Utilizing Real-Time Analytics to Drive Real-Time Decisions for Operators

Analyzing operators' fast data to drive improved operations and revenues

Publication Date: 08 February 2017

Adaora Okeleke

Executive summary

Catalyst

With the almost universal penetration and use of mobile broadband, telecom operators are finding that their networks and operations are generating more data, at faster rates, than ever before. In turn, consumption of digital services places demands on operators to deliver high-quality customer experiences. Exploiting live streams of data to derive insights that enable immediate and correct responses has become critical for telecoms application providers. However, one of the biggest challenges lies not in simply collecting the data, but correlating and analyzing it in real time to proactively respond to situations that present revenue opportunities or impact a subscriber's experience. Telecom operators and application vendors alike must understand how these live streams of data can be harnessed effectively using existing and upcoming technologies.

Ovum view

The telecommunications industry is facing multiple challenges; revenues are declining, yet traffic is growing exponentially while competition is intensifying among peers and over-the-top (OTT) service providers. While operators have adopted several strategies to deliver and monetize new services, as well as transform existing operations and network infrastructure, they must do a better job of harnessing insights generated from live, real-time data streams.

While much has been said about big data (historical stored data generated across the business), there is also significant potential that can be derived from the analysis of fast data. First, real-time analytics will enable operators to deliver the personalized service their digital subscribers expect. Second, operator applications can deliver proactive capabilities that use insights based on the current context of each subscriber and the services they consume, opening the way for responses such as targeted offers or delivery of just-in-time care to subscribers.

Traditional business intelligence (BI) and data warehousing approaches that rely on stored data and bulk analysis are incapable of delivering timely insights. Conversely, fast data technologies can enable the processing of live streaming data to drive immediate, actionable, and effective business results. Real-time analytics, using fast data technologies, is becoming essential for operators who are embracing innovations such as software-defined networking (SDN) and network functions virtualization (NFV) to support rollout of new services such as Internet of Things (IoT) and machine-to-machine (M2M) communications.

Key messages

- Declining revenues, increasing network traffic, and rollout of new technologies such as SDN/NFV are driving the need for operators to make quicker decisions.
- Fast data technologies enable operator applications to make real-time decisions based on the insights obtained from the analysis of live data streams.
- Key real-time analytics use cases include targeted marketing, churn prediction, and real-time billing and charging.

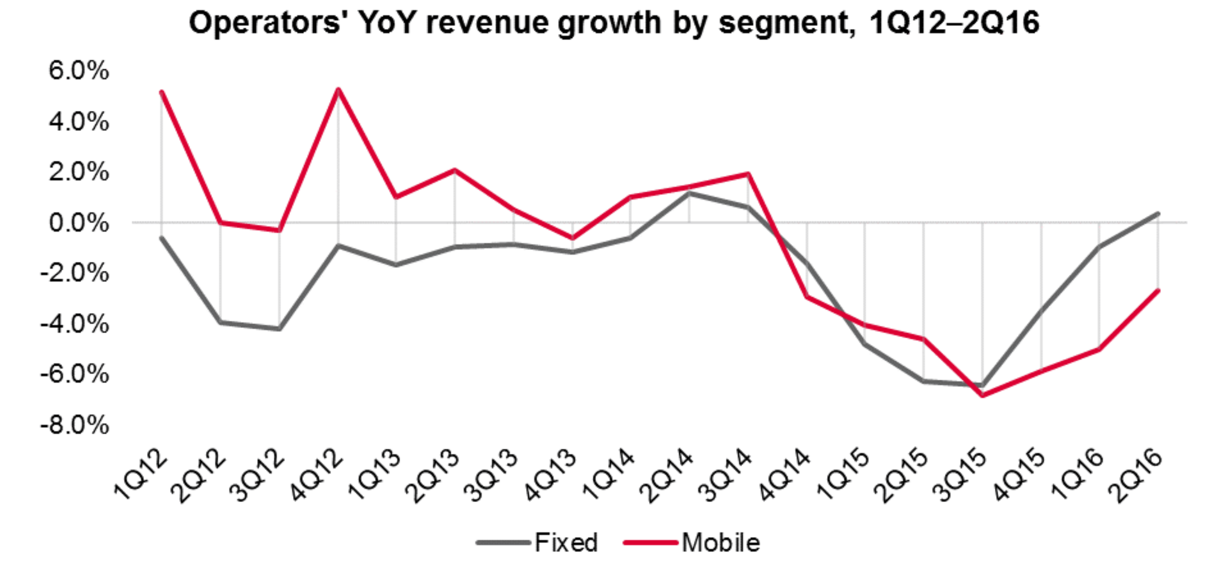
- To take advantage of new opportunities from real-time analytics and fast data, operators should consider implementing unified database architectures that enable delivery of data collection, analysis, and decisions as quickly as possible.

