



Experience of Using Pro2Oracle + OE CDC

The Big Data in OpenEdge

Who Am I?

- With Progress since 2003
- Training “Fundamentals of OpenEdge RDBMS Administration” (78 new DBA since 2008).
- Training “Advanced DBA for OpenEdge”
- Books (in Russian):
 1. “Fundamentals of OpenEdge database administration”
 2. “OpenEdge platform for beginners”
 3. “OpenEdge Table Partitioning for DBA”
 4. “OpenEdge Table Partitioning for programmers”
 5. “Database data encryption (Transparent Data Encryption)”
 6. “After-Imaging in OpenEdge”
 7. “Learning OpenEdge Replication”



Who Am I?

- <https://rupug.pro>

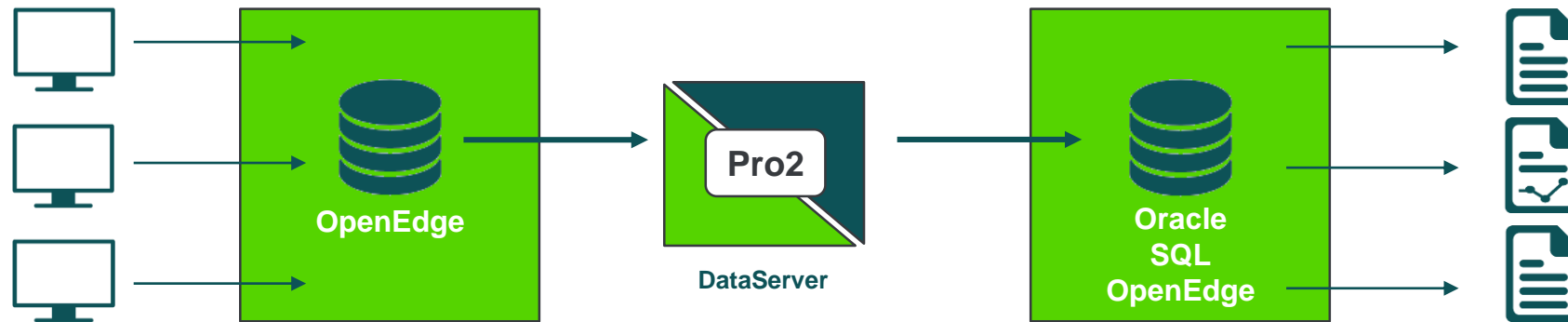
RuPUG Education Center

Own your future learning new skills online



What is Pro2

- Pro2™ is a solution for live, near real-time replication of OpenEdge data to a MS-SQL Server, Oracle or OpenEdge target.



- Delivers quick and easy access to mission-critical data from your OpenEdge system
 - Without disrupting normal business operations or risking transactional database performance and a stability*
- Data replication: **not** disaster recovery

The Bank X & OpenEdge

- 15 000 - active users in the database

The Bank X & OpenEdge

- 15 000 - active users in the database
- 17.5 TB - size of the largest database

The Bank X & OpenEdge

- 15 000 - active users in the database
- 17.5 TB - size of the largest database
- 55 TB - total size of 15 databases of the largest branch

The Bank X & OpenEdge

- 15 000 - active users in the database
- 17.5 TB - size of the largest database
- 55 TB - total size of 15 databases of the largest branch
- 600 TB - total size of all OpenEdge databases

The Bank X & OpenEdge

- 15 000 - active users in the database
- 17.5 TB - size of the largest database
- 55 TB - total size of 15 databases of the largest branch
- 600 TB - total size of all production databases
- ~120 TB - growth of data for the last year

The Bank X & OpenEdge

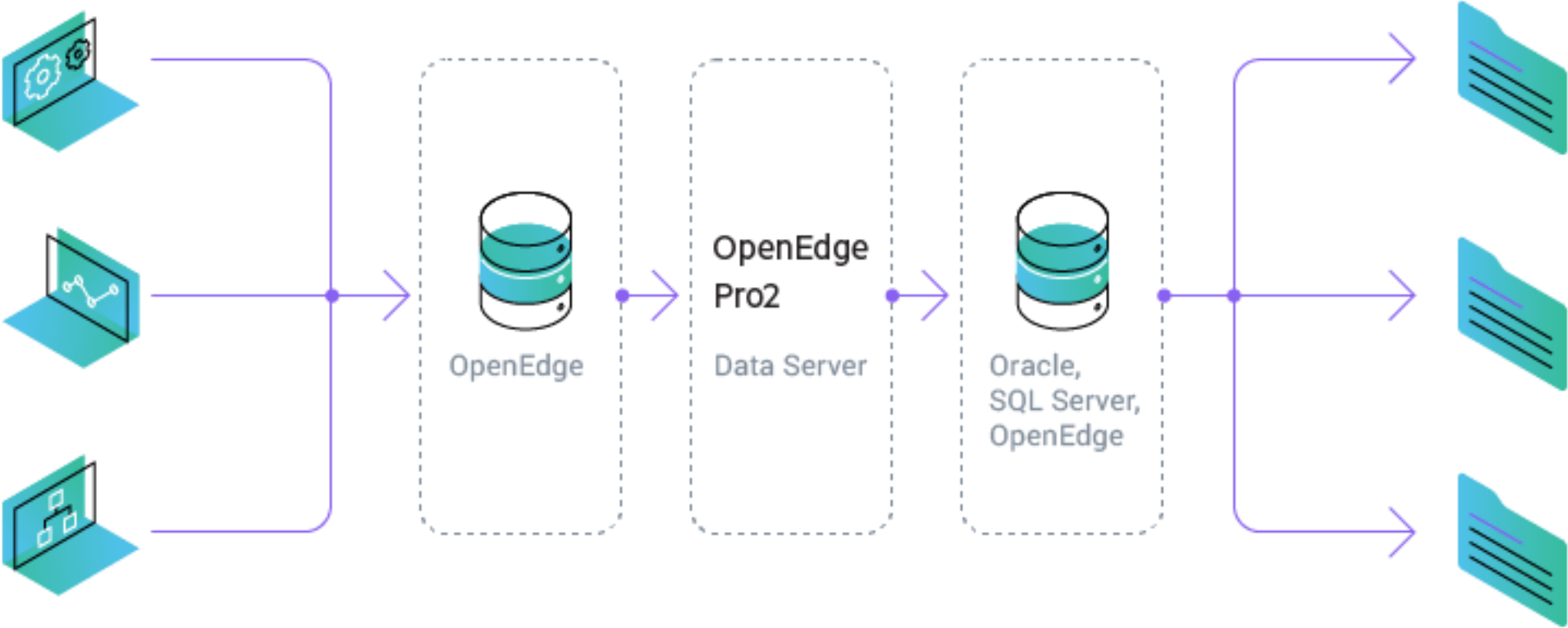
- 15 000 - active users in the database
- 17.5 TB - size of the largest database
- 55 TB - total size of 15 databases of the largest branch
- 600 TB - total size of all databases (include standby)
- ~120 TB - growth of data for the last year
- 1 million CRUD operation per second, without performance degradation

The Need For Pro2

Poor performance of an existing solution for uploading data to a data Warehouse

Reconciliation with other systems takes too long

The Bank X & Pro2Oracle



The Big Data – The Big Problems

Some largest tables for Pro2Oracle replication

Table	Db Analyze Recs	Size (Gb)
Table #1	4 223 377 285	499.4
Table #2	3 161 770 239	217.3
Table #3	2 441 435 657	169.9
Table #4	2 399 959 772	178.9
Table #5	2 384 362 565	165
Table #6	2 242 546 551	68.3
Table #7	2 095 401 605	128.8
Table #8	2 052 205 505	98.4
Table #9	1 989 082 913	78.9
Table #10	1 906 007 478	78.9
Table #11	1 753 559 790	42.3
Table #12	1 709 223 562	83.8
Table #13	1 671 115 196	78.7
Table #14	1 661 420 151	78
Table #15	1 573 281 277	43.4
Table #16	1 500 882 369	58.2
Table #17	1 347 514 073	32.1

Pro2Oracle - what did we encounter?

