DATA
PRO
— INSIGHTS —

DISCOVER THE NEXT GENERATION OF ETL

Enterprise data made easy

equalum

Contents

The next generation of ETL
Enterprise data made easy

In today's increasingly digital world, data has become a key factor in business success, with how you process it now a critical question facing many. This series of e-guides has been designed specifically to supply data pros with the very latest information on data management technology, providing them with the knowledge needed to support a truly data-driven, digital business.

In this e-guide, we'll examine the current and projected challenges with ETL and other forms of data processing, before exploring the next generation of ETL and learning how this can overcome these issues. We'll also look at the problems faced by data-driven business leaders, which will drive data pros like you to find new, enterprise grade solutions to data management.

equalum

The future is here

With 90% of the world's data having been created in the past two years, conventional extract-transform-load (ETL) technology simply can't keep up, and data pros are increasingly having to retrofit traditional batch-based ETL processes in an attempt just to keep pace. With next generation ETL technology like data streaming and data beaming though, businesses can now seamlessly move operational data to real-time analytics environments and digital customer touchpoints.

The age of big data

ETL, the process of pulling data out of one data source and placing it into another, is still the most common approach to data ingestion. But as data volumes explode - fueled in part by business users increasingly demanding continuous access to insights - those working with traditional batch ETL process have begun to face a number of technical challenges, including:

❯ **Working with legacy software, like IBM InfoSphere DataStage or Informatica**, that doesn't integrate with new technologies used within businesses (including open source software)

❯ **Inability to properly leverage the performance benefits of CDC** (change data capture) technology, which minimises the data that must be moved

❯ **Quickly migrating on-premises data systems to the cloud** without data loss

You're not alone

It isn't just data pros that are facing a challenging future though. To enable real-time analytics, operations and customer experiences, data-driven leaders everywhere are craving faster, richer data than ever before. As a result, **enterprises are increasingly being forced to address data-based issues, including:**

Deciding when it's time to let go of legacy technologies

How to speed up time-to-insight when making critical business decisions

Keeping both the business and technical sides of the company happy

equalum

The promise of possibility

The opportunities that big data provides are almost endless - from more efficient industrial processes to more personalised customer experiences at scale, and better fraud monitoring and detection - yet companies around the world are struggling to make the most of them because of legacy ETL processes.

ETL and its technical challenges

At the heart of data movement and processing for over 30 years, traditional ETL solutions are being pushed to their limits.

For instance:

- Increasing data volume can make previously reliable ETL process unstable or unacceptably slow
- Increasing batch frequency (e.g. from daily to hourly) can introduce overhead on the underlying database or app
- Schema evolution can lead to costly maintenance and custom coding needs
- Legacy ETL solutions only provide limited support for newer database technology like PostgreSQL and Cassandra, and non-database stores like IIoT control systems

However, answers to these challenges have now been found.

equalum

**02
About
ETL**

Next generation ETL

What exactly is the next generation of ETL? Well, while traditional implementations are unwieldy, cost-ineffective and don't provide end-to-end solutions, next gen ETL is not only infinitely scalable, but also allows users to move data in real-time from any source to any target, reducing the reliance on cumbersome implementation and maintenance that defined previous processes.

Key features

The next generation of ETL is characterised by some or many of the following:

- **Leveraging open source data frameworks and technologies** like Hadoop, Spark and Kafka to move, process or stream data in real-time

- **Minimising coding and setup time** thanks to easy-to-use interfaces

- **Reducing data being moved or processed** through CDC technology

- **Enterprise grade management and monitoring**

- **Flexibility to work across a huge array of data formats**, sources and target applications

Examples

There are some next gen ETL processes that bring together two or more of these new features. Data streaming for instance is an emerging approach that leverages open source technology like Kafka and Spark to move high volumes of data at speed, in effect at real-time. Another solution is end-to-end ETL, or data beaming. This combines real-time data streaming with CDC and a full technology, zero-coding wrapper that optimises setup, management and monitoring of multiple data pipelines.

