White paper

Tour de France and Tour de France Femmes avec Zwift 2023

The technology behind the world's largest connected stadium

Tim Wade
Vice President, Advanced Technology Group | NTT

As the Official Technology Partner for the Tour de France and Tour de France Femmes avec Zwift, our primary objective is to enhance digital engagement, facilitate deeper insights and optimize operational efficiency for Amaury Sport Organisation (A.S.O.), the race organizers. We strive to achieve these goals within the unique context of the race while constantly pushing technological boundaries and driving innovation.

In 2023, we continue to expand our capabilities in new and exciting ways. While refining fan experiences from previous years, we have utilized cutting-edge technology to create the world's largest connected stadium by developing a digital twin of the event. This digital model combines existing data with real-time information to accurately replicate critical elements of the event, including the NTT DATA Tech Truck, which features live data and complex calculations visualized in easy-to-consume ways.

Our solution integrates site plans, real-time asset locations, manually captured data, local weather measurements, real-time streaming analytics, artificial intelligence (AI) and machine learning (ML) models, all supported by a sophisticated network infrastructure. This enables us to provide innovative services that cater to the demands of A.S.O. and their fans.

For the 2023 races, our global operations team relies on advanced collaboration tools to ensure seamless coordination with both internal and external partners.

Our managed services – including Edge as a Service – are becoming increasingly crucial as we adopt scenarios involving edge computing and user devices in race vehicles. In these situations, an event operations view is essential.

A new, dynamic architecture necessitates a secure-by-design approach that involves distributed teams, services and hardware. We prioritize implementing the right security controls, visibility, reporting and processes, an approach that's been fundamental to every decision we've made leading up to the 2023 events in France.

This white paper delves into the specifics of our strategy and execution.

Contents

Components of the 2023 solution

The 2023 Tour de France solution has three main components. All utilize the same core elements to achieve distinct business outcomes and support fans, event organizers and our operational teams.

These components are:

1. **Transforming the fan experience through data-driven insights:** Since 2015, we have utilized AI and ML to revolutionize the fan experience by offering unprecedented insights into the world's most prestigious cycling race, whether fans are watching the broadcast or engaging through other digital platforms.
2. **World's largest connected stadium:** In 2023, we are expanding the world's largest connected stadium by developing a digital twin of the Tour de France. The digital twin harnesses all available data to create a digital replica of the event, allowing organizers to meticulously plan each stage and monitor the race in real time.
3. **Intelligent event operations:** We are drawing on our global expertise to develop and support the technology underpinning the Tour de France solution. Our Adaptive Cloud to Edge Managed Services Platform grants our support team full visibility of edge computing, IoT devices, the network and cloud-based data analytics platform. This offers A.S.O. a new level of insight into race operations, enabling them to proactively plan each stage, map critical areas, and swiftly respond to changes in the dynamic race environment.

Data-driven insights and reporting have been the foundation of the Tour de France solution since 2015 when we began partnering with A.S.O. on their digital transformation journey.

In 2023, we continue to innovate by incorporating additional data sources, facilitating intelligent event services and creating a connected stadium experience for A.S.O. and future roadside experiences for fans.

Transforming the fan experience through data-driven insights

We divide the fan experience solution into three primary areas:

- Sensors, transmission and data
- Real-time analytics, AI and ML
- Making sense of the data

1. Sensors, transmission and data

Our data journey begins with capturing latitude, longitude and speed from each rider. The information is transmitted from a transponder under the bike saddle every second during the race. We needed to consider environments with high radio-frequency (RF) traffic, such as the technical zone at each stage's end, which houses numerous Wi-Fi networks, mobile devices and TV broadcasters. To protect sensitive components, we enclosed the core microcontroller unit within a Faraday cage.

One of our **considerations needed to be that some of the environments where the devices need to work** have high levels of RF traffic.

The data collected is transmitted via a two-stage transmission network:

• The primary transmission network, a wide wireless area network (WWAN) based on the 3GPP 802.15.4 standard, creates a mesh network between telemetry devices and relay points. This moving mesh network enhances location coordinate accuracy and uses the best relay points nearby, sending the data to the airplane (<10km away).

• The secondary transmission network transmits data to the race's end using three licensed microwave frequencies. The data is multiplexed into the signal and sent along with camera footage to the end-of-race receiver placed on a mobile lift 20-50 meters above the technical zone.

Our data-collection endpoints are designed for redundancy to ensure service continuity in case of hardware or software failure.

In 2023, our new data sources incorporate a diverse array of transmission types and protocols, enabling the digital twin of the race (see Table 1).

Data source	Technology type	Network	Protocol
Smart event manual inputs	HTML 5 UI to input information	3G/4G/5G and Wi-Fi	HTTPS
Key locations at event (VIP, merchandise, WC, trucks, zones)	NTT DATA Trackers	3G/4G/5G	Renard
	NTT DATA Trackers	3G/4G/5G	Renard
People	NTT DATA Trackers	3G/4G/5G	MQTT
Environment	Env Sensors	2G	HTTPS
Crowd counts	Video analytics	Fibre optic	HTTPS
Crowd movement	Mac address	Wi-Fi	HTTPS
Weather	Weather stations	868Mhz	HTTPS
In-race car tracking	NTT DATA Trackers	3G/4G/5G	MQTT
In-race cameras	Video encoding	TCP/IP	RTMP/RTSP
Maps and geo overlays	KMZ/PDF	TCP/IP	HTTPS

Table 1: Tour de France data sources

The complex interactions between the various elements that make up the solution can be seen in the following diagram:

