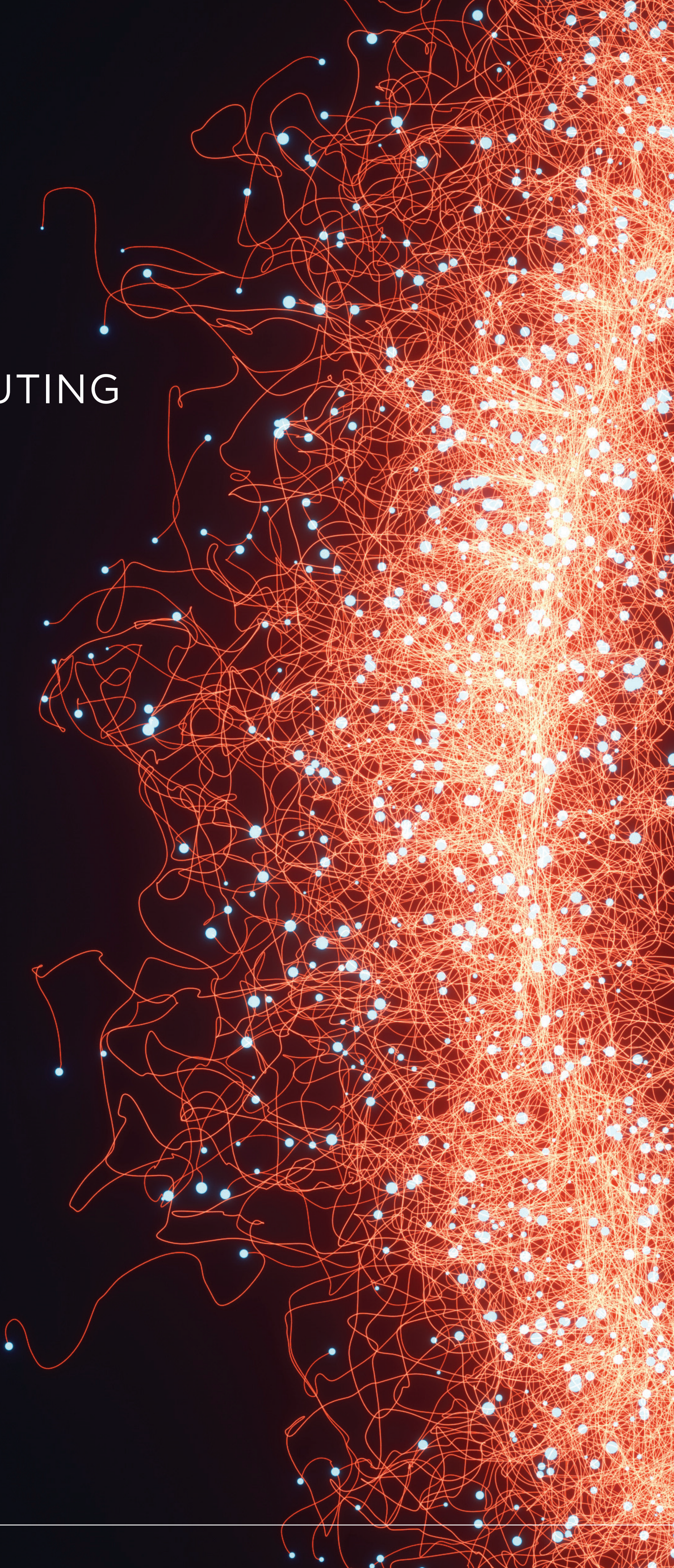
IoT EDGE COMPUTING REQUIREMENTS

JULY 2019



Introduction

The edge offers a multitude of new service offerings and improved outcomes for IoT deployments across all vertical markets. Industries have been busy digitising their services for a number of years, but edge computing designed specifically for IoT allows them to take this digitisation to the next level. New smart services that rely on high volumes of data to be collected, and automated local analysis and decision making, can be designed and deployed in a wide range of environments. Operators need to understand their customers constraints and how edge computing can take their IoT deployments to the next level, allowing their customers to offer not just improved levels of service and accuracy in the field, but wholly new ways of undertaking currently labour intensive tasks.

