dimension data

white paper

Tour de France 2016 – *A technology solution in action*

Abstract

As a primary technology partner for the Tour de France, a guiding principle of our technical solution was to deliver the business outcomes set out by the race owners, Amaury Sport Organisation (A.S.O.). Our ambition was to deliver these outcomes in the unique environment of the race and at the same time deliver more data insights than we did in the previous year's race.

With a multitude of data – collected by each rider's sensor and processed in real-time – harnessing the power of the internet of things (IoT), and creating a platform for this, underpinned the overall strategy for the race. This includes a dual transmission network, defined data segregation, as well as a process for filtering and data analytics, which are highlighted in this paper.

For the Tour de France 2016, we were taking data from tracking devices, equipped with progressive technology. Similarly, our big data truck was also fitted with enhanced technology. In fact, the big data truck is the nerve centre of the technical solution for the Tour de France. As a mobile workspace, it contains a cogent mobile data centre, connected to Dimension Data's global cloud platform, creating a cloud and hybrid IT framework to ensure a comprehensive solution. Outside the truck, the team also kept up a 24-hour test/development cycle to ensure the solution remained robust, reliable, and available to deliver real-time analytics on the race.

To ensure that we met the stringent business requirements of the A.S.O., we embraced a cloud and hybrid IT framework, as this would satisfy many of the application requirements and data protection laws of the European Union (EU). We also implemented a business continuity and disaster recovery plan to the highest ISO-standard to focus on possible risks and mitigating plans.

Putting the technology solution in action at the Tour de France

The internet of things

The Tour de France presents our technical teams with a unique challenge by having to reliably transmit data from a large cluster of devices to a central platform across various terrains along the Tour and in varying levels of radio frequency-(RF) heavy environments. We used the following architectural designs and techniques to overcome these challenges:

Device and transmission:

Transmission networks

Our technical experts use multiple transmission technologies for a dual-network configuration to transmit the data from the cyclist to the big data truck.

- The *primary transmission network* is a wide wireless area network – or WWAN - based on the third-generation partnership project standard for 3G (3GPP 802.15.4). The standard is used to create a mesh network between telemetry devices and relay points. In turn, this establishes 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.
- The *secondary transmission network* is used for the relay points to send data to the primary relay point in an aeroplane or helicopter and on to the end of the race. This network uses a licensed TV broadcast RF frequency. The data is multiplexed into the signal and transmitted to the end of the race in a near-line-of-sight manner. The end-of-race receiver is placed on a mobile lift about 40 metres above the technical zone at the end of the race.

Data transmission diversity

Data transmission diversity is critical for the solution as the challenging terrain can cause a multitude of errors. Device data is sent via relay points to the primary relay point and then on to the end of the race. This creates a more diverse transmission path. Information is de-duplicated to remove copies of data from the source before being processed.

Shielding

Some of the environments in which the telemetry devices have to work include heavy RF traffic. For example, the technical zone at the end of each stage has hundreds of WiFi 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 within its own Faraday cage. Unit antennae, of course, aren't located within the cage.

Redundancy

We designed the data collection end-points 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.

Data transmission diversity is *critical for the solution* as the *challenging terrain* can cause a multitude of errors.

Data

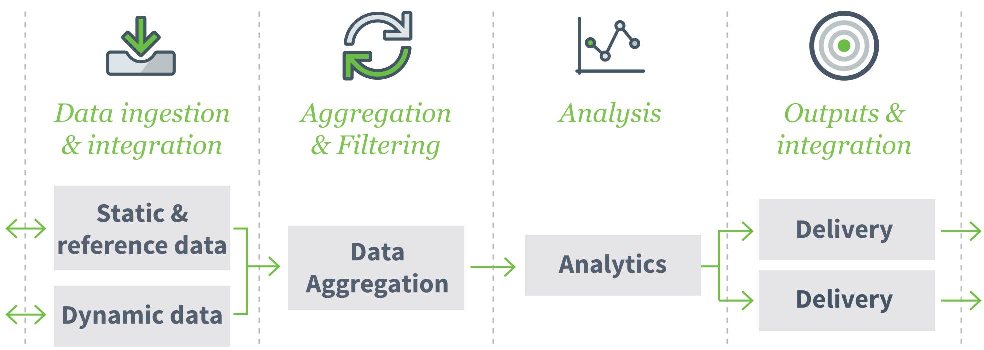
The data can be split into six respective segments:

- ***Telemetry data:*** This is the data that comes from rider-tracking devices and ingested into the IoT platform in the format of a race-situation JavaScript Object Notation – or JSON – file. Within this open-source file, we can see the timestamp, latitude, longitude and speed.
- ***Race data:*** At the beginning of each stage, the platform requires two key pieces of information. The stage data: this gives accurate location-based information for the whole stage – such as locations of sprints, start, finish and so forth. The race data: this details which riders are in the race, their current classification, bib number, rider name, team name, and so forth.
- ***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 hone in the accuracy of the output. This includes ordinance survey information, which gives details of the terrain, in particular gradient and height-above sea-level information. It also includes hyper-localised weather data, using the telemetry data to look up localised weather data for each rider.
- ***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, pictures, and so forth.

