



Powered by Intel®

## Turning data into actionable insights

Four steps for leveraging Smarter Data to becoming  
an analytics-driven, Data-Centered organization.

# Contents

## The Data Pipeline pg. 3

### **Step 1: Ingest** pg. 4

### **Step 2: Prepare** pg. 5

### **Step 3: Analyze** pg. 6

### **Step 4: Act** pg. 7

## **Conclusion** pg. 8

**We all know analyzing data helps us make more informed business decisions. But it's easier said than done — especially given the volume of data that exists, and how quickly it's expanding. The good news: we're here to help.**

This step-by-step guide will teach you how to make your business more data-driven by incorporating analytics and AI to drive more consistent, more actionable business insights about your customers, your competitive landscape, and even your internal processes.

 **Only 8% of companies have been able to productionalize and scale their analytics programs, according to McKinsey.**

Because while it can feel like a tall (and stressful) task, automating processes, increasing efficiency, and continually innovating can help you stay ahead of your competitors and seize opportunities in this new data economy.

## INGEST

Creation

Transmission

Ingestion

## PREPARE

Integration

Staging

## ANALYZE

Cleansing & normalization

Modeling & experimenting

## ACT

Training & distillation

Visualization & deployment

Archival

## Understanding your data pipeline

First, let's talk about the bigger picture. (Quite literally, the big picture on this page.) This is a data pipeline, sometimes also known as a data lifecycle, as defined by our partners at Intel®.

### ⬡ Your analytics strategy should include:

- Defining the business problem(s) you want to solve with analytics
- Making sure you have the right team members, experts, and resources
- Obtaining the right data for the specific business problem you want to address
- Securing infrastructure that supports data processing in a timely manner
- Outlining the steps necessary for solving the problem
- Continuously improving your practices to achieve desired results

Finding the most valuable business insights requires an end-to-end analytics strategy across your own data pipeline.

So, it's important to consider every stage of your pipeline when forming that strategy — making adjustments and optimizations where they'll have the most impact.

So, where should you start?

**Note:** This process is not linear, or even one-dimensional. There are usually subroutines and data morphing along the way, and even some overlap between stages. But for the purpose of this document, we will go step-by-step.

## INGEST

Creation

Transmission

Ingestion

## PREPARE

Integration

Staging

## ANALYZE

Cleansing & normalization

Modeling & experimenting

## ACT

Training & distillation

Visualization & deployment

Archival

## STAGE 1: Ingest

### Getting to the sources

It seems obvious, sure, but the first step in analyzing data is collecting it. That means extracting it from various sources, detecting what's changed, and moving it from those sources into a system where it can be stored and analyzed.

**TIP:** The path to Data-Centered isn't always easy, and this stage can quickly become very time-consuming. So, allow yourself some extra time.

There are many approaches you can take to collect raw data, based on its size, source, structure, and latency:

### Data from application events

This data, such as log files or user events, is typically collected in a push model, where the app calls an application programming interface (API) to send the data to storage. It can reveal user trends and provide valuable business insights.

### Streaming data

This data is made up of a continuous flow of small, asynchronous messages. With telemetry, data can be collected from geographically dispersed devices, like edge sensors. It's often used as input for machine learning tasks.

Collecting data from user devices, like when an app is opened or when it crashes, can provide valuable information about usage and code quality.

We're seeing a trend toward more real-time analytics of streaming data, like cybersecurity monitoring, which demands higher compute performance at lower power.

### Batch data

This data tends to consist of large numbers of files transferred and stored in bulk, or in relational or NoSQL databases. It can be located on-premises or on other cloud platforms, and ingesting it requires high aggregate bandwidth between the sources and the target.

AI workloads are often needed to digest this type of data.

## INGEST

Creation

Transmission

Ingestion

## PREPARE

Integration

Staging

## ANALYZE

Cleansing & normalization

Modeling & experimenting

## ACT

Training & distillation

Visualization & deployment

Archival

## STAGE 2: Prepare

Bringing all that data together

Data's often complex and layered and messy. It comes in all different structures, sizes, and speeds, and has to be cleansed, normalized, and prepared before it can be stored in a way that works best for your workload.