Operators need to discover how to simplify edge deployments for their customers, which today are largely bespoke implementations. New solutions and platforms that allow the operator to maintain their strengths around connectivity, security and device and application management will be needed to ensure success.

However, it's not just operators and their IoT customers that can benefit from new edge infrastructures. Other third parties also need to form part of the ecosystem to allow complete service delivery. Cloud providers will become more reliant on the edge to deliver low latency services, developers will be able utilise a new range of tools to create compelling new IoT applications for industrial users and, for end-users themselves, who will be able to seamlessly connect into a new class of services in their homes, cities and workplaces.

VERTICAL DIGITISATION OPPORTUNITIES AT THE EDGE

IoT edge computing deployments are not something which will be standard across all verticals in the way that network connectivity is. Each end-user or customer type may have different needs and opportunities, and operator tools should reflect the fact that these requirements will be diverse. Operators have a strong case for helping customers through the changes that edge computing brings to the IoT, and they need to offer relevant tools alongside edge infrastructure. The tables below show some of the different digitisation opportunities in different verticals, and the ways in which edge can help achieve better outcomes with the technologies available.

[1] Meaning action completed with minimal latency

A.) Smart Factory

Factories generate huge volumes of data from manufacturing machinery and warehouse operations, much of which is trapped inside unconnected systems. Factories are able to significantly benefit from the hosting of edge infrastructure either on their site or nearby, as this large volume of data can be accessed and processed locally in real-time[1] for localised insights and control. This opens up new opportunities to both better understand factory operations and predict where issues may occur as well as working towards improving product quality and minimising stock holdings to create more efficient operations.

	EDGE ENABLING OPPORTUNITIES	IMPROVED OUTCOMES
Manufacturing	Better planning using real-time data to minimise production time and maximise efficiencies Semi-autonomous robotics to undertake routine processes with precision Digital Twin modelling to drive better outcomes in manufacturing processes	Manufacturing efficiency optimised and cost to produce minimised. Improved quality and production times Optimise industrial processes and use of machinery in near real-time
Warehouse	Semi-autonomous robotics to undertake supply and fulfilment operations Augmented Reality to assist with fulfilment operations	Higher availability of automated operations Better information availability at point of distribution to ensure more accurate flows
Inventory Tracking	Real-time location tracking and asset management of components, completed goods and machinery	Analytics to optimise inventory, signal for replenishment at stations and improve security
Quality Control	Image analytics to enhance output from existing sensors Whole production line equipment monitoring to predict and address quality issues	Higher quality output with lower overall downtime due to previously unforeseen operational issues Improved predictive tools and digital twin pairing enable outcomes to be better understood before deployments
Equipment maintenance	AR fed from local sensor data to assist personnel in maintenance tasks Sensors to monitor maintenance needs of machinery including retrofit to legacy equipment	Knowledgebase available to engineers at all times whilst working on machinery Identifying new failure modes that are not reported by legacy equipment Predicting failures ahead of time Lower downtime due to better planning of maintenance windows
Health and Safety	Lone worker monitoring to ensure worker safety Restricted access to hazardous materials and equipment	Better enforcement of and compliance with health and safety policies

B.) Smart Grid

Electricity grids are undergoing a profound period of change as renewables and energy storage are added in more and more quantities to the energy mix. The vast majority of these new additions take place at the edge of the energy grid – in homes or in cities as opposed to traditional centralised power stations, leading to the introduction of new concepts such as virtual power plants for energy generation and consumption. As the grid control moves to the edge for some use cases, so network and data management also needs to move to the edge to offer the localised and low latency control that energy companies demand to maintain their quality of service for their customers.