- Code Page & Database Size
- Initial Sync
- 64bit ROWID
- Replication Triggers & Compression
- Splitting Replication Threads
- OpenEdge Change Data Capture (CDC)

Pro2Oracle - what did we encounter?

- Code Page & Database Size

“Ф” - Cyrillic upper case EF

OpenEdge

Oracle

CP866

UTF-8

HEX – 94

HEX – D0A4

BIN - **1001 0100** (1 byte)

BIN - **1101 0000 1010 0100** (2 byte)

Pro2Oracle - what did we encounter?

- Code Page & Database Size

OpenEdge

CHARACER = **31 Kb**

=>

ORACLE

VARCHAR2 = **4 Kb**

⇓

Error : catch :4212A column in this row being inserted or updated is too large (4212)

Pro2Oracle - what did we encounter?

- Code Page & Database Size

SUBSTRING(bfrSrcCustomer.Address,1,4000)

OpenEdge

ORACLE

CHARACER = 31 Kb

=>

VARCHAR2 = 4 Kb

||
v

Error : catch :4212A column in this row being inserted or updated is too large (4212)

Pro2Oracle - what did we encounter?

- Code Page & Database Size

SUBSTRING(bfrSrcCustomer.Address,1,2000)

OpenEdge

ORACLE

CHARACER = 31 Kb

=>

VARCHAR2 = 4 Kb

OR

Enable **Extended Data Types** in Oracle 12c

VARCHAR2 = 31Kb

Pro2Oracle - what did we encounter?

- Initial Sync

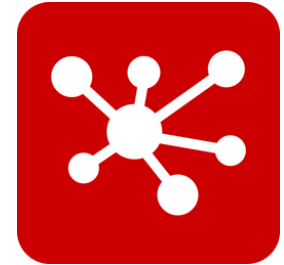
Old Pro2 Template: `tmpl_mreplproc_oracle.p`

*Processed 100,000 records so far. Time is 26/11/2017 13:37:26.311+03:00.
Total Elapsed Time is: **0h:5m:9s.845ms***

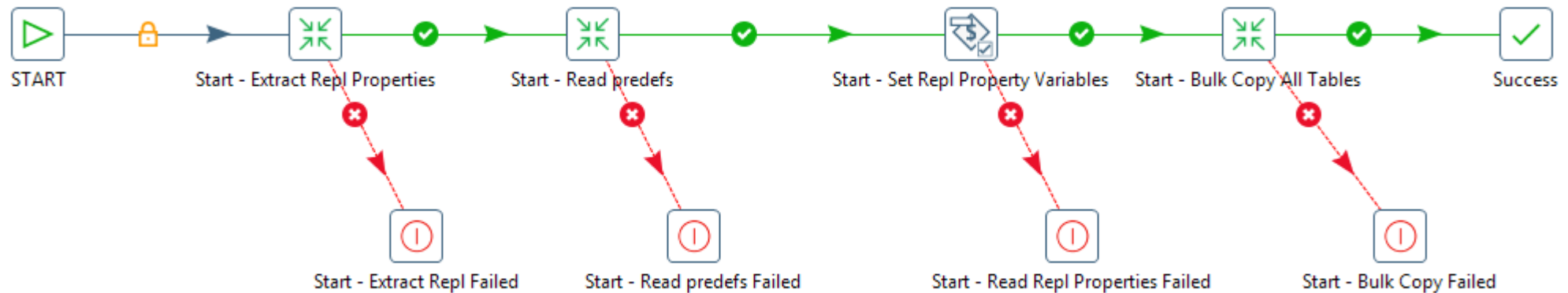
4 billion records = 20 weeks

Pro2Oracle - what did we encounter?

- Initial Sync

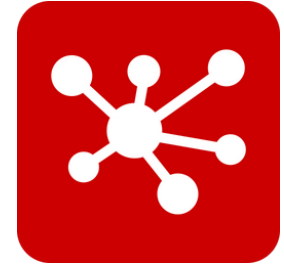


Third party solution: Pentaho Data Integration



Pro2Oracle - what did we encounter?

- Initial Sync



Third party solution: Pentaho Data Integration

2018/07/06 **13:25:34** - URT - replControl Status.0 – db1.table1: Row 100000 Status Update, Rows Completed: 1000000 ...

2018/07/06 **13:25:37** - URT - replControl Status.0 - db1.table1 : Row 100000 Status Update, Rows Completed: 1100000...

2018/07/06 **13:25:42** - URT - replControl Status.0 - db1.table1 : Row 100000 Status Update, Rows Completed: 1200000...

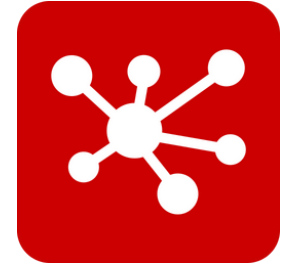
2018/07/06 **13:25:45** - URT - replControl Status.0 - db1.table1 : Row 100000 Status Update, Rows Completed: 1300000...

2018/07/06 **13:25:49** - URT - replControl Status.0 - db1.table1 : Row 100000 Status Update, Rows Completed: 1400000...

4 billion records = 4 - 5 days

Pro2Oracle - what did we encounter?

- Initial Sync



Third party solution: Pentaho Data Integration

2018/07/06 **13:25:34** - URT - replControl Status.0 – db1.table1: Row 100000 Status Update, Rows Completed: 1000000 ...

2018/07/06 **13:25:37** - URT - replControl Status.0 - db1.table1 : Row 100000 Status Update, Rows Completed: 1100000...

2018/07/06 **13:25:42** - URT - replControl Status.0 - db1.table1 : Row 100000 Status Update, Rows Completed: 1200000...

2018/07/06 **13:25:45** - URT - replControl Status.0 - db1.table1 : Row 100000 Status Update, Rows Completed: 1300000...

2018/07/06 **13:25:49** - URT - replControl Status.0 - db1.table1 : Row 100000 Status Update, Rows Completed: 1400000...

2018/07/07 **05:57:34** - URT - replControl Status.0 - db1.table1: Row 100000 Status Update, Rows Completed: **2 138 900 000** ...

2018/07/07 **05:57:34** - URT - Insert / Update replControl.0 - Values set for lookup: [BULKCOPY], [db1.table1]

2018/07/07 **05:57:35** - T - Source Table.0 - Couldn't get row from result set

2018/07/079 **05:57:35** - T - Source Table.0 - **[DataDirect][OpenEdge JDBC Driver][OpenEdge] Overflow error** (7485)

Pro2Oracle - what did we encounter?

- Initial Sync

NEW Pro2 Template: `templ_mreplproc_oracle.p` **Oracle_Bulk_Transaction_Count**

Using Oracle_Bulk_Transaction_Count of 10000

Processed 100,000 records so far. Time is 25/11/2018 07:14:53.302+03:00. Total Elapsed Time is: 0h:0m:23s.837ms

Processed 200,000 records so far. Time is 25/11/2018 07:15:14.150+03:00. Total Elapsed Time is: 0h:0m:44s.685ms

Processed 300,000 records so far. Time is 25/11/2018 07:15:35.997+03:00. Total Elapsed Time is: 0h:1m:6s.532ms

...

...

Processed 2,147,200,000 records so far. Time is 29/11/2018 03:35:27.443+03:00. Total Elapsed Time is: 3 days, 20h:20m:57s.978ms

Processed 2,147,300,000 records so far. Time is 29/11/2018 03:35:38.957+03:00. Total Elapsed Time is: 3 days, 20h:21m:9s.492ms

*Processed **2,147,400,000 records** so far. Time is 29/11/2018 03:35:49.904+03:00. Total Elapsed Time is: 3 days, 20h:21m:20s.439ms*

*Finished Mass Replication at 29/11/2018 05:00:40.764+03:00. Total Elapsed Time was: **3 days, 21h:46m:11s.298ms***

Pro2Oracle - what did we encounter?

