

GoldenGate

Zbigniew Baranowski

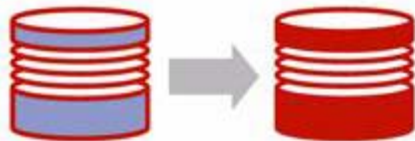
- What is GoldenGate?
- Architecture
- Performance
- GoldenGate vs. Streams
- Monitoring
- Summary

- Real-time data integration solutions
- Continuous data synchronization across heterogeneous environments
 - Oracle, DB2, SQL Server, MySQL and more
- Project started in early 90s
- Purchased by Oracle in late 2009

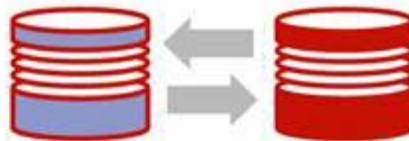
- Need of stable and reliable replication service
- Streams require frequent interventions (at least once per week)
 - Blocking sessions
 - Memory pools shortage
 - Human errors
- Streams administration is difficult
 - T1s administrators need our help
- **Streams project has been abandoned by Oracle**

UNIDIRECTIONAL

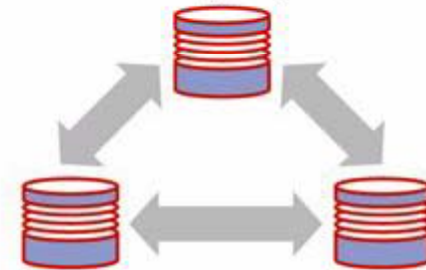
Reporting Instance

**BI-DIRECTIONAL**

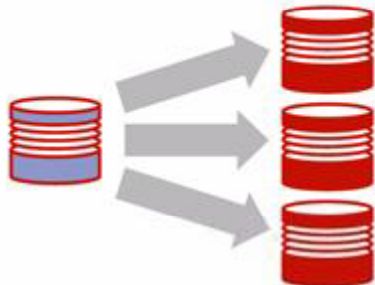
Instant Failover "Active-Active"