	EDGE ENABLING OPPORTUNITIES	IMPROVED OUTCOMES
Energy Generation	Monitoring of real-time renewables generation and load balancing	Improved generation mix with better information leading to reduced fossil fuel inputs
	Creation of virtual power plants through enabling microgrids and energy storage	AI at the edge can give improved load control and reliability for microgrids
Energy Transmission	Use of Drones to monitor high-voltage transmission lines with real-time analytics	Lower cost for identifying transmission line or pylon faults
	Real-time monitoring of transformer stations to provide analytics on energy load and voltage outputs	More efficient transfer from transmission to distribution network with resultant lower energy loss
Energy Distribution	Powering analytics to enable effective matching of energy consumption and provision for load control	Reduce energy loses and optimise links between generation and consumption
	Connect to third party devices, such as electric vehicles, to enable smarter energy consumption and storage	Better predictive demand for energy consumption and energy storage integration
Inventory Tracking	Real-time location of equipment, assets and field personnel	Improved worker safety and asset security
Customer Engagement	Smart Meters used to provide detailed analytics on energy consumption at every metering point	Improved meter analytics from more regular reads and links with other consumer services such as smart home and EVs
		Reduce fraud
		Improve grid efficiency
Equipment Maintenance	AR driven by IoT sensor data to assist personnel in maintenance tasks	Knowledgebase available to engineers at all times whilst working on machinery
	Sensors on equipment to monitor maintenance needs including adding monitoring to legacy equipment	Lower risk of blackouts due to better predictive maintenance

	EDGE ENABLING OPPORTUNITIES	IMPROVED OUTCOMES
Environmental	Advanced analytics for creating efficiencies in energy consumption to reduce carbon emissions	Reduce reliance on fossil fuels by better load balancing through use of AI.
Health and Safety	IoT sensor monitoring of high voltage hazardous equipment and areas Lone worker protection for engineers in the field	Better enforcement of health and safety policies. Reduced total hours for workers in dangerous environments

C.) Smart Cities

Cities are distributed over wide areas. As they deploy smart city services, so the use of centralised resources will become more unwieldy as the volume of data collected and the number of citizens introduced to those services increases. For the city it makes sense to ultimately decentralise many services and maintain local influence. Automation of many simpler tasks across a smart city such as monitoring streetlights or availability of parking spaces can be hosted within the city boundaries. As such operators should work with cities to understand appropriate points to place edge servers to collect, process and control data. Additionally, cities tend to collect a lot of data which they may want to retain control and ownership over, with data stored locally and not uploaded to the cloud. Local data centre and edge servers allow them to achieve this aim.

	EDGE ENABLING OPPORTUNITIES	IMPROVED OUTCOMES
Transportation	Control and analytics for public transport, driven by IoT services including V2X and on-board sensors Dynamic matching of transport demand with route scheduling	Improved responsiveness to performance issues; direct on-board status feed to driver Matching demand with route optimisations
Resource Management	Active management of traffic, pedestrians and parking spaces Processing of CCTV images to identify resource utilisation Improved water and waste water distribution, including water quality monitoring and leak detection	Better resource usage leading to lower future urban planning costs – through either reduction of traffic or active notifications of parking availability Less wastage of resources and better matching of demand with supply
Citizen Safety	Crowd management for crowd safety CCTV image recognition of threats e.g. fire, overcrowding	Prevention of dangerous overcrowding and analytics to allow better planning for pedestrian routes Immediate response to threats can be achieved

	EDGE ENABLING OPPORTUNITIES	IMPROVED OUTCOMES
Citizen Services	Enhanced citizen engagement with city services through linking IoT powered services to citizen dashboards	Improved identification capabilities to allow access to applicable services
	Enhancement of social services through IoT integration	Modelling city services and subscription services to citizens at point of use
Customer Engagement	Improved local statistics and analysis on city operations	Better information sharing with citizens on city performance criteria
Equipment Maintenance	AR driven IoT sensor data to assist personnel in maintenance tasks	Knowledgebase available to engineers at all times whilst working on public facilities
	Sensors on assets such as roads and city fleets to monitor maintenance needs	Lower downtime due to better planning of maintenance windows
Environmental	Improved air quality through better localised monitoring and control of pollutants	Better citizen health and cleaner local environment
	Improved waste collection and recycling services through use of IoT sensors and CCTV analytics	Improved recycling rates, reduction in fly-tipping and ability to share performance data with public
Health and Safety	Improved safety in public areas through sensors and real-time analytics	Identification of safety issues and potential city impacts passed to personnel for real-time management

IoT Edge Requirements

IoT uses edge in unique ways. IoT devices are not typically smart in their own right, and so the edge must perform a more in-depth supervision role than for many other non-IoT applications. This means that edge resources may need to be dedicated to IoT application in certain use cases, providing unique resources that are not required in non-IoT deployments. Additionally, many IoT use cases are mission critical and so edge deployments need to be designed to support these essential applications.