- 64bit ROWID

Recid: 4 547 935 090

Invalid ROWID in replqueue.srcrecord.

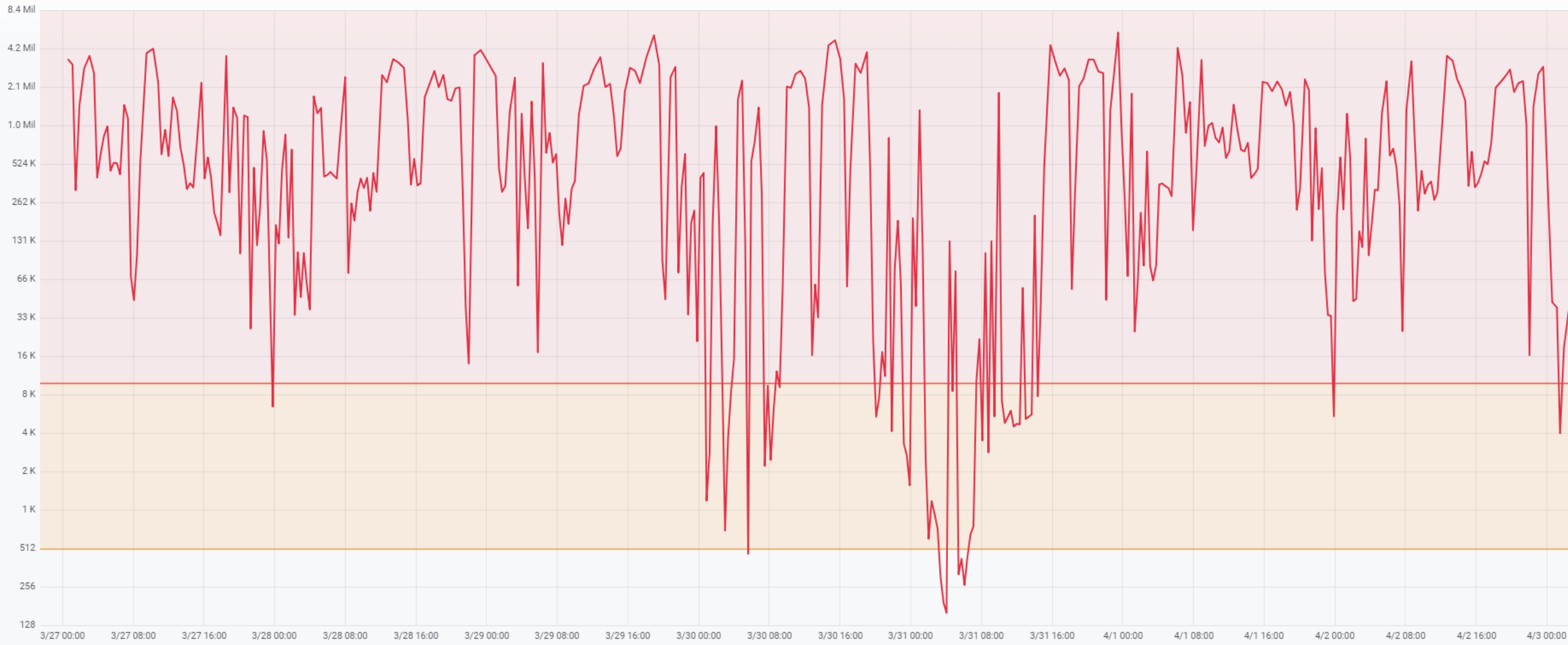
Aborting replication of this record. Sequence: 578,402,107 DB=db1 Table=table1

Rowid=0x10f13fb72 TnsID: 1,420,550,509

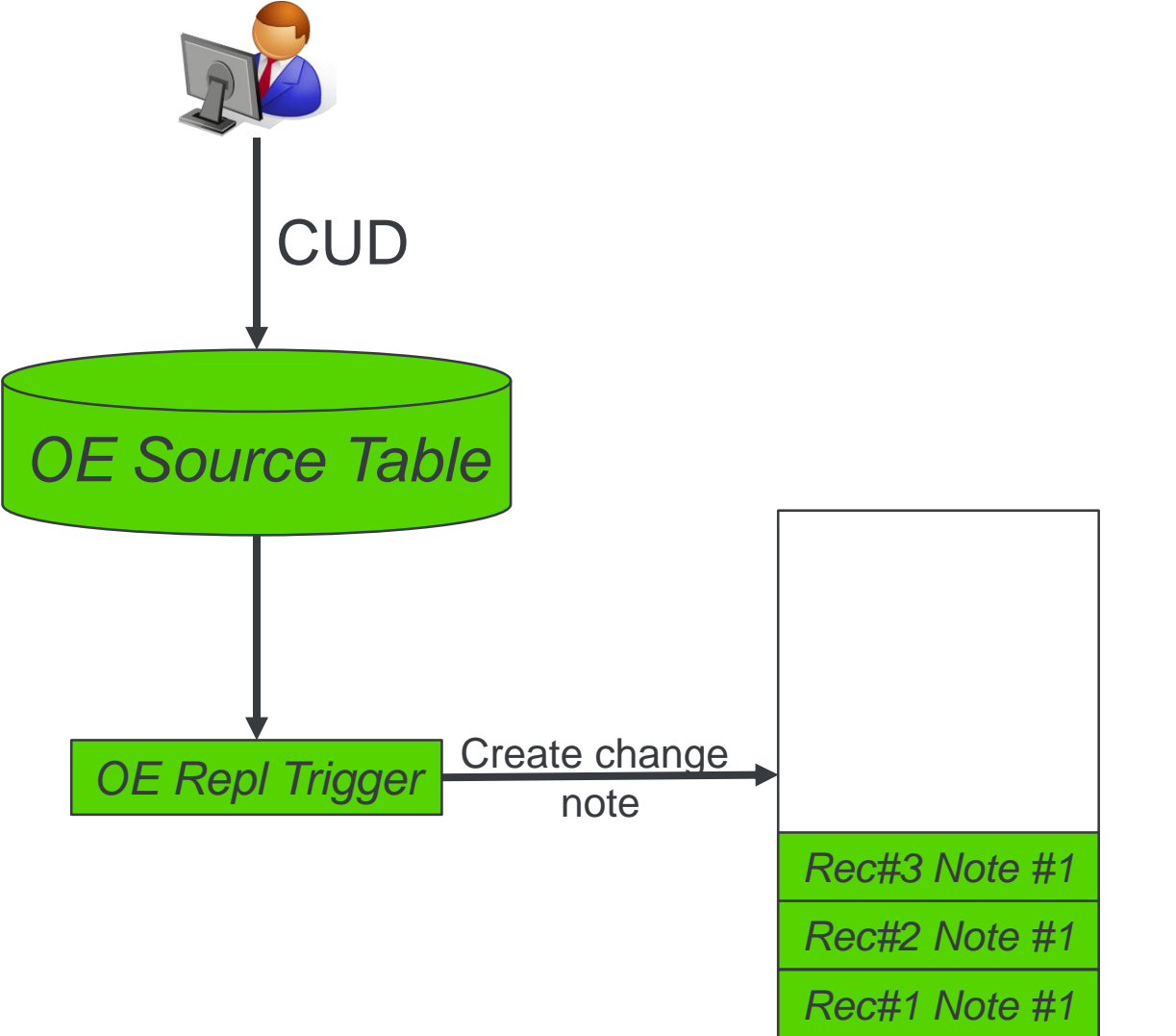
Must be: **0x000000010f13fb72**

Pro2Oracle - what did we encounter?