**PEER-TO-PEER**

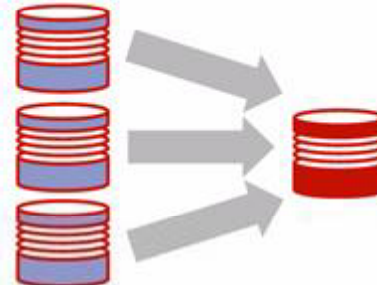
Load Balancing, High Availability

**BROADCAST**

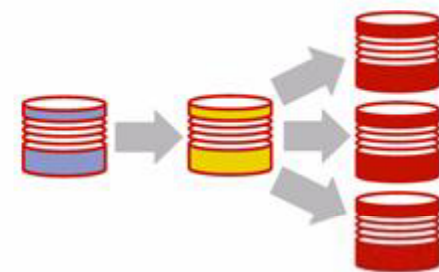
Data Distribution

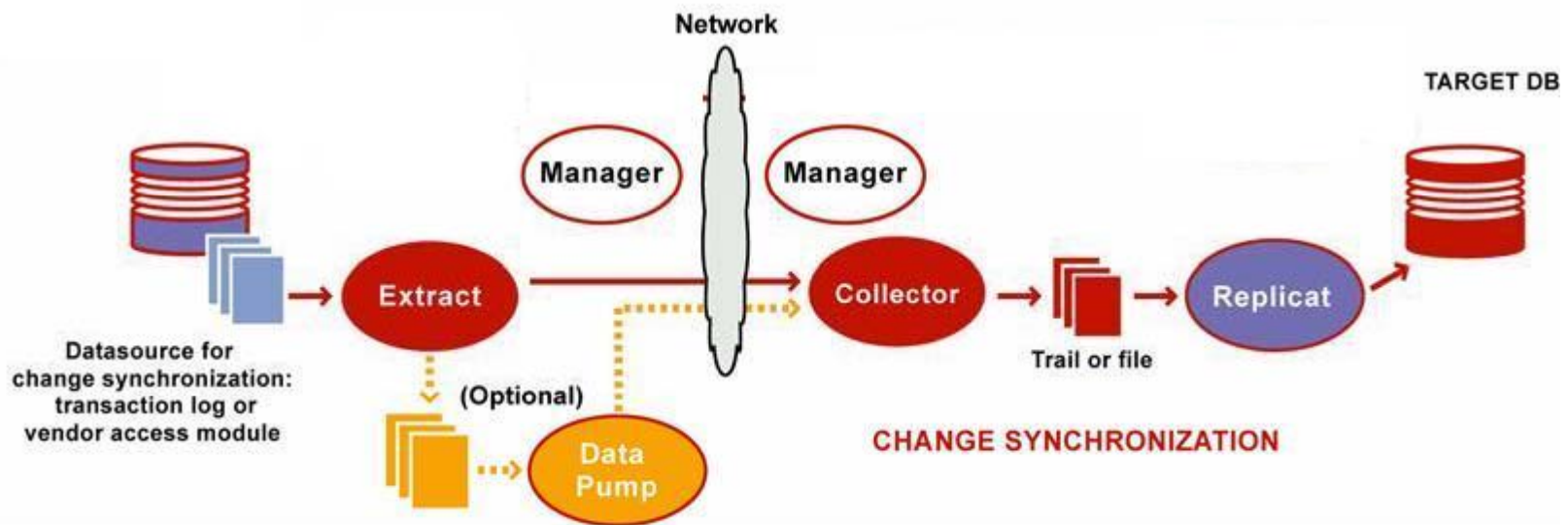
**CONSOLIDATION**

Data Warehouse/Mart/Store

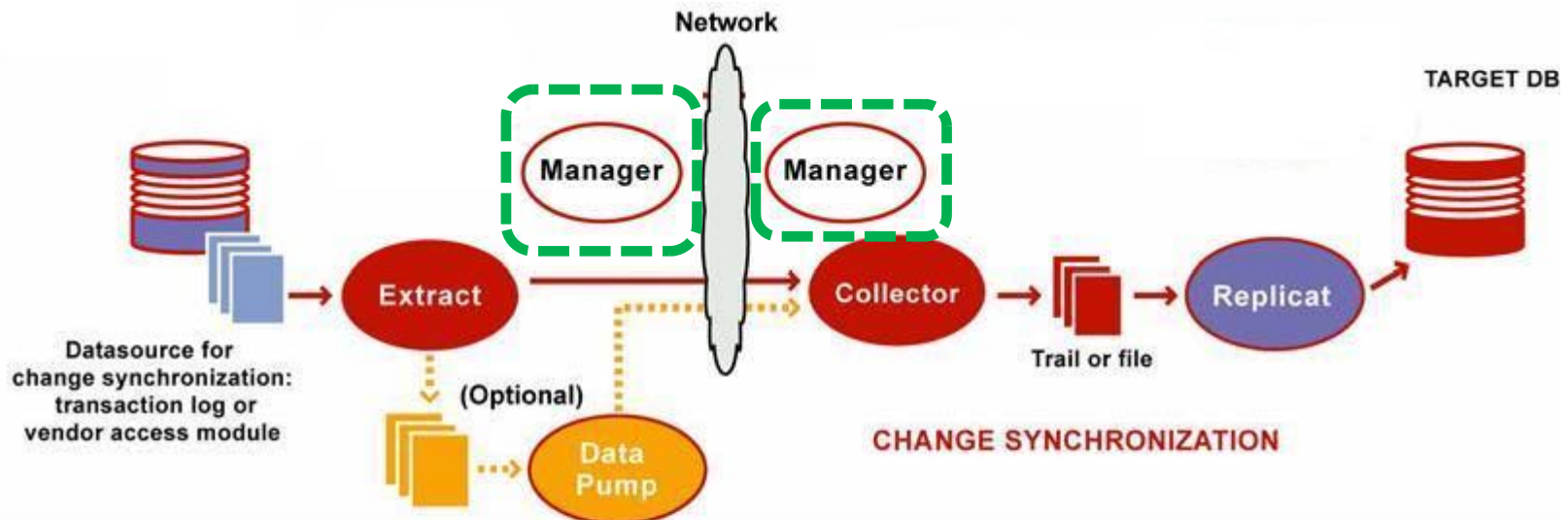
**CASCADING**

Scalability, Database Tiering

