Tour de France 2021 –

The technology used to create the world's largest connected stadium

Tim Wade
Vice President, Advanced Technology Group | NTT

As the official technology partner for the Tour de France, a guiding principle of our technical solution is to deliver the business outcomes of increased engagement across digital channels, enabling deeper insights and increased operational efficiency as set out by the race owners, Amaury Sport Organisation (A.S.O.). Our goal is to deliver these outcomes in the unique environment of the race and, at the same time, drive technology innovation, pushing the boundaries further every year.

2021 sees this extend into exciting new areas. We've made iterative improvements and updates to the fan experiences developed over the last few years, but now we're using this technology to create the world's largest connected stadium.

To do this we've created a digital twin of the event, using existing data and adding new layers of information to completely replicate all aspects of the race in a digital model.

We've integrated site plans, the real-time location of event assets, manually captured data, local weather measurement alongside real-time streaming analytical, artificial intelligence (AI) and machine learning (ML) models. This is underpinned by a complex network fabric allowing us to enable the innovative services that A.S.O. and fans are looking for.

For the Tour de France 2021, our operations team is distributed across the globe and to ensure that we're able to deliver on race day we need to leverage collaboration tools that work for our internal team and our external partners.

Our managed services are also becoming ever more important as we move into a scenario where use cases include edge compute units and end-user devices in race vehicles where it's impossible for technical personnel to physically access equipment, so an event operations view is imperative.

Ensuring this new, more dynamic architecture, with distributed teams, services and hardware, is secure by design has been a focal point. We have to have the correct security controls in place along with the right visibility, reporting and processes. This has been a fundamental building block and design choice across every decision we've made on the road to France in 2021.

This whitepaper delves into the detail of how we achieved this.

Contents

Putting the technology solution in action at the Tour de France

The 2021 Tour de France solution is made up of three main parts - leveraging the same core components to achieve different business outcomes - to support fans, event organizers and our operational support teams. These are:

Revolutionizing the fan experience through data-driven insights

Starting in 2015 we've leveraged data to create a revolutionary fan experience. Using AI and ML we've given fans, whether they're watching the broadcast or connecting through other digital platforms, unprecedented insight into the world's greatest cycling race.

World's largest connected stadium

In 2021 we're creating the world's largest connected stadium, developing a digital twin of the Tour de France. The digital twin leverages all the data we have access to and creates a digital replica of the race, allowing organizers to plan out every stage in detail and monitor the race in real-time with a view of all the riders, as well as any race vehicles.

Smart event operations

To support the Tour de France we're tapping into our pool of global experts to develop and support the technology underpinning the solution. Leveraging our Managed Services platform our support team has full visibility from edge computing and IoT devices through to the network and the cloud-based data analytics platform.

This provides A.S.O. with a new level of visibility into race operations, enabling them to proactively plan each stage, including mapping out critical areas and reacting quickly to any changes in this highly dynamic environment. Every aspect of the 3,400km race route has been mapped, giving their teams access to all the information they need to ensure the smooth running of the event.

Data-driven insights and reporting

This has been the cornerstone of the Tour de France solution, and since 2015 we've partnered with A.S.O. on their digital transformation journey. In pursuit of continuous innovation in 2021, we're adding additional data sources, enabling smart event services and creating a connected stadium experience for A.S.O. and the future connected roadside experience for fans.

This provides A.S.O. with a new level of visibility into their operations at the race, enabling them to proactively plan each stage, including mapping out critical areas and reacting quickly to any changes in this highly dynamic environment. Every aspect of the 3,400km race route has been mapped along with the Departure and Arrival Villages on each stage, giving their teams on the ground access to all the information they need to ensure the smooth running of the event. At the same time, the additional data allows them to create more interactive fan experiences, delivering an exceptional fan experience anywhere, anytime.

We split the solution into three key areas:

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

Sensors, transmission and data

Our data journey starts with three pieces of information that we capture from each rider: latitude, longitude and speed. This information is sent from a transponder under the saddle of each bike every second during the race. One of our considerations needed to be that some of the environments where the devices need to work have high levels of RF traffic.

For example, the technical zone at the end of each stage has hundreds of Wi-Fi networks, thousands of mobile devices and more than 50 TV broadcasters. The zone is, in short, the origin of the live TV feed for the race's global audience. This creates a high RF-noise environment right at the finish line of every stage.

The design principle was to shield the electronics from this static clutter by enclosing the core microcontroller unit and other sensitive componentry within their own Faraday cage. The data collected is then transmitted via a two-stage transmission network:

- The primary transmission network is a wide wireless area network – or WWAN - based on the 3gpp standard for 3G (3GPP 802.15.4). We create a mesh network between telemetry devices and relay points, this creates a moving mesh network with the ability to use the other telemetry devices as reference points to enhance the accuracy of the location coordinates and the ability to use the best relay points in the vicinity, this network then sends the data to the airplane (<10km distance away).
- The secondary transmission network is used to transmit that data to the end of the race. This network uses three licensed microwave frequencies. The data is multiplexed into the signal and transmitted to the end of the race in a near-line-of- sight manner along with the camera footage from around the race. The end-of-race receiver is placed on a mobile lift 20-50 meters above the technical zone at the end of the race.

