



ETL or ELT?

Choose the Right Process
for the Right Outcome

The evolution of data delivery.

Whether you have a well-established practice for data delivery or you're just starting to formalize your strategy, you share a common goal with everyone in your role: to provide data to your organization in an efficient, reliable, and repeatable way.

For years, there was a standard order of operations: ETL, or extract, transform, and load. Given the amount of data involved and the tools available for working with it, it made perfect sense to operate with that model. And it still works well for certain use cases today.

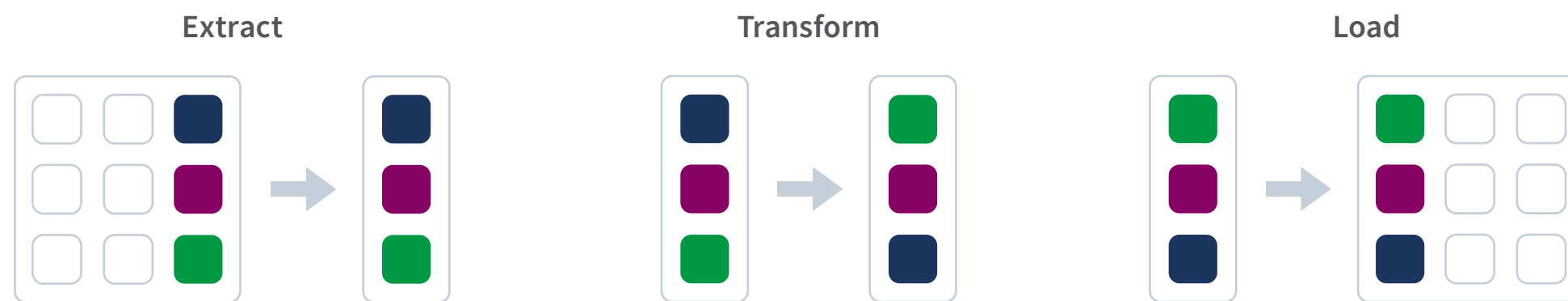
But as the volume and speed of data have multiplied, as the pace of business has increased, and as new integration tools have arrived on the market, another option has gained in popularity: ELT, or extract, load, and transform. The ELT model offers benefits that aren't available in ETL, especially for situations where speed is of the essence.

Which model should you use, and when? This eBook addresses that question.



The advantages of ETL.

ETL is the set of processes involved in extracting data from one system, transforming it, and loading it into a target repository. It's a traditional workflow for cleaning, enriching, and transforming data before integrating it for use in data analytics, business intelligence, and data science.



In ETL, transformation is performed in a staging area before loading it into the data warehouse. The entire data set must be transformed before loading – but once the data is loaded, systematic and accurate analysis can take place immediately.

The key benefits of ETL are:

- **Maturity.** ETL is a well-supported and well-understood function.
- **Faster analysis on a single, predefined use case.** Because the dataset has already been structured and transformed, analysis can happen immediately.
- **Transparency.** In some cases, understanding data lineage is more straightforward, because data transformation is performed inline, as opposed to at a later date.

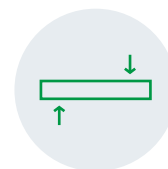


Where ETL falls short.

While ETL remains a solid option for certain use cases, it does have drawbacks, especially as the volume of data and the speed of business increase. ETL is:



Time-consuming, especially if you have developers write the code. The process of transformation is driven from a set of requirements, and those requirements can take weeks (or even months) to map to. After transformation, loading the data can be almost as time-consuming. True, there are tools for ETL that reduce or eliminate coding, and they speed up the process. But these tools don't offer much help tuning a running data integration pipeline.



Rigid. ETL doesn't support ad-hoc workflows. For ETL to work, all the parameters have to be mapped out and then implemented. Whenever a new use case arises, the entire ETL process has to be repeated.



Brittle. Altering any step in an ETL workflow can break other workflows, since there are often precedent dependencies. Put another way: When the first workflow needs to run and complete before the second can be started, if the first breaks, the second never happens.



Opaque. The transformation process is not transparent, since only the people transforming the data have access to it.

The trouble with batch processing.

There's another problem with the ETL approach: the limitations of batch processing. Generally, ETL processes operate on a predetermined schedule – every minute, hour, day, or week, depending on the use case. They can also run in response to an external trigger or event, but that's less common.