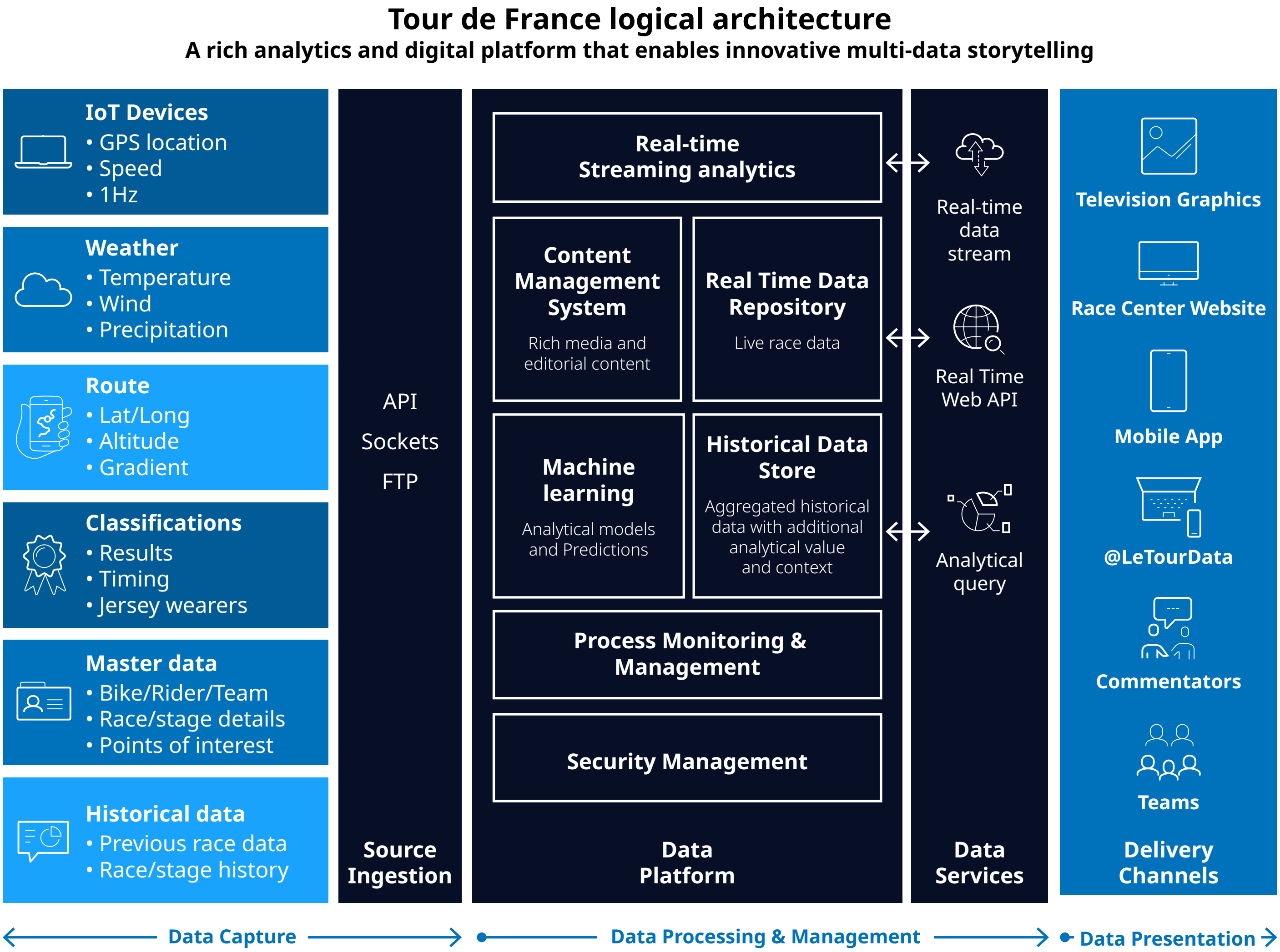
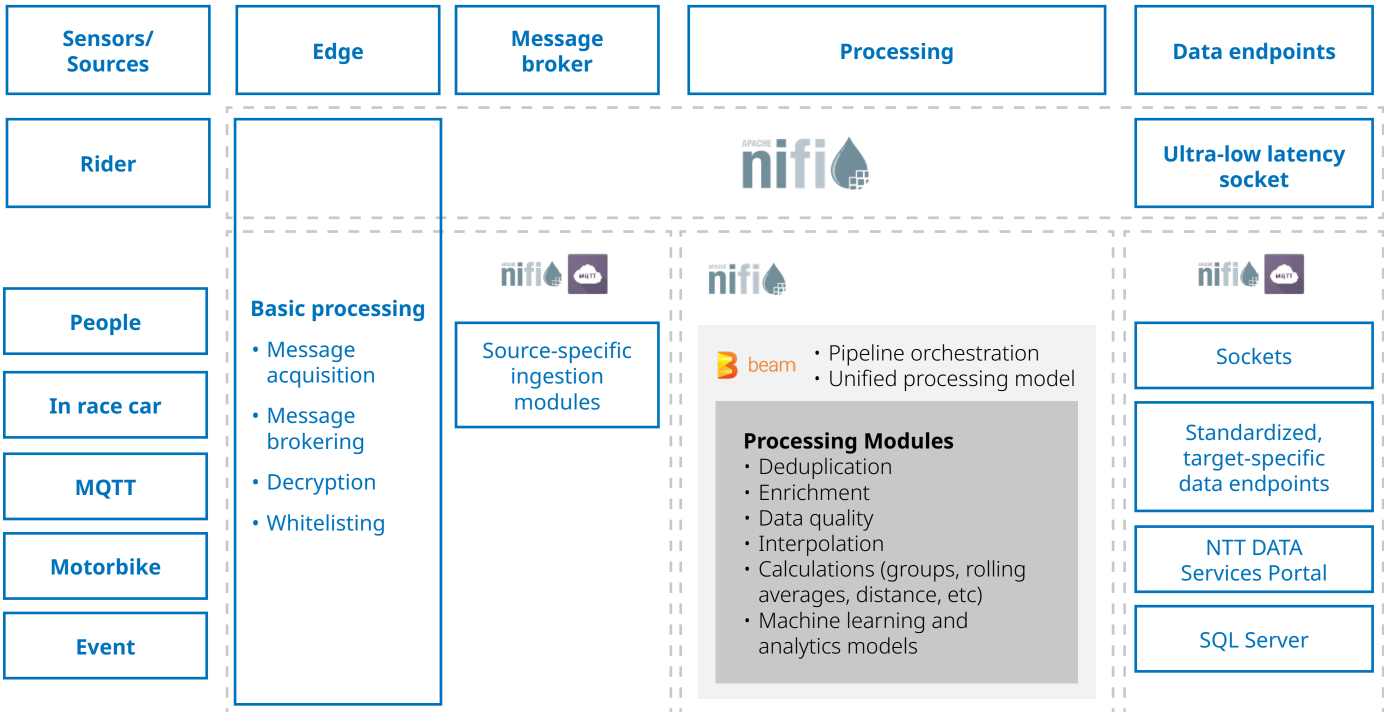


