dimension data accelerate your ambition

white paper

Tour de France 2017 – *A technology solution in action*

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 2017, 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 was the nerve centre for the operational aspects of the technical solution for the Tour de France. As a mobile collaboration and communications workspace, it contained a cogent mobile data centre, connected to Dimension Data's global cloud platform, Dimension Data Managed Services for Enterprise Networking, Data Centre and Manage Centre, creating a hybrid IT hub. 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 A.S.O., we embraced a hybrid IT and managed service strategy, 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 mitigation 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 the various terrains found along the route of the Tour and with 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 cyclists 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 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. 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 either directly or 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 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 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.

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

Data

The data can be split into seven 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 giving accurate location-based information for the whole stage including locations of sprints, start, and finish.

The race data detailing which riders are in the race including their current classification, bib number, rider name, and team name.

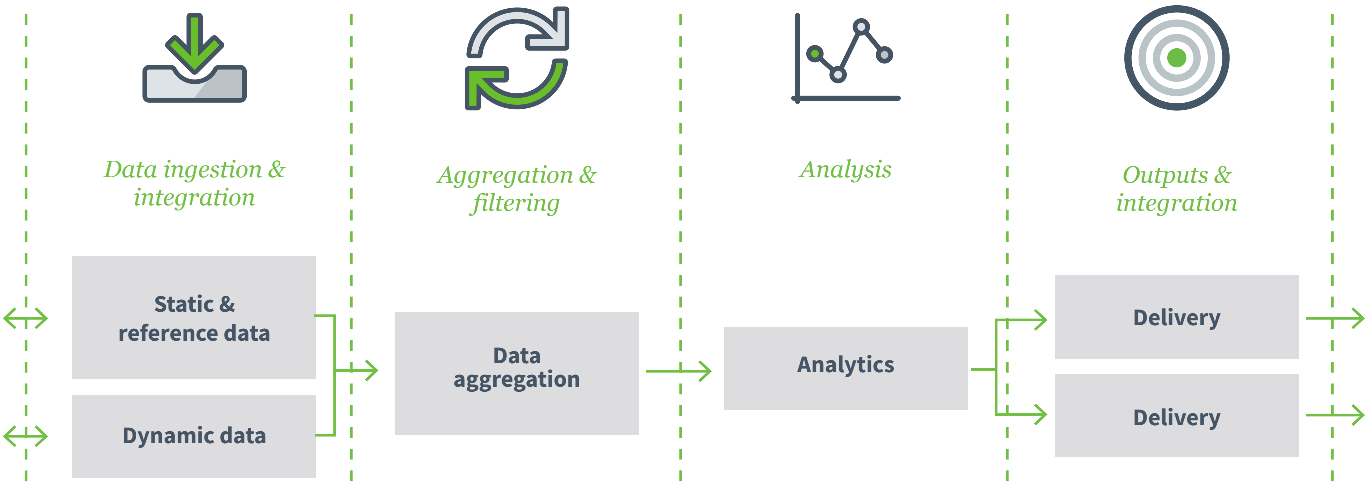
- *Timing data:* Data from the official timing provider, which provides official classifications, results, and photo finish information.
- *Environmental data:* The platform ingests third-party data sources to enhance and further refine the accuracy of the output. This includes ordinance survey information, which gives details of the terrain, in particular gradient and 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.

The stage data giving accurate location-based *information for the whole stage including locations* of sprints, start, and finish.

- *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 picture.
- *Machine learning data:* Machine learning data is created and is ingested via a direct database connection.

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
- Direct database connection

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 applying meaningful algorithms that deliver an understanding of the race situation and ultimately enable data journalists to tell the race 'story'. Some of the metrics we have created are:

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

Groups: speed, position, gaps, and composition

Machine learning

Building on the analytics platform and algorithms this year we used a combination of telemetry data, race results, rider data, course information, and conditions data to predict race outcomes. Dimension Data has access to more data, technology and expertise than has ever been available before; what did we predict?

Catch Predictor - #DDPredictor Will the breakaway be caught by the peloton before the end of the stage?

#DDEffortIndex How hard are the riders working at the moment? What is the difference in effort in different groups?

