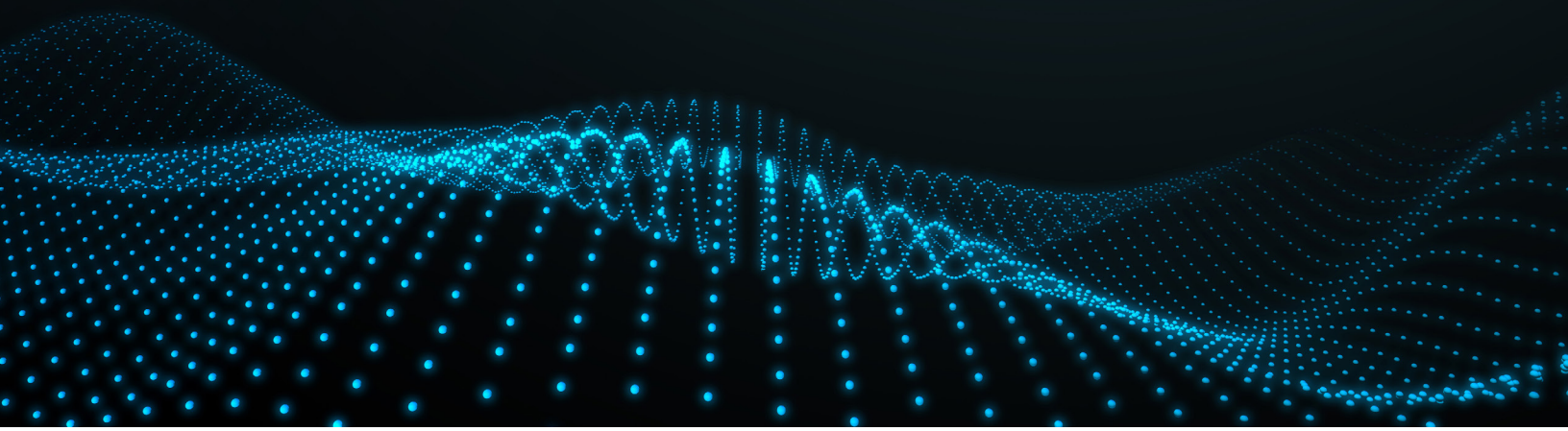
6 Common Problems with Vehicle Data Lakes (And How to Solve Them)



Vehicle manufacturers have invested heavily in engineering data lakes, but many aren't seeing the results they hoped to achieve.

Between 60% and 85% of all data lake projects will fail. Find out why — and how to make sure your data analysis efforts succeed.

Companies in all industries, including vehicle manufacturing, are investing heavily in data lakes. According to IDC, organizations spent $215.7 billion on big data and business analytics in 2021, and the researchers forecast that spending will grow by an average of 12.8% every year through 2025.

A separate study from NewVantage Partners found that 97% of organizations are investing in data initiatives, and 92% said that their rate of investment in data and AI is increasing. When asked about their investment priorities, data platforms and data lakes were the top vote getter among respondents.

But despite the amount of money they have thrown at the issue, organizations aren't seeing the kind of results they hoped to acheive from their data lake projects.

Depending on which report you read, between 60% and 85% of data lake projects fail. And those failures have downstream impacts on artificial intelligence (AI) and machine learning (ML) initiatives. By some estimates, only 13% of data science projects ever make it into production. In the NewVantage survey, less than half of those surveyed said that they were able to compete on data analytics.

Vehicle data lakes

Like companies in other industries, many vehicle manufacturers have set up data lakes for their engineering and validation data in the hopes of improving quality. More than three-quarters of manufacturers say they

APTIV

want to differentiate their products through high quality and reliability, and companies understand that their data holds the key to creating better products.

Vehicle makers are also looking to data lakes to help them bring products to market more quickly. As they transition to electric vehicles and introduce more advanced driver assistance systems (ADAS), they need to innovate much more rapidly than they have in the past.

Leaders know that making better use of their data could help them achieve their goals. But privately, many of them say that their efforts so far have achieved minimal results.