Figure 1: Tour de France logical architecture

The data can be split into the following segments:

- **Telemetry data:** This is the data that comes from rider-tracking devices. It's ingested into the real-time analytics platform in the format of a race-situation JavaScript Object Notation – or JSON – file.
- **Race data:** At the beginning of each stage, the platform requires two key pieces of information:
- **The stage data:** Gives accurate location-based information for the whole stage, including locations of sprints, climbs, start and finish.
- **The rider and team data:** Detailing which riders are in the race and including their current classification, bib number, rider name and team name.
- **Timing data:** Data from the official timing provider, which provides official classifications, results and photo-finish information.
- **Environmental data:** The platform ingests third-party data sources to enhance and further refine the accuracy of the output. This data includes ordinance survey information, which gives details of the terrain, in particular gradient and height-above-sea-level information. It also includes weather data, using the telemetry data to look up localized weather data for each rider and physical weather stations at certain locations in the race. We're also using air quality and environmental sensors in the NTT DATA Tech Truck to provide information about the working environment.
- **Social media data:** Data is taken from Twitter handles, for example, @letour and @letourdata.
- **Media data:** Media data is inputted into the system via a content management system. This data includes videos, data and images.
- **ML data:** ML data is created and ingested via a direct database connection.
- **Smart event data:** This data gives context and enables us to construct the digital twin of the event. It includes information on key route locations, vehicle tracking, caravan tracking, Arrival and Departure Village locations, NTT DATA Tech Truck location and crowd monitoring.
- **People movement data:** We're monitoring the movement of people at key locations, including the density of crowds and duration of stays.
- **AI data:** We also use an athlete-ranking system that focuses on results and considers the type of event and who the riders are competing against.

Tour de France real-time analytics architecture



2. Real-time analytics, AI and ML

The heart of the Tour de France solution is the real-time analytics architecture (see Figure 2).

To deliver a service that is transmitted to the entire world and aligns with a live television broadcast, we need low latency and reliability. Ingesting data and analyzing it with subsecond timing requires optimized data-processing pipelines and special architectures.

We've used a few key components in the architecture of this platform including Apache beam, Apache NiFi, Eclipse Mosquitto and Python. These are illustrated in Figure 2.

Because of the large number of integrations connecting into a range of data sources and data consumers, we have a few key integration types that we use:

- UDP/TCP
- HTTPS API (push and pull)
- MQTT
- File-based formats, which are agreed upon at each integration
- WebSockets over SSL/TLS (WSS) (Secure Web Socket
- SignalR
- Direct database connection

Data aggregation, filtering and analysis

Once data has been ingested into the platform, we have a series of algorithms to handle scenarios like:

- Data deduplication
- Data parsing
- Error correction

Once the data is cleansed, we have a clean data set that we can split into the correct streaming analytical flow, allowing us to apply a set of algorithms that give us an understanding of the race and event situation.

Some of the metrics we've created are:

- **Riders:** speed, position, gaps, direction, relative wind, gradient, distance from start, distance from finish
- **Groups:** speed, position, gaps, composition, distance from start, distance from finish
- **Event:** correlation, crowd information, estimated time of arrival (ETA)

The data streamed through the platform is not written to disk. This is done to ensure the subsecond latency that the Tour de France use case requires. Once the analytical models have been run, the output is stored in a relational database.

2.1. Artificial intelligence and machine learning for athlete analytics

To determine the capabilities of athletes who only compete in small groups in irregular competitions and in a mix of events that require significantly different skills, we've created an athlete-ranking system that focuses on results.

This model takes into account the type of event as well as the other riders in the race and is run in Python modules in the data-processing pipeline described above.

See Figure 3.

