Real-time replication vs ETL - How analytics requires new technologies designed for it, not '70s technology

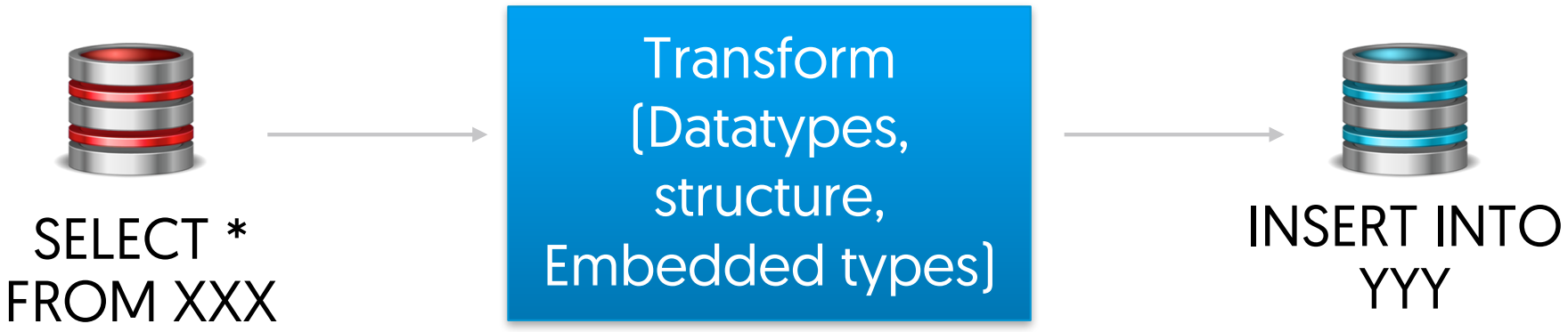
continuent

Topics

In today's webinar, we will discuss:

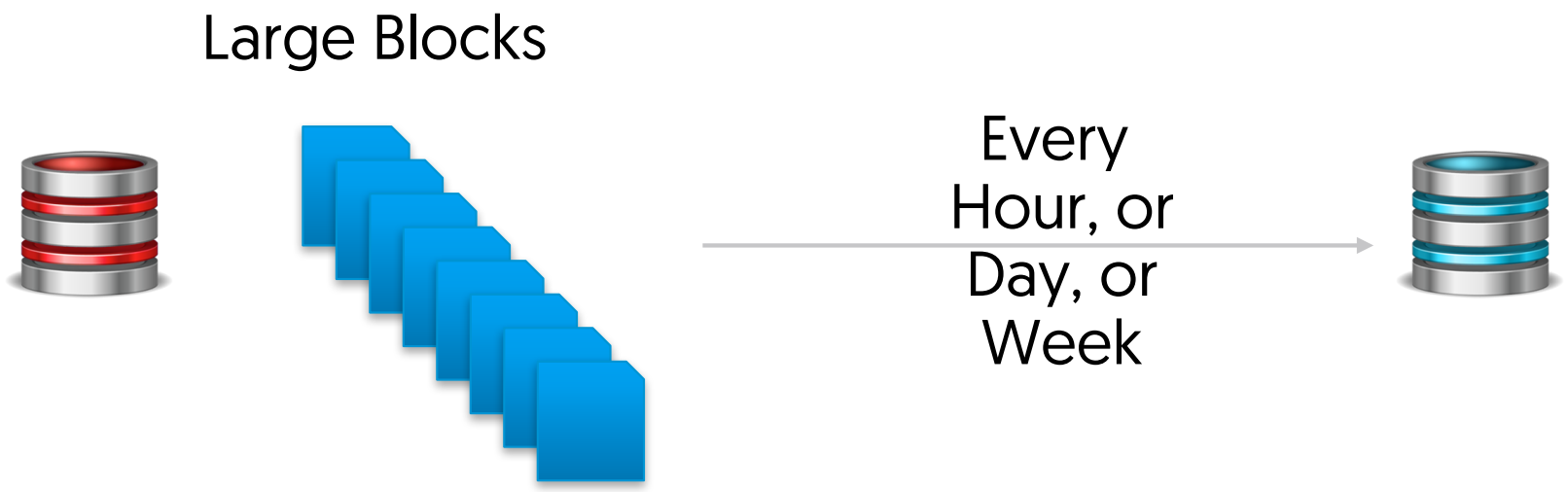
- What is ETL?
- What is CDC?
- What is replication?
- Differences