The platform ingests *third-party data sources* to enhance and further hone in the *accuracy of the output.*

The internet of things platform

The IoT platform is the critical element of the technical solution. The platform has been designed with extensibility and flexibility in mind. This enables it to be used across verticals to deliver specific client driven outcomes. The data processing is split into sub-sections each of which handle specific tasks.



Data ingestion, outputs, and integration

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

- Transmission/transport control protocol sockets
- Universal/user datagram protocol ports
- Hypertext transfer protocol secure listener/publisher
- Message queuing telemetry transport broker
- File-based formats, which are agreed at each integration

Data aggregation and filtering

Once data has been ingested into the platform, we have a rigorous series of steps to handle elements like data duplication, erroneous data, or missing data. We also look to combine the data into an efficient and manageable structure, which can have the appropriate algorithms applied to it.

The storage of this aggregated data is handled in a few ways, depending on the data type. For structured data, we use a typical relational database management system set-up (Microsoft structured query language). For unstructured data, we're using a document database solution (MongoDB). Data that is currently in 'flight' is stored in an in-memory database (Reddis).

Analysis

Taking the data we've ingested and creating meaningful analysis, which can help tell stories about the race situation, has been developed in the way of algorithms that produce a particular output. Some of the metrics we have created are:

Rider: speed, position, gaps, direction, and wind relation

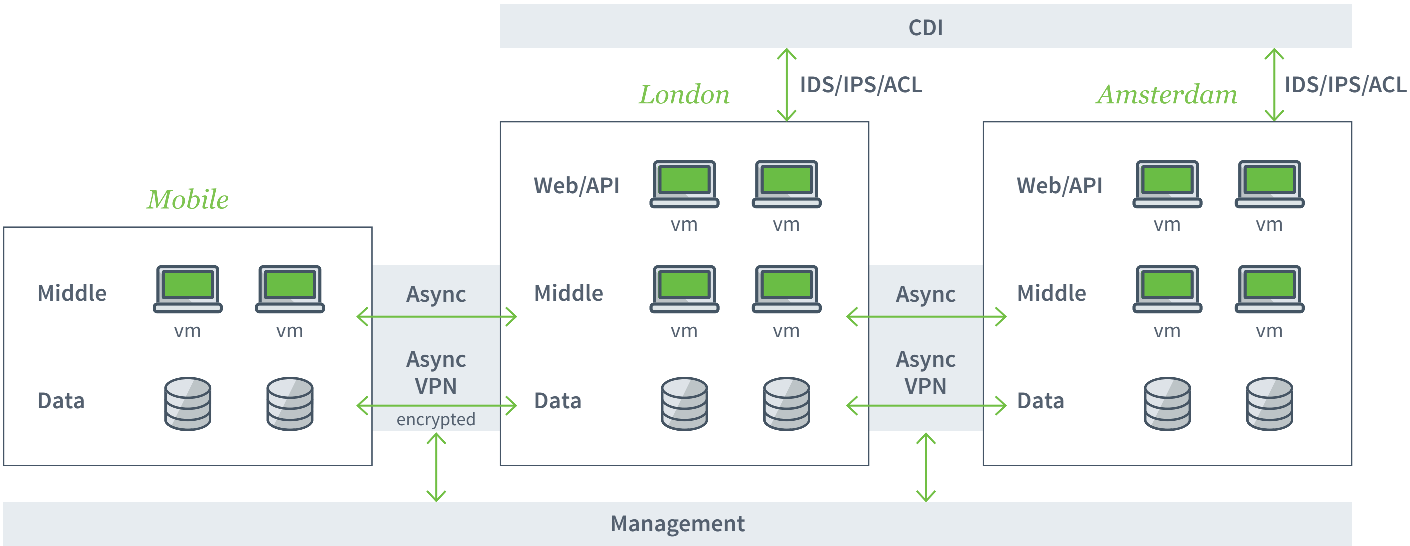
Groups: speed, position, gaps, and composition

Cloud and hybrid IT

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

- ***Large global audience:*** Public web services should be able to service hundreds of thousands of requests per second.
- ***Ability to monetise:*** Data must be available to other consumers in a secure and measureable 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.
- ***In-sync with other services:*** Some aspects of the solution should process data and make this data source available in data consumers in near real-time, which called for sub-second processing.

The conceptual solution architecture can be illustrated in the following manner. However, this doesn't give explicit detail around some of the technologies or the exact amount of virtual machines used in the live solution.