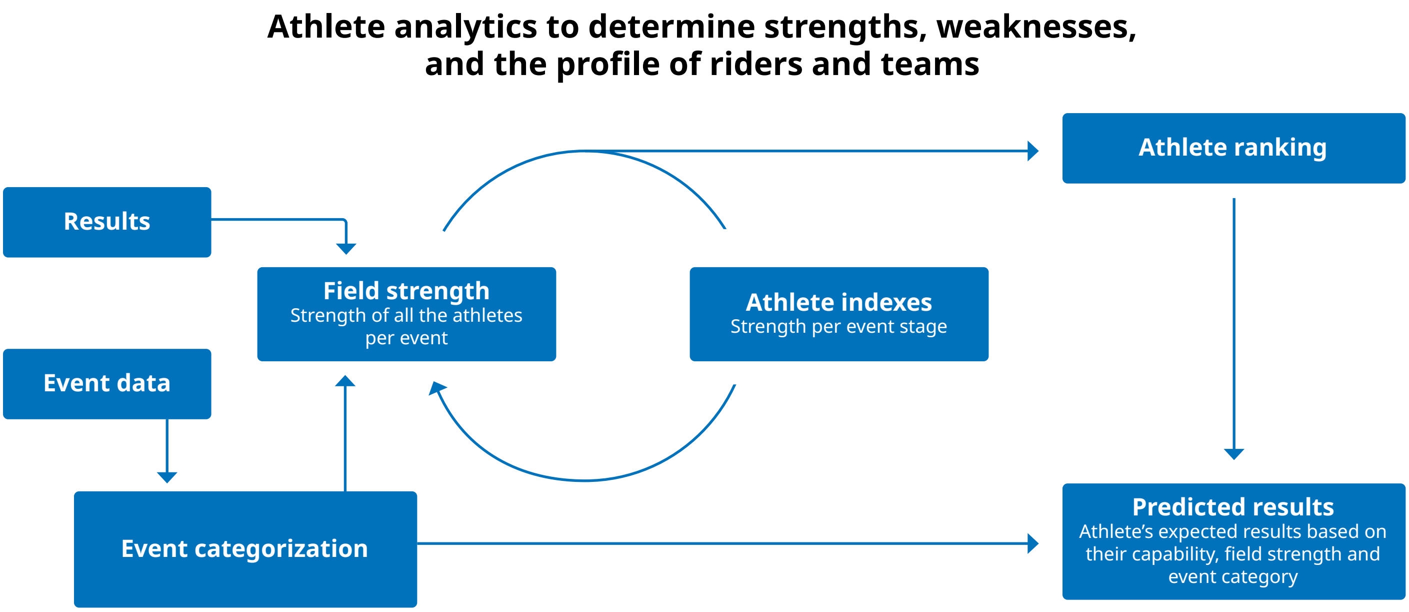


Figure 3: Athlete analytics

2.2. Machine learning

Building on the core analytics platform and algorithms, we've added a combination of telemetry data, race results, rider data, course information and conditions data to predict race outcomes. Access to more data, technology and expertise allows us to produce the following analyses an forecasts.

#NTTDATAPredictor

- **Catch predictor:** Will the breakaway be caught by the peloton before the end of the stage?
- **Stage favorites:** Which riders are likely to do well on a given stage based on their profile, results and the nature of the day's route?

#NTTDATAEffortIndex

- How hard are the riders working right now?
- What's the difference in effort within different groups?

Performance profile

- What are the attributes of different types of riders?
- What sorts of races or stages are they suited for?

People counting

- Based on video analysis at the edge.

We execute ML using three core pillars: data, team and technology.

Machine learning: data pillar

The richer the dataset available, the more factors the model can consider. We've created ML models based on Tour de France data and seven years of race results from an external source.

> This is a lot of information so the next challenge **we had to solve is how to visually interpret this data to allow us to tell stories** with it, make it engaging and easily digestible.

To enrich our data with additional features that provide greater context to the "raw numbers", we also use external third-party data services that incorporate factors such as the weather.

The example below shows the data that is applied to enable the ML algorithm to make the catch prediction.

Machine learning: team pillar

The skills required to deploy and train an ML solution are a critical component of its success. We've built a team of:

- **Data scientists,** who design and develop the ML models
- **Data visualization experts,** who present the data and predictions visually to make it easy to understand
- **Product owners,** who ensure the project is on track and focused on the agreed objectives
- **Business experts,** who bring deep knowledge of the business subject area and data
- **Data engineers,** who are experts in data integration, storage and analytics

Machine learning: technology pillar

We're using the information from the IoT platform along with the additional data described above and the new ML models to create two new sets of predictions:

- **Real-time predictions:** To get real-time predictions, we take the data from the live tracking platform and apply the ML models to it in real time. This enables us to effectively predict the effort estimate and catch predictor.
- **Batch predictions:** We also run overnight batch predictions that take into account the data from the previous stage. This enables us to predict stage favorites while applying the riders' performance profiles to allow us to model the next day's stage.

3. Making sense of the data

We take the initial three data points and turn them into over 50 data points. This means that for every stage of the Tour, we're creating 50 data points per rider, per second. This is a lot of information, so the next challenge we had to solve was how to visually interpret this data to allow us to tell stories with it; how to make it engaging and easily digestible for a wide range of users – and uses. These include:

Broadcast: The Tour de France is a global event, so the data we provide to these teams needs to be universally understandable and formatted in such a way that it can be seamlessly integrated into the broadcast feed.

This includes allowing the production team to track riders and speeds in real time, thereby enhancing the viewing experience.

Race Center: Fans across the world use the Tour de France Race Center to follow the race in real time. Using stage-profile and map-based visualizations, they can identify individual riders and follow their progress in real-time.

@LeTourData: Thanks to a dedicated team of data scientists, cycling experts and content specialists, @LeTourData provides insights into the state of the race and the performance of key riders, leveraging the full spectrum of data available. This content is delivered on social media and broadcast platforms.

Media Wall: To enhance the experience at the Departure and Arrival Villages, this year we're building on the success of the NTT DATA Media Wall which we introduced last year.

On a large screen, we integrate a curated set of race data, live race footage and insights into the riders and teams for fans on the ground. We also include geo-located content for fans on the roadside.

Digital human: Kiosks hosting our digital human, Marianne, will be installed in the NTT DATA Tech Truck and in VIP areas in the race villages. Marianne will incorporate basic and real-time race data along with a new ChatGPT integration this year.

Operations dashboard: The operations dashboard helps to enhance the operational efficiency of race organizers by giving them real-time data on the status of the race. Combined with footage from the race, this provides them with full visibility of all elements of the race, including riders and race vehicles in a 3d map view,. This enables them to make data-driven decisions to ensure that each stage proceeds smoothly.

A.S.O. Event App: In a highly dynamic environment, access to data on-the-go is critical. The A.S.O. Event App leverages our digital twin models to plan each stage, including the placement on the Departure and Arrival Villages. Personnel can monitor the status of all race assets and dynamically change plans when circumstances change. Clear lines of communication ensure that all users are immediately updated.

3.1 Architecture: hybrid IT

The technical requirements and architecture of the Tour de France solution are heavily influenced by A.S.O., the spectator/fan demographics and the event location. Our hybrid IT solution takes into account the following:

- **Right service, right location, right time:** We use four infrastructure deployment types – physical, cloud, containerized and serverless – to provide the best infrastructure for each part of the solution.
- **Large global audience:** Public web services should be able to service hundreds of thousands of requests per second.
- **Ability to monetize:** Data must be available to other consumers in a secure and measurable manner.
- **In-sync with other services:** Some aspects of the solution should process data and make it available to data consumers in near real-time, which calls for subsecond processing.
- **Business-critical service:** At this high-profile event, security is paramount and data protection laws must be enforced. Data has to stay in the EU.
- **Deployed as code:** The architecture is deployed as code to enable easy teardown and rebuild on different types of infrastructure.

