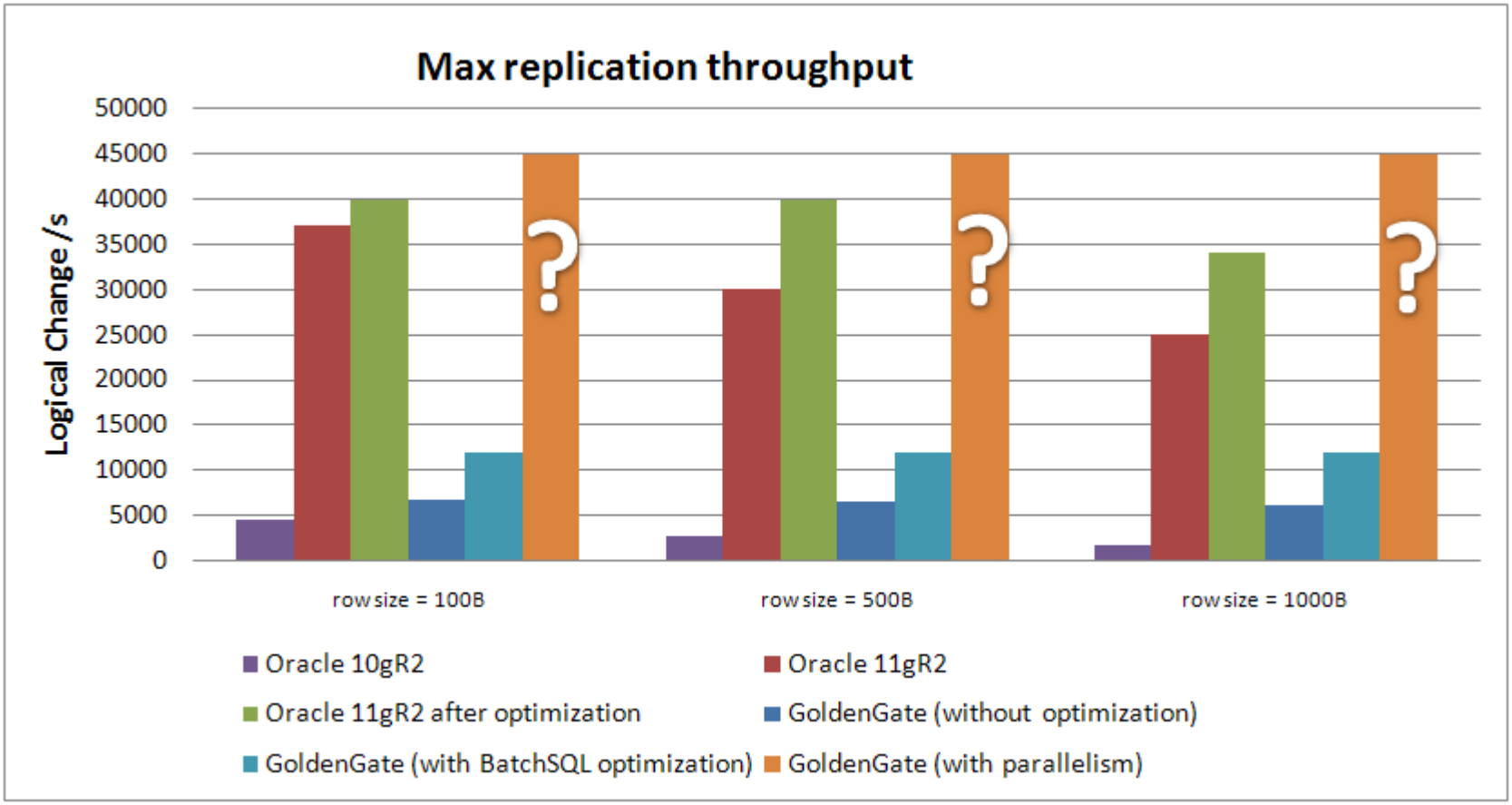
GoldenGate vs Streams (functionality and maintenance)