A scheduled ETL process operates in “batch mode,” where the frequency tends to be dictated by constraints like the timeliness of data required; the time it takes to extract, transform, and load the data; and the transaction volume between batches. Overall, this process works well. But when the volume of data reaches a certain point, or when it takes longer to perform the ETL than the desired batch schedule, things begin to break down.

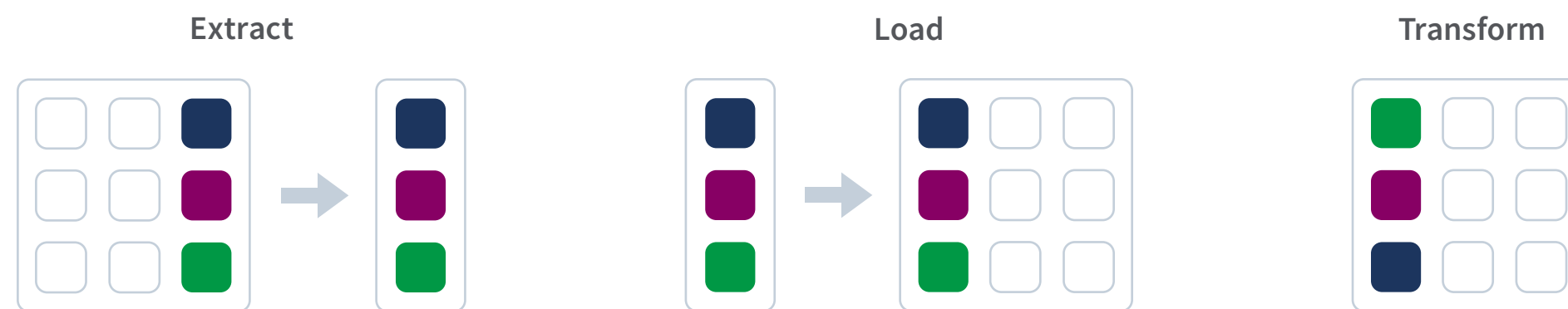
For example: A bank may need to update their data warehouse with a million transactions every 10 minutes. But what if it takes 15 minutes to extract, transform, and load the batch? Extending the frequency to 20 minutes isn't the answer. Because when 20 minutes have passed, the volume of data has also increased to two million rows – which will take even longer to extract, transform, and load. Clearly, in scenarios like this, another approach is required.



The arrival of ELT.

Today's organizations have to process a tremendous volume of data arriving in a wide array of forms. And if they want to compete in the real-time economy, they have to do it as quickly as possible. That's exactly the kind of scenario that ETL can't handle – and it's why the market is shifting toward ELT.

ELT stands for “Extract, Load, and Transform” and describes the set of data integration processes to extract data from one system, load it into a target repository, and then transform it for downstream uses such as business intelligence (BI) and big data analytics.



In the ELT process, data transformation is performed on an as-needed basis within the target system. First, a data extraction tool pulls data from a source or sources, sometimes storing it temporarily in a staging area in a database. Then the data is placed into the target system, typically a cloud data warehouse, where it's ready to be analyzed by BI tools or data analytics tools. The data is transformed – i.e., converted to the structure or format of a data set to match that of the target system – on an as-needed basis within the target system.

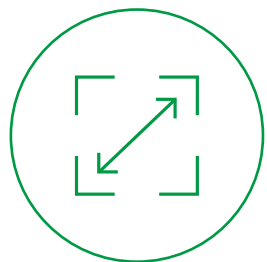
The advantages of ELT.

ELT makes the process of managing big data and cloud data warehouses more efficient and effective. Specifically, the key benefits are:



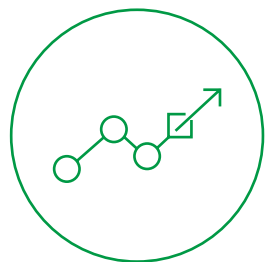
Speed, speed, and more speed.

The ELT process typically involves loading all your data into a cloud repository – and today’s cloud services allow you to quickly transform and process massive data sets.



Flexibility.

The ELT process enables you to replicate all your raw data on demand, as it’s needed for data analytics, BI, or other systems.



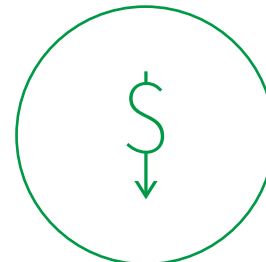
Scalability.