In this case, some industry-specific challenges are making it difficult to convert raw vehicle data into actionable insights that achieve the desired business results. Here are six of the most common problems with vehicle data lakes, along with some suggestions for how to fix them:

1. Vehicle data lakes are designed by data management experts.

If you're designing a data lake, it seems like common sense to bring in experts that have experience with other data lakes. In the NewVantage survey, 91.9% of those surveyed said that the biggest challenges to their data initiatives are cultural factors like not having the right people. Managers look at reports like that one and decide that they will avoid those cultural problems by building a new team with lots of IT and data management experience.

Unfortunately, this approach usually results in problems — not because automotive manufacturers put too many data management experts on the team but because they include too few vehicle engineering experts on the team.

Vehicle data is extremely complex — especially in its raw state. Transforming it into a format where engineers can actually use it to answer questions about vehicle issues can take days. And because the data lake stores this data in its raw state, engineers often need to redo this work every time they want to answer a new question.

The end result is an expensive data lake that vehicle engineers rarely use.

To avoid this kind of costly mistake, you need to involve your engineering staff in the earliest stages of your data lake project. Find out what kind of questions they need to answer on a regular basis and make sure your platform is architected to answer those specific questions.

Even better is to invest in an analytics solution specifically designed for vehicle engineering data. The people creating these solutions have already done the hard work of figuring out the best data storage and management platforms for vehicle data, and they were designed to answer exactly the kind of questions that your engineers have. This approach often saves time and money over homegrown solutions.

2. Data search takes too long.

A recent survey found that engineers waste 32% of their time on non-value-added work. Of that time, about 23% is spent searching for data that they can't find.

In theory, putting all your raw vehicle data into a data lake means that you'll always have everything you need.

In practice, putting all your raw vehicle data into a data lake is about as effective as throwing all the stuff you rarely use into your garage without organizing it in any way. It's all there, but finding it takes so long that it hardly seems worth looking.

The complexities of vehicle data can make it very difficult to find what you are looking for. For example, if you want to search your test drive data for examples of times when the lane centering feature deactivated on a sunny day while a vehicle traveled beneath an underpass, you probably won't be able to find them using a normal search tool. You either need very robust data tagging capabilities, or you'll need to create some sort of custom data science application tailored to answer this sort of question.

The best solution to this problem is to store this kind of data in an architecture designed to handle vehicle data and access it with an analytics tool designed specifically for vehicle data. You can build a solution of this type on your own (usually at a high cost in time and money spent), or you can purchase a tool with these capabilities built in.

Test Coverage



Aptiv Connect Qualifier makes it easy to find and analyze vehicle data, and it can integrate data captured with non-Aptiv data loggers

3. The data hasn't been standardized.

While every vehicle generates some standard data as required by law, every make and model also has its own unique set of proprietary signals. Even within a single manufacturer, various models can differ from each other quite significantly.

In addition, most vehicle manufacturers have many different engineering teams focused on different aspects of the vehicle and different models. Those teams might all use different data loggers to collect data, and those different data loggers may generate different kinds of files.

As a result, it's very difficult — if not impossible — to compare data across different vehicle lines or data captured by different hardware.

Aptiv solves this problem in a couple of different ways. First, Aptiv Connect Qualifier can use aliases for groups of similar signals, making it possible to run analytics across different models with different signal names.

Second, Qualifier has the ability to ingest files captured with nearly any brand of data logger. You can then analyze data from all those different tools with the Qualifier web application. If you already have an existing vehicle data lake, you can feed as many of those existing files as you choose into the application, which instantly makes them easier to search and analyze.

4. The data lacks reliable reference points.

Some of the files you have in your data lake may contain streaming data from an entire test drive. Others may have a snapshot of data from a single point in time. And others may have just a subset of data.

Some of the files may have associated GPS data. Others may not.

Some of the files may be time-stamped to GMT. Some of the files may have different types of time coding. And some of the files might have no time coding at all.

Your data lake needs to have a system for syncing files in time and space so that you can analyze them accurately. The best-case scenario is to capture GPS and time codes at the time the files are generated. If that isn't possible, then you need a tool like Aptiv's Qualifier Ingest that can take existing files and transform them for analysis in a way that provides meaningful context.