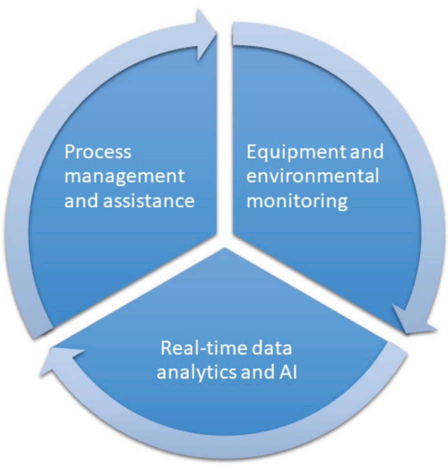
Generic design considerations that should be considered when building an edge deployment to support IoT services that need low latency, local data processing and local data storage include:

- Use of operating systems that support real-time applications and management techniques such as support for containerisation
- The ability to run image processing through the use of GPU acceleration.
- Compute and storage resource optimisation
- Communications frameworks to allow edge systems to co-operate on control and data processing
- High system availability
- Clear security frameworks and access controls

Beyond the technical, there are a range of IoT requirements which maybe best achieved by the use of an edge deployment. Offering lower latency and localised decision making naturally benefits some IoT requirements, which can be broadly grouped into three broad categories:

- Process management and assistance
- Equipment and environmental monitoring
- Real-time data analytics and AI

These three categories are interlinked, with an edge application potentially covering all three areas:



In a manufacturing environment, for example, an application may traverse these different areas by tracking components through a production process to the final product in order to understand if the production line is at full capacity or could handle more whilst maintaining product quality. In a smart city an application may be able to predict air quality based on transport usage patterns or divert traffic away from congested areas.

Within each of these categories there are specific requirements that operators need to address to ensure that developers and customers are able to benefit from these complex applications at the edge.

A.) Process Management and Assistance

Process management at the edge is a new way of managing both traditional business process management whilst introducing new possibilities creating by dedicated edge process management. In order to take full advantage of this capability at the edge, new and existing processes need to be designed to be adaptable and automated in order to intelligently use available edge resources and maximise operational efficiencies, whilst taking advantage of unique edge attributes such as low network latency.

There are a few major requirements that need to be considered when deploying process management at the edge.

Whilst maintaining data security, transparency of process needs to be established so that activities hosted at the edge are visible to members of the ecosystem, to enable them to best design and control related services.

- Dynamic process operation at the edge is required to ensure that operations remain agile in nature. Processes will need to be adapted and improved as the environment and data sources change around them.
- Modelling of processes at the edge will also require a different approach as to benefit from latency at the edge, autonomy and resilience needs to be baked in from the outset.

What this means in reality is that a great deal of choreography is required between different processes and services to ensure that the service entity as a whole is not undermined. In turn, this means that testing and QA become important attributes in edge delivery.

When other ecosystem players are factored in, then process design becomes a critical strength. Data sources and storage and device ontology will need to be clear to all players to ensure that wider systems and applications can be seamlessly integrated down to a granular level.

B.) Equipment and Environmental Monitoring

The key to success of any edge based service is the quality and timeliness of data available to it. Edge computing is able to radically transform the quality and quantity of data that can be collected from local devices, even from older machinery with proprietary protocols. Today's new generation of machines are more likely to be connected to the internet than not for data collection and control. The data that has been available for processing has typically been siloed and wider insights have been harder to achieve. Additionally, data has been limited in scope and quantity as the cost of collection has been high on many networks, combined with the fact that only limited bandwidth has traditionally been available for machine to machine type communications.

Edge computing turns this paradigm on its head by collecting, storing and processing data closer to its source. There are several key requirements for equipment and environment monitoring that for the first time can be realised by edge computing.