Modern cloud data warehouses, data lakes, and data lakehouses give you the scale to leverage all your raw data.



Transparency.

Data consumers can see what data is available almost as soon as it comes in the door.



Lower cost.

ETL workflows require an investment in on-premise hardware. Cloud-based ELT, on the other hand, usually offers a lower total cost, given that you pay for only what you use.

ETL + ELT: Why not do both?

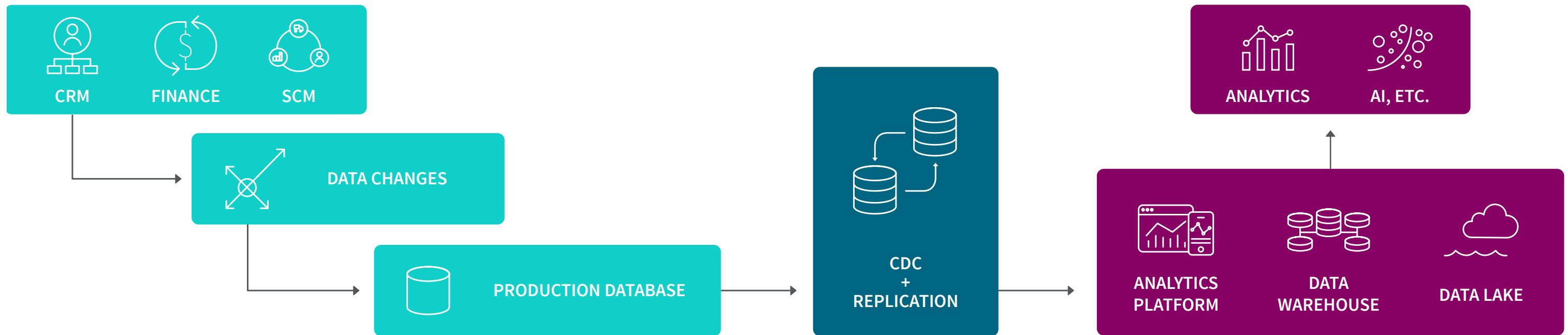
Generally speaking, the ETL process is effective for small data sets that require complex transformations and aren't needed quickly. ELT, on the other hand, is appropriate for larger structured and unstructured data sets and when timeliness is important.

Many organizations use both ETL and ELT to cover their wide range of data delivery needs, mixing and matching as business requirements dictate. For example, you could use ETL for legacy reporting requirements and ELT for real-time data pipelines based on change data capture technology (see page 9), feeding newer analytics processes like fraud reporting and machine learning.

But there's a reason that the market is heading toward ELT as the standard. The volume, velocity, and variety of data have grown so massively in recent years that ELT has replaced ETL in many instances as the de facto pattern for data movement, especially in scenarios like cloud data migration, data warehouse and lake ingestion, and MLOps (the continuous delivery and automation of data pipelines using machine learning).



Taking ELT further with change data capture.



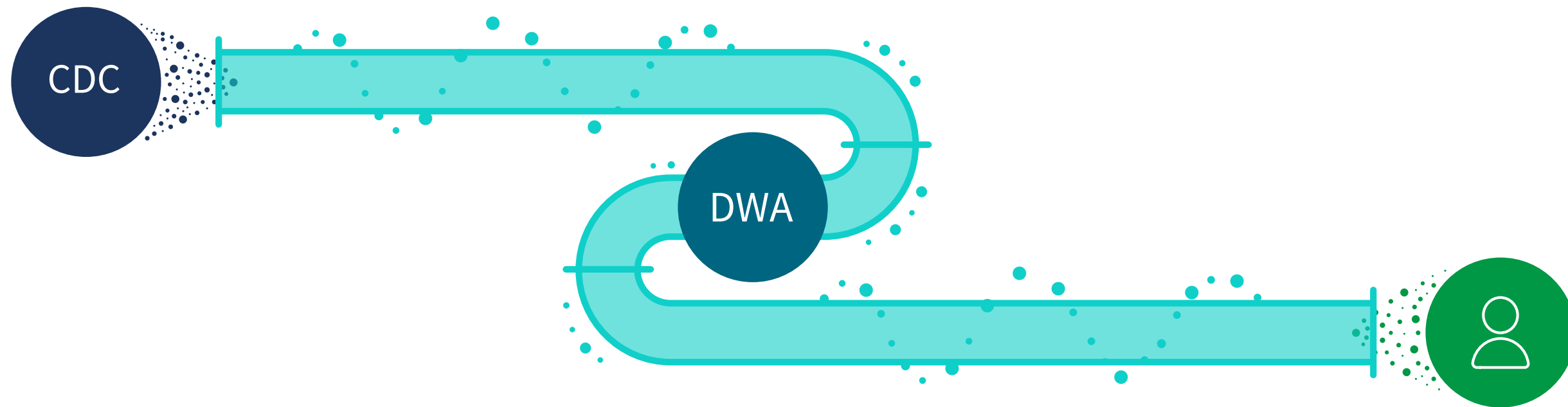
The ELT approach, as we mentioned, is notable for its speed. Let's go back to our bank example from page 5 and reimagine it with ELT instead of ETL. When the time to extract, transport, and load the data is 15 minutes, using ELT gets the data into the target much more quickly – even within seconds. The bank then has the option to schedule a bulk transform process (ideally when more resources are available) or defer the transformation until the data is consumed.