We designed the data collection endpoints to be redundant. This architecture was put in place so that in the event of hardware or software failure, the hot standby server can be activated to enable service continuity.

In 2021 our array of new data sources add a highly diverse set of transmission types and protocols which enable the digital twin of the race.

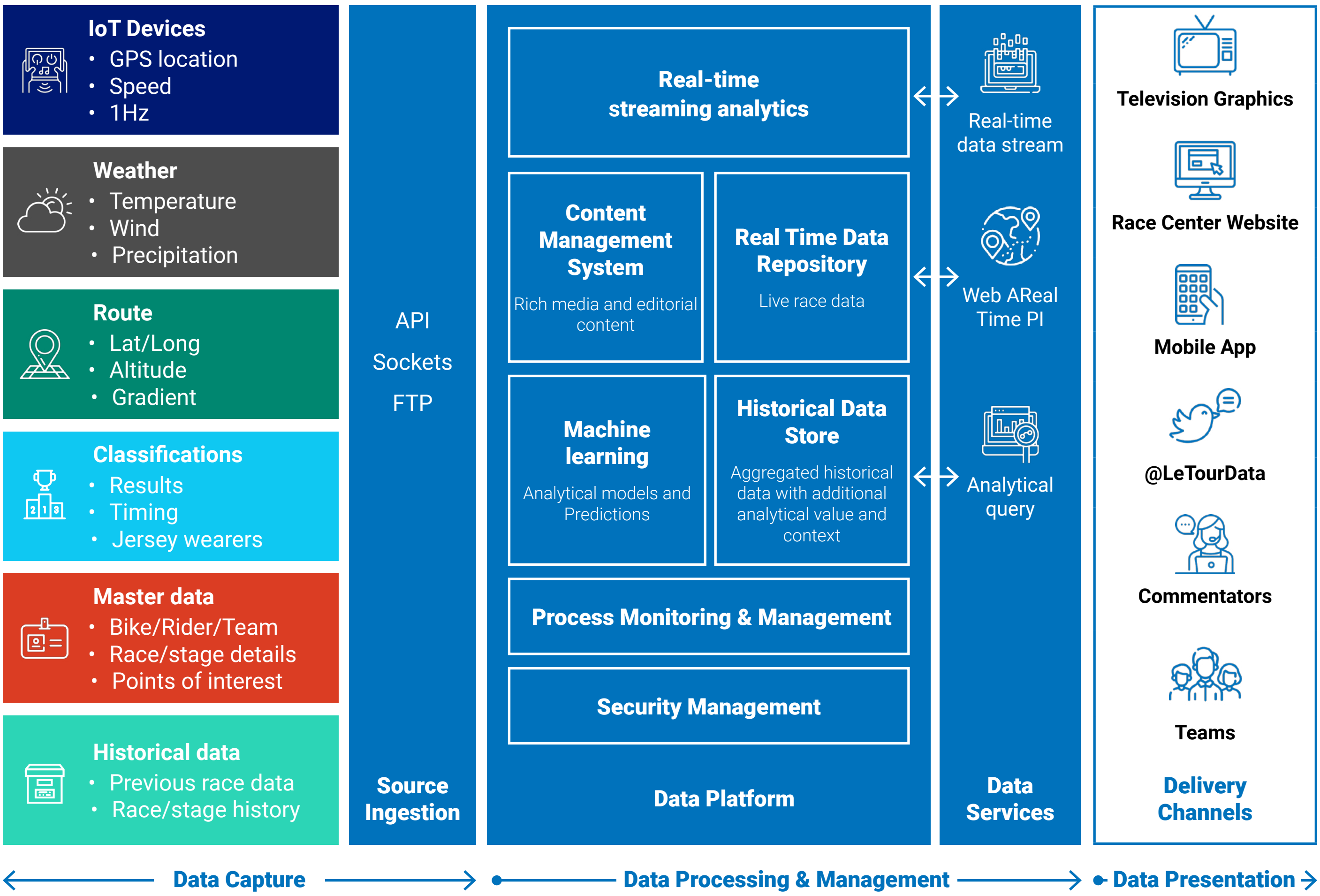
One of our **considerations needed to be that some of the environments where the devices need to work** have high levels of RF traffic.

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,	SigFox trackers	SigFox (Ultra narrowband BPSK)	Renard
Merchandise, WC, trucks, zones)	SigFox trackers	SigFox (Ultra narrowband BPSK)	Renard
People	Mobile application	3G/4G/5G and Wi-Fi	MQTT
Crowd counts	NTT Smart platform, Wi-Fi	3G/4G/5G and Wi-Fi	HTTPS
Weather	Weather stations	868Mhz	HTTPS
In-race car tracking	NTT Trackers	3G/4G/5G	MQTT
In-race cameras	Video encoding	TCP/IP	RTMP/RTSP
Maps and geo overlays	KMZ/PDF	TCP/IP	HTTPS

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

Tour de France logical architecture

A rich analytics and digital platform that enables innovative multi-data story telling