	Streams	GoldenGate
Schemas versioning	•Multi Version Object Dictionary	•Single Version Object Dictionary •Potential problems with recapturing of older changes
Resuming of replication	•Slow - capture process needs to re-init dictionary from last checkpoint	•Very fast •Using checkpoint files
Failing transactions	•Mechanism for re-execution of failed transaction	•No extra mechanism. •Rollback the sequence change number (SCN) and restart the delivery process

Performance

- Simple workload (inserts, updates, deletes)
  - With basic configuration (no parallelism): 7 LCRs /s
  - With BatchSQL optimization 12 LCRs /s and can be more
  - With parallelism: at least 15 LCRs /s (each parallel process increases throughput linearly)
- COOL workload
  - BatchSQL does not improve performance
- **Delivery process is the bottlenecks** due to limits of resource utilization by single session.

Where are we?


Max replication throughput
Logical Change /s
50000
45000
40000
35000
30000
25000
20000
15000
10000
5000
0
?
?
?
row size = 100B
row size = 500B
row size = 1000B
Oracle 10gR2
Oracle 11gR2
Oracle 11gR2 after optimization
GoldenGate (without optimization)
GoldenGate (with BatchSQL optimization)
GoldenGate (with parallelism)

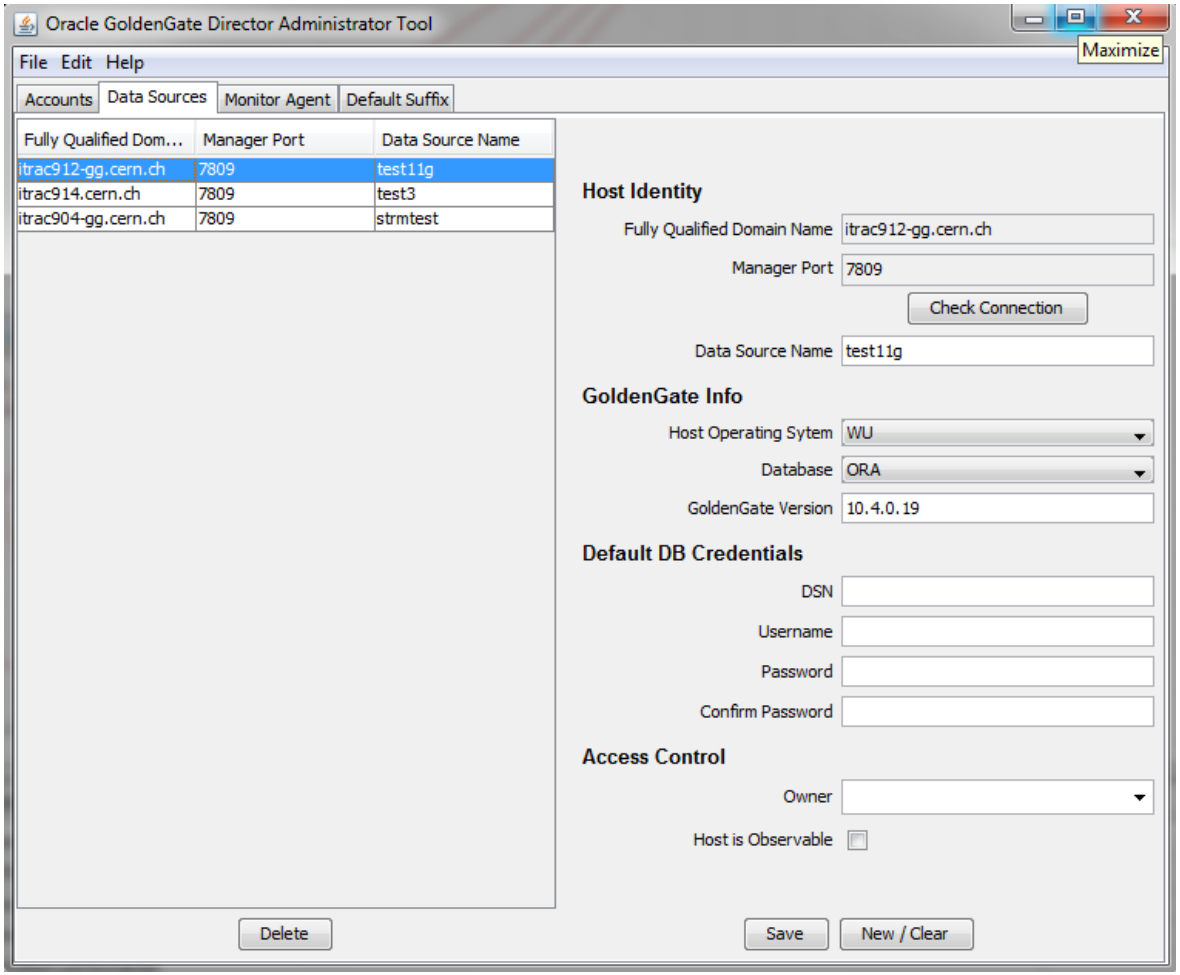
Discovered problems

- Supplemental logging groups are created to late to replicate the following DML operations
  - Workaround: add global supplemental logging for primary and foreign keys