- **Low latency.** Sensor data can be collected in near real-time by an edge server. The closer to the edge the server is located, the lower the latency. For services such as drone control or image recognition, edge servers can be located in very close proximity to the device, meeting customer's low latency requirements whilst offering the same level of control as centralised services.
- **Network independence.** IoT services do not care for the delivery mechanism of data. Customers just need the data to be made available by the most effective means possible. This in many cases will be mobile networks, but in some scenarios such as smart buildings or homes, wifi or local mesh networking may be the most effective mechanism of collecting data to ensure latency and other collection requirements can be met.
- **Security.** Data security and data privacy are requirements that must be met with the same rigour at the edge as they are at the core. However, data security at the edge has different challenges to the core, not least that data is spread across many more locations at the same time, and customers need control of security at each of these points. Additionally, physical security requirements are more prevalent at the edge, as servers may be located outside of highly secure operator data centres. Both customers and cloud providers will need security assurances backed up by tools such as constant monitoring of edge nodes, reliable access logs and appropriate authentication and others. The integration of equipment running proprietary or legacy protocols with other data sets from modern equipment. A secure edge can offer a point of protocol translation, allowing data from multiple sources to be combined and analysed at the edge.
- **Flexibility in future enhancements.** Additional sensors can be added and managed at the edge as requirements change. Sensors such as accelerometers and cameras can be added to equipment, with seamless integration and control at the edge.

C.) Real-time Data Analytics and AI

The edge can be used to introduce new concepts into IoT deployments, which rely on low latency and localised processing power not available at the core. Real-time analytics and AI both rely on access to large volumes of data and large amounts of processing power to deliver rich, real-time insights and decisions, which is only possible at the edge.

Developers, cloud providers and end customers all rely on the availability of these resources to ensure that service automation and application performance values can be met.

Requirements that for analytics and data at the edge can be quite specific as low latency requirements mean that decisions must be made rapidly. This in turns means that pertinent data needs to be readily accessible.

- Clearly defined inputs and outputs. AI at the edge will only work effectively if the customer is able to define precisely what outputs they need and what inputs are available to make those decisions.

- Environmental customisation. In order to facilitate decision making in the required timescales, optimisation of local operating environments at the edge need to be made, and flexibility in architecture is required. Edge servers may need to be located in close proximity to devices and sensors, or hardware acceleration used to identify data of interest. Some data, particularly from cameras, needs to have high bandwidth on demand to ensure image processing can take place with the required low latency.
- Customers require not just proven AI techniques, frameworks and mechanisms, but algorithms that are improved and refined over time to get more accurate outputs. This means that algorithms and other tools must be deployed within a framework in order to allow intelligent collaboration across edge devices and nodes where required so that they can be 'trained'. This means that parameters can be shared across an ecosystem and applications enhanced to enable learning activities.
- Where analytics and AI are used these are both data and processing 'hungry'. Edge devices supporting these will need substantial RAM, storage devices, and in many cases acceleration devices for machine learning such as GPUs.

User Requirements

There are three groups of users which these requirements need to be adapted for. Although each will have slightly different requirements, they will still fall into broadly similar groupings which give the operator opportunities to develop new propositions as the edge for each group.

A.) CLOUD PROVIDERS

Cloud providers will find edge/cloud orchestration for IoT challenging, as they will need to engage in a more distributed ecosystem that supports distributed data and distributed processing for the large volumes of IoT data generated. The co-ordination required to extend cloud processing across different sites and ensuring that data at the edge can accessed effectively will be key to ensuring that the cloud operator is able to offer a seamless experience to their own customers.

DECENTRALISED DATA STORAGE AND ACCESS

Deployment of applications and services at the edge requires a certain amount of decentralisation of data and data processing. This presents challenges to the cloud provider, who must make it easier for their customers to enable full or partial distribution of data and processing to the edge. This will require new frameworks to allow for detailed mapping of data sources and locations, ensuring that centralised queries are able to locate the correct resources at the edge. Access to data held at the edge is also important both in order to catalogue and archive it but also to provide better context for complex queries where place and time are important factors in determining the correct actions.

COORDINATION BETWEEN THE EDGE AND CORE

With real-time decision making occurring at the edge, it is likely that only archive data need to be stored in the cloud, and so any queries that require data from a contextual time close to an event will require access to data at the edge. This mix of archive and current data to present the correct context for cloud queries will require a level of co-ordination between the cloud and edge that may not be present today.

CONSISTENCY AND FLEXIBILITY

Large centralised systems are not good at the type of dynamic changes needed at the edge to support AI, automation and real-time decision making. Therefore, cloud providers need to work closely with edge infrastructure providers and operators to obtain consistent service levels in order to provide a dependable interface between edge and cloud. The challenge is to ensure that the flexibility required by applications and developers is matched to the consistency required by cloud providers. Application providers and IoT devices need flexibility in their deployments, and edge providers will need to have flexible approaches to how data is accessed, shared and stored to meet the requirements of different use cases.