Extract, Transform, and Load (ETL)



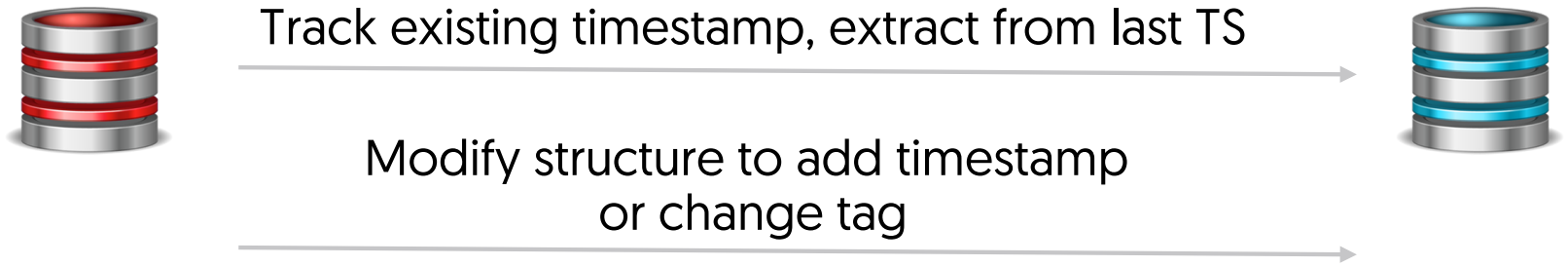
continuent

ETL Latency



continuent

ETL Incremental



continuent

ETL Parameters

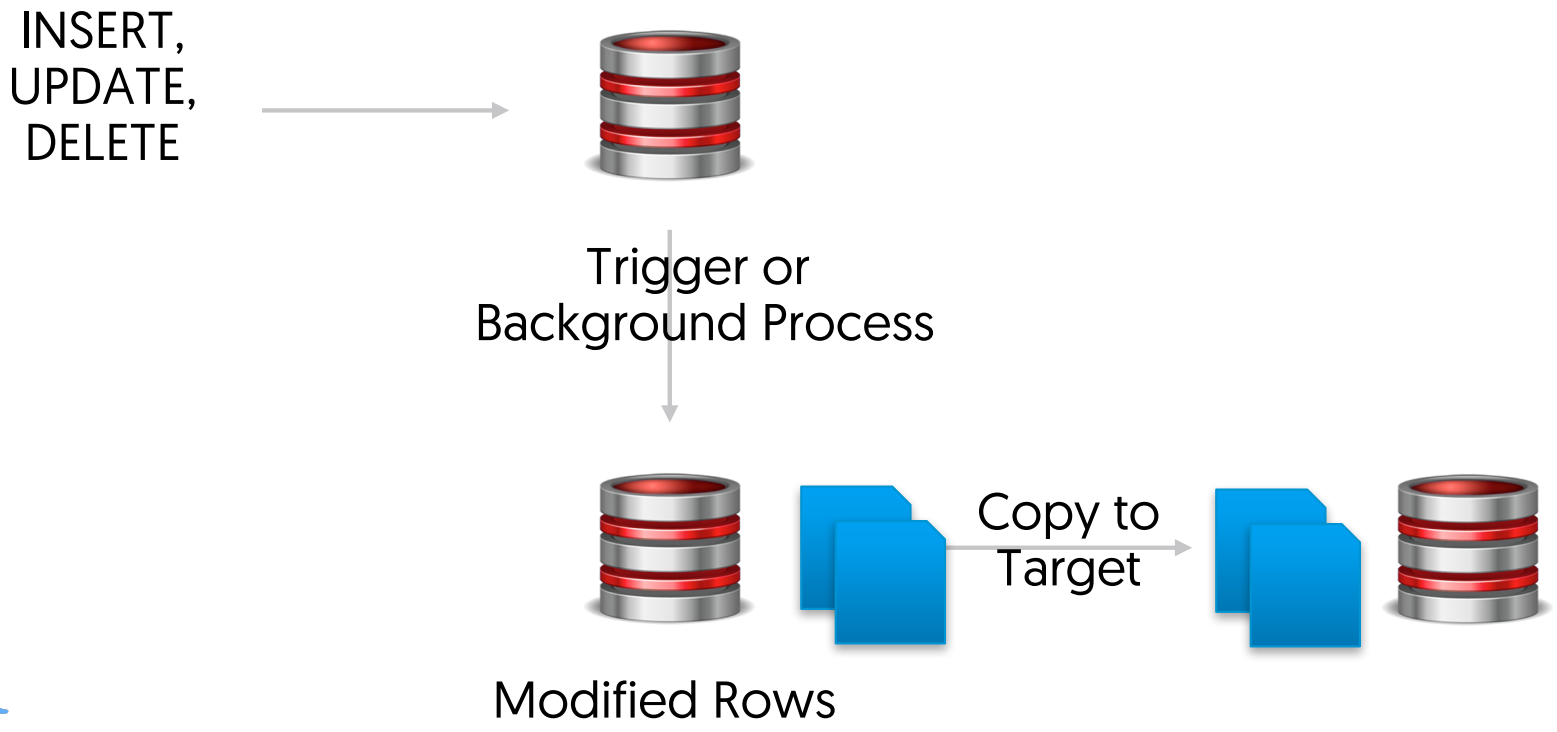
Advantages

- Loads large blocks of data efficiently
- Allows for complex and detailed transformation
- Allows for full data extraction
- Can be easy to implement
- Efficient data loading for some target environments