- Cannot use create as select – select statement is executed on the destination database as GGADMIN user
- Delivery parallelism: lack of DMLs and DDLs synchronization – problem with temporal tables

Monitoring

- 3 applications
  - Admin tool (defines manager location)
  - Director web application (status, logs, start/stop, email notifications)
  - Director desktop application (status, logs, start/stop, email notifications, replication configuration and management, setup topology)
- Written in JAVA
- Requires Java 1.6 SDK and Weblogic server
- Wraps command line interface

Admin tool


Oracle GoldenGate Director Administrator Tool
Maximize
File Edit Help
Accounts
Data Sources
Monitor Agent
Default Suffix
Fully Qualified Dom...
Manager Port
Data Source Name
itrac912-gg.cern.ch 7809 test11g
itrac914.cern.ch 7809 test3
itrac904-gg.cern.ch 7809 strmtest
Host Identity
Fully Qualified Domain Name
itrac912-gg.cern.ch
Manager Port
7809
Check Connection
Data Source Name
test11g
GoldenGate Info
Host Operating Sytem
WU
Database
ORA
GoldenGate Version
10.4.0.19
Default DB Credentials
DSN
Username
Password
Confirm Password
Access Control
Owner
Host is Observable
Delete
Save
New / Clear

Director - web app

ORACLE GoldenGate

Network

- Entire Network
- test11g
  - D_STRM2
  - P_TEST
  - D_STRM
  - C_TEST
  - D_STRM4
  - D_STRM3
  - D_STRM5
- test3
  - D_TEST
- strmtest
  - C_STRM
  - P_STRM
- Watch Lists
  - Configure Watch Lists
- My Email Alerts
- Live Monitoring is ON
- Log Out