Current market conditions in telecoms

Operators' declining revenue performance drives the need to grow revenues

Telecom operator revenues have been falling since 3Q14. Both fixed and mobile revenues are declining, with mobile revenues dropping faster than fixed revenues. Ovum predicts this decline will continue until at least 2021.

Figure 1: YoY quarterly performance shows how operators are struggling to grow revenues



Source: Ovum

While competition from peers has always existed, behavioral changes among subscribers, especially around the use of mobile devices to consume OTT services, such as Facebook, WhatsApp, Skype, and Viber, are further driving this revenue decline. Furthermore, regulatory stipulations are becoming stricter. Telecom operators are trying to fight off OTT services by providing value-added services (VAS) or by bundling OTT services with their services. Another approach is the broadening of service portfolios into media and content delivery through mergers and acquisitions (M&A), as seen in Verizon's acquisition of AOL and possibly Yahoo.

Impact of OTT services on operators

The introduction of OTT services into the consumer market has been the bane of telecom operators. OTT services are not only cannibalizing traditionally lucrative voice and SMS services, but are also increasing the volume of traffic within the network. Additionally, subscribers consuming these services are placing a high demand on operators for improved service delivery and customer experience.

For example, Ovum's *OTT Messaging Forecast: 2016–20* predicts that the strong growth in OTT messaging app (e.g. chat apps) users and traffic will continue over the next five years, as the network effect of these apps takes hold across the globe. Total mobile OTT messaging traffic (i.e. text, photo, and video) will more than double from 30.99 trillion messages in 2015 to 68.81 trillion messages in 2020. This growth will cause severe congestion on operator networks. To avoid impacting customer experiences, traffic must be processed quickly.

OTT players also have raised the bar regarding subscriber expectations for service satisfaction. They are delivering the personalized experiences and quick responses that customers expect. As a group, OTT service players have effectively leveraged their subscriber data to gain a better understanding of subscriber desires, using these insights to deliver their services.

OTT competition is forcing operators worldwide to reconsider their position in the changing ecosystem and put issues such as quality of service (QoS), pricing, network capacity, and collaboration at the forefront of their strategies.

IoT/M2M becoming dominant play for operators

While IoT and M2M present significant opportunities for operators to reclaim lost revenues, they also present multiple challenges. They enable operators to participate in the digital transformation occurring across multiple industries such as automotive, healthcare, transport, government, and smart cities. Operators are serving not only as connectivity providers, but also as management platform and VAS providers. Telecom operators active in IoT include AT&T, Vodafone, and Telefonica.

However, the traffic generated by IoT/M2M communications also contributes to network congestion. With more than 34 billion connected devices expected by 2020, bandwidth demand will be huge. Furthermore, to retain this new business, network operators must be able to guarantee instantaneous communications with very low latencies. For example, constant and uninterrupted communications are essential to support connected car service offerings. In addition, IoT services such as smart home and smart metering present opportunities for unauthorized access to device information and personal data. Consumers will express concerns as security breaches transcend varying boundaries: the device, the application, or the network. These issues are among many which telecoms operators must carefully consider as the IoT/M2M industry evolves.

Adopting SDN/NFV to achieve operational agility

Operators must transform how they run their networks to deliver and launch compelling digital service packages and assure a positive customer experience. Digital services have varying requirements for bandwidth and latency. For operators to deliver these requirements effectively, existing network infrastructure and services need to become agile and flexible so that they can adapt as quickly as possible to support these diverse requirements. SDN and NFV are two technologies that operators are adopting to enable these changes. SDN focuses on more-efficient programming of routers and switches to manage traffic. In turn, NFV is about the virtualization of network components; it enables components traditionally powered by proprietary hardware platforms to be implemented in software on commodity IT platforms. Operators such as AT&T, China Unicom, China Telecom, NTT, Softbank, Telefonica, Telstra, and Verizon have adopted these technologies.