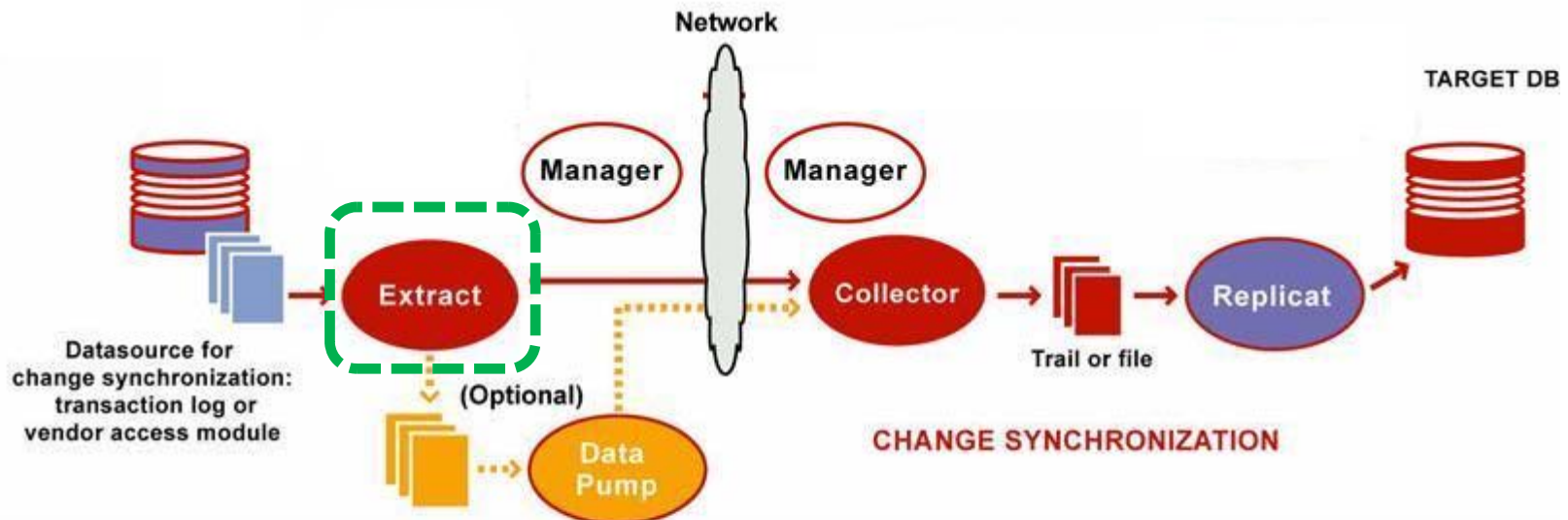




- Manager
 - Runs on both databases
 - Starts and monitors GG processes
 - Manages trail files
 - Reporting

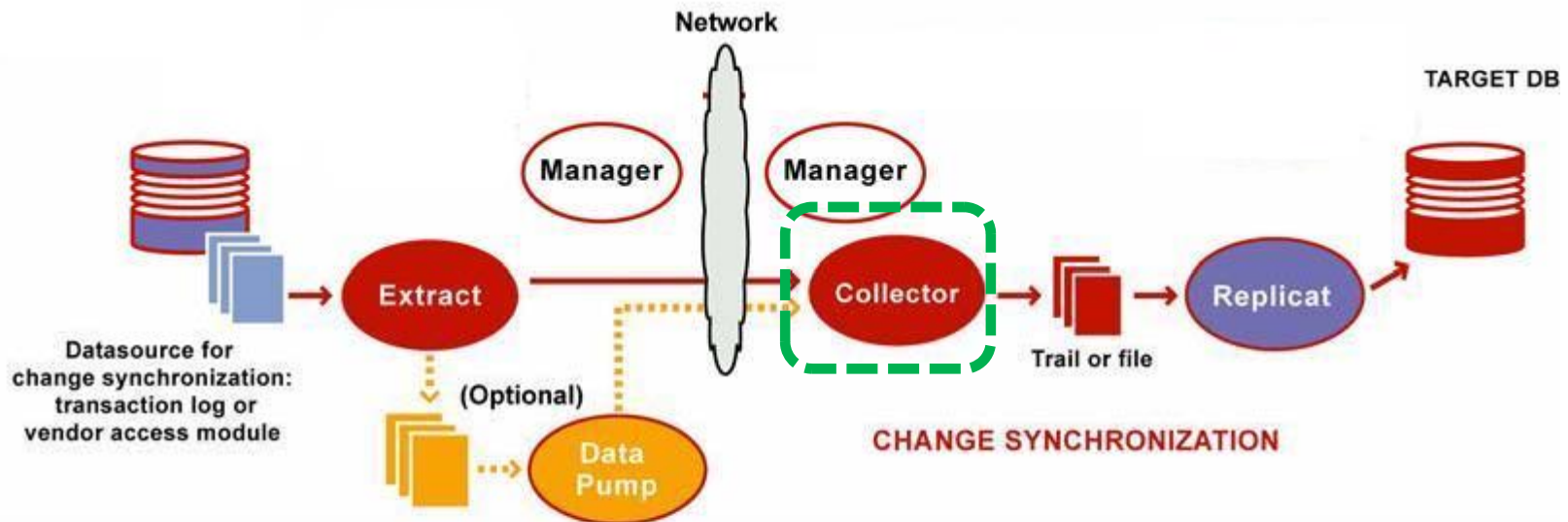


- Extract (capture process)
 - Runs on the source database
 - Extracts data changes from redo logs
 - Writes transactional and DDL changes (in common GG format) to trail files