The numbers don't lie

15x increase in efficiency relative to traditional ETL

Robust and fully-optimised, technology like data beaming boosts efficiency relative to a legacy ETL process by up to 15 times - that is, the same amount of data is ingested up to 15 times faster or requires 15 times less computing resources.
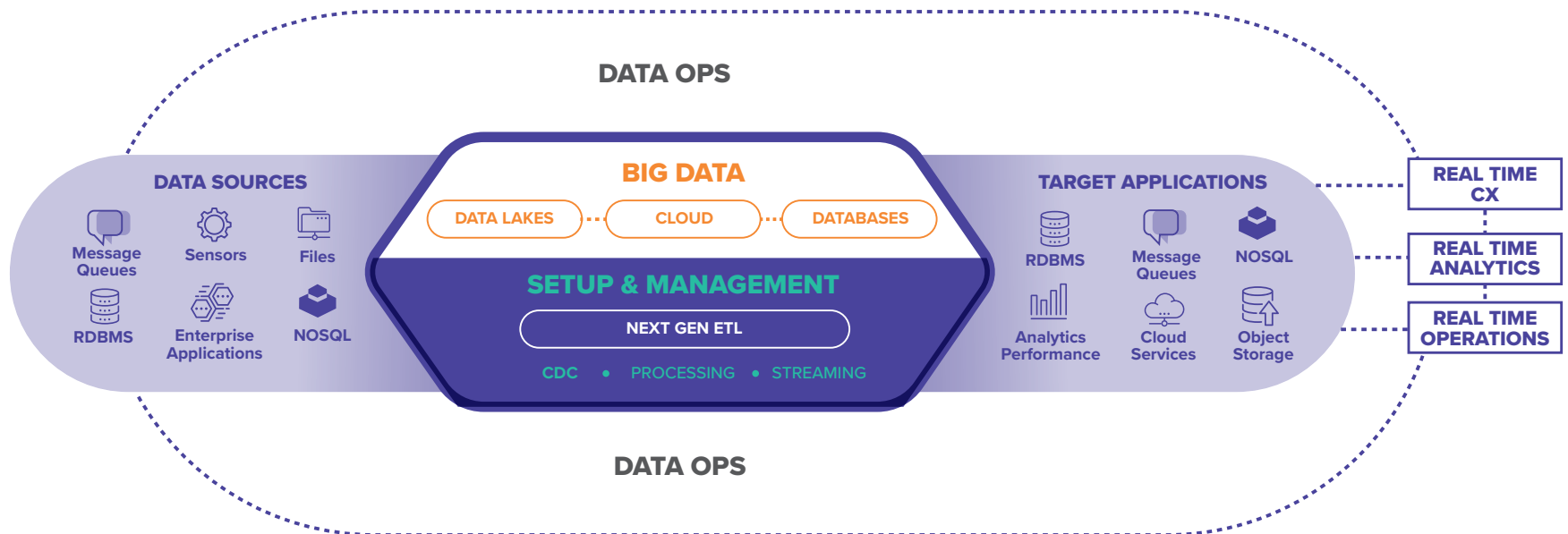
3x faster than traditional ETL tools

The **zero-coding** approach ensures that development and implementation time for similar logic is achieved 3x faster than with traditional tools.

equalum

How it works

Next gen ETL works in a similar way to its traditional predecessor,
but thanks to the very latest data tech, vital improvements have been made.

Ingestion

Traditional ETL processes are increasingly bug-prone, regularly require custom integration from every file type, often fail under high data volumes and, in addition to only supporting batch updates, don't efficiently handle complex schema. In contrast, next generation ETL streams data from any file format to any destination the instant it's created, meaning you can now correlate data sources from across your business for real-time insights.

Processing and delivery

As data volumes increase, traditional ETL can become unstable and slow. Next generation tech leverages the scalability of open source big data frameworks like Spark and Kafka to dramatically improve the performance of existing batch processes - which means your business can take advantage of increased data volumes to deliver improved performance while minimising impact to your systems.

Setup, management and monitoring

Engineered from the ground up for robustness and reliability, next generation ETL is the perfect plug-and-play solution for businesses in all industries. With intuitive, user-focused, drag-and-drop interfaces and a no-coding approach, it gives you the ability to configure, maintain, and derive insights quickly and easily.

Business applications

With best-in-class security, monitoring and fault tolerance, next generation ETL has been designed to help businesses get value from their data, fast. In fact, whether it's streaming mechanical sensor data to optimise just-in-time manufacturing great delivery or supporting real-time loan analytics, it's already proving its value for pioneers in manufacturing and financial services.

A brave new era

According to Gartner research, more than 50% of big data projects are currently falling short of expectations for business value. Built to address this challenge, next generation ETL technologies are designed specifically to deliver a complete, enterprise-ready solution.

Prepare for tomorrow. Succeed today.

Representing the future of data ingestion, the next generation of ETL technology harnesses the power of open source frameworks in end-to-end, enterprise-ready solutions, meaning data teams can now:

- Guarantee data accuracy
- Integrate new technologies and retire older ones
- Deliver data at scale and in real-time thanks to CDC and high velocity data streaming
- Build and maintain database infrastructure
- Leverage the latest cloud technologies and capabilities for storage and data availability
- Collaborate with technical and non-technical parties to achieve database goals
- Keep data secure and confidential

All this means that you can now satisfy the needs of your business - especially data-driven leaders in IT, operations, marketing and other areas - by becoming truly data-driven, not only powering real-time analytics, but also speeding up time to insight and driving more personalised, customer centric experiences.

> "In media and entertainment, visibility into the impact of digital marketing spend is crucial for driving ROI. Equalum's Data Beaming platform helped us connect the dots by correlating our display and social media investments with TV viewership — in real-time. The result is significantly more data-driven and dynamic allocation of our marketing dollars. Equalum powered for us in a single week what would have taken years to build in-house."
>
> **Zachary Mannon, VP Product, CW**

Develop your ETL knowledge

As well as being a great alternative to more traditional technologies, by adding in CDC and zero-coding setup and management, next gen ETL solutions from organisations like Equalum can also help solve any problems you may be having embracing new open source technologies or dealing with demands for more and more data streaming pipelines.

To learn exactly what this might mean for your business:

Check out the other e-guides
in our 'Enterprise data made easy' series

....

Learn more about the 'Death of ETL'

....

Connect with one of our data experts to get a live demo of data beaming in action and find out just how the technology works

Get in touch today to find out more

Visit: **https://equalum.io**
Email:

equalum