The reality is that these innovations will not be implemented overnight, but will be steadily phased in. Virtual and physical network infrastructure will coexist side by side. Careful orchestration will be critical for operators to benefit from SDN and NFV while continuing to meet customer QoE expectations as these technologies are implemented over time.

Advanced real-time decision-making to drive real-time actions

Evolution of 3GPP specifications requires adoption of real-time capabilities

In the early days of 2G, telecoms networks used connection-oriented technology (where a communication session is established before any useful data is transferred). Protocols and network changes were applied centrally, with settings remaining fairly static. Network carriers typically collected call detail records (CDRs) at intervals ranging from five minutes to an hour, with as much information as possible, to analyze these services.

However, the move to 4G/LTE has transformed operators' networks to packet networks based on Ethernet and IP. Instead of determining predefined paths for the traffic, IP networks define and update paths depending on the network conditions. While IP networks can react faster to changing network conditions, there is less certainty in predicting paths for traffic flow. IP networks are also "bursty" by nature, meaning bandwidth consumption can vary rapidly within short periods, and they support much higher speeds. That means that the definition of "real-time" network monitoring or response grows far more stringent when the backbone can transit individual packets or Ethernet frames within nanoseconds.

The onset of 5G networks will further compound the challenges for network operators. They will be expected to deliver traffic at even lower latencies and higher throughput compared to 4G, and serve larger populations of users. Analysis of xDR (call, packet, or event data records) must occur in real time – measured in microseconds or seconds, not minutes – so that operators can become more agile and responsive in managing networks and meeting QoS promises.

Fast data as the enabler of real-time analytics and decisions

The definition of real time for the operator does vary from seconds in applications such as the Policy Control Enforcement Function (PCEF) and Policy Charging and Rules Function (PCRF), to microseconds in mobility management applications. Keeping charging functions accurate and reliable is also vital. Consequently, modern telecom applications need database systems to ingest, analyze, store, and make decisions on live data streams while at the same ensuring the validity of these decisions. Databases with multilayered architectures will not deliver the performance necessary to support these processes. Fast data approaches that deliver actionable insights in real time will be essential for operators as they seek to optimize their networks and minimize churn.

Essential requirements of a real-time analytics and decision database engine

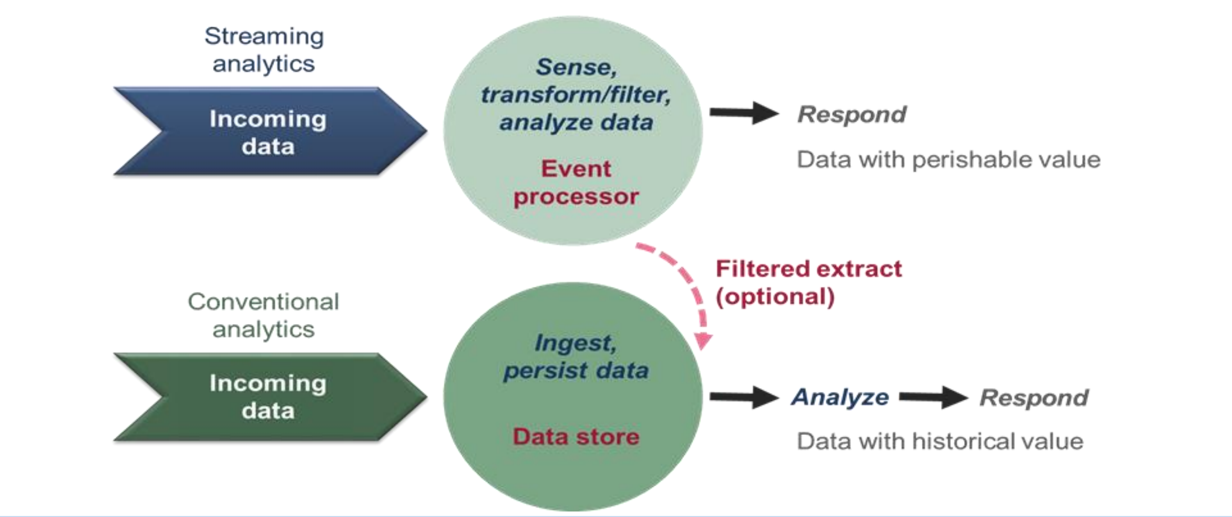
Databases supporting fast data capability must satisfy several key functional and nonfunctional requirements. Functional requirements include streaming analytics and event processing, while nonfunctional requirements involve in-memory database technologies built on scale-out infrastructures to ensure performance, scale, and reliability.