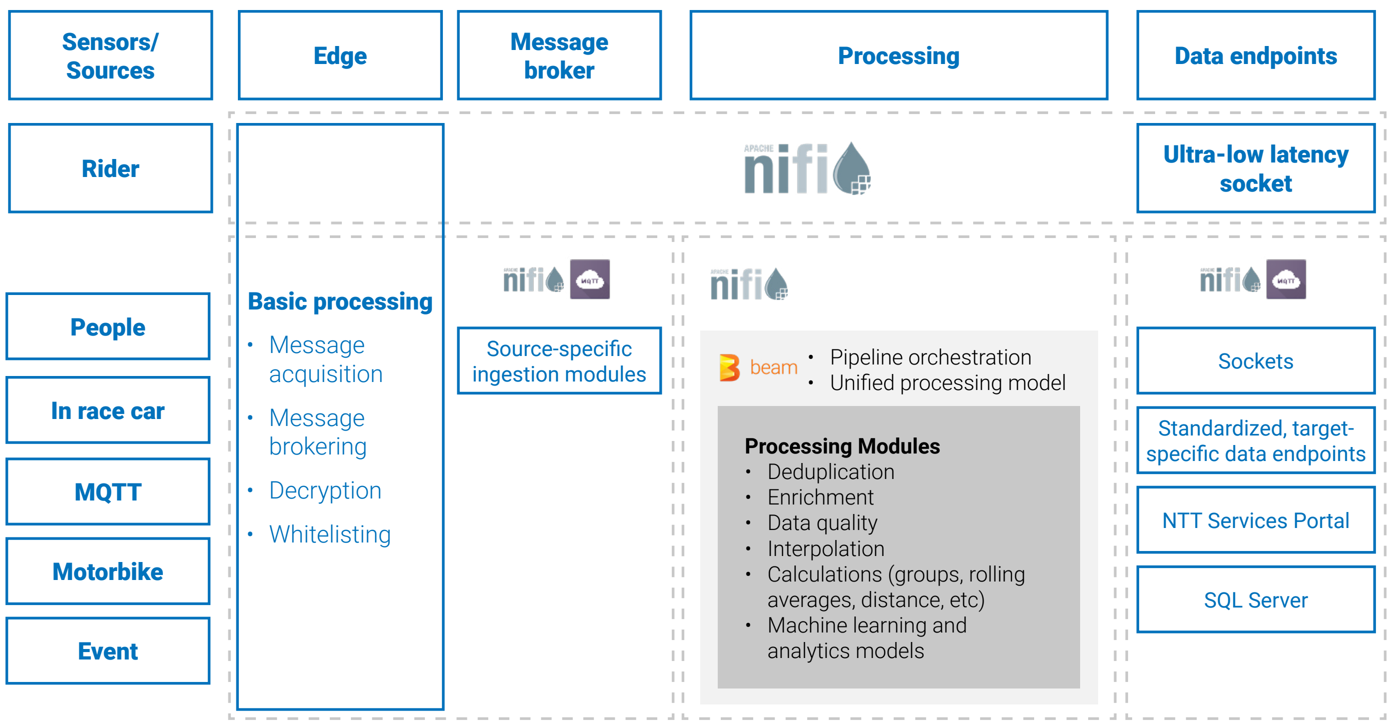
The data can be split into segments:

- **Telemetry data:** This is the data that comes from rider-tracking devices and is 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 giving 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 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 includes ordinance survey information, which gives details of the terrain, in particular gradient and 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.
- **Social media data:** Data is taken from Twitter handles, for example, @letour or @letourdata.
- **Media data:** Media data is inputted into the system via a content management system. This data includes videos, data, and images.
- **Machine learning data:** Machine learning data is created and ingested via a direct database connection.
- **Smart event data:** Data that gives context and enables us to construct the digital twin of the event. For example locations, map overlays and environmental information.
- **AI data:** An athlete ranking system that focuses on results and considers the type of event and who you are competing against.

Real-time analytics, AI and ML

The heart of the Tour de France solution is the real-time analytics architecture. The requirements of delivering a service that is transmitted to the entire world and aligned with live television broadcast are low latency and reliability. Ingesting data and analysing it with sub-second 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, Mosquitto and Python which are illustrated below:

Tour de France real-time analytics architecture



Due to 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
- 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 de-duplication
- 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 providing us with an understanding of the race and event situation.

Some of the metrics we have created are:

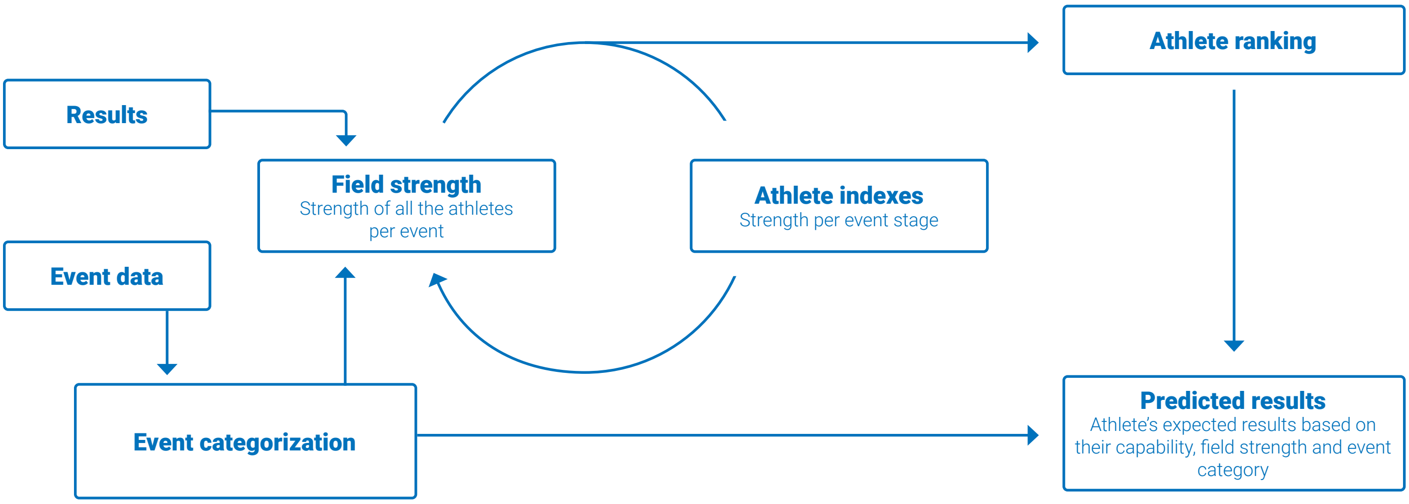
- Riders: speed, position, gaps, direction, relative wind, gradient, distance from start, distance from finish.
- Groups: speed, position, gaps, composition, distance from start, distance from finish.
- Event: correlation, crowd information, ETA.