CLOUD AGENTS

Most cloud providers today make use of edge agents to ensure that the edge resources are allocated appropriately and that the core and edge remain co-ordinated.

In order to make use of any edge infrastructure, cloud providers will need support for these agents to be offered natively at the edge. This means that the edge operating environment needs to be designed with support for these agents in mind.

B.) CUSTOMER

In reality, end-customers may have little say over the shape of the edge unless they have very specific requirements that will dictate dedicated and bespoke edge resources to support their operations. Most applications will utilise shared infrastructure that is available across a range of use cases. There are several core requirements that need to be met regardless of the class of infrastructure deployed at the edge.

DATA QUALITY

Data stored at the edge is subject to the same needs as data stored elsewhere – it must be secure and relevant to the customer applications operating at the edge. Data quality at the edge is a key requirement to ensure that customer operations are able to operate in demanding environments. To maintain data quality at the edge, applications must ensure that data is authenticated, replicated as required and assigned into the correct classes and types of data category.

DATA SECURITY

End-customers require data at the edge to be kept to the same security requirements as when it is stored and used elsewhere. This presents challenges due to the larger vector and scope for attacks at the edge. Security policies need to be adapted to take this into account but otherwise data authentication and user access is just as important at the edge as it is on the device or at the core. Additionally, physical security of edge infrastructure needs to be considered, as it is likely to held in less secure environments than equipment hosted in dedicated data centres.

LOW LATENCY

Taking advantage of low latency applications at the edge is one of the key benefits of edge deployments and customers will require applications to benefit from this. Any edge deployment needs to ensure that these benefits are not lost through poor development practice or inadequate processing resources at the edge. Maintaining data quality and security at the edge whilst enabling low latency is a challenge that operators need to address whilst building their edge architectures.

DATA ACCESS

Ownership of data at the edge is always retained by the customer or data generator . This means that they need a level of control over the data to ensure that the customer is able to access it and handle it as required. This presents challenges when data is distributed by the very nature of the edge. Therefore, edge services need to allow scope for customers to understand how their data is held and how they can allow third parties to access it.

REPORTING AND OTHER TOOLS

Customers also want key edge performance statistics available to them. Reporting on criteria such as system availability, performance trends and resource utilisation will be needed to highlight performance issues and consistency of quality of service A range of edge enablement tools are likely also required to enable the customer to administer for example the network configuration or audit the system for having up to date patches. Customers will also likely have existing tools and reporting platforms, and integration with these will need to be considered.

C.) DEVELOPERS / DEVOPS

In reality, end-customers may have little say over the shape of the edge unless they have very specific requirements that will dictate dedicated and bespoke edge resources to support their operations. Most applications will utilise shared infrastructure that is available across a range of use cases. There are several core requirements that need to be met regardless of the class of infrastructure deployed at the edge.

DEVELOPMENT SUPPORT

Developing edge applications will likely necessitate a fundamental shift in development mindset towards tooling and programming languages designed to span multiple systems, rather than a traditional programming model which is solely concerned with creating logic that runs as a single process on a single machine and distributes work via a shared-memory model.

New programming paradigms like Unison, efforts to extend the Erlang process model to planetary scale like Lasp, and the EU LightKone project are some current examples of attempts to shift the edge compute programmatic model in this direction.

MAPPING

Developing an application for the edge can be unwieldy compared to developing for centralised systems. Developers require access to a 'map' of the edge which allows them to understand key edge parameters such as edge server locations including location of data storage and data processing capabilities, device types and device IDs, network bandwidth availability and message pathways. Holding all of this information will allow the developer to create applications which use appropriate edge resources and create outputs not constrained by the environment that processes them.

API ACCESS

There are a large number of edge APIs available to developers, partly dependent on the reference architecture deployed at the edge and the platforms being used to support edge functionality. Support of these standard APIs is beneficial as speeds up development time and reduces the prospect of errors. However, customer requirements may necessitate the use of bespoke APIs to manage data and devices at the edge, particularly where complex, data intensive tasks are being deployed, such as in industrial IoT or medical use cases.