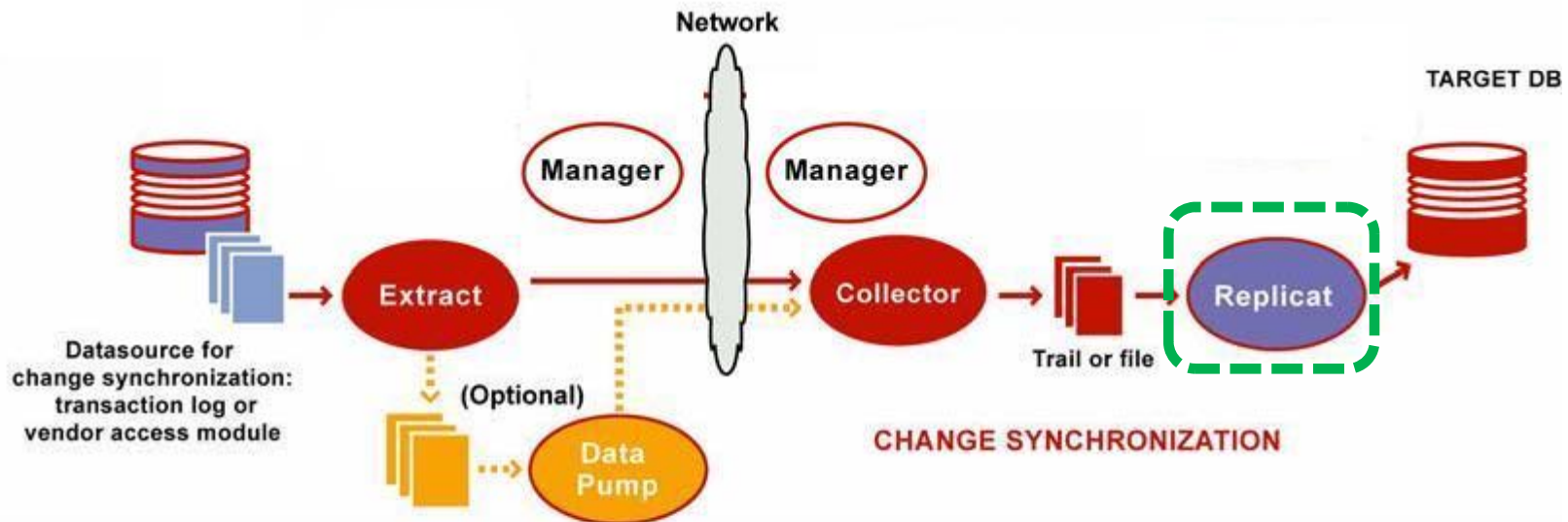


- Collector

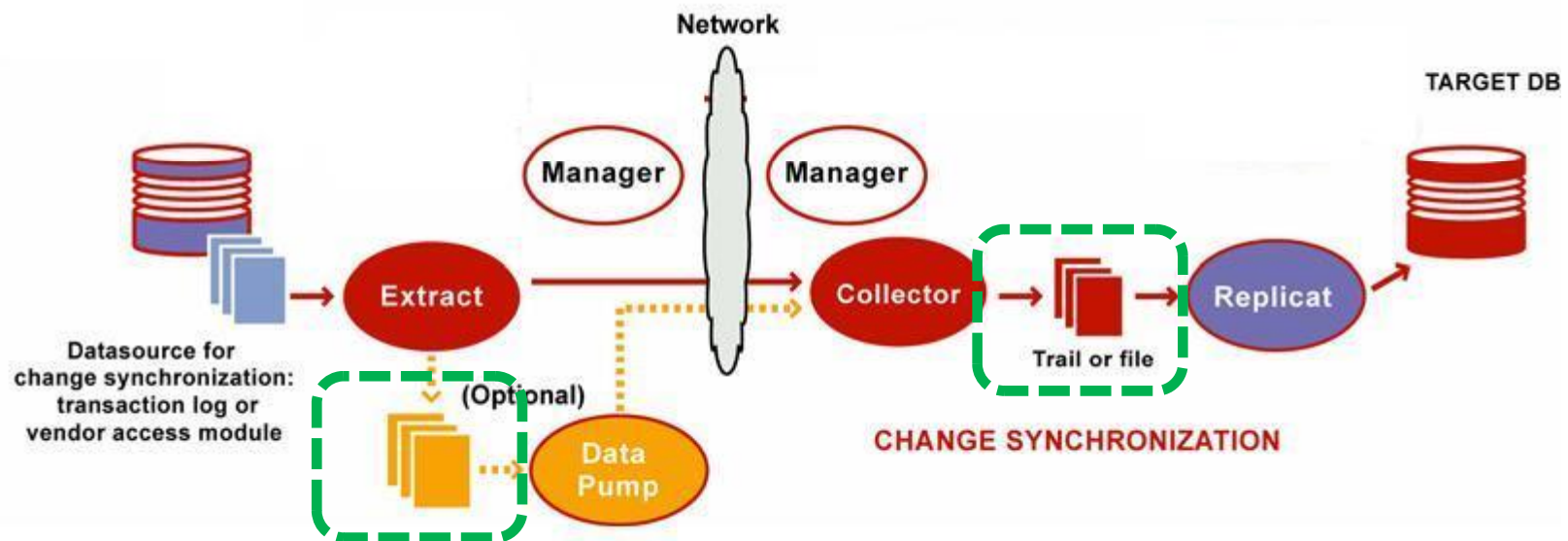
- Background process which runs on the target system
- Two modes: static and dynamic (managed by manager)
- Receives data changes from TCP/IP and writes to remote trail files