Functional requirements

Streaming analytics

Streaming analytics typically analyze data as it is being ingested, performing operations such as filtering, correlation, aggregation, and pattern detection. In some cases, filtered (or "stripped") extracts might subsequently be persisted to a back-end database. Unlike a conventional analytics procedure where ingested data is stored before being analyzed, with streaming analytics, the analysis is done as the data is ingested, enabling instant responses to be generated.

Figure 2: Streaming analytics versus conventional analytics



Source: Ovum

Streaming analytics can involve running *continuous* queries, such as who has been connected to the network over the last 10 minutes and what he or she is doing on the network, with continual updates of results. Streaming analytics are critical when executing closed-loop automated processes that trigger actions based on the satisfaction of specific rules. That action can be reactive, such as offering mobile operator subscribers instant promotions when they have suffered from bad service, or prescriptive, prompting operators to schedule preventive maintenance earlier than planned.

Event processing

Event processing, the core of all fast data solutions, enables the processing of real-time data, resulting in the determination of actions that need to be taken in response to the event. Event processing can be delivered as either complex event processing (CEP) or event stream processing (ESP). CEP combines data from multiple sources to identify events or patterns that suggest more complicated circumstances which need to be responded to, while ESP enables the processing of multiple event streams to identify and react to important or critical events in such streams.

Machine learning

Machine-learning capabilities are increasingly being used to make sense of and respond more adaptively to situations that could not be envisioned with static rules. Machine learning enables analytics software to build knowledge based on experience using a toolbox of algorithms that are trained to adapt to a particular function or model. For instance, specialized machine-learning algorithms can be used for predictively segmenting customers, mitigating fraud, and optimizing network performance.

Nonfunctional requirements

Table 1 provides an overview of nonfunctional requirements that must be satisfied by database systems that deliver fast data capabilities.

Table 1: Nonfunctional requirements for fast data solutions

Nonfunctional requirements	Overview of requirements
In-memory database	Tables and key information are held in main memory and not on disk, reducing the read/write time to/from a disk. It also eliminates the "seek time" spent in querying data on a disk.
Ingest and interact with live stream data directly from the source	Real-time analytics and decision engines need to connect directly to data sources such as network nodes, BSS, OSS, and home subscriber systems (HSS); process the data instantly; embed a certain level of intelligent analysis; and trigger actions right away. As a result, connectors are required to easily feed data streams into the database engine.
Scalability	Distributed operations using a scale-out architecture are becoming very important to ensure continued processing as the volume of the data streams or the complexity of the operations increases. In-memory databases also must scale to support extreme event rates at operators' busy times.
Security	A stream-processing system must ensure high availability to preserve the integrity of mission-critical information and avoid disruptions in real-time processing. For example, if a failure occurs (in the network or hardware), the real-time analytics function needs to failover to backup hardware to allow operations to continue.
Reliability	There are different degrees of reliability for real-time analytics, with error-tolerance levels depending on the use case. In essence, the choice of reliability for the use case depends on the goal: if it is getting the big picture, where some errors (e.g., dropped or duplicate data events) are permissible, or if an exact picture is required. Levels of reliability include: • At most once: The least demanding form of reliability requires processing of the data set be done no more than once. Use cases include subscriber engagement and location-based marketing. • At least once: This requires that the system registers every event; in effect, incoming data must be cached until it is processed. An example is security use cases where the goal is to detect and identify an intrusion event. • Exactly once: The most stringent form of reliability requires that every single event be processed. Examples include low-latency applications such as network optimization.
Support a virtualized environment	With operator networks evolving to support virtualization, a real-time analytics function needs to take into consideration data operations within a hybrid environment.

