Cloud Customer Architecture for IoT

Executive Overview

The Internet of Things (IoT) is one of the most exciting and most dynamic areas of IT at the present time. IoT involves the linking of physical entities ("things") with IT systems that derive information about or from those things which can be used to drive a wide variety of applications and services which may be directly or indirectly connected or related to those things. IoT covers a very wide spectrum of applications, spanning enterprises, governments and consumers and represents the integration of systems from traditionally different communities: Information Technology and Operational Technology. As a result, it is important for IoT systems to have architectures, systems principles, and operations that can accommodate the interesting scale, safety, reliability, and privacy requirements.

Some examples of the application of IoT include:

- Logistics applications, fleet telemetry and supply chain management; the tracking of physical objects such as packages and containers
- Manufacturing and Industrial applications, involving control and operation of industrial equipment and smart production lines
- Asset management and smart shelving. Connected storage and vending devices.
- Building automation or "smart buildings" where monitoring and control systems are applied to all the systems within a building, facilitating smooth operation of the building and the proactive management and maintenance of the equipment and facilities
- Intelligent transportation systems in particular the management of road and rail transport
- Connected vehicles, involving such capabilities as information feeds to drivers about road status or the use of "black boxes" which assess insurance risks/premiums dynamically
- Smart cities, where monitoring and control of city-wide systems are handled automatically for greater efficiency and to serve citizens better
- Smart grid systems, involving instrumenting the electrical grid at all scales for better management and maintenance of equipment, for optimising the use of power in the grid and dealing with intermittent power sources such as wind
- Consumer applications, typically based on the use of smartphones and wearables
- Medical applications, such as remote monitoring and treatment of patients
- Retail and "intelligent shopping" – making use of information about the consumer to make offers and to direct the consumer to items of interest
- The smart home – autonomous management of domestic premises, including control of heating systems, the operation of consumer appliances and extending to automation of maintenance and ordering of consumables (food, etc.)

Fundamental to IoT are electronic devices that interact with the physical world; sensors that gather information about objects and human activities; actuators that can act on objects. Sensors can take many forms. Devices such as thermometers and accelerometers measure real world characteristics and generate numerical information, whereas cameras and microphones create streams of video

and audio information containing more complex information about the real world. Beacons and load sensors are also part of the IoT category. Actuators also take different forms – for example relays which can switch on or off equipment, such as a heater, a piece of manufacturing equipment, or alternatively displays which can be used to inform people such as drivers.

There are several aspects that apply to IoT systems that affect their architecture and implementation, as follows:

- **Scalability**: Scale for IOT system applies in terms of the numbers of sensors and actuators connected to the system, in terms of the networks which connect them together, in terms of the amount of data associated with the system and its speed of movement and also in terms of the amount of processing power required.
- **Big Data**: Many more advanced IoT systems depend on the analysis of vast quantities of data. There is a need, for example, to extract patterns from historical data that can be used to drive decisions about future actions. The extraction of useful information from complex data such as video is another example of analysis requiring large amounts of processing. The ability to mine existing data for new insights and the need to combine different datasets in novel ways are characteristics likely to be part of an IoT system. IoT systems are thus often classic examples of "Big Data" processing.
- **Cloud computing**: IoT systems frequently involve the use of cloud computing platforms. Cloud computing platforms offer the potential to use large amounts of resources, both in terms of the storage of data and also in the ability to bring flexible and scalable processing resources to the analysis of data. IoT systems are likely to require the use of a variety of processing software – and the adaptability of cloud services is likely to be required in order to deal with new requirements, firmware or system updates and offer new capabilities over time.
- **Real time**: IoT systems often function in real time; data flows in continually about events in progress and there can be a need to produce timely responses to that stream of events. This may involve stream processing; acting on the event data as it arrives, comparing it against previous events and also against static data in order to react in the most appropriate way. There is a parallel need to ensure that corrupted data is detected and not used – whether introduced by faulty sensors or malicious action – since the use of corrupted data could cause harm and damage to humans, equipment, and the environment.
- **Highly distributed**: IoT systems can span whole buildings, span whole cities, and even span the globe. Wide distribution can also apply to data – which can be stored at the edge of the network or stored centrally. Distribution can also apply to processing – some processing takes place centrally (in cloud services), but processing can take place at the edge of the network, either in the IoT gateways or even within (more capable types of) sensors and actuators. Today there are officially more mobile devices than people in the world. Mobile devices and networks are one of the best known IoT devices and networks.
- **Heterogeneous systems**: IoT systems are often built using a very heterogeneous set of. This applies to the sensors and actuators, but also applies to the types of networks involved and the variety of processing components. It is common for sensors to be low-power devices, and it is often the case that these devices use specialized local networks to communicate. To enable internet scale access to devices of this kind, an IoT gateway is used.
- **Security and Privacy**: The question of the security and trustworthiness of distributed heterogeneous IoT systems is a hard problem whose solutions must scale and evolve with