5. Vehicle data lakes have too much data.

In the early days of the big data trend, many organizations were focused on getting as much data as possible into their data lakes. But they soon realized that cleaning, transforming, and adding context to data to make it meaningful takes a lot of time and effort. Put simply, the big data wasn't as valuable as they thought it would be.

Today, companies are pivoting to a slightly different data strategy. According to Gartner, "By 2025, 70% of organizations will shift their focus from big to small and wide data, providing more context for analytics and making artificial intelligence (AI) less data hungry."

We'll discuss wide data a little farther on in this paper, but small data is data that humans can understand. It can be a single data point, or it can be the results of big data analysis.

APTIV CONNECT QUALIFIER

Used by more than 70% of the world's leading vehicle manufacturers, Aptiv Connect Qualifier dramatically reduces the time it takes to find and fix vehicle problems, helping save millions of dollars. It includes Aptiv's Edge hardware that collects more signals in finer detail than any other solution and the Connect Edge software that begins analyzing that data right in the vehicle.

Qualifier transmits this pre-processed data to the cloud for additional processing before making it available in the Qualifier web application, which offers issue management, reporting, root-cause analysis, visualizations, and advanced analytics. Optional capabilities include video and integrated road and weather data. Qualifier can even ingest files captured with other data loggers, allowing you to perform comprehensive analysis on your vehicle data.

To set up a demo, contact Aptiv's sales team at sales@aptivcs.com.

The important thing is that it is comprehensible enough that you can take action as a result.

For vehicle engineering teams, one effective way to move away from big data to small data is to use data loggers that can do analytics at the edge. You can then store just the results of that analysis — the small data, or as Aptiv likes to call it, the rich data. It's much easier — and more cost effective — for vehicle engineers to extract valuable insights from this rich data than from a vast lake of raw data.

> "By 2025, 70% of organizations will shift their focus from big to small and wide data, providing more context for analytics."
>
> — Gartner

6. Vehicle data lakes don't have enough data.

This might seem like a contradiction to the last point, but it's not. Vehicle data lakes often have too much raw data from the vehicle but not enough data from external sources to provide context.

The "wide data" the Gartner quote references is data that comes from a variety of different sources. In a vehicle engineering context, this could mean integrating all your data from different engineering teams and different data loggers, like Aptiv does through Qualifier Ingest.

But wide data can also involve bringing in data from external sources to supplement the vehicle data. For example, when validating autonomous vehicles, ADAS features, or electric vehicles, data about weather conditions, road characteristics, and traffic can be extremely helpful. Some engineering teams also rely on video — preferably with object recognition capabilities — to ensure that they have met all the desired test conditions and to determine if issues arise in certain outlier conditions.

Aptiv provides these capabilities through its Environmental Insights feature, which integrates weather and road data into Qualifier. It also offers an ADAS Solution that includes video with object recognition and Qualifier Analytics for advanced reporting. It helps ensure appropriate test coverage and streamlines troubleshooting.

Extract more value from your data.

Vehicle manufacturers that are trying to become more data-driven have the right idea — but executing on that idea has proven extremely difficult. The automotive industry has some unique challenges that make meaningful analytics far more complex than in other industries.

Aptiv Connected Services has been overcoming these challenges for the world's largest vehicle manufacturers for more than a decade. For more information on how we can help you extract more value from your vehicle data, visit www.aptiv.com/connected-services.

Additional Resources

- 5 Things You Should be Doing with Your Pre-Production Vehicle Data
- Aptiv Connect Qualifier vehicle data collection and analysis solution
- Aptiv Connect Ingest centralized repository for vehicle data from multiple types of loggers
- Environmental Insights integrated road and weather data
- Qualifier Overview Video

About Aptiv

Aptiv is a global technology leader with more than 190,000 people across 127 manufacturing facilities and 12 major technical centers worldwide. With a presence in 46 countries, we address mobility's toughest challenges through our deep software and systems integration expertise, delivering market-relevant solutions for our customers.

Our mission is to make the world safer, greener, and more connected than ever before. As these trends converge and the world of mobility changes, we will lead this change.

Mobility has the power to change the world, and we have the power to change mobility.