Disadvantages

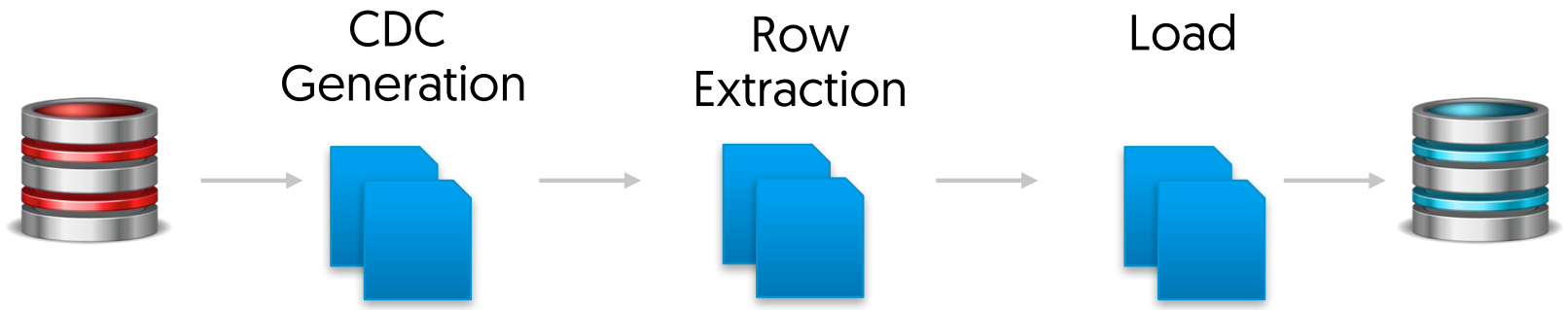
- Can be time consuming
- Places extreme load on source database
- Transformation can take excessive time
- Bulk load can create data ingestion performance problems on the target database
- Incremental loading difficult or complex
  - Either timestamp based
  - Data/structure changes

Change Data Capture (CDC)



