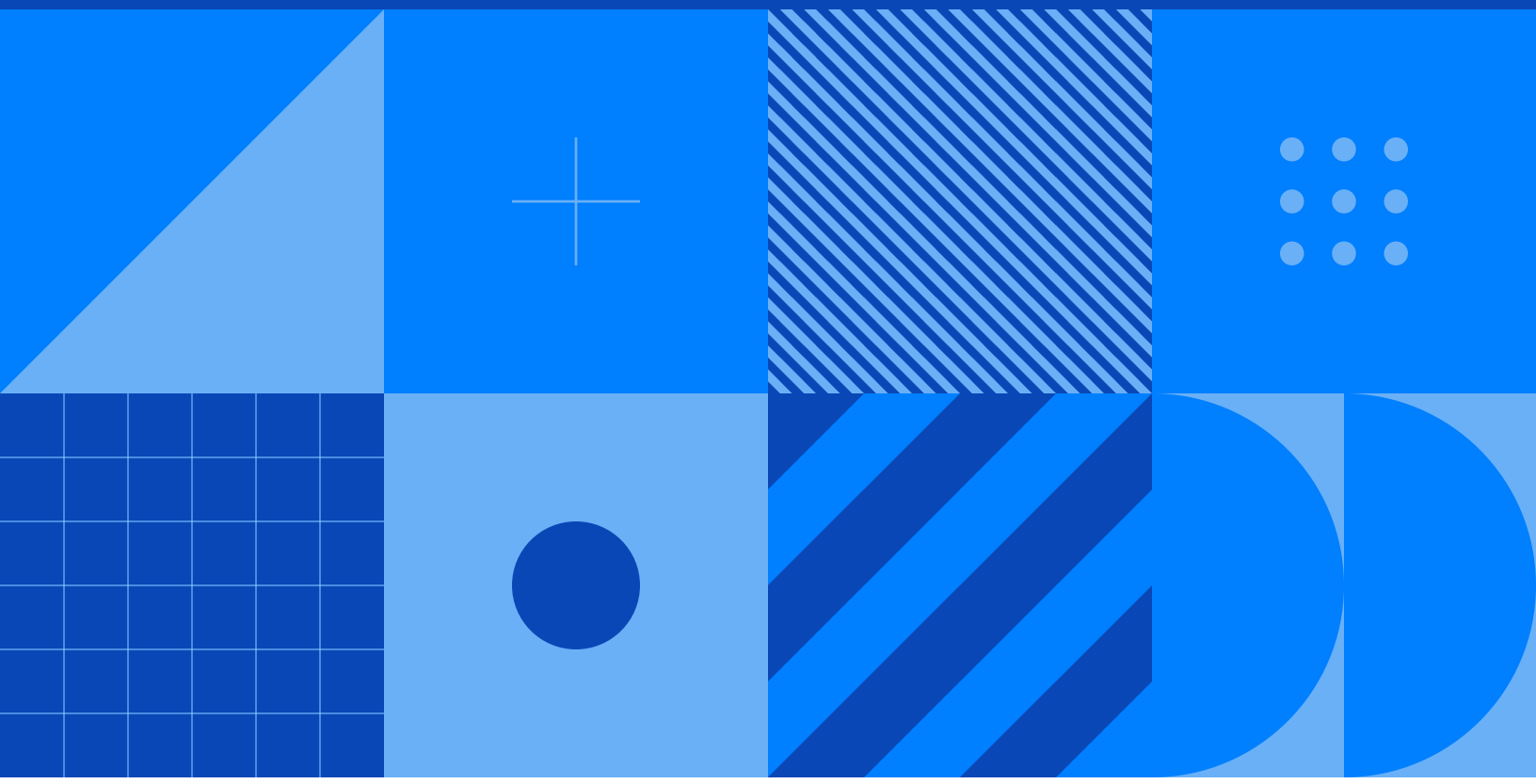
Get Started With Automated Data Integration

Struggling with data integration?
Consult the following seven steps to better analytics.



Fivetran

Learn how Fivetran data integration powers business intelligence at fivetran.com

Automated data integration, combined with a cloud-first, fully managed data stack, promises not only reliable data connections but also savings of time, labor, money, and morale. Furthermore, automated data integration offers the possibility of democratizing data fluency and building a data-driven culture. The following steps discuss practical measures for beginning your organization's path to better and more accessible analytics.