But we can do better than that by adding change data capture (CDC), a technology for replicating changes to data as they occur. CDC doesn't operate on a batch schedule like ETL. Instead, it's triggered every time a change occurs to the data source. In our bank example, the EL (no T yet) process would run for every transaction, with a minimal amount of data being transmitted over the wire. There would be no waiting to process a million rows. Effectively, the extract and load processes would happen in real time.

What can change data capture do for your business?

Get an overview on our [CDC page](#).

Building a real-time data pipeline with data warehouse automation.



Just as CDC automates the “extract” and “load” portions of the data pipeline, data warehouse automation (DWA) handles the “transform” part of the process. DWA eliminates error-prone manual coding and automatically generates the commands, data warehouse structures, and documentation necessary for designing, building, and maintaining a data warehouse program – helping you save time, reduce costs, and reduce project risk.

When you have seamless integration between your change data capture solution and your data warehouse automation solution, you enable real-time ELT. In other words, you create a real-time data pipeline. That sets you up to achieve Active Intelligence, a state of continuous awareness where everyone in your organization is empowered to act on data in the moment, as it changes. That’s always been the end goal (and, for decades, the pipe dream) of business intelligence – and it’s becoming more and more necessary for competitive edge in today’s market. Now the technology is here.

How does data warehouse automation revolutionize data delivery?

Get the details at our [DWA page](#).

ELT SUCCESS STORY

Jaguar Land Rover delivers data-ingestion-as-a-service with Qlik®.

THE CHALLENGE

In a complex international market with varying regulations – compounded by supply chain disruptions – Jaguar Land Rover needs access to real-time data to make immediate, informed decisions. But valuable data was locked inside multiple mission-critical source systems, both on-premises and in the cloud.

THE SOLUTION

The enterprise data team decided to adopt a data lake and use change data capture/streaming technology to fill it. After evaluating multiple solutions, Qlik Replicate™ was identified as the most efficient (and indeed the only) way to transfer data to a broad variety of storage layers within the lake.

THE RESULTS

Ingestion-as-a-service is now a capability for use throughout the business, with a new replication taking only 15 minutes to establish. SAP data is now exposed in external-facing JLR websites for different countries, allowing a quick response to the regulatory demands of global markets.

[Get the full story.](#)



“We need to have the data available to be able to quickly answer challenges when they arise. Qlik was the only solution that could provide breadth of coverage on all the key systems within JLR’s estate.”

Michael Cockbill
Technical Product Manager

Speed, volume, and competitive edge: It's time for ELT.

As the economy changes, as data evolves, and as new tools come on the market, data strategies have to shift, too.

While there are still good use cases for ETL, and while it makes sense to keep it in your mix, ELT offers important new benefits, especially when it comes to integrating massive volumes of data at speed. And by enabling you to create a pipeline for acting on data in real time, ELT can be an important factor in positioning your organization for competitive edge.

Interested in trying your hand at moving data in real time using an award-winning ELT solution? Test-drive Qlik Replicate, our solution for change data capture.

[Free Trial](#)

Qlik's vision is a data-literate world, where everyone can use data and analytics to improve decision-making and solve their most challenging problems. Our cloud-based Qlik Active Intelligence Platform delivers end-to-end, real-time data integration and analytics cloud solutions to close the gaps between data, insights and action. By transforming data into Active Intelligence, businesses can drive better decisions, improve revenue and profitability, and optimize customer relationships. Qlik does business in more than 100 countries and serves over 38,000 active customers around the world.