**TIP:** Plan to build in the storage, compute, and network capabilities you'll need both now, and in the future, to avoid a potential bottleneck.

It might help to set criteria for tiering your data depending on its urgency and usefulness, so you can ensure the most cost-efficient memory or storage technology for each tier.

The goal here is to have your data in a format that allows it to be easily accessed and queried. This means your compute performance, memory capacity, and network bandwidth have to be aligned, so the data your application needs is as close and as accessible as possible.

**TIP:** Work with data gravity as much as possible — bringing compute to the data, rather than always moving it to a centralized compute resource.

Because if your applications and programs don't have the data they need when and where they need it, they won't be able to run your analytics properly.

## INGEST

Creation

Transmission

Ingestion

## PREPARE

Integration

Staging

## ANALYZE

Cleansing & normalization

Modeling & experimenting

## ACT

Training & distillation

Visualization & deployment

Archival

## STAGE 3: Analyze

### Making sense before you make change

This might go without saying, but we'll say it anyway: focus on the use cases that will deliver the most value for your organization.

That means taking it easy, at least at first. There's no need to go straight for deep learning when simpler analytics methods have yet to be used to find your data's potential. So, take a deep breath and start small.

### Preparing your infrastructure for analytics

The most important part of preparing for larger analytics and AI workloads is making sure your storage and network capabilities can keep up.

Larger volumes of real-time data for analysis generally need scale out, rather than scale up infrastructure to keep pace with performance as data grows, so it's best to think not just about what you need today, but how you'll need to scale your capabilities over time.

If you're used to CPUs, you may start to feel pressure to invest in GPUs to support your heavier workloads. So, keep in mind that while GPUs work well running large numbers of computations on small amounts of data, they're not suited to database operations that require a lot of data to be moved back and forth. This type of workload – common in analytics – may be best served by a CPU-based environment like the one you're already running.

And don't forget about memory. Many real-time analytics use cases rely on in-memory databases that can quickly run out of capacity.

## INGEST

Creation

Transmission

Ingestion

## PREPARE

Integration

Staging

## ANALYZE

Cleansing & normalization

Modeling & experimenting

## ACT

Training & distillation

Visualization & deployment

Archival

## STAGE 4: Act

### Finding the insights that drive smarter action

At last, you've made it to the pipeline stage you've been waiting for — the one where you can pull meaningful insights from the data you've worked so hard to integrate, store, clean, and process.

**TIP:** Stay on top of data maintenance and governance over time. Monitor your entire pipeline and data usage to track trends, bottlenecks, and opportunities, and make upgrades regularly to ensure optimal function, and of course, ROI.

Here, you can use in-depth exploration and visualization to better understand your results, and turn them into something you can actually use to make an impact.

It's best to begin with your existing infrastructure and business intelligence (BI) tools, then evolve as your capabilities and resources expand. Data scientists and business analysts can help you make the most of your findings, as well.

Data visualizations, dashboards, and reports can help business users understand these results. Your software should be flexible enough to evolve and develop over time to meet changing needs, as well as any new analytics and AI use cases.

And in this way, the end of this journey comes back to the beginning: choosing the right use case, right team, and right data to find even more value within.

INGEST



PREPARE



ANALYZE



ACT



Conclusion

As your data grows, the opportunities you have to make the most of it all grow, too. By taking the approach we’ve outlined — by considering your data pipeline and your needs holistically — you can go from data center, to Data-Centered.

Let’s get Data-Centered

- Lenovo’s data and analytics solutions deliver high performance data management at a faster time to value. So you can take control of the growing volume, variety, and velocity of your data, and take your business forward without leaving opportunities behind.
- Increase visibility and security
- Improve resource management
- Identify actionable insights faster
- Gain efficiency and control

[Explore products](#)

[Get in touch](#)

Lenovo and the Lenovo logo are trademarks or registered trademarks of Lenovo.

Intel and Intel Inside is a trademark of Intel Corporation or its subsidiaries in the U.S. and/or other countries. Other company, product and service names may be trademarks or service marks of others.

© Lenovo 2020. All rights reserved.