

Real-Time Analytics with Griddable.io

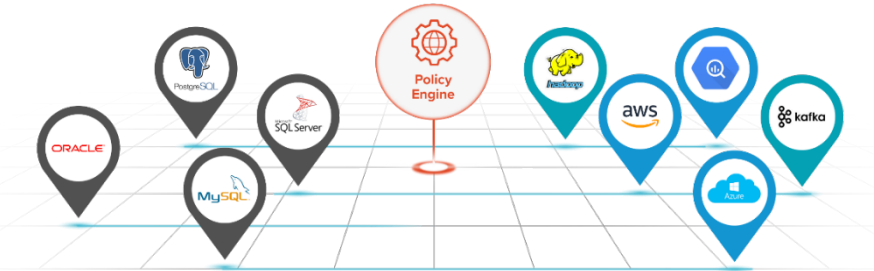
Digital businesses are adopting real-time analytics three times faster than traditional analytics¹. Typically use cases are automating fraud detection, creating digital customer experiences, and improve global collaboration. Griddable.io provides a SaaS platform to synchronize enterprise data into analytics platforms to complement high velocity data from streaming systems. The platform is a cloud-first approach to data integration that is uniquely suited to the new hybrid cloud reality.

Key Benefits

- *Synchronized consistent clones eliminate copy sprawl. Guaranteed continuous delivery of transactionally consistent clones from source databases eliminates the copy sprawl and lag-time of ETL or snapshots.*
- *Flexibly transform and mask regulated data*
Data masking policies can be applied at the source to mask private or secure data before it goes into the cloud. Further data transformations applied at the destination for ingestion into relational or non-relational analytics platforms.
- *Automated bulk load and incremental update*
Policies define database clones that include only relevant data from each source. Clones are bulk loaded into the destination, then incremental changes are continuously applied to keep analytics platform up-to-date.

A fresh approach is needed

Over 60% of enterprises now rely on hybrid cloud or public cloud for big data analytics² providing flexible low-cost processing power to drive sub-second automated decisions.



The Griddable.io smart grid for enterprise data guarantees transaction consistency across any number of clouds, databases, and schemas.

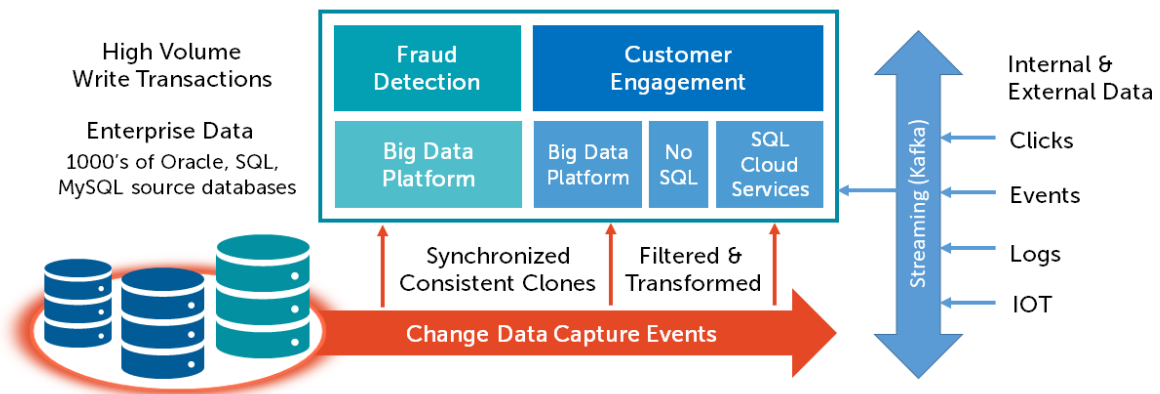
This real-time need requires new approaches to data integration replacing legacy ETL, snapshot, and messaging busses.

