miso

WHY IS MANAGING DATA LIKE CHECKING FOR RED CARS?

A Miso guide to ETL, Batch, Orchestration

Contact
0121 232 8000

Website
www.misoportal.com

E-mail
info@misoportal.com

WHY IS MANAGING DATA LIKE CHECKING FOR RED CARS?

Imagine your data pipeline is a busy motorway carrying thousands of vehicles from one place to another. Based in a motorway service station, you've been given the task to check all the red cars hurtling down the road while minimising the disruption to the rest of the traffic.

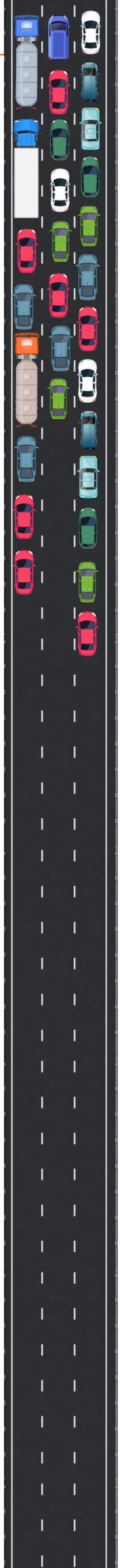
How would you do it?

STOP THE MOTORWAY (ETL)

Well, you could stop the entire motorway and turn all three lanes into a checking area. Every vehicle comes to a halt. You inspect them one by one, whether they're red cars or not. Then, once everything has been checked, you restart all the traffic.

This is classic ETL behaviour. It works, it's simple and only requires one process, so there are no integration requirements.

The fundamental problem with this approach is that all traffic has had to stop. The entire motorway is gridlocked while you're checking. With the speed of the checking process depending entirely on how much power you have. This is particularly troublesome if your data is interconnected or time-dependent.

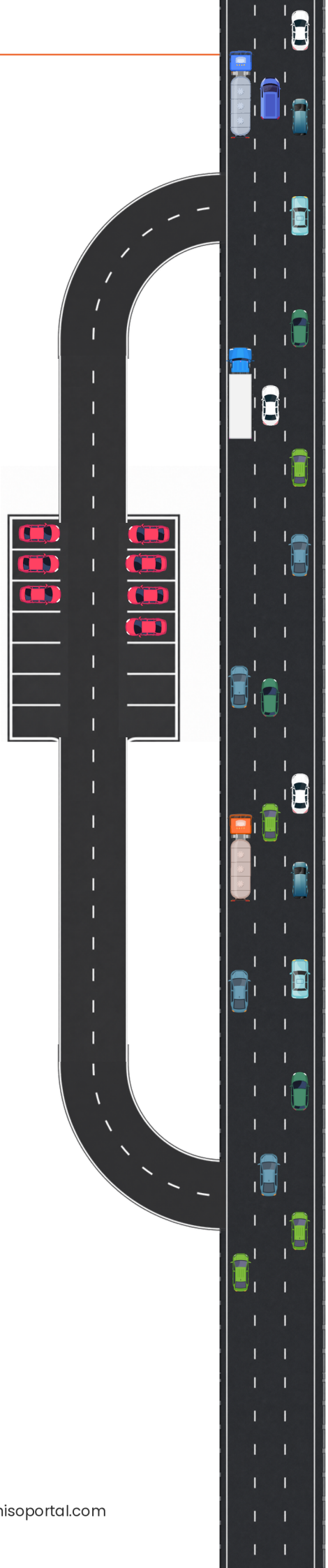


PARK THE CARS (BATCH)

As traffic flows down the motorway, a Highway Officer could spot the red cars and flag them off to your service station. The red cars get parked in a large, dedicated car park. Later, often overnight, you inspect them in one focussed session. Once they're all inspected and processed, they're released at the same time back onto the motorway

This is batch or bucket processing. The advantage is the rest of the traffic isn't affected, and the processor does less work because it only looks at red cars, instead of every vehicle on the road.

Unfortunately, there's a delay in processing. You need another service to do the flagging (the Highway Officer) and you need a big storage area (the car park) and critically, while the batch is being processed, the red cars are sitting still. It works well when you need to see where everything is, but the delay to the red cars is a challenge in a real-time world.



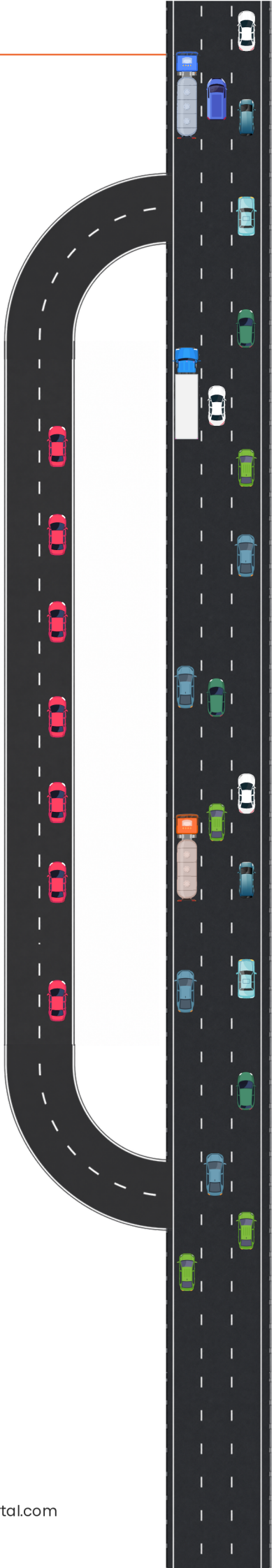
ROLLING INSPECTION (ORCHESTRATION)

A third option is to keep the red cars moving. This time a Highway Officer could spot red cars and flag them off the motorway. The red cars travel slowly down the service station exit without stopping. As they drive by, they're inspected automatically. The moment the inspection is complete, they rejoin the motorway and continue their journey.

This is an example of orchestration, where the main pipeline keeps moving while specialist processes are triggered only when they are needed.

That makes it faster, more flexible and less disruptive, especially when only certain data needs checking, transforming or routing. Instead of stopping everything like ETL, or parking work up for later like batch processing, orchestration allows specific tasks to be handled in parallel and returned to the flow once complete.

The challenge of this approach is ensuring close cooperation between different components i.e. the Highway Officer, the inspection service, the rejoin process.



THE EVOLUTION OF FME

These three processes mirror the journey of FME. It started as a traditional ETL platform where data pipelines were batch-oriented and predictable.

Then the world changed with real-time data becoming more important, data pipelines getting faster and everything growing more complex. Over the years, FME has evolved dramatically with this shift in the data landscape.

It still supports batch and traditional ETL but FME's orchestration capabilities have become remarkable. Through data virtualisation, Event Grid integration, and other native orchestration tools, FME has become built for the drive-by world.

WHEN YOU'RE ASKED TO COUNT ALL THE RED CARS

You have choices now. You can stop the motorway when that makes sense. You can park the cars when batch processing is the right approach or you can keep everything moving with orchestration.

Most organisations need a combination of all three. FME gives you that flexibility, and it works alongside the data infrastructure you have already invested in, whether that is Azure Data Factory, AWS Glue, Google Dataflow or something else entirely.

You do not have to tear out your existing motorway. You do not have to choose between stopping traffic and parking cars. You can use the right approach for the right situation, and when the job calls for it, you can keep the data flowing.

miso

www.misoportal.com