The data streamed through the platform is not written to disk to ensure the sub-second latency that the Tour de France use case requires, but once the analytical models have been run the output is stored in a relational DB.

AI and ML 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.

Athlete analytics to determine strengths, weaknesses, and the profile of riders and teams



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. We have access to more data, technology and expertise and this allows us to predict:

- **Catch predictor** - #NTTPredictor: will the breakaway be caught by the peloton before the end of the stage?
- **#NTTEffortIndex** - 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, and what sorts of races or stages are they suited for?
- **Stage favourites** - #NTTPredictor: Which riders are likely to do well on a given stage based on their profile, results and the nature of the day's route?
- **People counting** – Based on video analysis at the edge.

We executed machine learning using three core pillars.

Data:
The richer the dataset available, the more factors the model can consider. After creating machine learning models based on Tour de France data, we use seven years of race results from an external source, and also made use of external third-party data services incorporating factors such as the weather to enrich our data with additional features that provided greater context to the raw data. The example below shows the data that is applied to enable the machine learning algorithm to make the catch prediction.

> 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.

Team:
The skills required to deploy and train a machine learning solution are a critical component of its success. We built a team consisting of:

- **Data scientists** – design and develop the machine learning models.
- **Data visualization experts** – present the data and predictions visually to make it easy to understand.
- **Product owner** – ensures the project is on track and focused on the agreed objectives.
- **Business expert** – brings deep knowledge of the business subject area and data.
- **Data engineers** – experts in data integration, storage, and analytics.

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

- **Real-time predictions** – The real-time predictions take the data from the live tracking platform and apply the machine learning models to it in real-time. This enables us to effectively predict the effort estimate and catch predictor.
- **Batch predictions** – Overnight batch predictions are run, considering the data from the previous stage. This enables us to predict stage favourites while applying the riders' performance profiles to allow us to model the next day's stage.

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 second per rider. 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.

Broadcast – The Tour de France is a global event and the data we provide to these teams needs to be universally understandable and formatted in 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, enhancing the viewing experience.

Race Center – Race 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 fans can identify individual riders and follow their progress in real-time.

@letourdata – Through 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 platforms and broadcast.

3D Race Tracker – Increasing the opportunities for fans to follow the race is a key objective for A.S.O. Our augmented reality app allows race fans to track the progress of all the riders and groups, giving roadside fans a new view into the progression of the race.

Media Wall – To enhance the experience at the Departure and Arrival Village, this year we're piloting the NTT Media Wall. Delivering an integrated experience on a large screen, this provides a curated set of race data, combined with live race footage and insights into the riders and teams for fans on the ground.

Operations dashboard – To enhance the operational efficiency of race organizers the operations dashboard provides them with real-time data on the status of the race. Combined with footage from the race this provides them with full visibility into all elements of the race including riders and race vehicles. This enables them to make data-driven decisions to ensure that each stage proceeds smoothly.