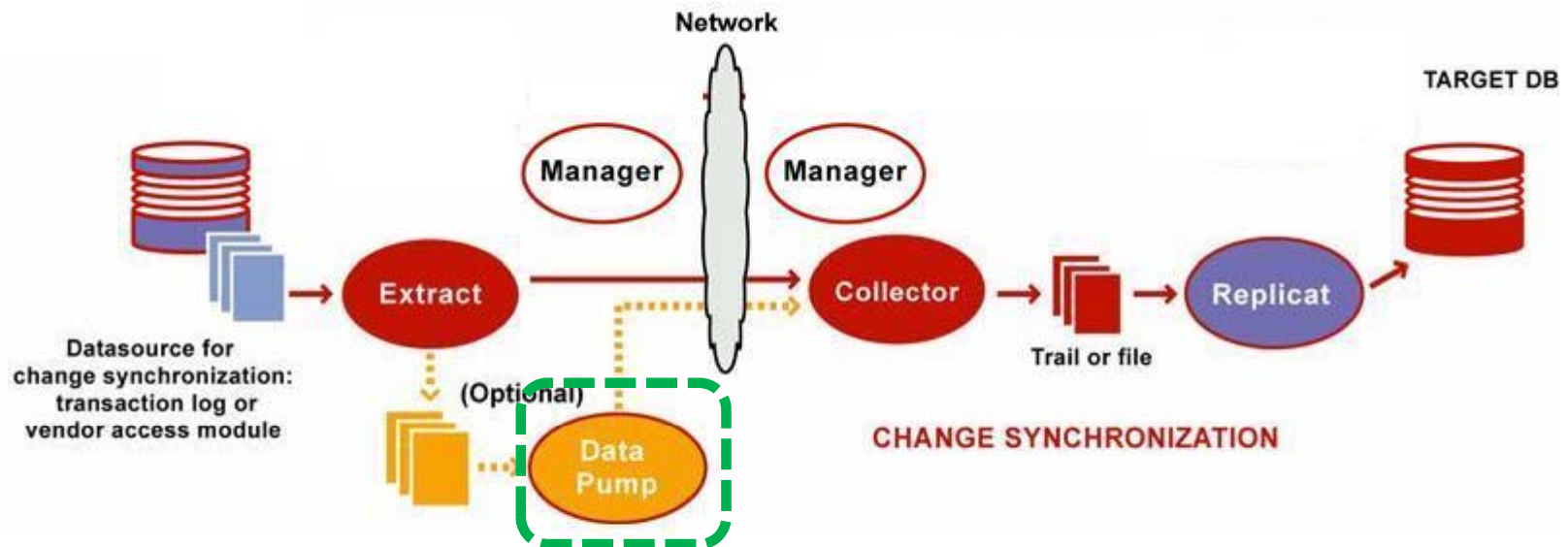
- Replicat (apply process)
 - Runs on the target system
 - Reads transactional data from trail files and applies them on target database
 - Can be parallelized