continuent

CDC Latency



continuent

CDC Parameters

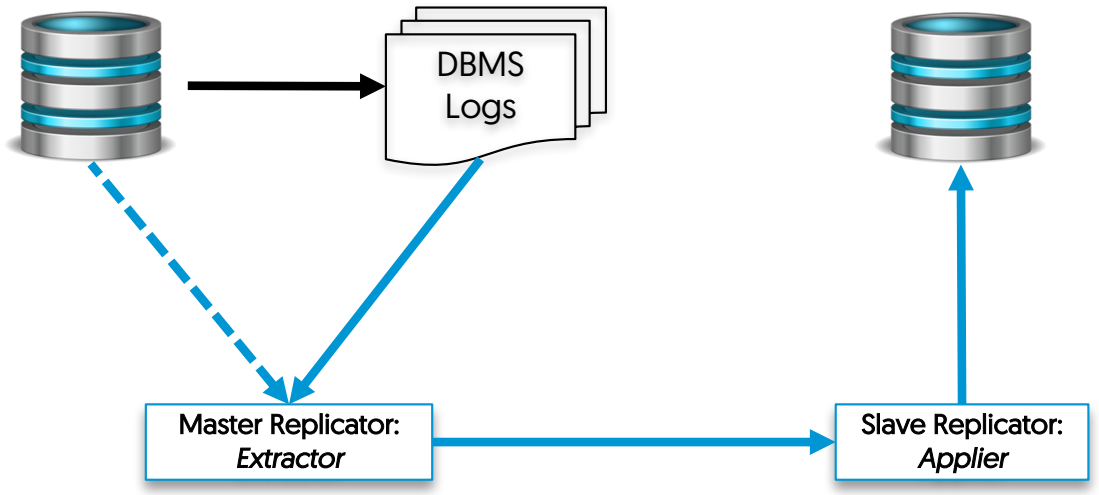
Advantages

- Easy method of identify or extracting data
- Requires no schema modifications to the source database
- Medium latency
- Medium data loading cadence

Disadvantages

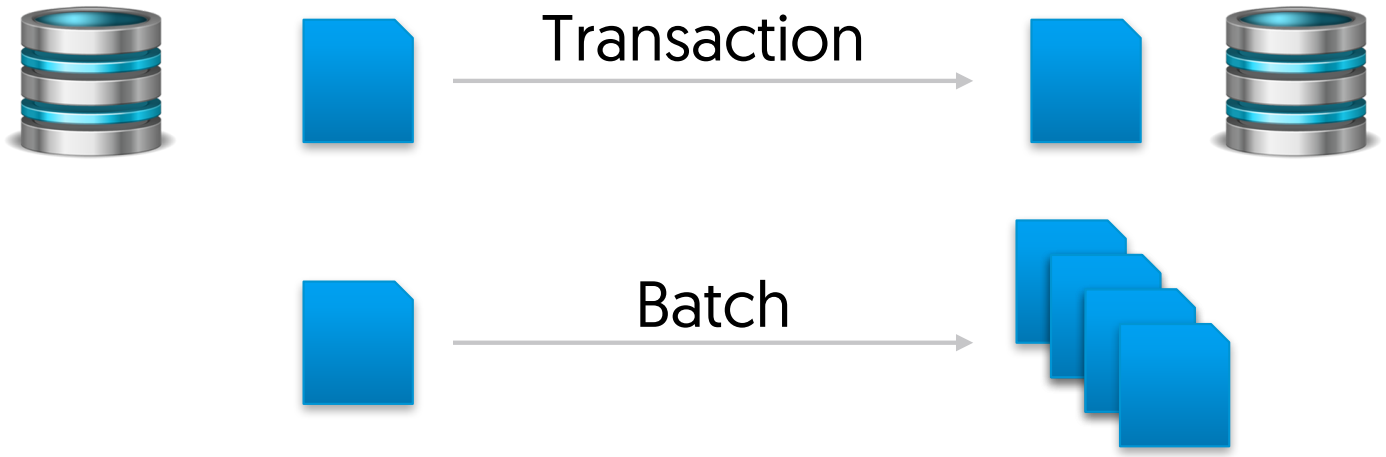
- Requires database level support
  - Either custom CDC or Triggers
- Requires additional database and table space
- Implies overhead on database (both for generation and extraction)
- Extraction latency can increase as data volumes increase
- Requires source access to DB
- No or few transformations

Replication



continuent

Replication Latency



continuent

Replication Parameters

Advantages

- Low latency replication
- No source database access
- No source database load
- No source database changes