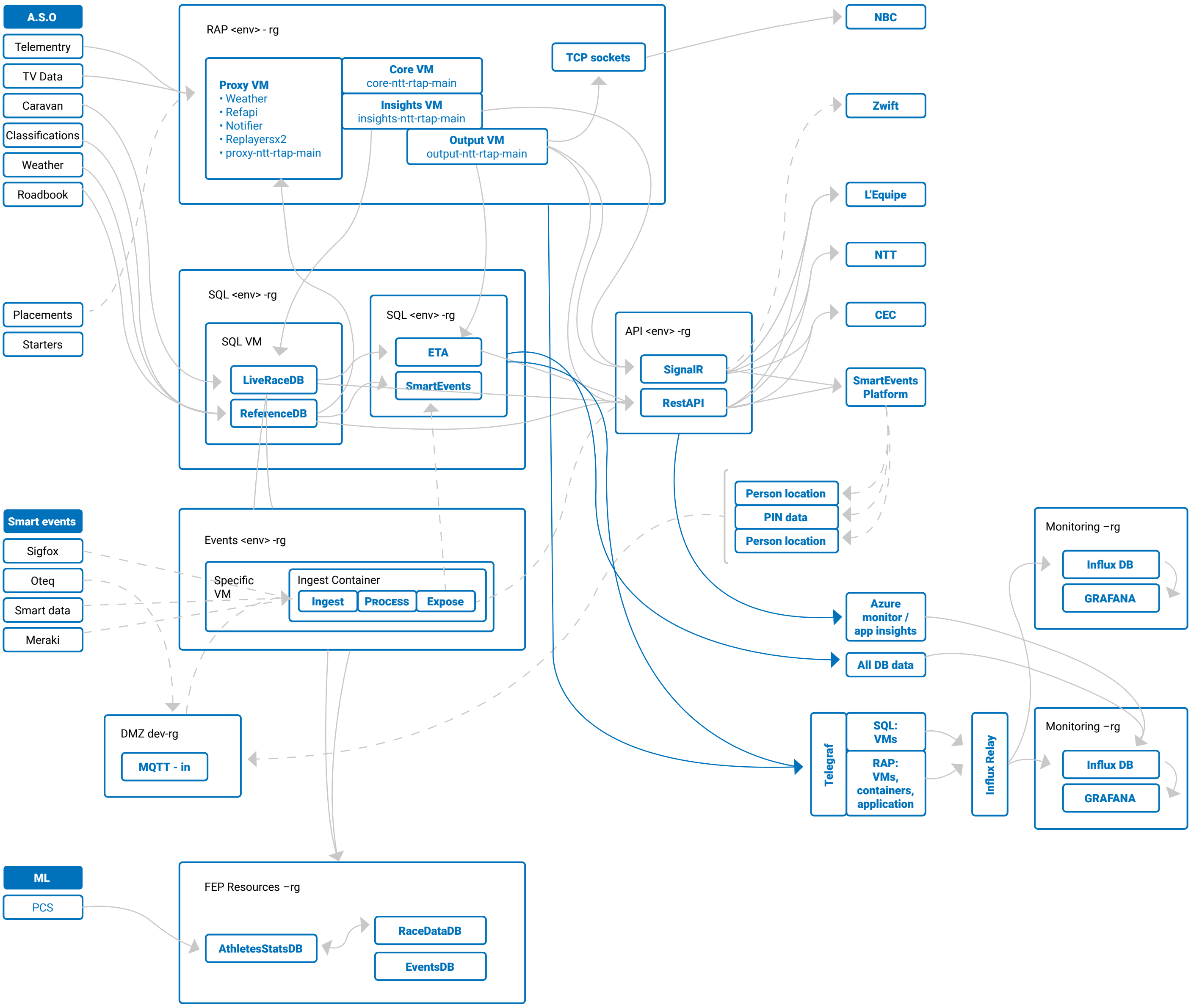
Event app – In a highly dynamic environment, access to data on-the-go is critical. The A.S.O. Event App leverages our digital twin models to plan each stage, including the placement on the Departure and Arrival Village. Personnel can monitor the status of all race assets and dynamically change existing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

Architecture

Hybrid IT

When creating the architecture for the Tour de France the technical requirements and architecture of the solution were heavily influenced by A.S.O., the spectator/fan demography and the event location.

- Right service in the right location at the right time: Utilizing the 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.
- Business-critical service: A high-profile event where security is paramount and data protection laws must be enforced.
- Data had to stay in the EU.
- Deployed as code to enable easy teardown and rebuild on different types of infrastructure.
- In-sync with other services: Some aspects of the solution should process data and make this data source available to data consumers in near real-time, which called for sub-second processing.



The conceptual solution architecture can be illustrated in the following manner:

Unique solutions for a unique event

Event edge computing and networking

Some of the challenges of transporting IT equipment over 3,400 km in a month are not 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 on how to do field maintenance are critical elements to keeping data flowing.

Networking and wireless

- Cisco Meraki MX67/64
- Cisco Meraki MR52 Wi-Fi AP
- High gain 4G LTE antenna

Edge compute and services

- Intel NUC i9 64Gb with Nvidia CUDA GPU
- AJA Helo video encoders
- AXIS 4k camera's

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.

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

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; we utilized the standard in the creation of the overall strategy.

The high-level process is outlined below:

1. Understanding stakeholders and requirements.
2. Understanding legal and regulatory requirements.
3. Define organizational requirements.
4. Define the business continuity plan (BCP) and its scope.
5. Establish the BCP plan 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, which 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 multi-region incident take place.

Management

To give an overarching, single pane of glass view for the services and infrastructure used to support the Tour de France we implemented the NTT service portfolio. The three areas we leveraged were:

- network and compute with MHIS
- communications with MCS
- holistic view of everything with customer portal

Support from the virtual zone technique

Our solution for A.S.O. is delivered using a distributed global team working across five countries.

Cloud platforms and leading collaboration platforms allow us to place workloads in the best places and provide 24/7 coverage.

- Customer portal – UK
- Observability platform - UK
- Analytics platform, data services, API – SA
- AR App – Australia
- Letourdata – France/SA/Australia/UK

The trust, accountability, planning, tools, technical skills and, most importantly, pro-active communication all drive successful operational delivery.

Enabling this team to support the Tour de France in the model where we have moved from onsite support to a globally distributed team, we have had to ensure, using some of the technologies we have previous discussed in this article, that the team have the correct communication channels, connectivity and observability of the solution.