Source: Ovum

Applications of real-time analysis and decision-making for telecom operators

Existing use cases

Business support systems

Business operations such as marketing, sales, and support must align to subscriber needs if operators are to deliver on their customer experience objectives. Telecom operators need to monitor subscribers' lifecycles to identify the best time to deliver offers and any relevant information.

Campaign management

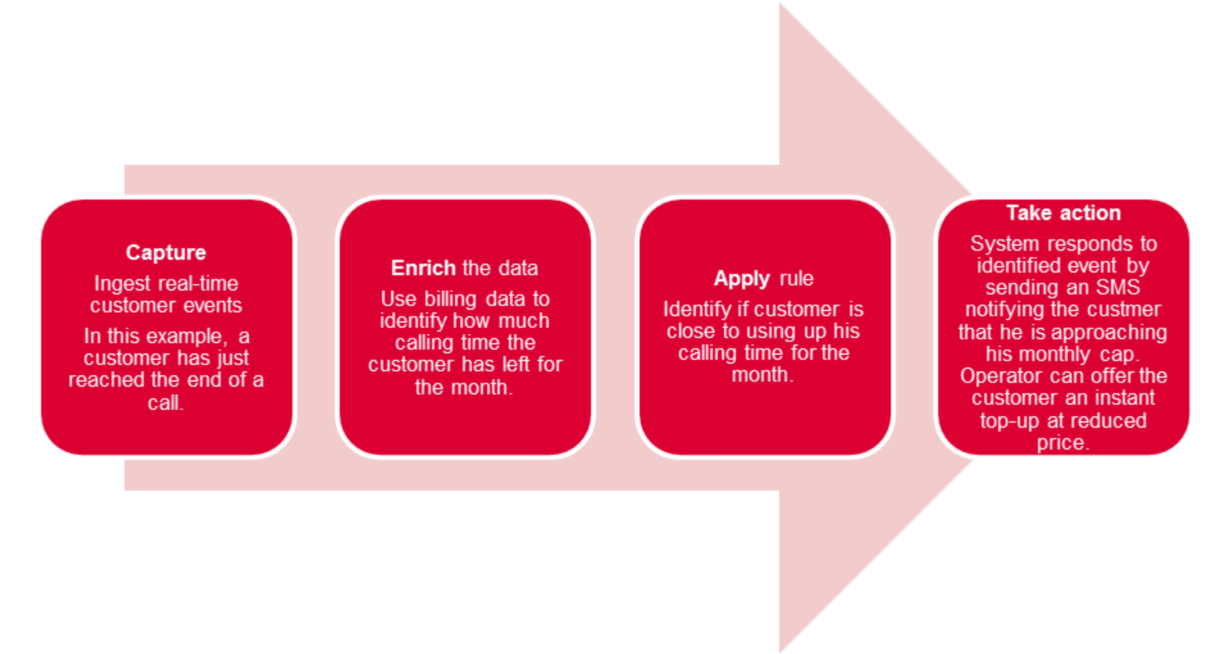
Timing and relevance determine the success of any campaign. Operators have had disappointing track records because the insights were typically drawn from batch processing that produced results that were often out of date. However, by using real-time data such as browsing data, enriched with subscriber profile information, relevant and timely campaigns can go live based on the *current* perceived needs of the subscriber.

Real-time results are especially relevant for preventing churn. By using real-time data and predictive analytics, emerging problems that place subscribers "at risk" (e.g., a sudden spike in dropped calls) can be nipped in the bud. Actions to be taken by the operator could include discount offers, loyalty points, or quick fixes sent remotely to the subscriber's device if required.

Subscriber engagement

Telecom operators are engaging their subscribers more proactively to establish positive experiences. The challenge is, however, in using the right channel(s) to deliver the right message/experience to the user at the right time. Given the degree to which digital mobile experiences have become ingrained in subscribers' lives, real-time insights will be critical for operators to stay engaged. Figure 3 illustrates the process flow and shows the role that real-time analytics plays.