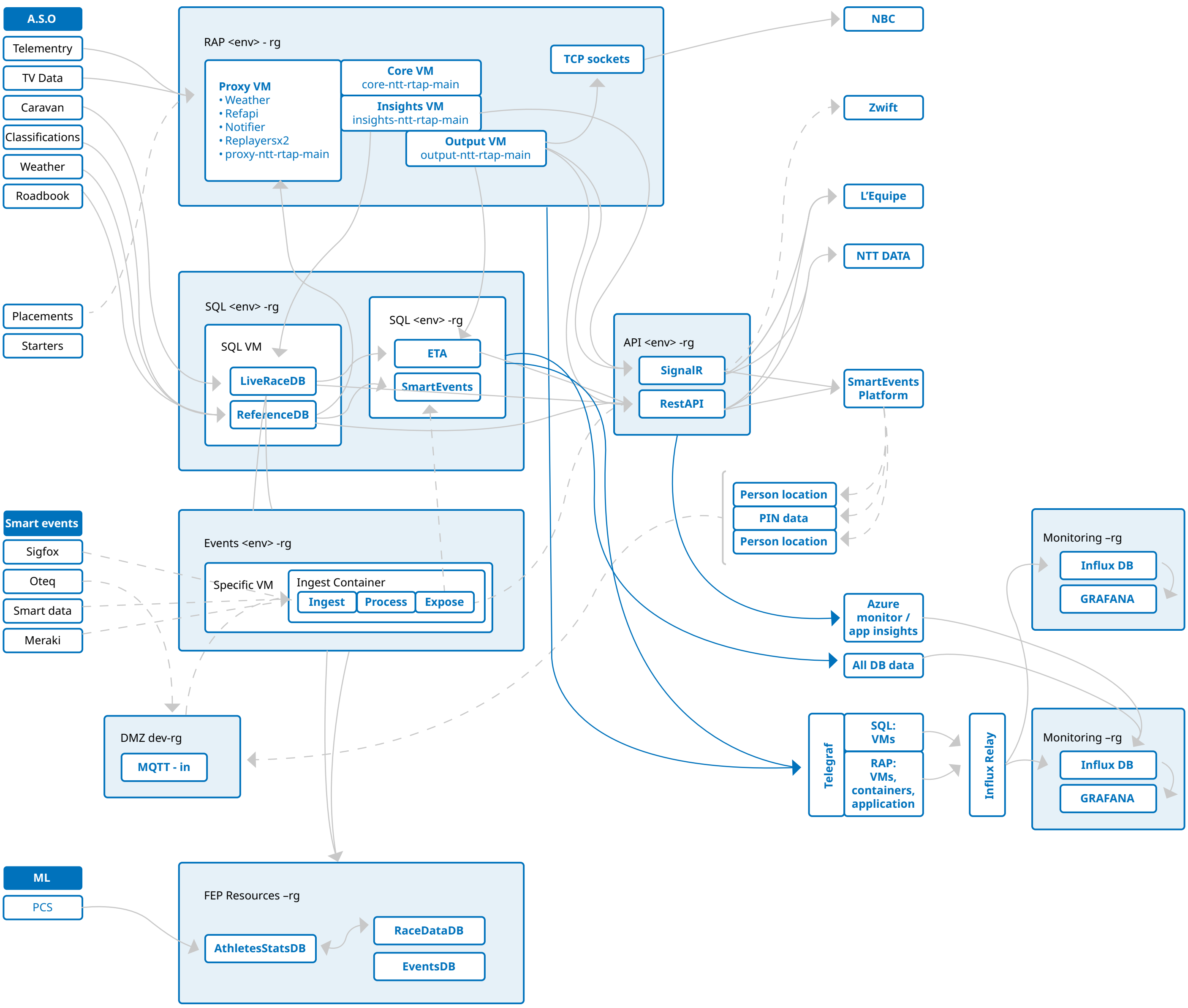


Figure 4: Illustration of the conceptual solution architecture

4. Unique solutions for a unique event

Some of the challenges of **transporting IT equipment over 3,400 km in a month** are not always obvious.

4.1. Event edge computing and networking

Some of the challenges of transporting IT equipment over 3,400km in a month aren't always obvious. For example, the outside temperature can affect the inner cooled area when it varies by 50 degrees, depending on the location and day, so keeping the architecture as simple as possible is key.

Vibration plays a huge role in the type of technology that is used. Moreover, ensuring that spares are available and having a remote-monitoring solution and a plan for field maintenance are critical elements to keeping data flowing.

End-of-race equipment is deployed in flight cases to ensure that it's protected from the daily build and tear down that can knock and damage kit that isn't permanently installed.

4.2. Business continuity and disaster recovery

Keeping business continuity in mind during every stage of the design process was a core part of the successful implementation of the technical solution.

We took the applicable requirements from the ISO 22301:2012 standard in the creation of the overall strategy.

The high-level process is outlined below:

1. Understand stakeholders and requirements
2. Understand legal and regulatory requirements
3. Define organizational requirements
4. Define the business continuity plan (BCP) and its scope
5. Establish the BCP and process
6. Operate
7. Evaluate

As part of the business continuity element of the solution, we focused on key perceived risks and mitigation plans.

Having no single point of failure in the solution, the analytics solution deployed in a high-availability configuration and the API deployed on a redundant service gave us a robust solution. Having a final fallback of being able to deploy a new software-defined solution from scratch using code gives us a way to mitigate single-provider risks should a multiregion incident take place.

4.3. Management

To give an overarching, single-pane-of-glass view for the services and infrastructure used to support the Tour de France, we leveraged threeareas of the NTT DATA service portfolio:

- Network and compute with MHIS
- Communications with MCS
- Holistic view of everything with customer portal

4.4. Support from the virtual Zone Technique

Our model has evolved from on-site support to using a globally distributed team, working across five countries, to deliver our solution for A.S.O.:

- Australia: augmented reality (AR) app
- South Africa: Analytics platform, data services, API
- UK: Customer portal and observability platform
- Australia, France, South Africa, UK; @letourdata

Cloud platforms and leading collaboration platforms allow us to place workloads in the best places and provide 24x7 coverage.

In addition, we use NTT DATA's technologies and services to ensure that the teams have the best communication channels, connectivity and observability of the solution.

Trust, accountability, planning, tools, technical skills and, most importantly, proactive communication drive successful operational delivery.