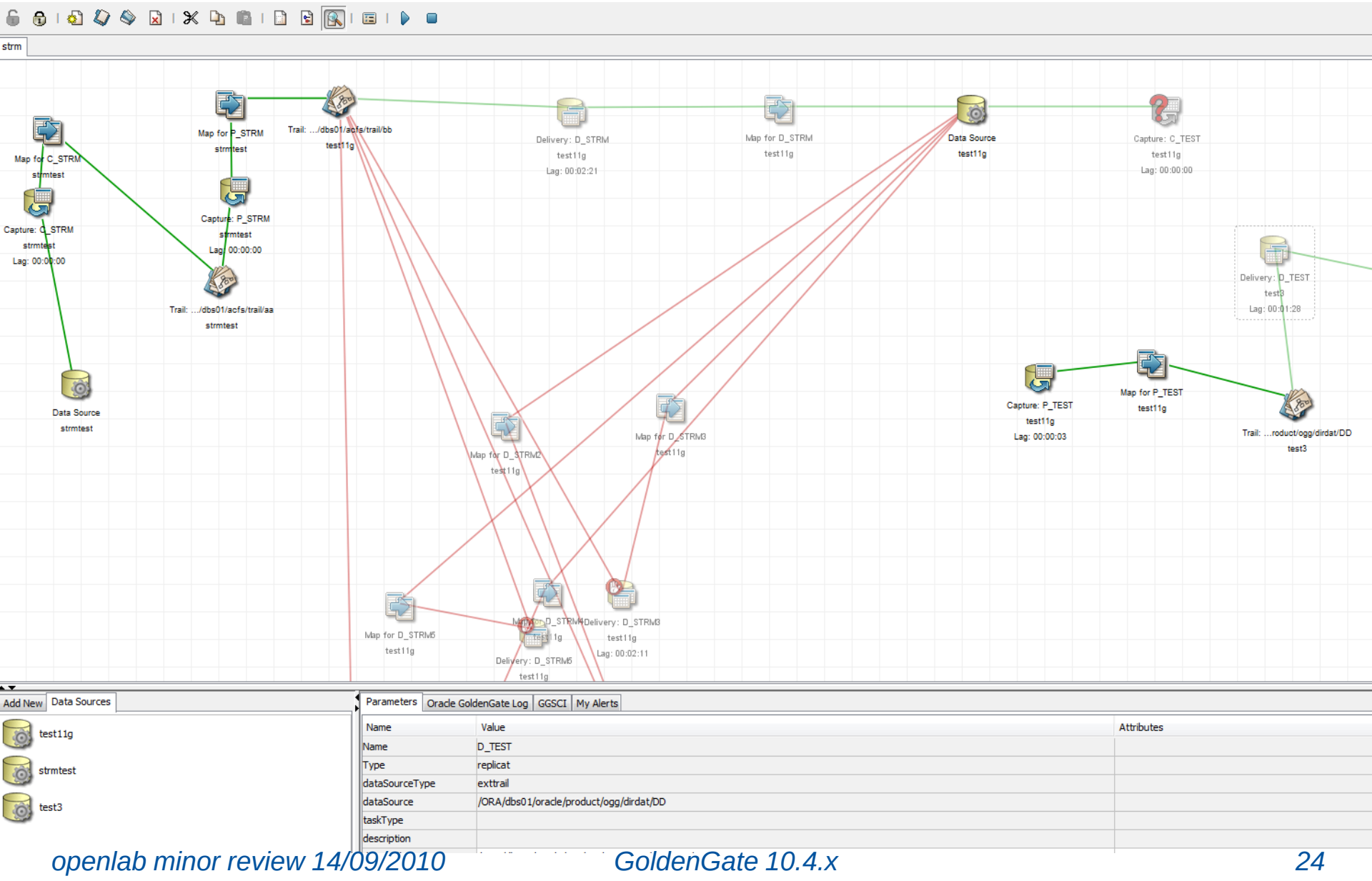
Processes | Tasks

Manager	Process Name	Type	Status	Lag	Data Source - Type
test11g (WU - ORA)					
	D_STRM2	Deliver	Abended	00:02:10	...1/acfs/trail/bb - GGS Log Trail
		More Info... Last Record: 6/16/10 2:55:03 PM	Last Status: 6/16/10 2:57:13 PM		
GGSCI	P_TEST	Capture	Running	00:00:00	...1/acfs/trail/CC - GGS Log Trail
		More Info... Last Record: 6/18/10 3:45:13 PM	Last Status: 6/18/10 4:01:42 PM		
	D_STRM	Deliver	Running	00:00:00	...1/acfs/trail/bb - GGS Log Trail
		More Info... Last Record: 6/18/10 3:42:50 PM	Last Status: 6/18/10 4:01:45 PM		
	C_TEST	Capture	Running	00:00:00	...1.268.721324923 - Oracle RAC Redo Log
		More Info... Last Record: 6/18/10 4:01:42 PM	Last Status: 6/18/10 4:01:43 PM		
	D_STRM4	Deliver	Abended	00:02:12	...1/acfs/trail/bb - GGS Log Trail
		More Info... Last Record: 6/16/10 2:55:03 PM	Last Status: 6/16/10 2:57:15 PM		
	D_STRM3	Deliver	Abended	00:02:11	...1/acfs/trail/bb - GGS Log Trail
		More Info... Last Record: 6/16/10 2:55:03 PM	Last Status: 6/16/10 2:57:14 PM		
	D_STRM5	Deliver	Abended	00:02:12	...1/acfs/trail/bb - GGS Log Trail
		More Info... Last Record: 6/16/10 2:55:03 PM	Last Status: 6/16/10 2:57:15 PM		
test3 (WU - ORA)					