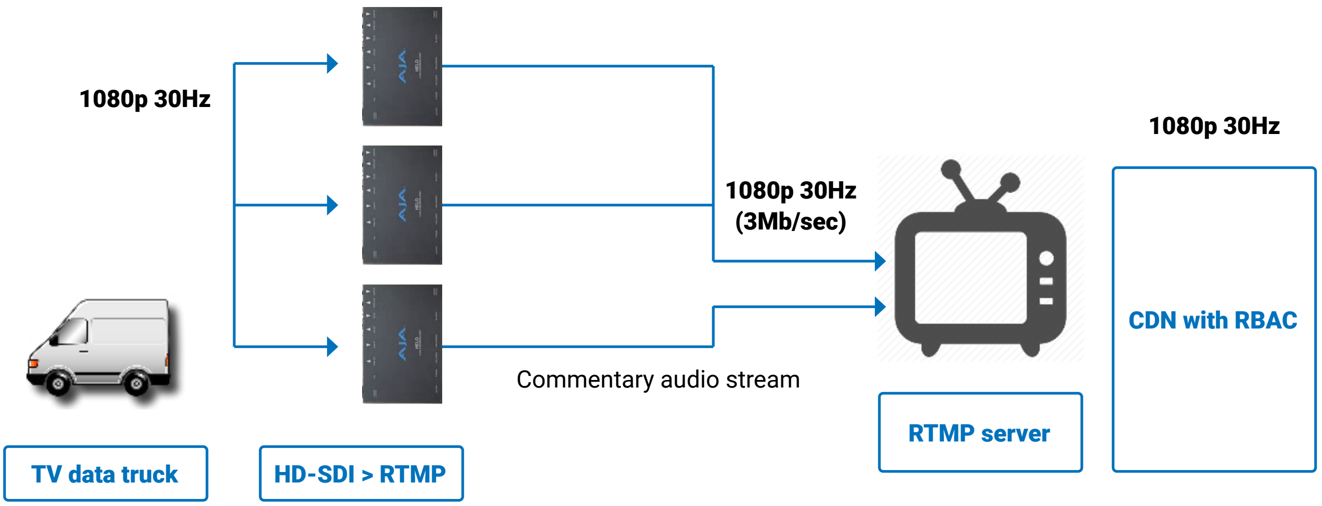
Communication and observability

During live race hours, we use live video conferencing supported by the NTT MCS service. This allows us to have real-time communications during very time-critical moments.

Live video

To enable the teams to see what is happening live during the event, we have live video encoding, which is performed in the following way:

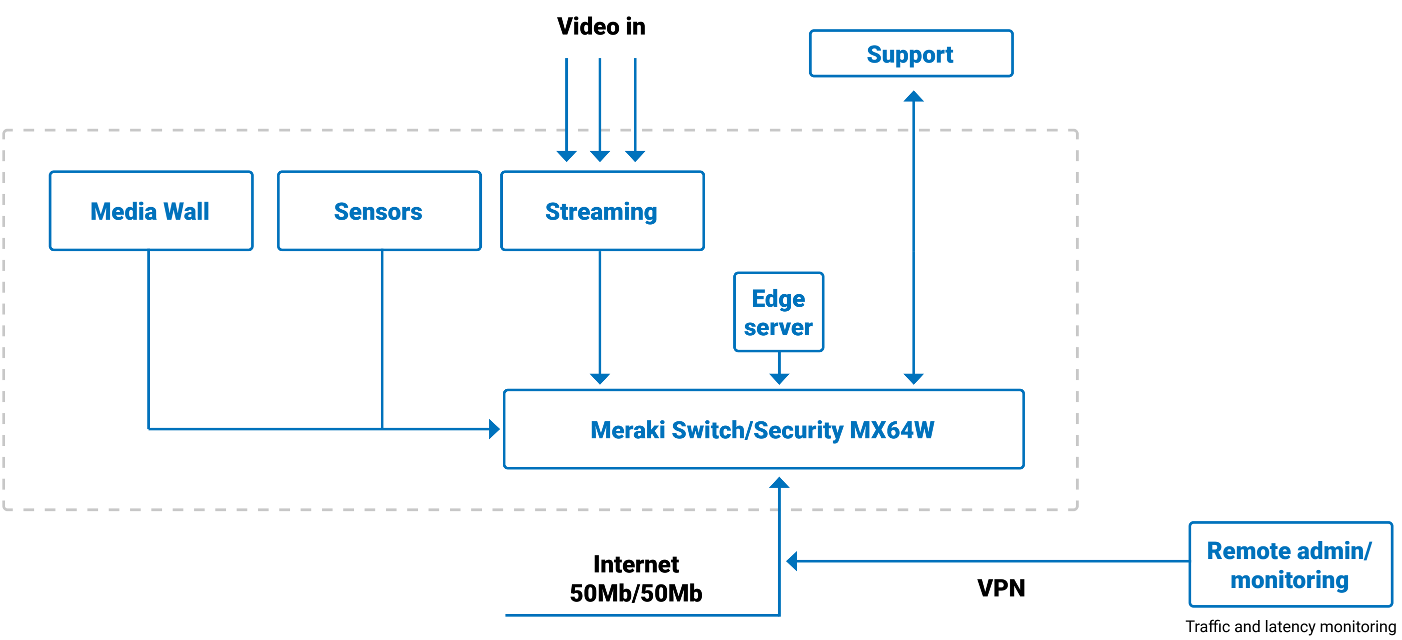
Live video architecture



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:

End of race architecture

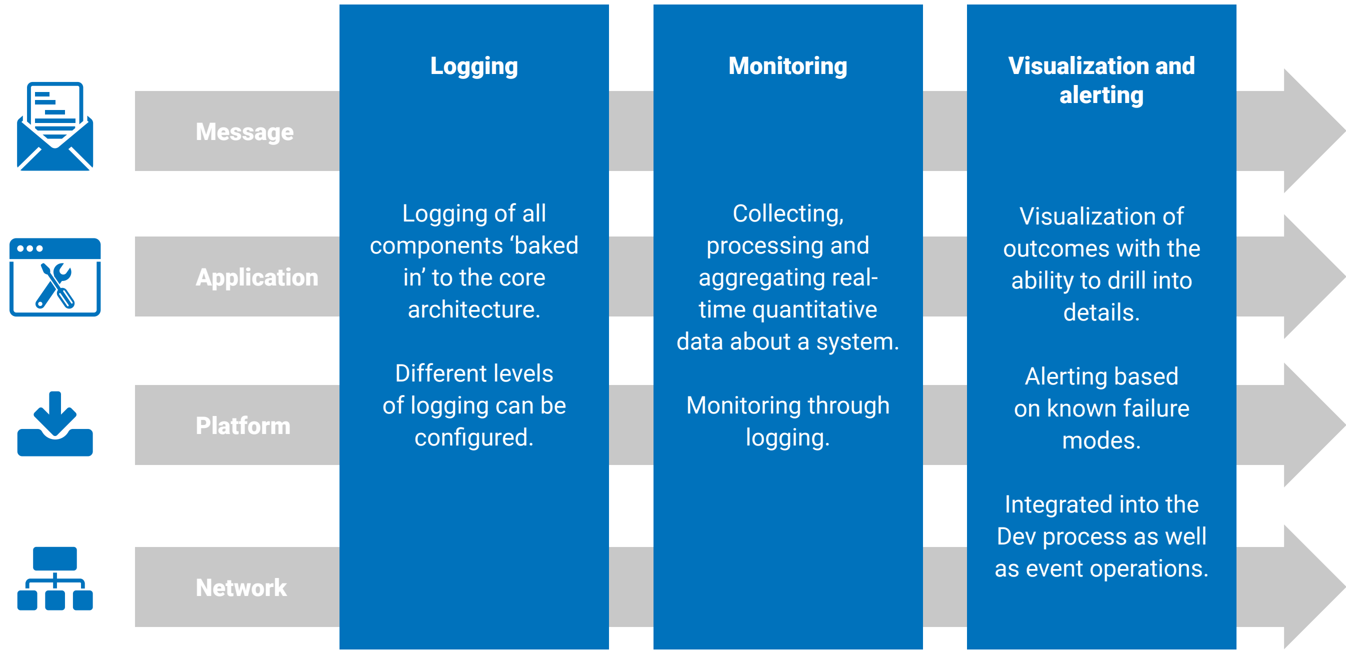


Observability

We have broadly followed the Golden Signals methodologies to drive the observability process and provide a consistent experience for the end-user. The collection of the right metrics is key in providing useful information:

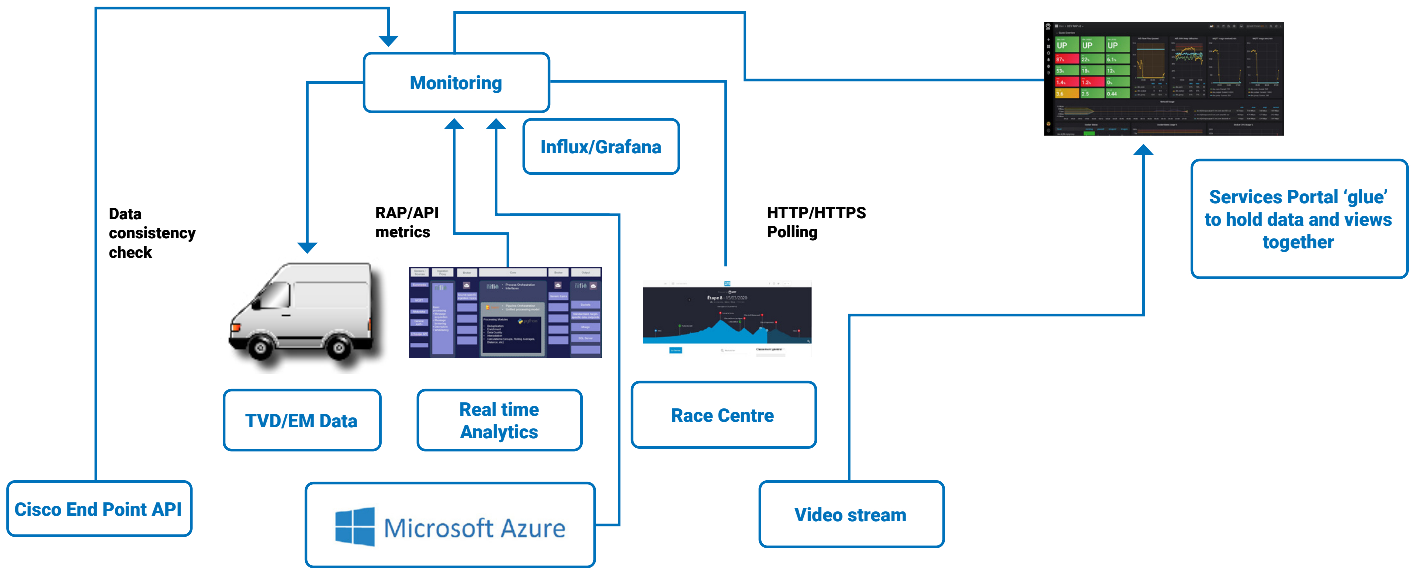
- **Latency** – The time it takes to service a user request.
- **Traffic** – A measure of how much demand is being placed on the system (HTTP requests, network throughput, transactions).
- **Errors** – The rate of requests that fail.
- **Saturation and utilization** – Consumption of constrained resources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