4.5. Communication and observability

During live race hours, we use live videoconferencing supported by NTT DATA's MCS service. This allows us to have real-time communications during time-critical moments.

4.6. Live video

Live-video encoding enables the teams to see what's happening live during the event (see Figure 5).

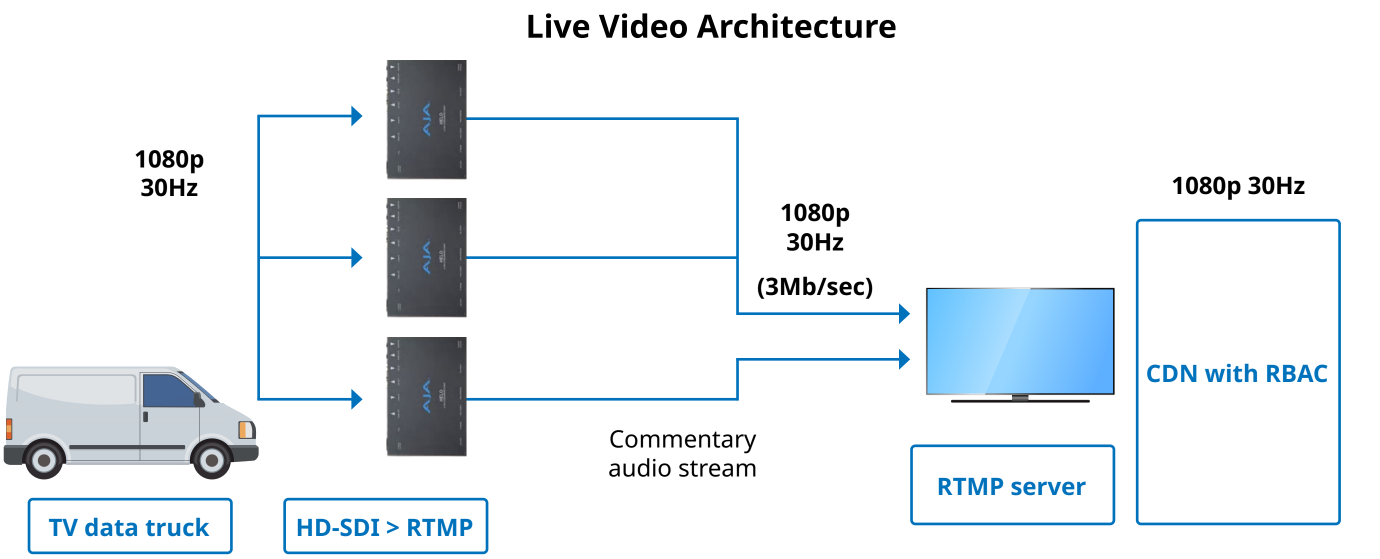


Figure 5: Live-video architecture

4.7. End of race

As we have equipment deployed at the end of the race, robust remote connections have been established to enable remote access and support:

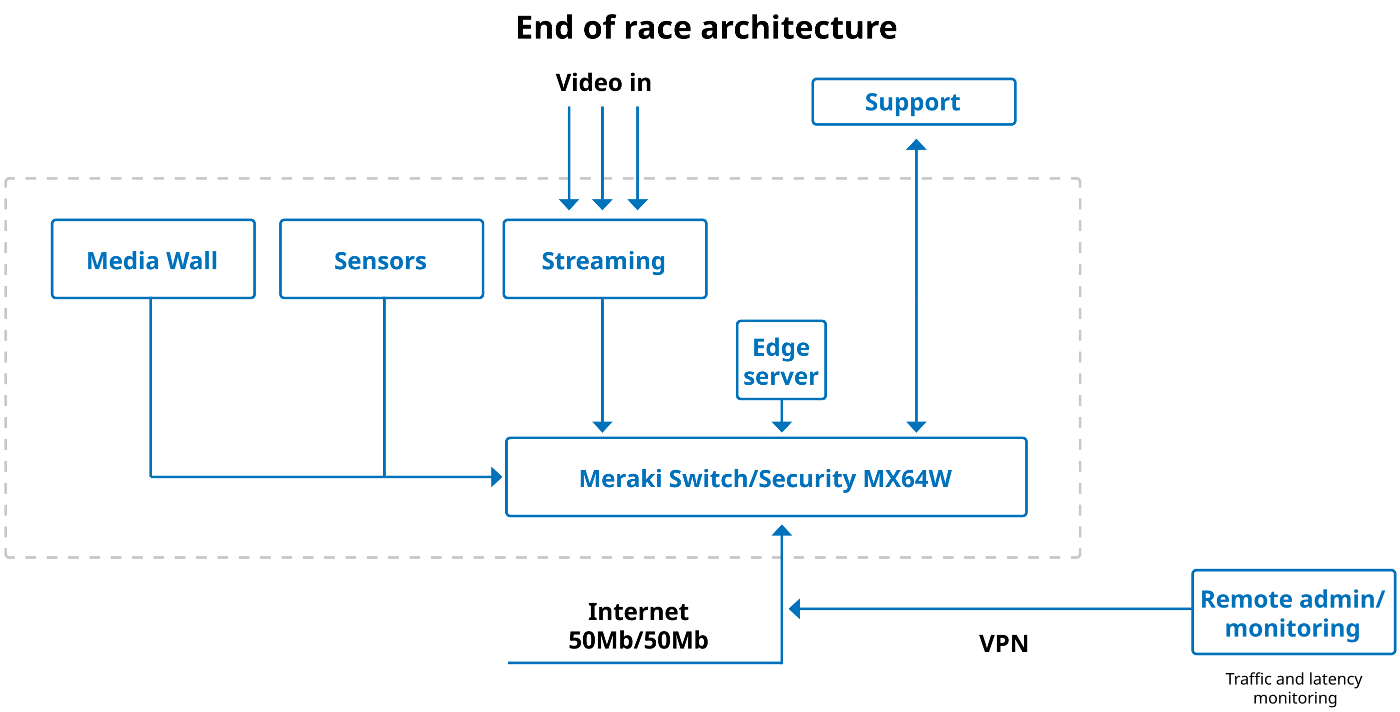


Figure 6: End-of-race architecture

4.8. Observability

We have broadly followed Google's golden signals approach for monitoring to drive the observability process and provide a consistent user experience. The collection of the right metrics is key in providing useful information on:

- **Latency:** The time it takes to service a user request
- **Traffic:** A measure of how much demand is being placed on the system (HTTP requests, network throughput, transactions)
- **Errors:** The rate of requests that fail
- **Saturation and utilization:** Consumption of constrained resources (memory, input/output, CPU slices, etc.)