Event Log

Event	Manager	Process	Type	Date/Time
INFO	strmtest	MANAGER	MANAGER	6/18/10 3:58:05 PM
	302 - Oracle GoldenGate Manager for Oracle, mgr.prm: Manager started 'ggsci' process on port 7844.			
INFO	strmtest	MANAGER	MANAGER	6/18/10 3:58:05 PM
	301 - Oracle GoldenGate Manager for Oracle, mgr.prm: Command received from GUI on host 137.138.132.166 (START GGSCI).			
INFO	test11g	P_TEST	CAPTURE	6/18/10 3:45:02 PM
	224 - Oracle GoldenGate Capture for Oracle, p_test.prm: Rolling over remote file /ORA/dbs01/oracle/product/ogg/dirdat/DD001200.			
INFO	test11g	C_TEST	CAPTURE	6/18/10 3:45:02 PM
	224 - Oracle GoldenGate Capture for Oracle, c_test.prm: Rolling over remote file /ORA/dbs01/acfs/trail/CC001205.			
INFO	test3	PORT 7858	SERVER	6/18/10 3:45:02 PM
	370 - Oracle GoldenGate Collector, port 7858: Opening /ORA/dbs01/oracle/product/ogg/dirdat/DD001200 (byte -1, current EOF 0).			
INFO	test3	PORT 7858	SERVER	6/18/10 3:45:02 PM
	371 - Oracle GoldenGate Collector, port 7858: Closing /ORA/dbs01/oracle/product/ogg/dirdat/DD001199.			
INFO	test11g	P_TEST	CAPTURE	6/18/10 3:44:26 PM
	224 - Oracle GoldenGate Capture for Oracle, p_test.prm: Rolling over remote file /ORA/dbs01/oracle/product/ogg/dirdat/DD001199.			
INFO	test11g	C_TEST	CAPTURE	6/18/10 3:44:26 PM
	224 - Oracle GoldenGate Capture for Oracle, c_test.prm: Rolling over remote file /ORA/dbs01/acfs/trail/CC001204.			
INFO	test3	PORT 7858	SERVER	6/18/10 3:44:26 PM
	370 - Oracle GoldenGate Collector, port 7858: Opening /ORA/dbs01/oracle/product/ogg/dirdat/DD001199 (byte -1, current EOF 0).			
INFO	test3	PORT 7858	SERVER	6/18/10 3:44:26 PM
	371 - Oracle GoldenGate Collector, port 7858: Closing /ORA/dbs01/oracle/product/ogg/dirdat/DD001198.			

Director – desktop app



Monitoring

- Intuitive interface layout (fast information display – no nested views)
- Access to process statistics and reports
- Easy troubleshooting
- Email notification
  - Lag (floods mail box – sends every minute)
  - Event (errors, warnings)
- Difficult to monitor many replication setups
- From time to time problems with refreshing
- Not all operations works after first attpempt (retries are needed)
- No history data
- No plots except lags

Conclusions

- External software
  - Administration from command line tool
  - Setup require more afford (attaching to DB)
  - Not a black box
  - Minimal resource utilization
  - Minimum impact on the database
- Golden Gate is a very stable, efficient.
- Technology focused on data modification operations
  - CERN's data profile is definitions + modifications

Conclusions

- Performance (safe mode) better than Streams10g
  - There is big potential of improvements but we need to avoid ‘driving without breaks’
- Errors not well documented
- Lack of experience (administration)
  - Errors handling
- Poor monitoring
- **Strategic replication solution for Oracle**

Future plans

- Per schema delivery parallelism – performance
- Real PVSS's and Condition's data replication tests
- Establishing contact with GG development team
  - Arranged meeting at Oracle Open World next week
- Evaluation new features on GG 11g (new version just has been released)

GoldenGate Keeps Evolving

- 2013 update:
- Next Generation GoldenGate vs Streams for Physics Data
  - OGG 12.1.2 comparison with Streams
  - Focus on Integrated Capture and Delivery features
- We are planning to migrate from Streams to GG12c next year (2014)