- Replication Triggers & No Compression

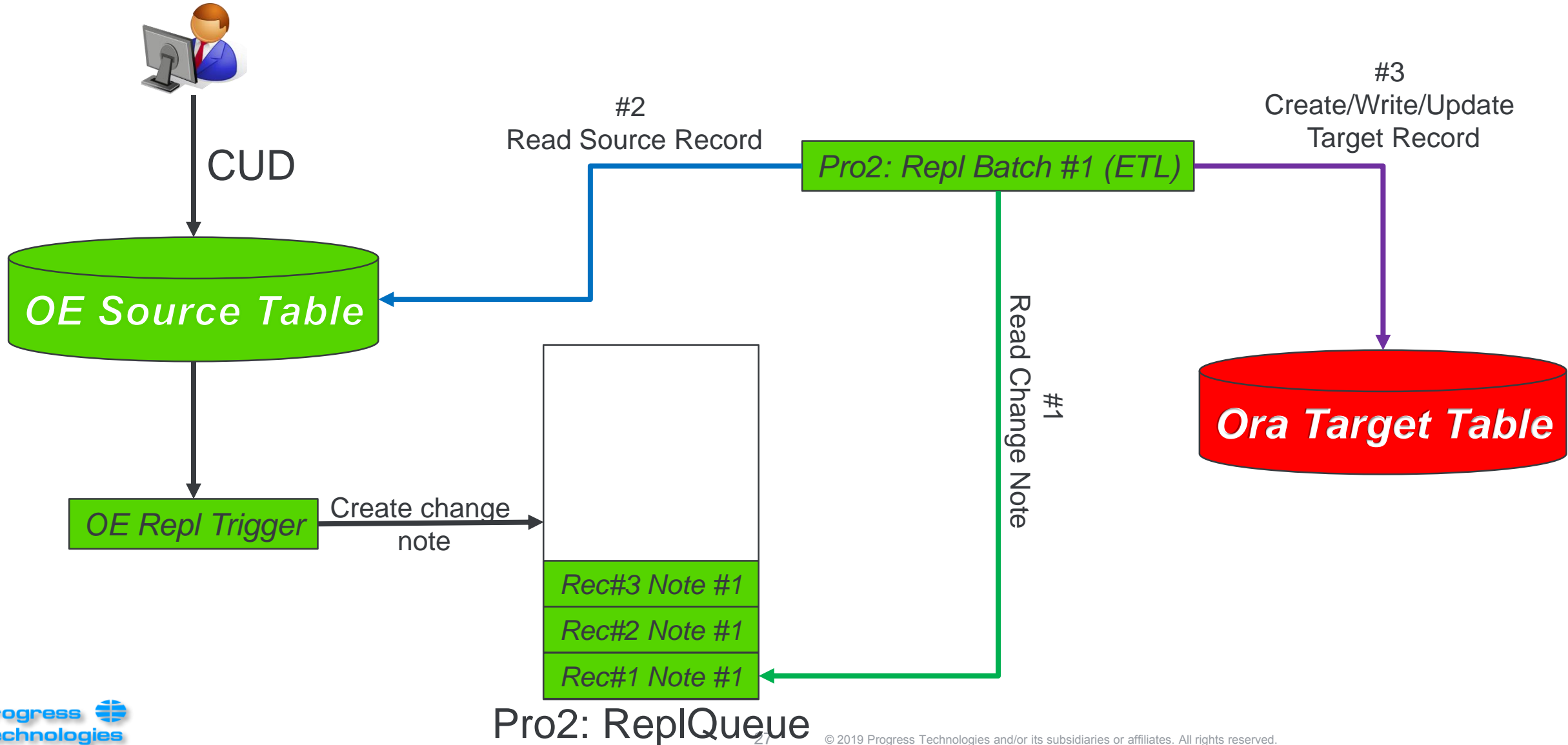


Pro2: How Repl Trigger works

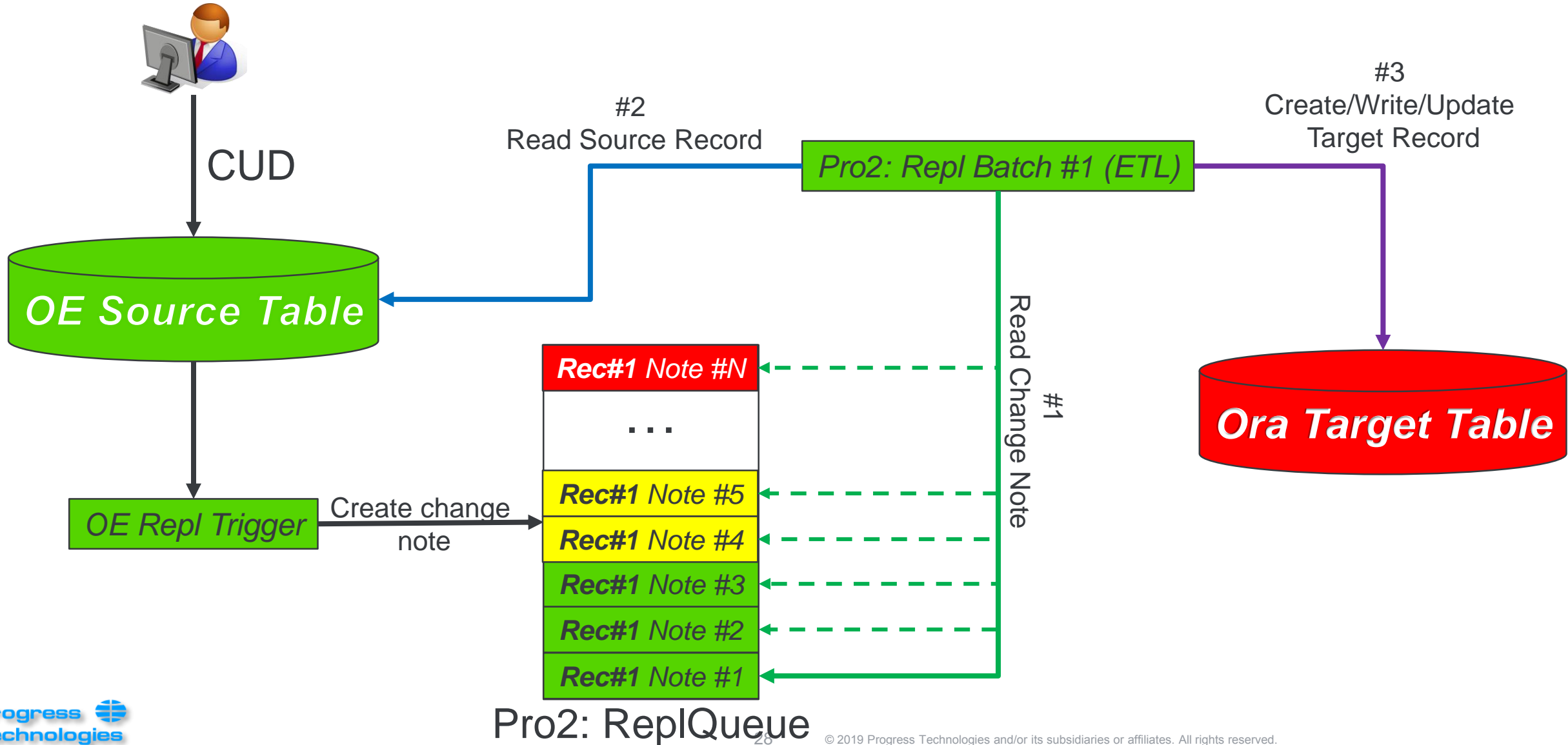


Pro2: ReplQueue

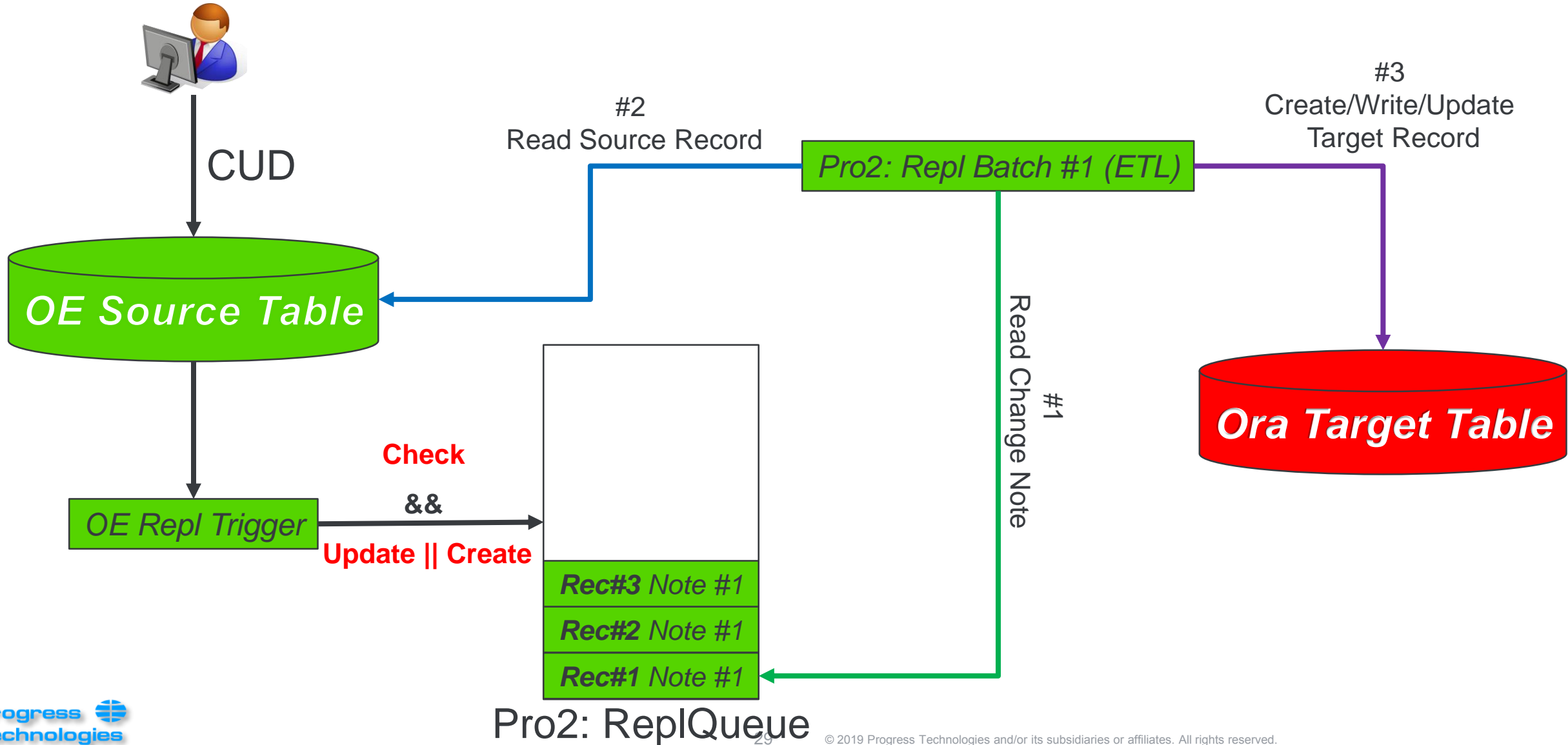
Pro2: How Repl Trigger works



Pro2: How Repl Trigger works

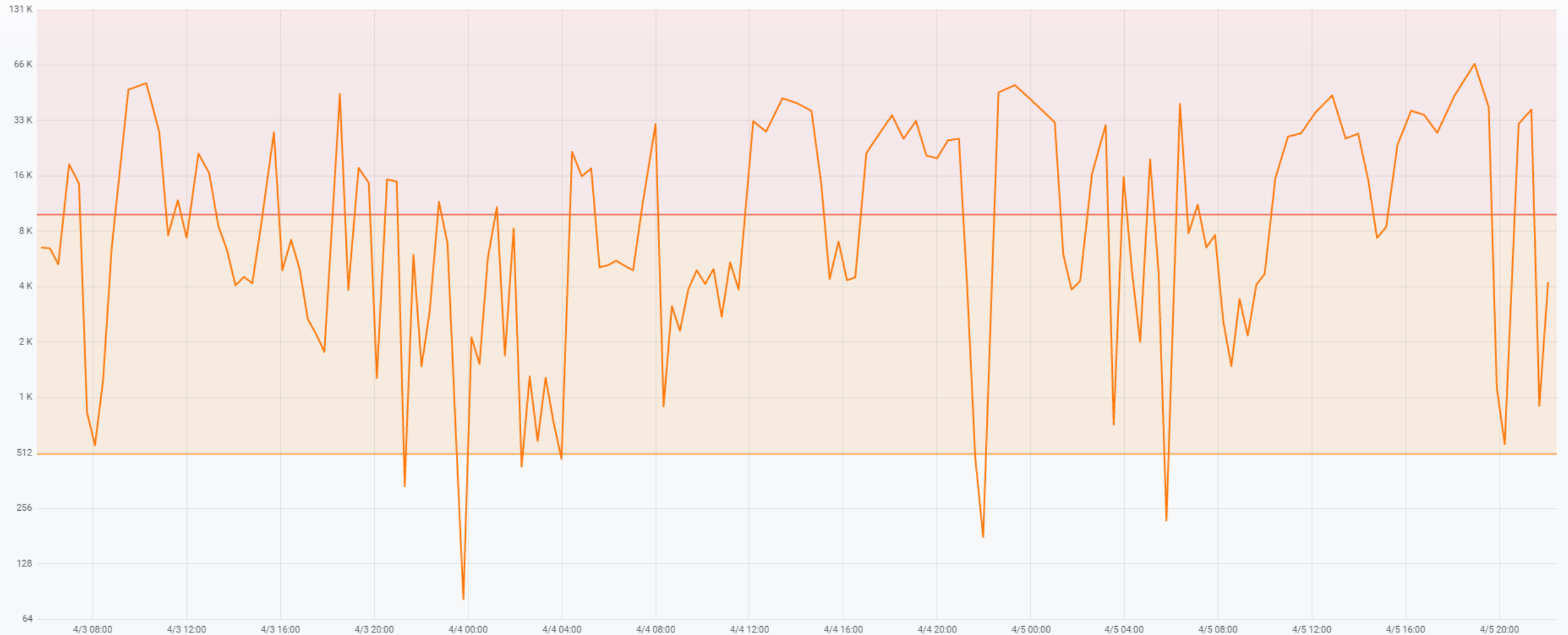


Pro2: How Repl Trigger with **Compression** works