the systems. Data protection is necessary, including significant privacy concerns regarding data that relate to individuals. Gaining assurance that these systems are safe, secure, resilient and uphold their stakeholders expectations about privacy is especially challenging.

- **Compliance**: Providing confidence about the operation of these IoT systems is necessary both due to the regulations of specific industries, sectors and verticals and also the norms and expectations of the stakeholders of the IoT systems.
- **Integration**: IoT systems do not exist on their own, but need to connect to existing operational technology systems like factory systems, building control systems, and other types of physical management systems as well as existing enterprise systems including enterprise applications and enterprise databases.

Understanding IoT architectures builds on understanding mobile, web application hosting, and big data and analytics capabilities, please refer to the CSCC's Cloud Customer Reference Architecture papers for Big Data and Analytics, Web Application Hosting and Mobile [1] [2] [3] for a thorough discussion and best practices on each specific topic.

Figure 1 shows the elements that may be needed for any IoT solution across five domains: user layer, proximity network, public networks, provider clouds, and enterprise networks.

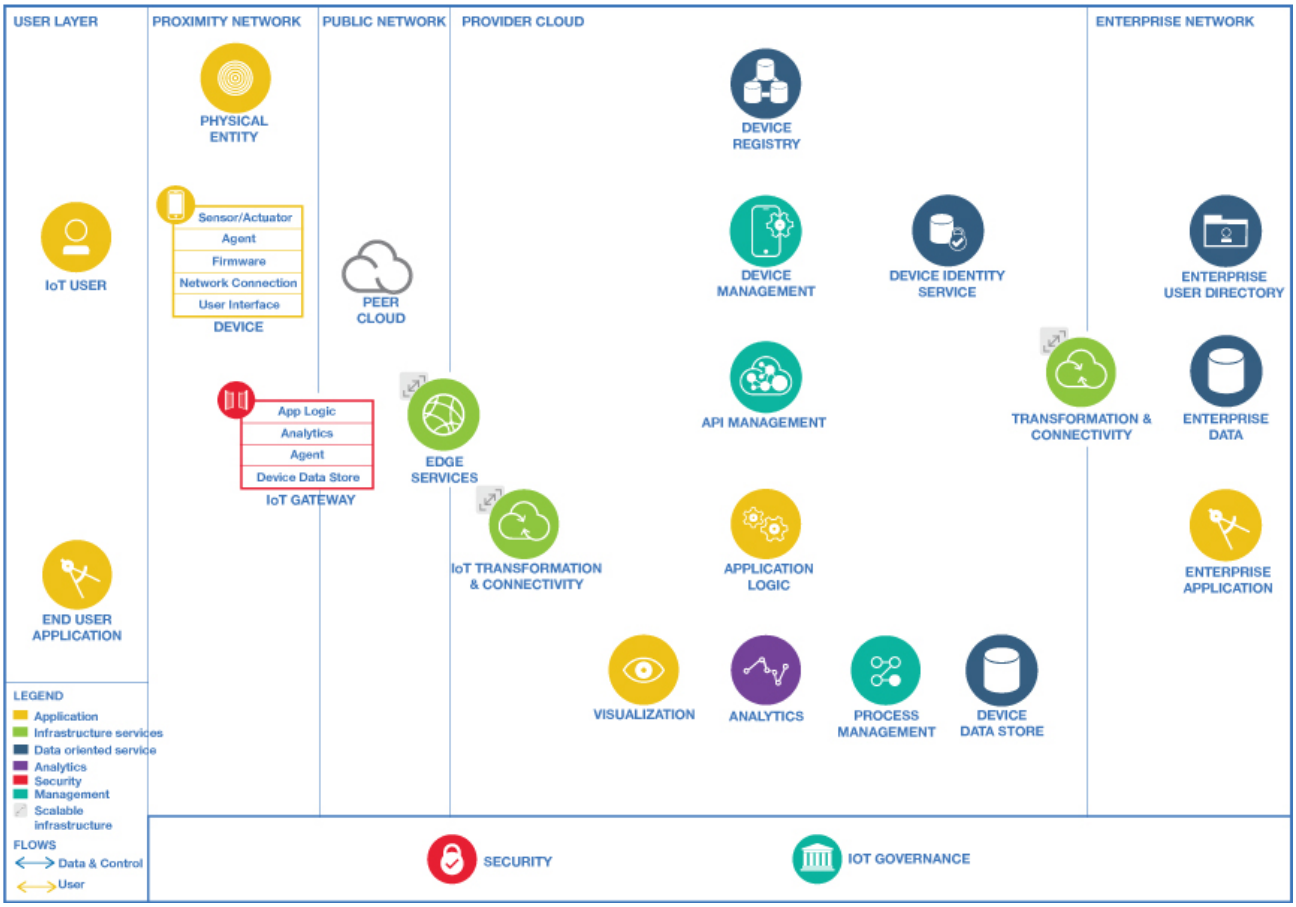


Figure 1: Elements of IoT Solutions

Aspects of the architecture include:

- The user layer is independent of any specific network domain. It may be in or outside any specific domain.
- The proximity network domain has networking capabilities that typically extend the public network domain. The devices (including sensor/actuator, firmware and management agent) and the physical entity are part of the proximity network domain. The devices communicate for both data flow and control flow either via an IoT Gateway and edge services or directly over the public network via edge services.
- The public network and enterprise network domains contain data sources that feed the entire architecture. Data sources include traditional systems of record from the enterprise as well as new sources from Internet of Things (IoT). The public network includes communication with peer clouds.
- The provider cloud captures data from devices, peer cloud services and other data sources (for example Weather services). It can use integration technologies or stream processing to transform, filter and analyse this data in real time and it can store the data into repositories where further analytics can be performed. This processing, which can be augmented with the use of Cognitive and Predictive analytics, is used to generate Actionable Insights. These insights are used by users and enterprise applications and can also be used to trigger actions to be performed by IoT Actuators. All of this needs to be done in a secure and governed environment.
- Results are delivered to users and applications using transformation and connectivity components that provide secure messaging and translations to and from systems of engagement, enterprise data, and enterprise applications.

Cloud Customer Reference Architecture for IoT

Figure 2 shows the capabilities and relationships for supporting IoT using cloud computing.