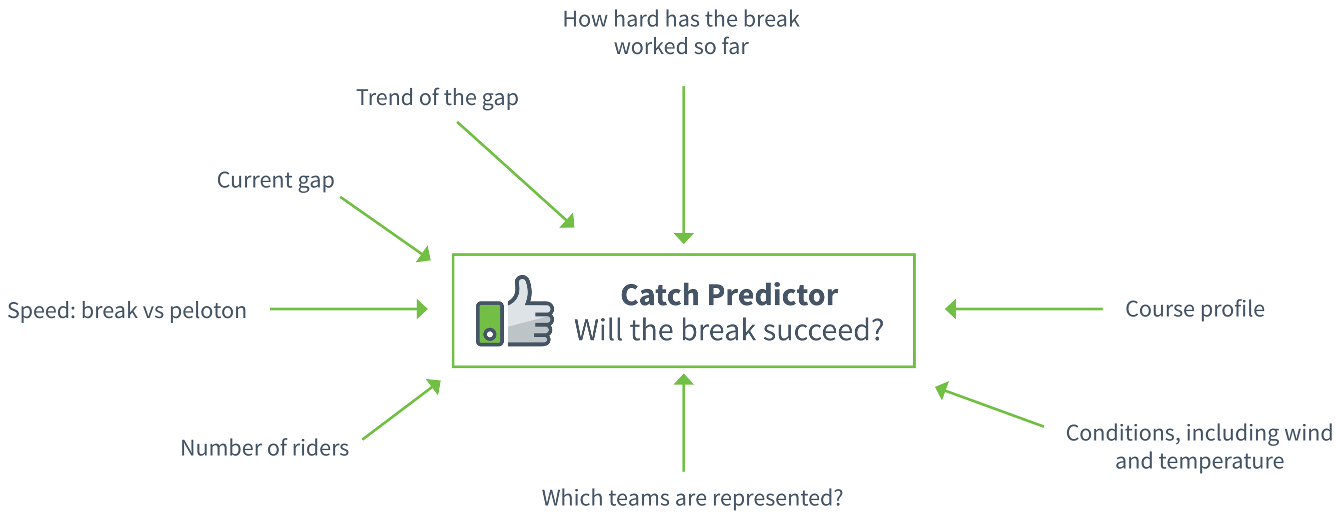
Performance Profile What are the attributes of different types of riders, and what sorts of races or stages are they suited for?

Stage Favourites - #DDPredictor Which riders are likely to do well on a given stage based on their profile, results, and the nature of the day's route?

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 TDF data, we used five years of race results from an external source, and also made use of external 3rd 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.



Team

The skills required to successfully deploy and train a machine learning solution are a critical component of the success, we built a team consisting of:

- *Data scientists* – design and develop the machine learning models
- *Data visualisation 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

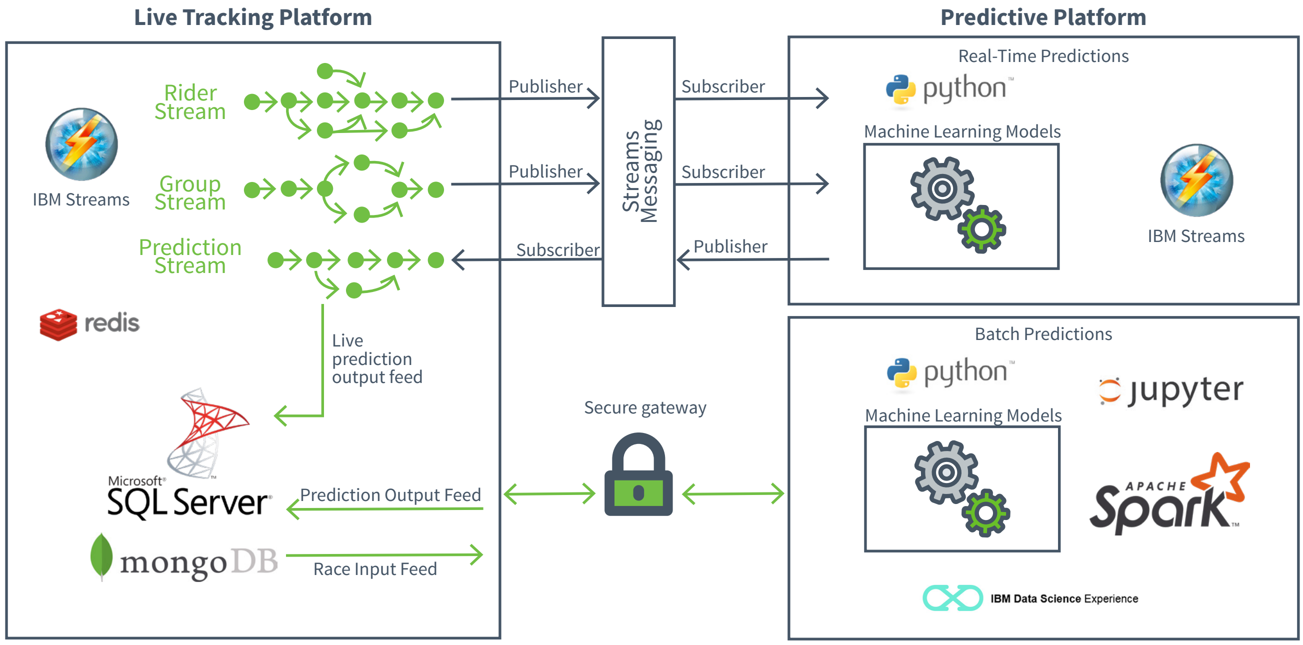
How does it work?