Information architecture



This has led our design process from collection to alerting, providing essential and useful information to empower the user, forming one of the core components that enables the virtual zone technique.

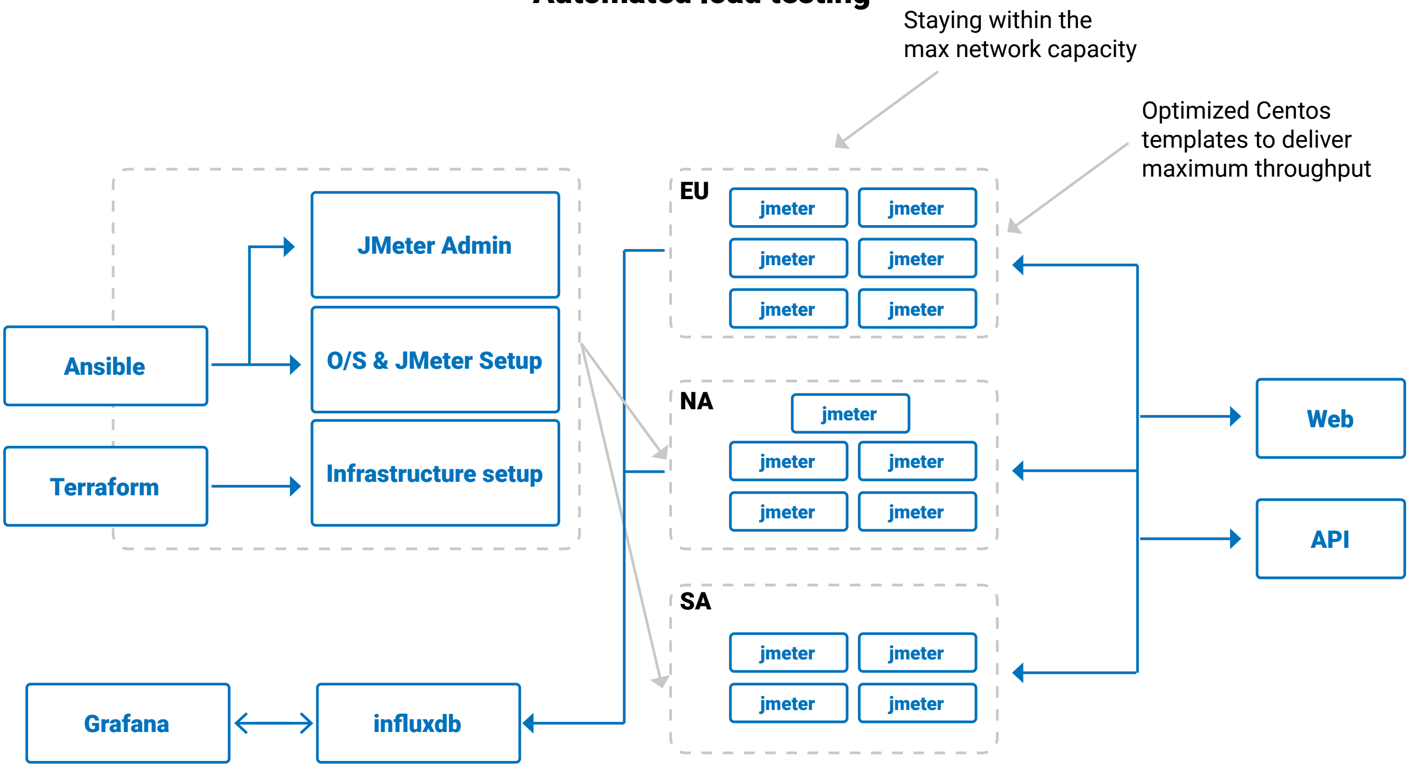
Virtual Zone Technique observability



Testing

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

Automated load testing



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

- review code scanning tools settings
- conduct static
- current scanning scope and methodology

Manual analysis and findings triage

- manual analysis of static code scanning results
- identify false alarms
- identify repetitive issues

Findings prioritization

- classification of the criticality of applications
- multiple factors prioritization based on application criticality, issue rating, business priority, etc.

Understanding of program design

- understanding of program logical, architectural and functional design as well as business logic and flow

Recommendations on closing issues

- 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

Together we do great things