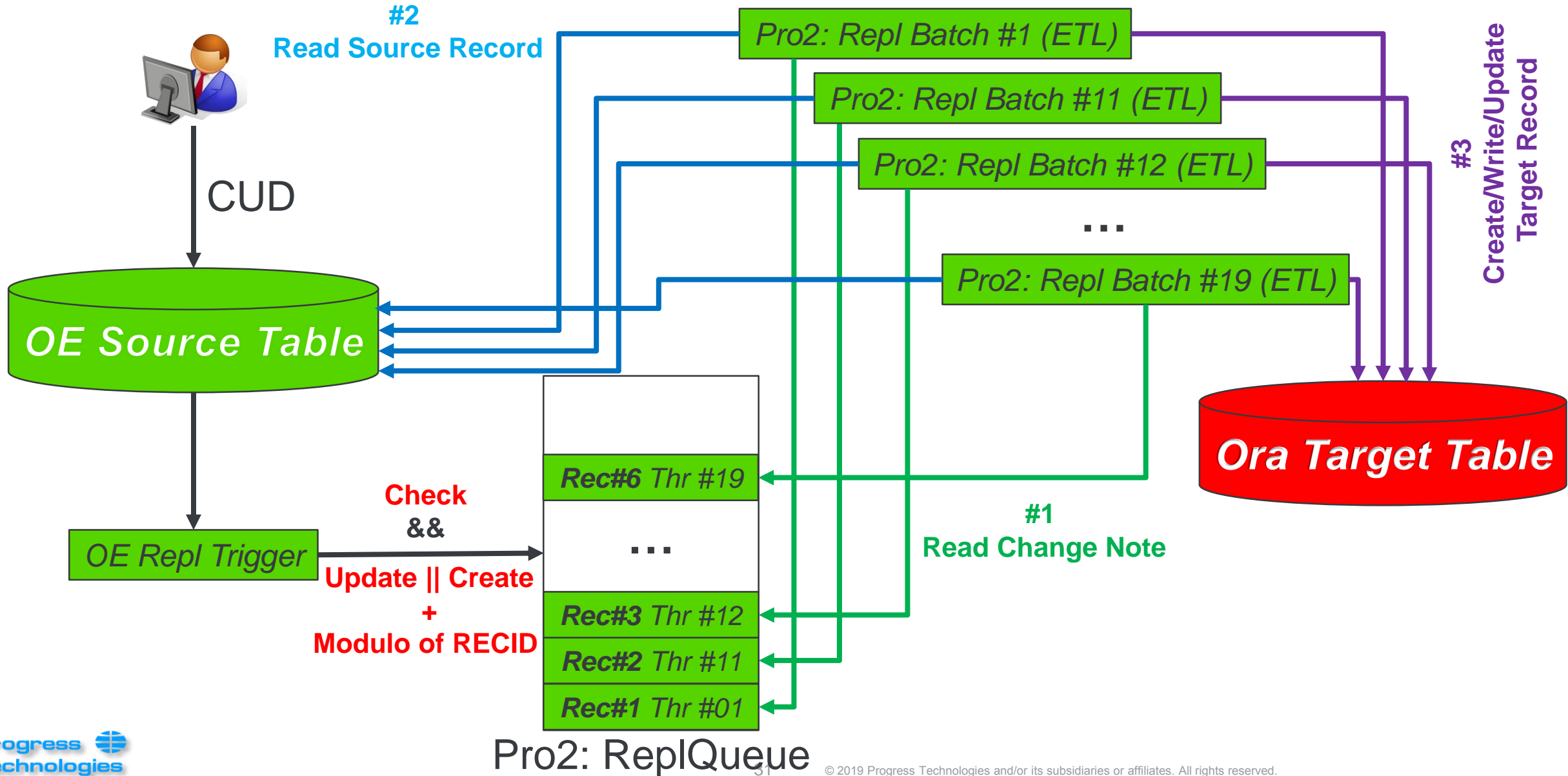


Pro2Oracle - what did we encounter?

- Replication Triggers & Compression

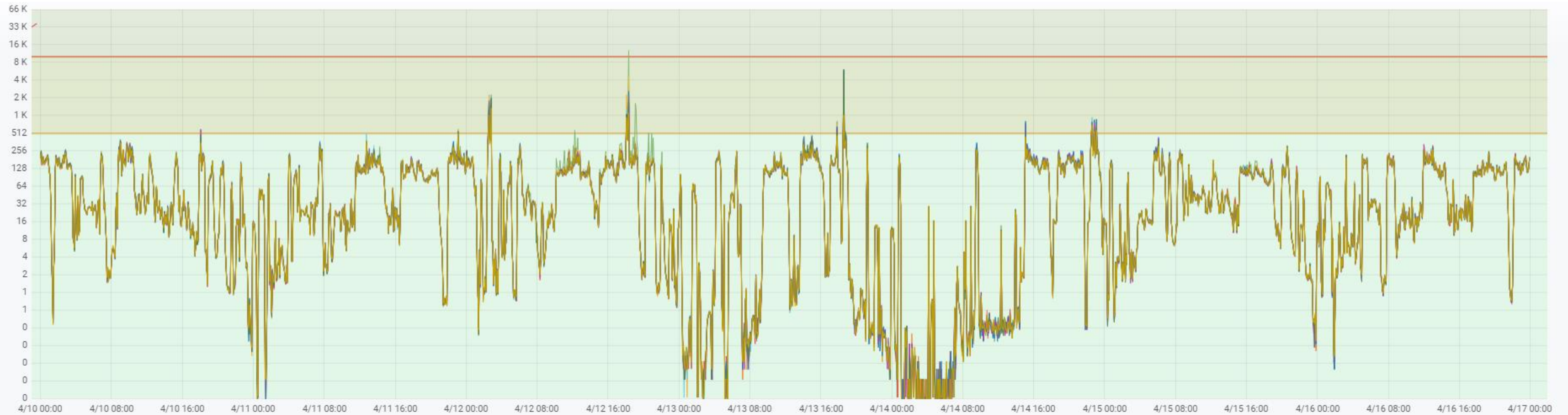


Pro2: How Repl Trigger with **Splitting Replication Threads** works



Pro2Oracle - what did we encounter?

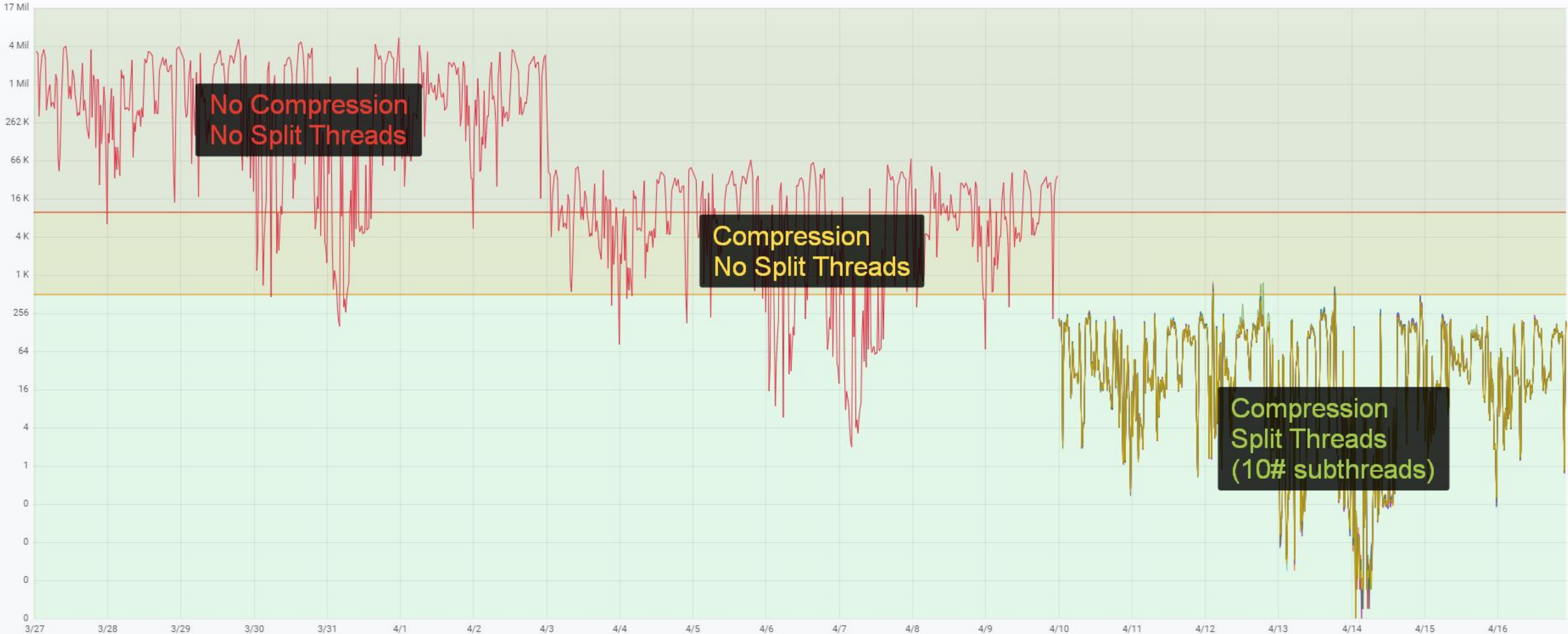
- Splitting Replication Threads



	max	avg	current	total
Thread11: recs	13.15 K	92	197	184.85 K
Thread12: recs	5.94 K	83	198	168.10 K
Thread16: recs	6.22 K	81	188	164.01 K
Thread13: recs	3.95 K	78	197	157.30 K
Thread17: recs	3.97 K	78	203	156.38 K
Thread19: recs	5.98 K	77	198	155.74 K
Thread15: recs	2.19 K	75	190	150.85 K
Thread18: recs	2.48 K	74	190	148.41 K
Thread1: recs	1.37 K	73	198	148.04 K
Thread14: recs	2.48 K	73	188	146.92 K

Pro2Oracle - what did we encounter?