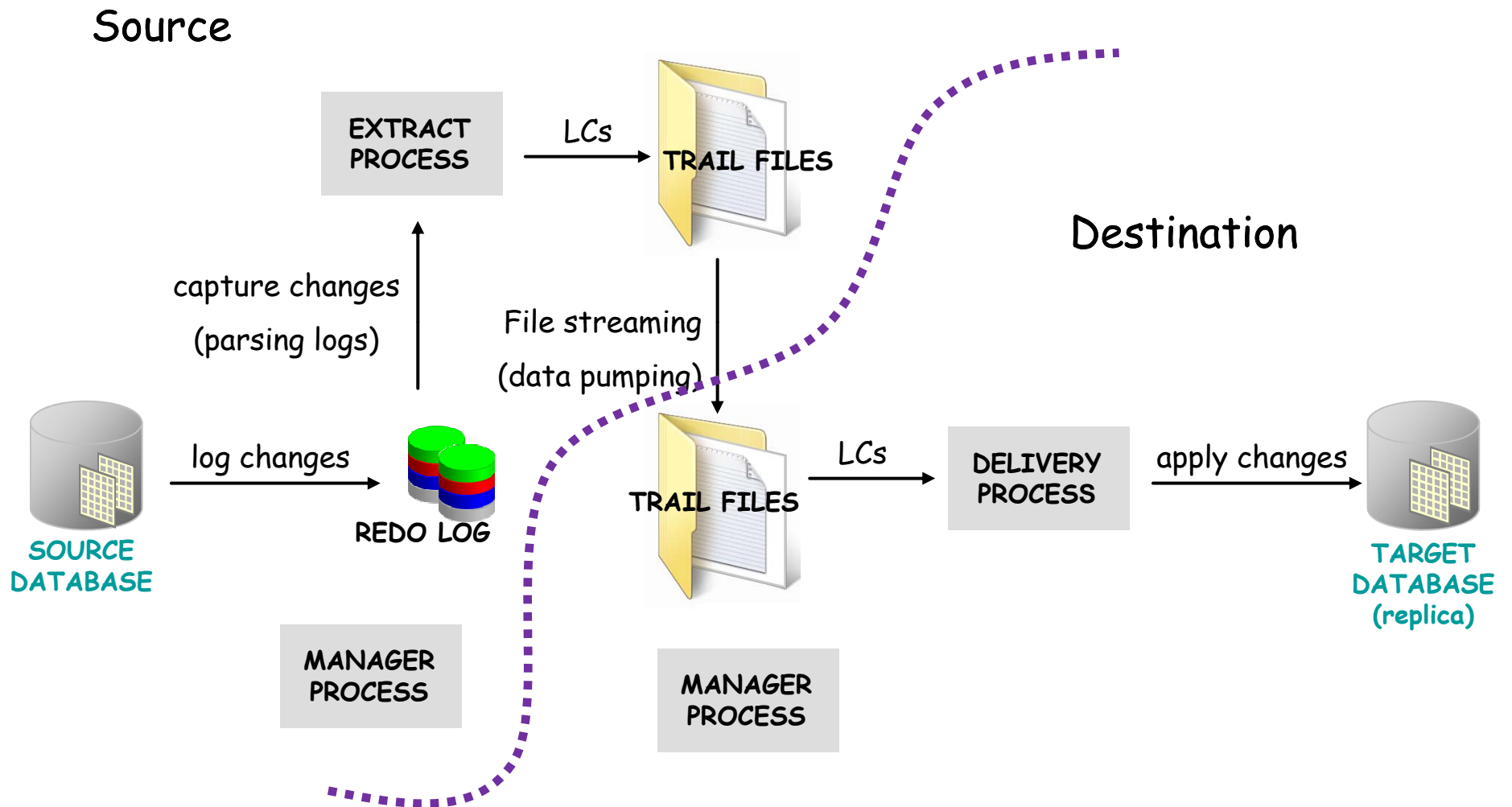


- Trail files
 - Contains data changes written in GG common format
 - Series of files that GoldenGate temporarily stores on disks
 - Two types
 - Extract Trail (located on source - optional)
 - Remote Trail (located on destination)

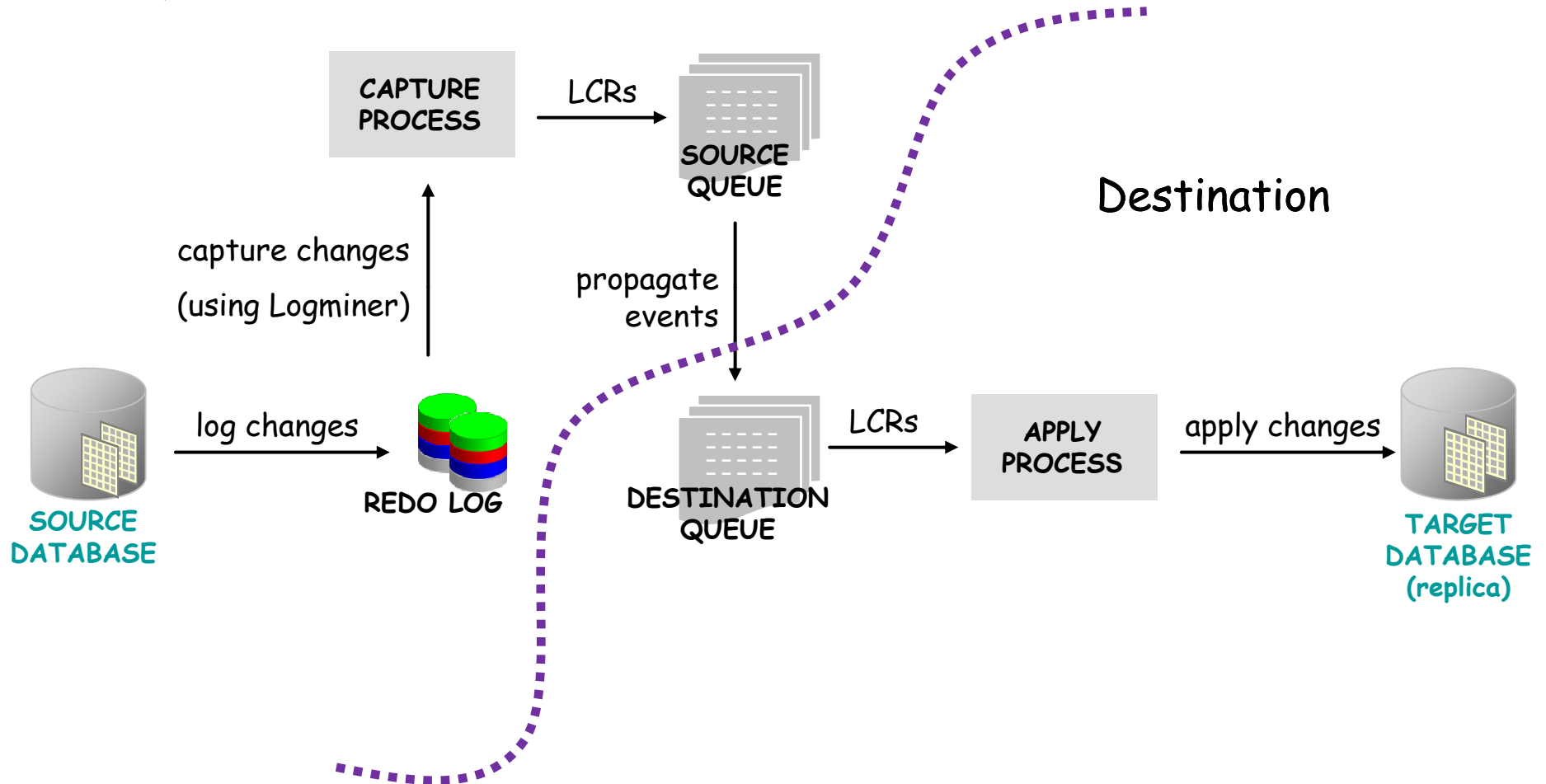


- Data pump
 - Sends data changes from extract trail via TCP/IP to remote trail on the target
 - Optional component
 - Extract process can send changes directly to remote trail
 - Optional encryption and compression over TCP/IP