NETWORK ACCESS – MEC

Network connectivity is essential at the edge, but unlike traditional centralised systems, network access can be from a variety of different bearers, each with different characteristics and bandwidth availability on both uplink and downlink. Developers need access to this information along with device support for each type of connection, as bandwidth and thus message delivery times and formats will change accordingly. Making this information available means that applications will be able to corelate different service classes into a single experience.

Modelling operator IoT propositions to match these requirements

The edge can be defined by its ability to make local decisions quickly. Without such local decision making, the power of the edge is limited to data storage and routing. Decision making power at the edge derives from the availability of processing power, clear use cases, machine learning and digitisation of complex systems. Operators must think about how they can enable their customers to make more effective decisions through the use of edge and edge-based tools. Simply providing connectivity and edge infrastructure is not enough for customers to benefit from the edge.

Operators are uniquely placed to embrace and manage both the edge infrastructure but also the IoT tools that allow customers to benefit from the edge.

Operators need to view their propositions at the edge as enabling customer digitisation, which means providing the required tools or embracing new ways of analysing and building intelligence from customer data. There are 3 steps that operators can go through to ensure that they are delivering tools which benefit their customers at the edges.

1. Understand customer change requirements and the impact of edge

Operators need to build a holistic view of a customer's operations. By beginning to understand the areas where data already exists in an organisation or areas where new data can be extracted, understanding new straightforward ways of adding value to the organisation through the deployment of IoT and edge computing will begin to emerge.

2. Introducing new opportunities

The use of edge opens up new IoT and data analytics opportunities, that would not be possible with traditional 'centralised' cloud-based systems. By putting data collection, storage and processing close to or right where the customer needs it, new concepts become achievable. For example. the use of Digital Twin systems can be effectively enabled by edge computing, enabling better predictive maintenance and improved product quality in industrial IoT.

3. Delivery and development

How operators deliver the edge and the services on it will depend very much on the tools and skills available to them. Its clear that IoT will have to share edge infrastructure with other services for the majority of customers, and its also likely that technical skills will cross these boundaries also. Therefore, operators need to ensure that edge infrastructure deployments are designed to meet IoT requirements from the outset, not added as an afterthought. This means that high quality, operator-native data analytics at the edge are more likely to be achieved.

Research conducted by GSMA and IDC in 2018 identified the deployment of edge services as an empowering IoT role for operators. A mix of skills from inside and outside the organisation, acquired through partnerships or acquisition, will likely be needed to pursue new business models and markets. Equipped with a strong set of partners and technology enablers, operators will be in a position to drive new business models forward and create new products and services which will help customers meet their ambitions through the use of IoT.

This report is available at https://www.gsma.com/iot/iot-opportunities-report/

Conclusion

New IoT services at the edge offer considerable opportunities for operators as vertical IoT customers look to further digitise their operations. Additionally, third parties such as cloud service providers and developers will need access to edge infrastructure to optimise their own offerings. It makes sense for the operator to work with them as part of a larger ecosystem to retain a level of control over edge deployments.

Edge today is a complex proposition for many end-users and they are reliant on bespoke offerings from specialist providers to meet their needs. Operators need to work out how to meet these customer requirements for edge. This will likely be by offering a solution or platform which minimises customer complexity and plays to operator strengths of connectivity, security, device management and application integration. Working across the wider ecosystem such as with cloud providers, and helping customers adapt their services across areas of digitisation focus such as data analytics and process automation will let operators create new service opportunities and potential revenue streams.

Edge infrastructure for IoT is something that will be deployed by many operators for other services that are non-IoT related, and so IoT teams need to work closely with their network teams to ensure that they are able to offer services based around these core requirements in the future.

About the GSMA

The GSMA represents the interests of mobile operators worldwide, uniting more than 750 operators with over 350 companies in the broader mobile ecosystem, including handset and device makers, software companies, equipment providers and internet companies, as well as organisations in adjacent industry sectors. The GSMA also produces the industry-leading MWC events held annually in Barce-lona, Los Angeles and Shanghai, as well as the Mobile 360 Series of regional conferences.

For more information, please visit the GSMA corporate website at www.gsma.com. Follow the GSMA on Twitter: @GSMA.