We are 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, taking into account the previous stages data. This enables us to predict the stage favourites and apply the riders' performance profile to allow us to model the next day's stage.

We use the following technology stack to deliver these outcomes:



Cloud and hybrid IT

When creating the architecture for the Tour de France, 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 measurable manner.

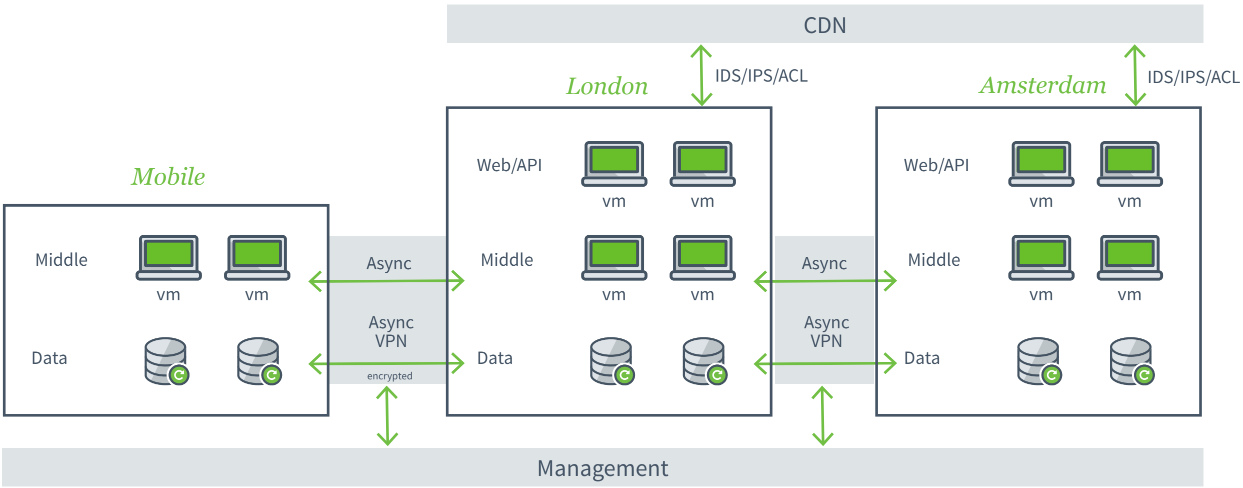
When creating the architecture for the Tour de France, A.S.O., and the spectator/ fan demography, *heavily influenced the technical requirements* **and the architecture of the solution.**

- *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 (WebAPI) tier, and create a higher performing global user experience.

All cloud-based virtual machines were integrated into the Dimension Data managed 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 were 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 km 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 Dimension Data partners:

Networking

- Cisco Meraki MX84 Security Appliances
- Cisco Meraki MS350-48 Switches
- Cisco Meraki MR52 Wi-Fi AP

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 mitigation plans. Having no single points of failure in the solution, the analytics solution triplicated and the WebAPI 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 since our experience at Mount Ventoux. During Stage 12 2016, where 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 then activated our business continuity plan, which involved a switchover of technical zone integrations, to be purely cloud-based with redundant public interfaces. We have since moved to a model where the connectivity can be easily switched and the issues become more of a people and space problem that can be overcome with the use of other trucks and premesis.

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

> **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.**

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.

Management

To give a real overarching single pane of glass view for the services and infrastructure used to support the Tour de France we implemented the Dimension Data service portfolio, the four areas we leveraged were:

- Network Management with MSEN
- Data Centre management with MSDC
- Communications with MSVC
- Holistic view of everything with Manage Centre

Conclusion

The hybrid, comprehensive, over-arching architecture and orchestration of the solution proved its mettle in the Tour de France 2017. We achieved exceptional uptime while processing over 3 billion data records.

Tim Wade

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