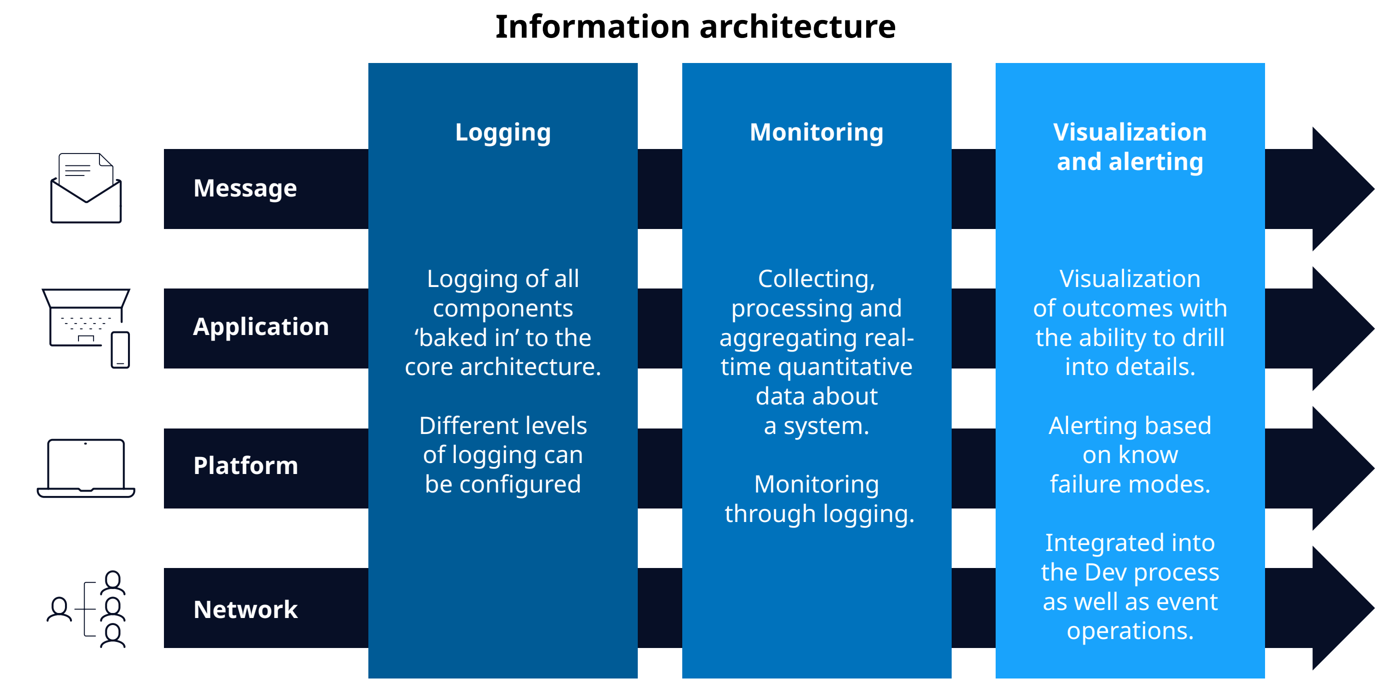


Figure 7: Observability information architecture

This approach has driven our design process, from collection to alerting, providing essential and useful information to empower the user. It's one of the core components that enables the virtual and physical Zone Technique.

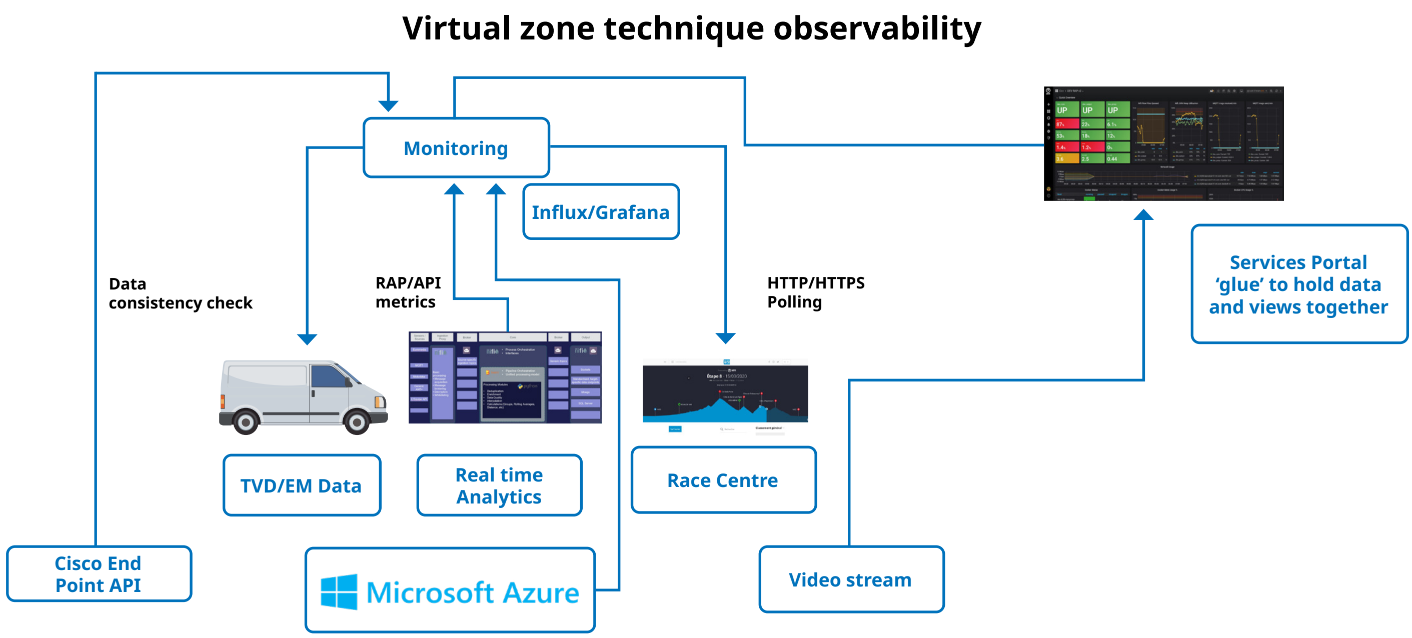


Figure 8: Zone Technique observability

4.9. Testing

Before we roll into a new season, we perform automated platform testing, allowing us to simulate user traffic quickly and easily leverage open-source tooling.