- ETL batch and micro-batch operations create time lag for ingested data measured in hours or longer. Micro-batch can reduce this to under 1 hour but creates processing overhead and does not capture incremental changes rapidly.
- Sixty percent of all data storage is a copy of some kind, creating a \$50 billion problem for the IT industry³. The traditional approach of expensive centralized storage arrays with de-duplication prevents distribution of data to analytics projects in the cloud.
- Streaming is rapid replacing legacy messaging busses for high velocity event, click-stream, log, and IoT data but these technologies do not guarantee transaction consistency or reliable delivery of enterprise data.

Synchronized consistent clones

Continuous synchronization of operational data into analytics

Griddable.io automatically and continuously synchronizes data across heterogeneous combinations of clouds and databases. Griddable.io data synchronization is performed in real-time, resulting in a grid of databases which are all time synchronized, consistent, and providing a specific data set appropriate to the needs of each use case.



The grid data pipeline is optimized for transactional data by reading change data capture logs at the source, converting them to a neutral timeline consistent data format.

For analytics use cases, a Griddable.io grid can be configured for N:1 or N:M topologies across clouds, enabling database teams to flexibly connect operational data sources in the cloud or on-premises into relational and non-relational analytics platforms in the cloud

The grid data pipeline is optimized for transactional data by reading change data capture logs at the source, converting them to a neutral timeline consistent data format, and applying changes at the destination in the correct transaction sequence. The cloud-first architecture enables the pipeline to scale-up as needed for performance and recover from network failures or delays.

Policy Based Clones for bulk load and continuous updates

Policies define the relevant data from each source to create a database clone using a parallel read operation. This clone is ingested into the destination using a bulk load and then automatically cutover to incremental change updates of the relevant source data on a continuous basis. This minimizes network utilization into cloud platforms and replaces separate back-up or snapshot processes followed by manual operation to replace change logs.

Continuously synchronize into any cloud analytics platform

The cloud provides an ever-expanding number of analytics platform options from relational data warehouses, non-relational big data platforms, and cloud-native services. The Griddable.io destination adaptors convert timeline consistent data into destination format in real-time providing flexibility to connect into any analytics platform or multiple platforms concurrently.

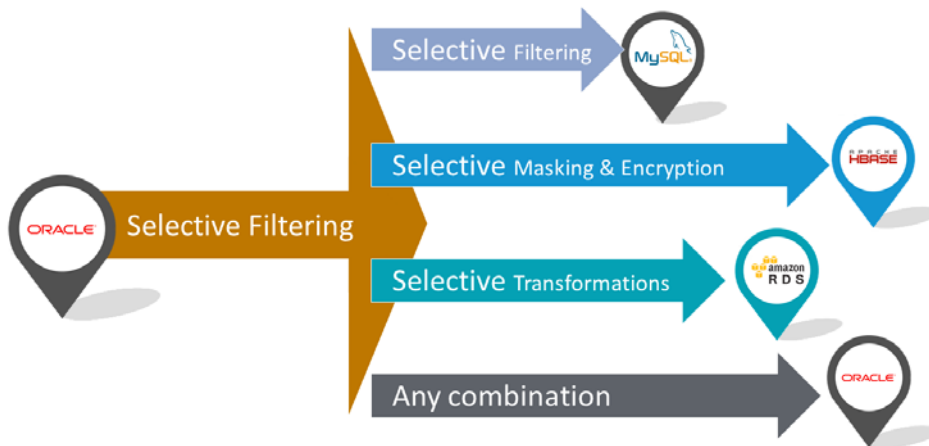
Flexibly transform and mask regulated data

Griddable.io automatically transforms data across heterogeneous combinations of source and destination databases. All data transformations are performed in grid

services without impacting the source data, eliminating the need to manually copy and transform.

Declarative Policy Language radically simplifies scripting

The Griddable.io platform provides a powerful and expressive JSON-like language to include everything needed to configure the grid, filter data, or define data transformations and



Griddable.io performs data filtering and transformations at the source or destination, based on policies. This allows selective filtering and masking at the source to protect private data and further transformations as needed at destination analytics platform.

masking. Policies are declarative, allowing the data architect to focus on the results and not the mechanics of copying, filtering, masking and encrypting data. This saves substantial time and effort, replacing customized database or ETL scripts for single function integration tools.