Figure 3: Adopting real-time analytics for effective subscriber engagement



Source: Ovum

In the example shown above, the subscriber's mobile voice and data usage is tracked in real time using a real-time subscriber engagement tool. A rule is defined to set a trigger to identify the exact time at which the subscriber will approach his data cap or use up his calling time. Once these conditions are met, the operator can instantly send out an SMS indicating the subscriber is about to exhaust his or her calling time, and then provide a pay-as-you-go offer to help the subscriber extend service at a reduced price. By doing this, the operator is able to ensure that the subscriber feels cared for.

First-time call resolution

Most service issues from subscribers are rarely resolved on the first call. Delays trigger a death spiral: the longer it takes to resolve the issue, the costlier it is for the operator, and the more dissatisfied the customer becomes. When responding to an incoming service call, the customer care agent should know before answering the call who the subscriber is, why he or she is calling, and be able to identify the quickest way to resolve the problem. This is achievable through constant monitoring of factors influencing the subscriber's service such as:

- current location and the current state of the network serving that location;
- state of the subscriber's device and applications running on the device;
- subscriber's billing history; and
- current state of his plan.

Using event processing or streaming analytics, alerts to the contact center can be generated when problematic scenarios occur, providing critical advanced notice.

Mediation

Real-time mediation platforms are essential for handling the digital service issues of customers. Billing and charging information from CDRs must be immediately accessible, up-to-date, and accurate to reflect current account, service status, and amounts to enable subscribers to understand when they are close to reaching their current caps. Such real-time insight not only helps avoid bill shock for customers, but also enables operators to offer new services. For example, an operator may consider providing hybrid plans allowing post-paid subscribers traveling to a different location to purchase roaming data on demand with the lower price points of a prepaid plan. To fulfill these complex functions, mediation platforms must move from the store-and-forward mechanism (batch processing) to real-time processing to determine when these changes need to occur.

Policy Charging and Rules Function (PCRF)

PCRF is dedicated software that determines the policy rules required to control access, resource, and necessary QoS levels for every user data session initiated on the network, thereby enabling service monetization and assurance. For example, service providers can use PCRF to charge subscribers based on their usage of high-bandwidth applications, charge extra for QoS guarantees, and limit app usage while a user is roaming, or lower the bandwidth of wireless subscribers using heavy-bandwidth apps during peak usage times. The PCRF must operate in real time, aggregating information to and from the network, OSS, and other sources (such as portals). It must support creation of rules for closed-loop policy decisions for each subscriber active on the network.

Fraud detection and prevention

Fraud has become a pervasive and constantly evolving problem for operators. As global connectivity grows, fraudsters have a larger array of channels to attack and more opportunities to commit fraud – making prevention of subscriber, usage, and payment fraud more complex. Furthermore, the potential for revenue leakage, particularly through fraudulent activities, will continue to increase as service providers add more noncore products and services to their portfolios. Services that are acquired from third parties are of particular concern because there is the possibility of leakage at the cost side of the equation as well as revenues.

Instead of identifying fraud at the network enterprise data warehouse level, where latencies would be introduced, fraud detection must be done in the moment, using real-time analytic and decision infrastructure. The fraud detection system would instantly receive xDRs for every billable event from the network and then enrich the data with subscriber information to identify if he or she is associated with previous fraud activities. A fraud alert can then be sent to the fraud prevention system. This process must be performed on all event records, which for a typical network ranks in multiples of 100 million records a day.

Operations support systems

Network planning and optimization systems

Given declining revenue performance, operators must take extra care when allocating investment. As applications generate network traffic, optimization tools must allow operators to analyze the traffic to understand its impact on the network elements and what decisions must be made to avert congestion or degradation of operational performance. Network-planning tools must provide a real-time view of an operator's distributed network to proactively identify and allocate resources to network locations where they are required.

Mitigating network security threats

Operators' networks must be protected from malicious attacks such as distributed denial-of-service (DDoS) and malware attacks. While the core network may be considered extremely inaccessible given the difficulty of and expertise required for break-ins, it is fast becoming a very desirable point for cyberattacks. Recent incidents with Dyn (a DNS infrastructure company in the US), DT, TalkTalk, and KCOM indicate how vulnerable operators' networks are to cyberattacks. Although some of these attacks are devastating, it may take a while before users and internal service-monitoring systems notice their effects. Clearly, an approach that includes a real-time decision engine is necessary. First, the decision engine enables detection of suspicious traffic upon entry into the network and provides the ability to set off real-time triggers to alert network engineers. Second, it can enable the redirection of suspicious traffic to a mitigating device for remediation actions with subsecond mitigation responses, addressing the problem before the damage is done.