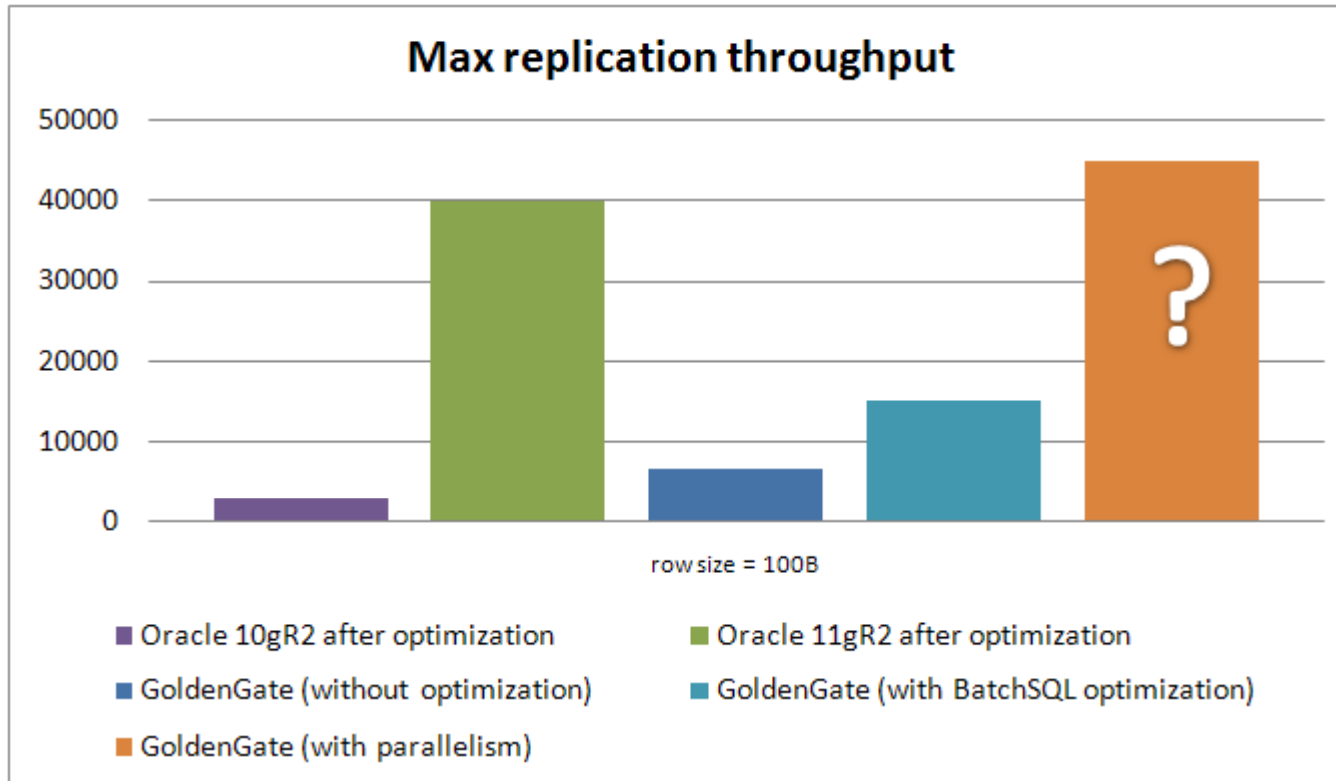


Source



- Supplemental logging groups are created too 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

- Simple workload (inserts, updates, deletes)
 - With basic configuration (no parallelism): 7K LCRs /s
 - With BatchSQL optimization 15K LCRs /s and can be more
 - With parallelism: at least 15K LCRs /s (each parallel process increases throughput linearly)
- COOL workload
 - ~3K LCRs /s
 - BatchSQL does not improve performance
 - Unable to use parallelism
- **Delivery process is the bottlenecks** due to limits of resource utilization by single session.
 - Possible solution: per schema delivery parallelism



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

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

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

	Streams	GoldenGate
Schemas versioning	<ul style="list-style-type: none">•Multi Version Object Dictionary	<ul style="list-style-type: none">•Single Version Object Dictionary•Potential problems with recapturing of older changes
Resuming of replication	<ul style="list-style-type: none">•Slow - capture process needs to re-init dictionary from the last checkpoint	<ul style="list-style-type: none">•Very fast•Using checkpoint files
Failing transactions	<ul style="list-style-type: none">•Mechanism for re-execution of failed transaction	<ul style="list-style-type: none">•No extra mechanism.•Rollback the sequence change number (SCN) and restart the delivery process

- 3 applications
 - Admin tool (defines managers 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 GGSCI command line interface

Oracle GoldenGate Director Administrator Tool

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

Data Source Name: test11g

GoldenGate Info

Host Operating System: WU

Database: ORA

GoldenGate Version: 10.4.0.19

Default DB Credentials

DSN:

Username:

Password:

Confirm Password:

Access Control

Owner:

Host is Observable:

ORACLE GoldenGate

Network

Entire Network

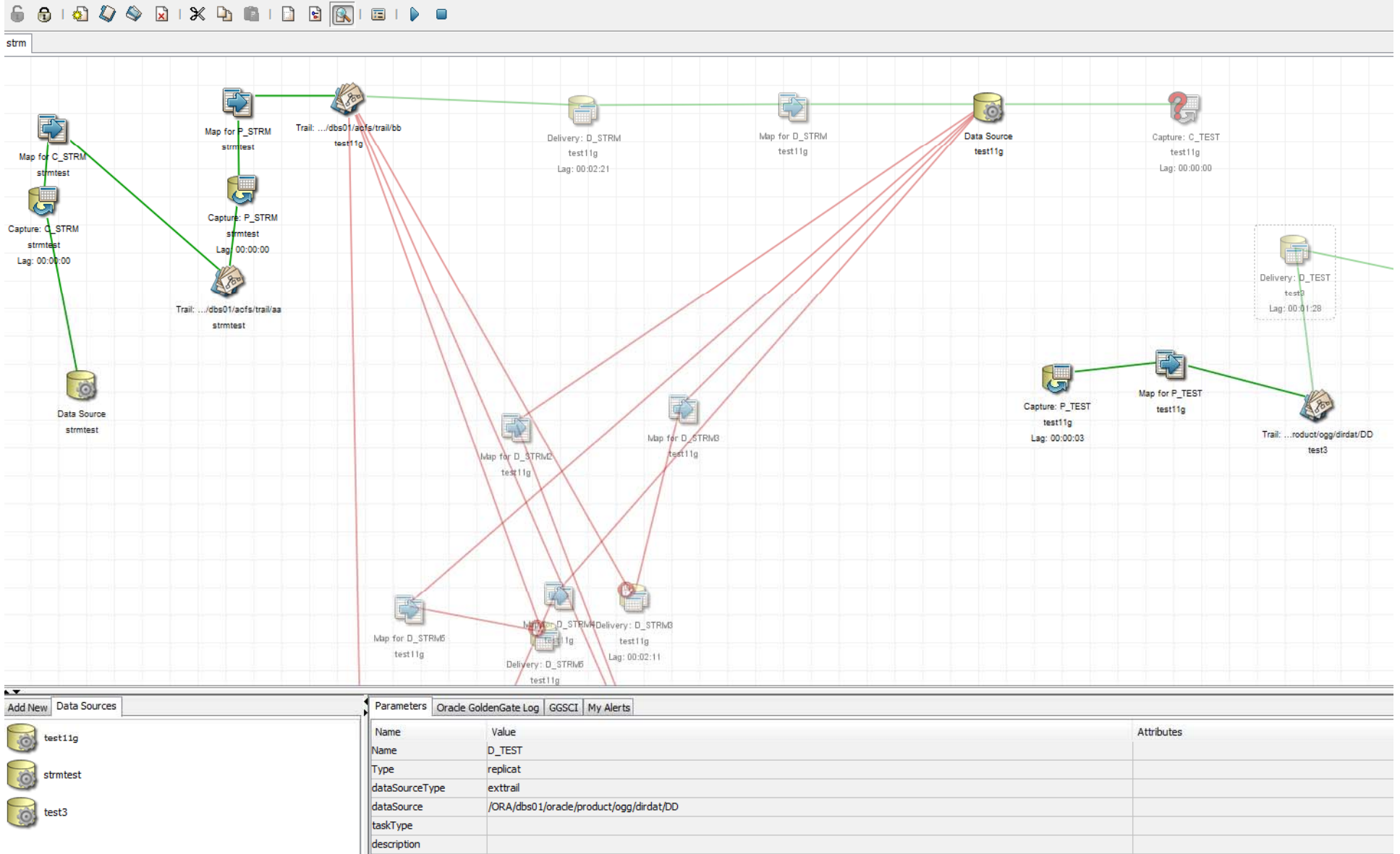
- test11g
 - D_STRM2
 - P_TEST
 - D_STRM
 - C_TEST
 - D_STRM4
 - D_STRM3
 - D_STRM5
- test3
 - D_TEST
- strmtst
 - C_STRM
 - P_STRM
- Watch Lists
 - Configure Watch Lists
 - My Email Alerts
 - Live Monitoring is ON
 - Log Out

Processes

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	
	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	strmtst	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	strmtst	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.			



- Intuitive interface layout (fast information display – no nested views)
- Access to process statistics and reports
- 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 attempt (retries are needed)
- No history data
- No plots except lags

- DBMS independent software
 - Deployment requires more afford ('attaching' to DB)
 - Minimal machine resource utilization
 - Minimal impact on the database
- Golden Gate is stable and efficient.
- Technology focused on data modification operations
 - CERN's data profile is definitions + modifications

- 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
- Fair monitoring (in multiple setups context)
- **Strategic replication solution for Oracle**

- More performance tests
 - Per schema delivery parallelism
 - 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)