The infrastructure solution was split into two sections:

- cloud-based services
- mobile data centre and networking

Cloud-based services:

We used Dimension Data cloud-based services to deliver a 100%-available platform for the Tour de France. This was a geo-diverse and fully redundant architecture, which integrated into the mobile data centre deployment. To further facilitate localised delivery models, we integrated a content delivery network from a Dimension Data partner to provide some offload to the web application programming interface – or WebAPI – tier, and create a higher performing global user experience.

All cloud-based virtual machines were integrated into the Dimension Data cloud monitoring services, which initiated the required levels of failover, redundancy, and monitoring services to meet our service level agreements. Stateful machines were configured to use cloud backup in the event of data loss.

The development, deployment, and operations of the platform was performed in a development and operations (DevOps) manner. We leveraged tools like Confluence to enable effective design collaboration, Jira for software development planning, and Octopus Deploy to facilitate clean, repeatable, and automated installation and roll out of new code and builds into the platform.

Mobile data centre and networking

Some of the challenges of transporting a platform 5,000 kilometres in a month are not always obvious. For example, outside temperature can affect the inner cooled area when it varies by 50 degrees, depending on the location and day. Vibration plays a huge role in the type of technology that is used. Moreover, ensuring that spares are available and having an understanding of how to do field maintenance are critical elements to keeping data flowing.

The on-site physical solution was provided by a blend of Dimension Data partners:

Networking

- Cisco Nexus 5k: providing core switching and segmentation capabilities
- Cisco SGH 500: for management and external WiFi connectivity
- F5 Bigip 5050: providing load balancing and security; this was where main internet connectivity fibre was integrated

Compute

- Dell EMC VCE VXrail 120: hyper-converged virtualised platform used for core services
- Intel Nuc: used for remote mobile compute for example, press centre installation and dashboarding

All cloud-based virtual machines were ***integrated into the Dimension Data cloud monitoring services,*** **which initiated the required levels of** ***failover, redundancy, and monitoring services*** **to meet our service level agreements.**

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 utilised the standard in the creation of the overall strategy. Additionally, we took Dimension Data Cloud ISO 27011:2015 certifications as an underpinning set of assurances around the cloud services we utilised.

The high-level process is outlined below:

1. understanding stakeholders and requirements
2. understanding legal and regulatory requirements
3. define organisational 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 mitigating plans.

Having no single points of failure in the solution, the analytics solution triplicated and the web/api duplicated, which gave us a robust solution. However, the approach to high availability in any architecture can change slightly when some elements, by their nature, have to be a single point of failure. Our big data truck is a good example of this.

Big data truck

Parked in the Tour de France technical zone as a mobile workspace, the big data truck was where our technical, communications, and live data reporting teams came together every day of the event. It was a nerve centre.

Overall, this solution is the most critical element and yet the most difficult for which to create a plan. Some questions dominated our strategy. What happens if the truck wasn't there? Or was involved in a crash, or caught on fire, and so forth?

Our plan was simple yet effective. Should we have an issue with the truck, we would provide internet-facing interfaces to connect to, and utilise a templated instance of the IoT platform to deliver the service, albeit with a slightly higher latency due to the extra distance the data would had to travel.

This was put to the test at Mount Ventoux. During Stage 12, the stage route was changed due to severe weather forecasts. As a result, the technical zone had to be split into two separate areas, with no interconnectivity. We activated our business continuity plan, which involved a switchover of technical zone integrations, deploying the truck's built-in virtual memory system and, in the end, managed to achieve 100% service availability according to our service level agreements.

If we went one step further and lost all connectivity and primary truck infrastructure, we had a scaled-down version of the truck set up and a long-term evolution connectivity solution to further mitigate against the full failure of the solution.

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

Technical team

The technical team was a serious concern when it came to service delivery. If they didn't make it to the end of the race, the solution would not work. If this happened, it would be for multiple reasons:

- the daily set-up and on-the-ground infrastructure integration and deployment
- the check and starting of the platform
- in-race monitoring and local interfacing

We mitigated these risks with a few approaches. Whenever possible, we ensured that the technical team travelled in separate cars; members of each car have technical capability and competence to set up and start up the solution in the absence of the other team.

Cloud data centre connectivity loss

In the event that we lost full connectivity to either London or Amsterdam Dimension Data clouds, we had architected the solution to be highly available between the two clouds. Moreover, the switch over would be handled automatically and not result in an outage. Leveraging the cloud exchange networks to keep the geographically diverse sites in sync made this approach a lot easier.

Conclusion

The hybrid, comprehensive, over-arching architecture and orchestration of the solution proved its mettle in the Tour de France 2016. We achieved exceptional uptime while processing 127.8 million data records.

The rider data was processed in our hybrid cloud. More than 100 virtual machines leveraged 300 unique cloud compute services, delivered at 99.999% availability to serve 55,000 requests per second. Furthermore, the live-tracking website supported 17.8 million viewers with over 2,000 page requests per second.

Tim Wade

Group Senior Director – Technology and Architecture: Digital Practice
Dimension Data