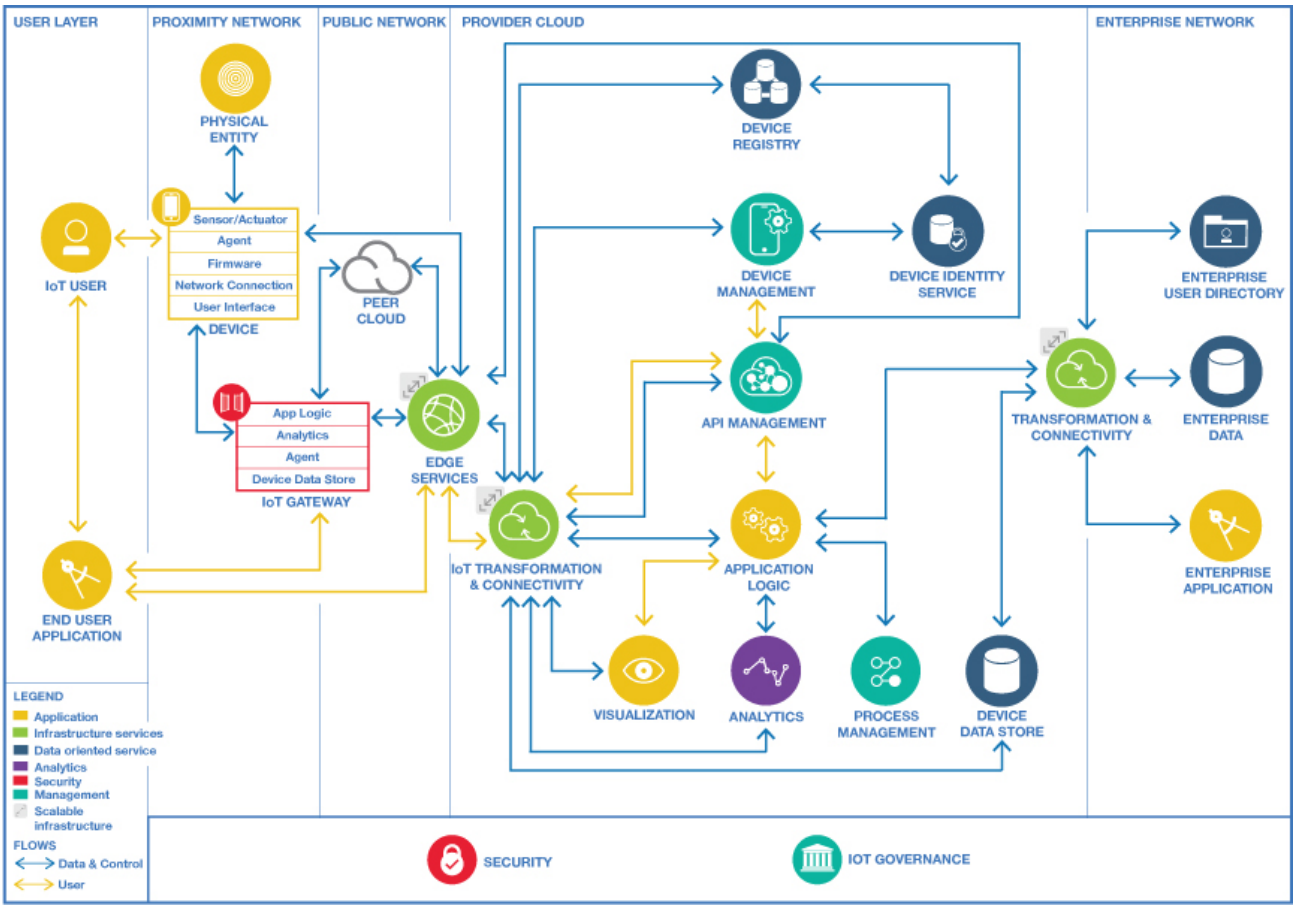


Figure 2: Cloud Components for IoT

The cloud components of IoT architecture are positioned within a three-tier architecture pattern comprising edge, platform and enterprise tiers, as described in the Industrial Internet Consortium Reference Architecture [4].

- The edge-tier includes Proximity Networks and Public Networks where data is collected from devices and transmitted to devices. Data flows through the IoT gateway or optionally directly from/to the device then through edge services into the cloud provider via IoT transformation and connectivity.
- The Platform tier is the provider cloud, which receives, processes and analyzes data flows from the edge tier and provides API Management and Visualization. It provides the capability to initiate control commands from the enterprise network to the public network as well.
- The Enterprise tier is represented by the Enterprise Network comprised of Enterprise Data, Enterprise User Directory, and Enterprise Applications. The data flow to and from the enterprise network takes place via a Transformation and Connectivity component. The data collected from structured and non-structured data sources, including real-time data from stream computing, can be stored in the enterprise data.

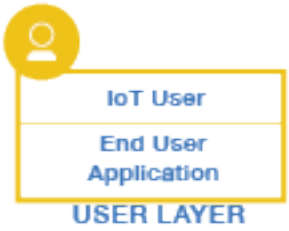
One of the features of IoT systems is the need for application logic and control logic in a hierarchy of locations, depending on the timescales involved and the datasets that need to be brought to bear on the decisions that need to be made. Some code may execute directly in the devices at the very edge of the network, or alternatively in the IoT Gateways close to the devices. Other code executes centrally in the provider cloud services or in the enterprise network. The term "edge computing" is sometimes applied to the case where code executes in the IoT Gateways or the devices. This is sometimes alternatively called "fog computing" to contrast with centralised "cloud computing", although fog computing can also contain one or more layers below the cloud that each could potentially provide capabilities for a variety of services like analytics. This design allows flexibility in how connectivity and services are designed for optimization and resiliency.