- Splitting Replication Threads



Pro2Oracle + OE CDC

1. OE CDC embedded into RDBMS engine.
2. Can be enabled online.
3. CDC policies can be activated online. (replication triggers can not)
4. Faster than standard replication triggers.
5. Captures ABL and SQL changes.
6. OE CDC ABL API
7. Component of the Advanced Enterprise Edition RDBMS

Overview of Data Capture and ETL Processes

Source Table (Customer)

User Data

Record operations: create, update, delete, ...

CDC Internal Database Triggers

CDC Cache (CDC Policy Tables: *_Cdc-Table-Policy*, *_Cdc-Field-Policy*)

Change Tracking Table (*_Cdc-Change-Tracking*)

_Policy-Id | *_Tran-id* | *_Time-Stamp* | *_Change-Seq* | *_Op*

Change Table (*CDC_Customer*)

Meta-data Columns | User Data (captured data)

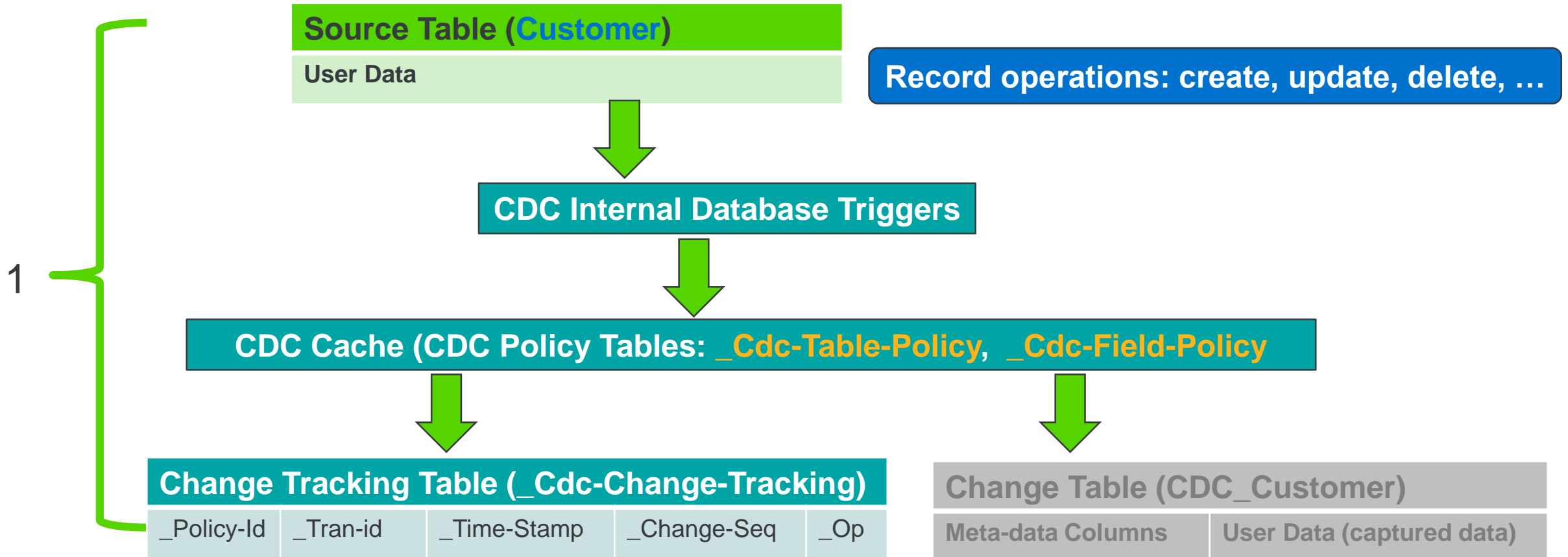
Pro2 ETL Process

Data Warehouse

1

2

Overview of Data Capture and ETL Processes



OpenEdge Tables

Source Table (Customer)

User Data

Source tables: OpenEdge database user tables

- Identified in a CDC policy as the source captured data
- Source tables exist prior to implementing a CDC policy
- Source tables are not changed when the CDC policy is implemented

CDC Policy Tables

CDC Policy Tables: `_Cdc-Table-Policy`, `_Cdc-Field-Policy`

- Policy tables are created when CDC is enabled on an OpenEdge database

CDC Table Policy Table – `_Cdc-Table-Policy`

- Contains one record for each policy defined on a source table
- There may be multiple policies defined for a source table (pending, current, previous)
- Relationship: A table policy record is parent of zero or more field policy records

CDC Field Policy Table – `_Cdc-Field-Policy`

- Contains one record for each tracked field of a CDC table policy
- Note if a table policy is set to level 0 there are no field policy records

CDC Change Tracking Table

Change Tracking Table (_Cdc-Change-Tracking)

_Policy-Id	_Tran-id	_Time-Stamp	_Change-Seq	_Op
------------	----------	-------------	-------------	-----

CDC Change Tracking Table : _Cdc-Change-Tracking

- Table created when CDC is enabled on the database
- Contains a record for each CDC tracked operation on a source table
- Operations are captured for the current policy on the source table when that policy is active

CDC Change Tables

Change Table (CDC_Customer)

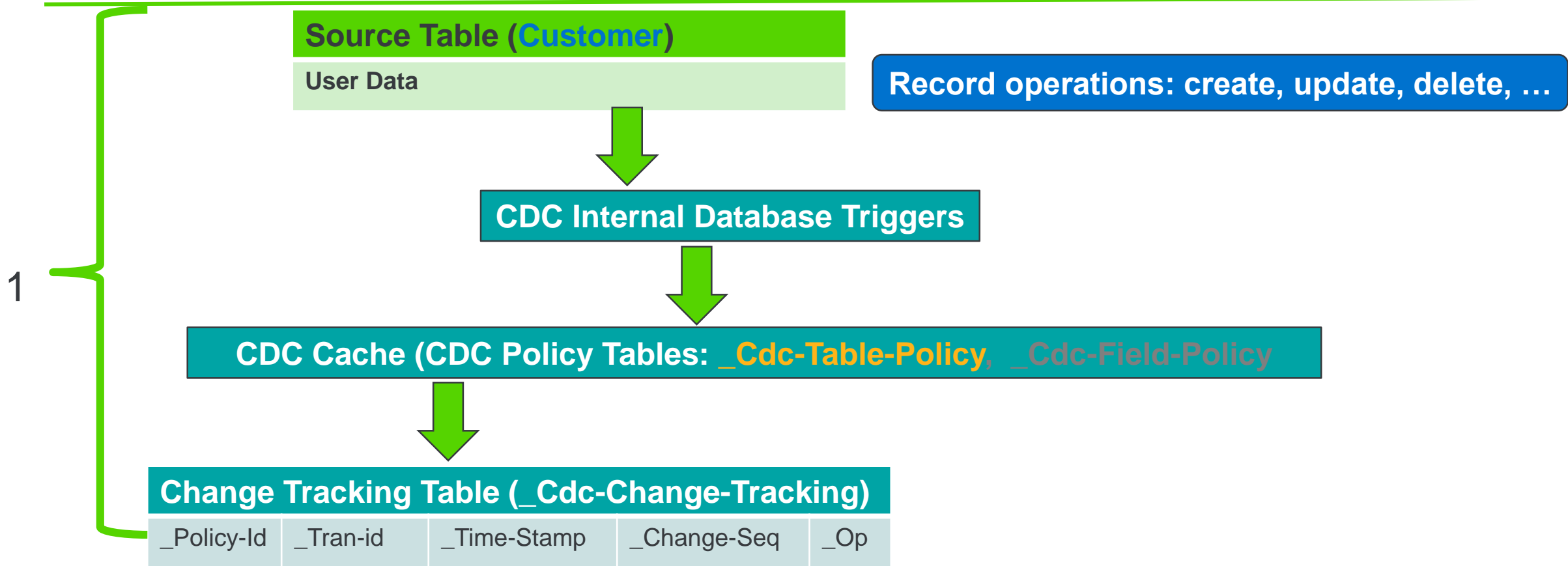
Meta-data Columns

User Data (captured data)