The Visual Policy Designer UI provides an easy way to browse source schemas, select data filters for synchronization by table, row, or column, and apply transformation or masking. The policy code is automatically generated to be applied as-is or further modified for a seamless experience between CLI and UI operation.

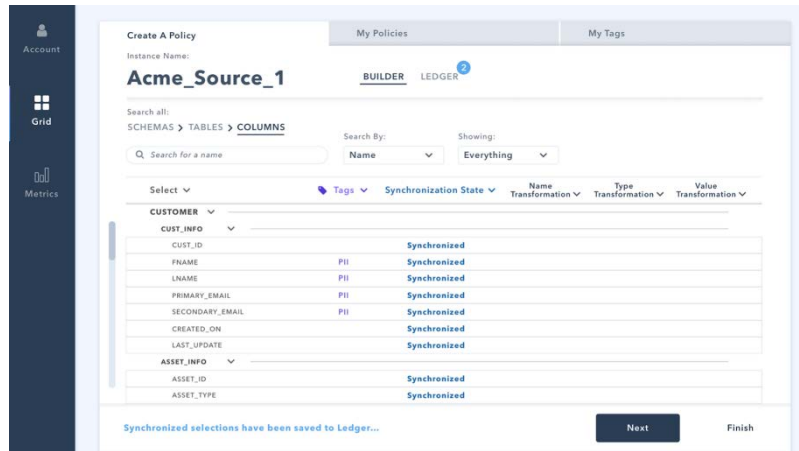
Griddable's cloud-first architecture

Griddable.io uses a loosely coupled distributed architecture of relays and consumers to synchronize and transform data. Relays interface with source databases, posting changes to an in-memory pub/sub buffer. Consumers read changes from relays and enter them in their destination datastores.

Between relays and consumers is the Griddable.io Change History Server, which archives changes and makes them available later. If a consumer falls behind due to load or network delay, the Change History Server provides required events without imposing delay on the relay, other consumers, or the source database.

The benefits of this loosely coupled architecture are reliability and scale. Consumers can pull changes from any relay, or the Change History Server, and relays are free to operate at optimal speed. There are no limitations in the system like the capacity of the file system to hold change data files. Both relays and consumers scale by simply adding additional relays or consumers and partitioning the load across them.

For further information on Griddable's cloud-first architecture, see the Griddable Architecture white paper, available on the Resources tab of company web portal.



Griddable includes a visual policy designer to choose the exact schemas, tables, and columns to be synchronized and transformed.

Using policies to meet privacy and regulatory requirements

Many emerging privacy regulations such as GDPR require personal or sensitive data to be anonymized or remain in the locale of origin. Isolating personal data throughout the enterprise and limiting its distribution is extremely difficult and when accomplished with traditional database ETL and backups is nearly impossible to demonstrate to auditors.

Griddable.io provides complete control on how personal data is synchronized throughout the enterprise without reorganizing or disrupting source databases by using policies to mask or encrypt private data. Policy tags can be defined to apply these controls to data subsets that match certain criteria. Policy definitions are hierarchical which enables data governance teams to set master controls that are then inherited by policies defined database administrators or data integration specialists.

griddable.io

2540 North First Street, Suite 201
San Jose CA 95131 USA
Phone 669.284.2143
www.griddable.io

© 2018 Griddable, Inc. All rights reserved. Griddable is a registered trademark of Griddable in the United States. All other company and product names may be trade names or trademark

- 1 100 Data And Analytics Predictions for 2021, Gartner Group and Tealium, 2017
- 2 State of Cloud Analytics in The Enterprise, Louis Columbus , Forbes, March 11, 2017
- 3 Data Protection Insights, Peter Eicher, Computerworld, July 11, 2013