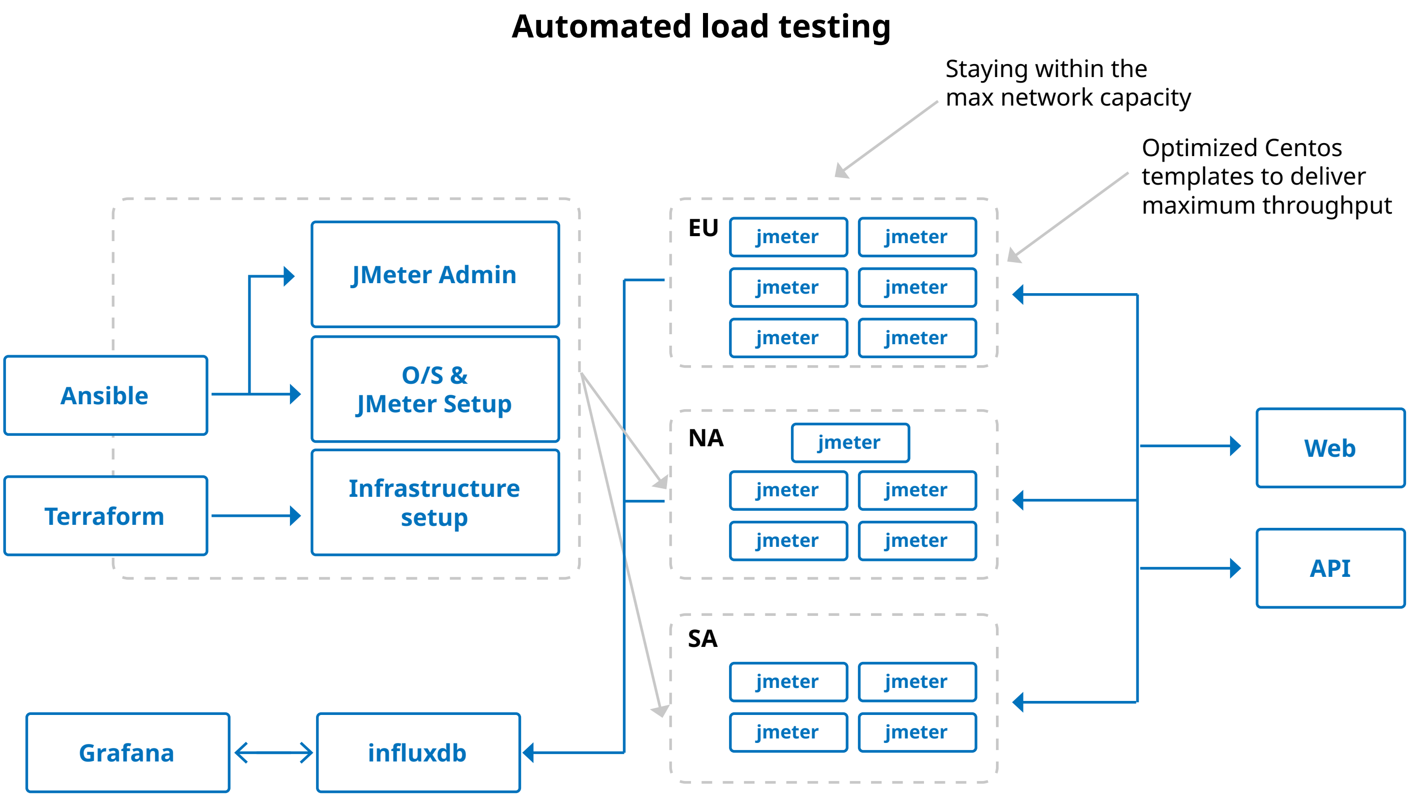


Figure 9: Preseason automated load-testing model

We also perform security testing to identify potential vulnerabilities so we're able to address them prior to the event.

Development assessment

Static source code scanning	Manual analysis and findings triage	Findings prioritization	Understanding of program design	Recommendation on closing issues
• review code scanning tools settings • conduct static • current scanning scope and methodology	• manual analysis of static code scanning results • identify false alarms • identify repetitive issues	• classification of the criticality of applications • multiple factors prioritization based on application criticality, issue rating, business priority, etc.	• understanding of program logical, architectural and functional design as well as business logic and flow	• coding level of remediation advice • best practices advice for repetitive issues

Platform assessment

TCP/IP stack vulnerabilities	Missing patches	Buffer overflows
Unsupported services	Unnecessary services	Backdoors
Denial of service	Information leakage	Default configurations
Weak configurations	Open administrative interfaces	Anonymous or guest access
Unencrypted protocols	Peer-to-peer services	Remote file access
RPC services	SNMP access	Open SMTP relay detection
Default passwords	CGI abuses	SSL certificate issues
SSL protocols supported	SSL ciphers supported	Timing evasion
Server and general HTTP	Data injection and manipulation	Sessions and authentication
Server misconfiguration	Cross-site scripting (XSS)	Anonymous or guest access
SSL cipher and certificate	SQL and blind SQL injection	Brute force attacks
Default configurations	Parameter redirection	Cookie and session weakness
Canonicalization attacks	Command injection	Web service attacks
CGI and filesystem abuses	Code injection and file uploads	Cross-site request forgery (CSRF)
Information leakage	Buffer and integer overflows	Open administrative interfaces
HTTP response splitting	Client side technologies	Privilege escalation

NTT DATA