CDC Change Tables :

- Contains change data as prescribed in the CDC policy tables
- Includes both meta data and user data (which are identified in the CDC policy table)
- A CDC change table is created the first time a policy with level 1 or higher is defined
- There is at most one CDC Change Table for a CDC source table
- May be associated with different CDC policies over the course of its lifetime

Overview of Data Capture

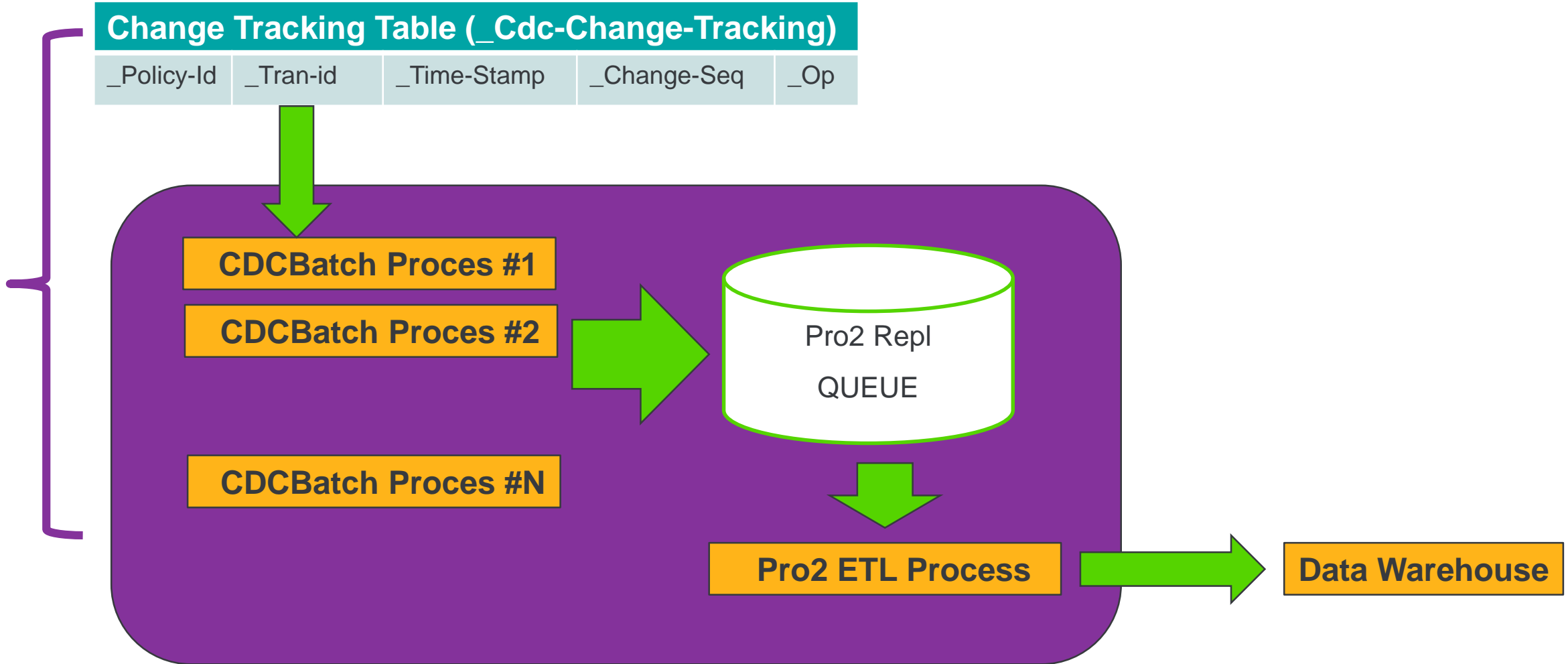


Data Capture and Pro2 ETL Processes

Change Tracking Table (_Cdc-Change-Tracking)

_Policy-Id	_Tran-id	_Time-Stamp	_Change-Seq	_Op
------------	----------	-------------	-------------	-----

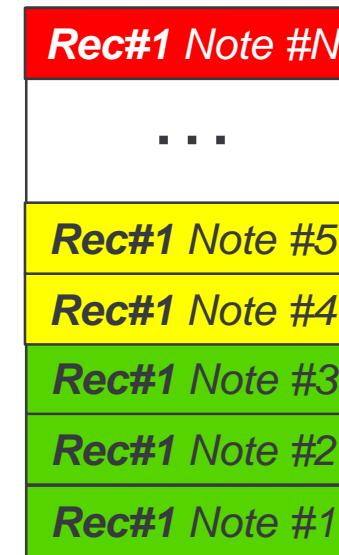
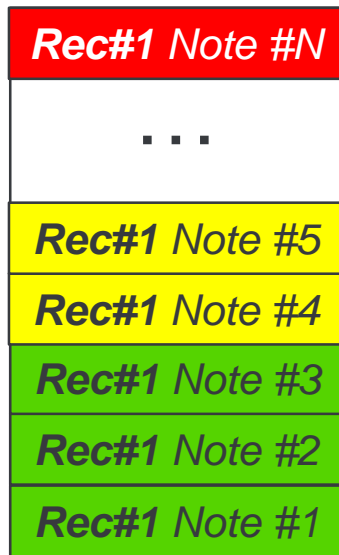
2



Pro2Oracle + OE CDC - what did we encounter?

There is no internal compression for CDC Level 0!

OE CDC:
_Cdc-Change-Tracking



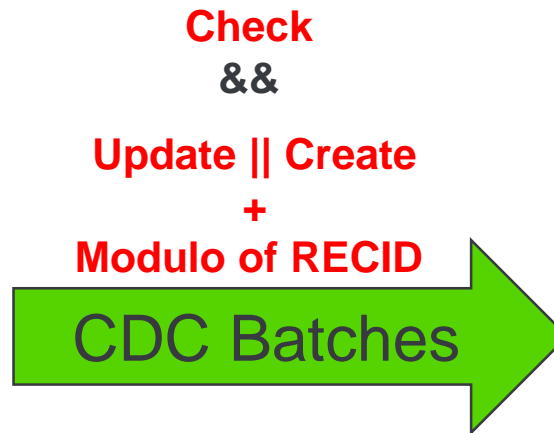
Pro2: ReplQueue

Pro2Oracle + OE CDC

Solution: CDCBatch with Compression & Split Replication Threads

OE CDC:
_Cdc-Change-Tracking

Rec#1 Note #N
...
Rec#1 Note #5
Rec#1 Note #4
Rec#1 Note #3
Rec#1 Note #2
Rec#1 Note #1



Rec#6 Thr #19
...
Rec#3 Thr #12
Rec#2 Thr #11
Rec#1 Thr #01

Pro2: ReplQueue

You need to implement Pro2?

Just ask me:

v.bashkatov@progress-tech.ru

Find me:

Progress - OpenEdge Zone

<https://www.facebook.com/groups/ProgressOEZone/>