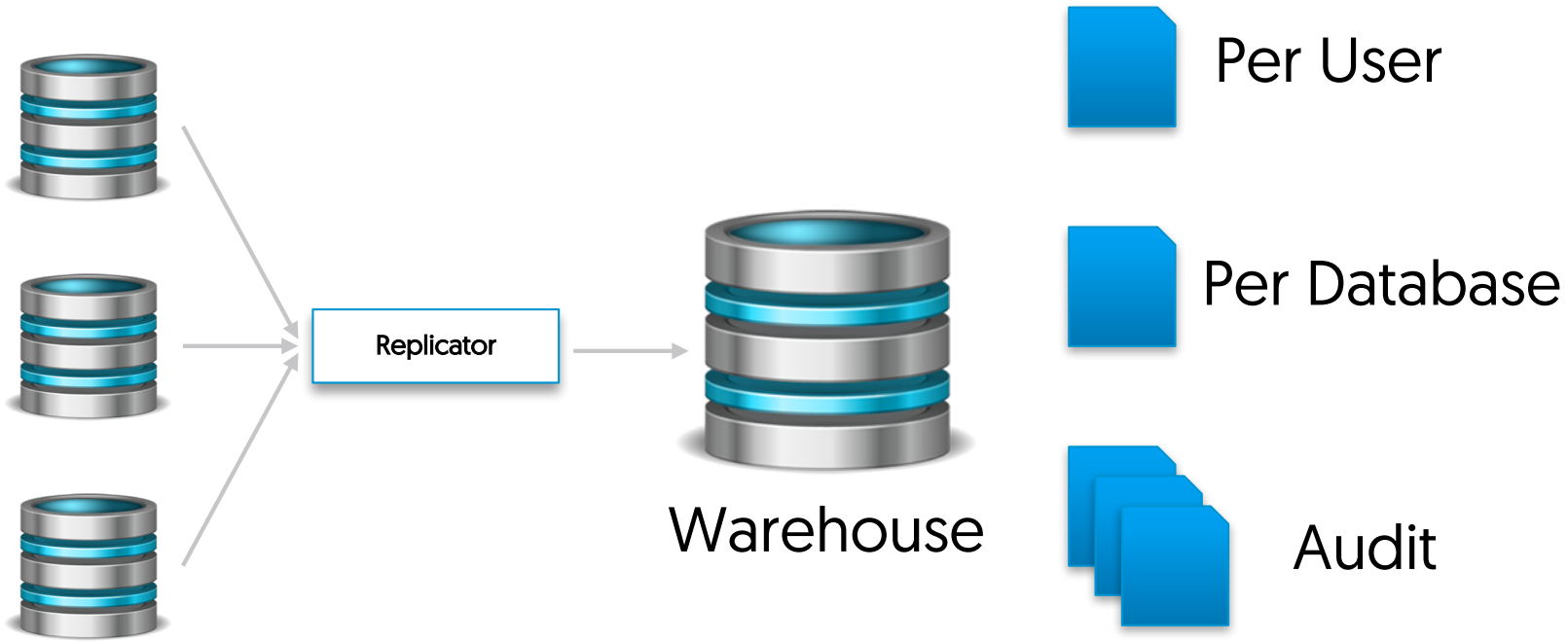
Disadvantages

- Limited transformation and combination of data
- Per-transaction loads slower
  - Mitigated by batching or parallel apply

Comparison

Feature	ETL	CDC	Replication
Data Rate	Low, periodic	Periodic	High, real-time
Transformations	All data and modifications possible before load	Very low complexity only	Low complexity only
Combination	Complex combinations possible	Very limited	Very limited
Source DB Load Impact	High to Very High	Medium	Low
Incremental Support	Difficult, or requires DDL changes	Fully supported	Fully supported

Example of modern Heterogeneous Deployments



continuent

Conclusions

- ETL
  - Efficient and simple system
  - Slow, and not usable for modern deployments
  - Incremental Complex
- CDC
  - Capable with database load
  - Lower latency
  - DB dependent
- Replication
  - Very low latency/High performance
  - Limited transformations and combinations

Next Steps

- If you are interested in knowing more about Tungsten Replicator and would like to try it out for yourself, please contact our sales team who will be able to take you through the details and setup a POC – sales@continuent.com

- Read the documentation at http://docs.continuent.com/tungsten-replicator-5.2/index.html

- Subscribe to our Tungsten University YouTube channel! http://tinyurl.com/TungstenUni

For more information, contact us:

Eero Teerikorpi
Founder, CEO
eero.teerikorpi@continuent.com
+1 (408) 431-3305

Eric Stone
COO
eric.stone@continuent.com

MC Brown
VP Products
mc.brown@continuent.com

Chris Parker
Director, Professional Services EMEA & APAC
chris.parker@continuent.com