---

1. Assess Your Needs

You might not want to outsource your data operations to a third party or a cloud.

The first and most obvious reason is that your organization may be very small or operate with a very small scale or complexity of data. You might not have data operations at all if you are a tiny startup still attempting to find product-market fit. The same might be true if you only use one or two applications, are unlikely to adopt new applications, and your integrated analytics tools for each application are already sufficient.

A second reason not to purchase a modern data stack is that it may not meet certain performance or regulatory compliance standards. If nanoseconds of latency can make or break your operations, you might want to avoid third-party cloud infrastructure and build your own hardware.

Otherwise, if your organization is of a sufficient size or maturity to take advantage of analytics, and data refresh cycles of a few minutes or hours are acceptable, proceed.

2. Migrate or Start Fresh

Data integration providers should be able to migrate data from old infrastructure to your new data stack, but the task is a notorious hassle because of the intrinsic complexity of data. Whether your company decides to migrate or simply start a new instance from scratch depends heavily on whether historical data is important to you.

If your organization has already purchased or contracted for products or services, it may be costly to end those contracts. Beyond money, familiarity with and preference for certain tools and technologies can be an important consideration.

Take care that prospective solutions are compatible with any products and services you intend to keep.

3. Evaluate Cloud Data Warehouse and Business Intelligence Tools

You will have to compare and contrast solutions for every part of the data stack. Start a little downstream and think about what features you will need in a cloud data warehouse and business intelligence tool.

Cloud data warehouse features to consider include:

- Centralized vs. decentralized data storage
- Elasticity – can the data warehouse scale resource use up and down quickly? Are compute and storage resources independent or tightly coupled?
- Concurrency – can the data warehouse accommodate multiple simultaneous tasks?
- Load and query performance
- Data governance and metadata management
- SQL dialect
- Backup and recovery support
- Resilience and availability
- Security

Business intelligence tool features to consider include:

- Seamless integration with cloud data warehouses
- Ease of use and drag-and-drop interfaces – especially helpful if you want to create a data-driven culture across your company
- Automated reporting and notifications
- Ability to conduct ad hoc calculations and reports by ingesting and exporting data files
- Speed, performance and responsiveness
- Modeling layer with version control and development mode
- Extensive library of visualizations

Make sure any data warehouses and BI tools you evaluate are compatible with each other. It also pays to carefully review a range of perspectives on different tools. Publications like Gartner often aggregate such information. Read before you leap!

4. Evaluate Data Integration Tools

There are many important characteristics to consider with regard to data integration tools.

A short list of what you should look for:

- Customization and configurability vs. ease of use and accessibility
- Reliability and performance of the software
- Quality and responsiveness of customer support teams
- Number and type of data sources covered
- Costs and payment plans

Many publications offer aggregate reviews and ratings of data integration tools, as they do for data warehouses and business intelligence tools. Be sure to comparison-shop, and make sure all parts of your proposed data stack are mutually compatible.

5. Calculate Total Cost of Ownership and ROI

The modern data stack promises substantial savings of time, money and labor. Compare your existing data integration workflow with a range of possible candidates.

Calculate the cost of your current data pipeline, which might require a careful audit of prior spending on data integration activities. You'll need to consider the sticker price, costs of configuring and maintaining, and any opportunity costs incurred by failures, stoppages and downtime. You should also consider the costs of your data warehouse and BI tool.

On the other side of the ledger, you will want to evaluate the benefits of the potential replacement. Some may not be very tangible or calculable (i.e., improvements in the morale of analysts), but others, such as time and money gains, can be readily quantified.

6. Establish Success Criteria

What should your analytics practice look like if you have successfully implemented a modern data stack?

Key criteria include:

- Time, labor and monetary savings compared with the previous solution
- Expanded capabilities of the data team
- Successful execution of new data projects, such as customer attribution models
- Reduced turnaround time for reports
- Reduced data infrastructure downtime
- Higher rates of business intelligence tool adoption within your organization
- New metrics that are available and actionable

7. Set Up a Proof of Concept

Once you have narrowed your search to a few candidates and determined the standards for success, test the products out in a low-stakes manner. Most products will offer free trials for a few weeks at a time.

Set up connectors between your data sources and data warehouses, and measure how much time and effort it takes to sync your data. Perform some basic transformations. Set aside dedicated trial time for your team, and encourage them to stress-test the system in every way imaginable. .