5G/Industrial IoT

While development of 5G network standards is still under way (with commercial deployments expected to start by 2020), the applications expected to run on these networks are fast evolving. IoT, for example, is already a part of our lives, including the workplace, places of leisure, and our homes.

5G networks must meet several targets, including delivering latencies of about 1 millisecond for mission-critical applications, and provide support for a diverse array of session requirements. For instance, HD videoconferencing in a telemedicine application will have a very different session profile than a wireless M2M device communicating its hourly status. To manage a 5G network, operators will need to slice their network resources to support different types of services, each with unique requirements around quality of service, security, latency, etc. Real-time applications (powered by a real-time decisioning engine with the capability to scale to support billions of messages in about a second or less) that are service aware and capable of responding to the varying requirements of services thus become essential. These real-time applications will drive decisions regarding which network resources are best positioned to handle subscriber services as they are initiated.

NFV orchestration

NFV promises to improve the speed and flexibility of processes such as service creation and activation; orchestration will be essential. An NFV orchestrator coordinates the resources and networks needed to set up NFV-based services and applications. It allows operators to quickly deploy services, or virtual network functions (VNFs), using software rather than specialized hardware networks. In addition, an NFV orchestrator would be required to automate the dynamic delivery of virtualized network services. That will entail capabilities such as rapid configuration, provisioning and chaining, intelligent service placement, dynamic and elastic scaling of services, and full management of VNFs.

A real-time analytics and decision engine will be required to support this real-time orchestration of network services. Metrics such as average utilization of resources (e.g. compute, storage, network, and energy of VNFs, VMs, hardware, and network elements) must be calculated on the fly – providing real-time visibility into the operating point of all the VNFs in the NFV node. The NFV orchestrator can then use the average utilization information from the real-time analytics engine to determine the appropriate time to scale up/down the running of software instances. These real-time capabilities would ensure that operators meet their business objectives of achieving agility, reducing costs, and providing faster service delivery.

Benefits of real-time analytics and decisions

The ability to collect data, analyze it, and make decisions in real time provides several advantages to operators and their vendor partners. These include:

- enabling the development of context-aware telecoms applications that generate immediate value to operators;
- delivering proactive response to subscriber- or network-related operational issues;
- promptly identifying events that lead to fraud to avoid impact on operators' bottom lines;
- enhancing customer loyalty as real-time and personalized engagement with subscribers drives improved customer experience; and
- efficiently utilizing network infrastructure. Operators also can achieve a more predictable and optimized return on investment (ROI) by deploying additional network services without unnecessary equipment costs.

Recommendation

Adopt a unified database architecture that supports real-time analytics

The realization of true real-time capabilities will be dependent on the database architecture being adopted. The speed at which the collection, storage, analysis, and decision-making processes occur will be dependent on the structure of the database.

Conventional database architectures have different components that handle the ingestion, storage, analysis, and decisioning processes. Once these nonintegrated components perform these processes, higher latencies are introduced, which then makes the delivery of real-time functions difficult. However, a database architecture that integrates all of the mandatory functions and components into a single component makes lower latencies possible and so is able to deliver true real-time functionality.

Appendix

Author

Adaora Okeleke, Senior Analyst, Telecoms Operations and IT

adaora.okeleke@ovum.com

Ovum Consulting

We hope that this analysis will help you make informed and imaginative business decisions. If you have further requirements, Ovum's consulting team may be able to help you. For more information about Ovum's consulting capabilities, please contact us directly at consulting@ovum.com.

Copyright notice and disclaimer

CONTACT US

www.ovum.com

analystsupport@ovum.com

INTERNATIONAL OFFICES

Beijing

Dubai

Hong Kong

Hyderabad

Johannesburg

London

Melbourne

New York

San Francisco

Sao Paulo

Tokyo