IoT governance and security subsystems span all elements of the architecture to ensure controls and policies for all data and applications are defined and enabled across the system. Compliance is tracked to ensure controls are delivering the expected results.

The remainder of this section describes the various components in detail.

User Layer - contains IoT users and their end user applications.

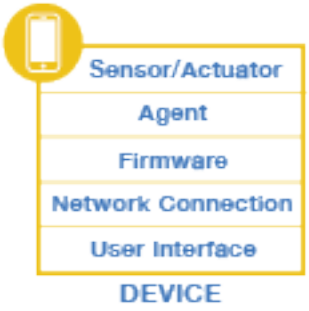
- **IoT User** (people/system) - a person or alternatively an automated system that makes use of one or more end user applications to achieve some goal. The IoT User is one of the main beneficiaries of the IoT solution.
- **End User Application** - domain specific or device specific application. The IoT user may use end user applications that run on smart phones, tablets, PCs or alternatively on specialised IoT devices including control panels.



Proximity Network - contains the physical entities that are at the heart of the IoT system, along with the devices that interact with the physical entities and connect them to the IoT system.

Physical Entity - the physical entity is the real-world object that is of interest – it is subject to sensor measurements or to actuator behavior. It is the "thing" in the Internet of Things. This architecture distinguishes between the physical entities and the IT devices that sense them or act on them. For example, the thing can be the ocean and the device observing is it a water temperature thermometer. Another example is a depot shipping parcels: the parcels are the physical entities and there are devices with sensors capable of observing and identifying each parcel – e.g. via RFID tags or via Barcode readers. It is clear that the RFID Tag reader is one thing and the parcels are something completely different - the identity of the parcel is the physical entity here.

Device - contains sensor(s) and/or actuator(s) plus a network connection that enables interaction with the wider IoT system. There are cases where the device is also the physical entity being monitored by the sensors – such as an accelerometer inside a smart phone.



- **Sensor/Actuator -** senses and acts on physical entities. A sensor is a component that senses or measures certain characteristics of the real world and converts them into a digital representation. An actuator is a component that accepts a digital command to act on a physical entity in some way.
- **Agent -** provides remote management capabilities for the device, supporting a device management protocol that can be used by the Device Management service or IoT management system.
- **Firmware -** software that provides control, monitoring and data manipulation of engineered products and systems. The firmware contained in devices such as consumer electronics provides the low-level control program for the devices.
- **Network Connection** - provides the connection from the device to the IoT system. This is often a local network that connects the device with an IoT gateway – low power and low range in many cases to reduce the power demands on the device. However, there are cases where the network connection is direct to the public network and no IoT gateway is required. In IoT systems, a wide range of alternative communication mechanisms are used which include local area networking using low-power, low-range methods, such as Bluetooth, Bluetooth Low Energy (BTLE), and others. It may also include local area networking using WiFi, to wide area networking using 2G, 3G, and 4GLTE.

User Interface - allows users to interact with applications, agents, sensors and actuators (optional – some devices have no user interface and all interaction takes place from remote applications over the network).

IoT Gateway - acts as a means for connecting one or more devices to the public network (typically the Internet). It is commonly the case that devices have limited network connectivity – they may not be able to connect directly to the Internet. This can be for a number of reasons, including the limitation of power on the device, which can restrict the device to using a low-power local network. The local network enables the devices to communicate with a local IoT Gateway, which is then able to communicate with the public network. The IoT Gateway often has other capabilities, including the ability to filter and intelligently react to data, the ability to send and receive data or commands to and from the Internet, the ability to run application or service logic locally (processing data and executing control logic without the need to communicate to a central location). It can also provide operational efficiency by allowing multiple devices to share a common connection. The IoT Gateway contains the following components:

- **App Logic -** provides domain specific or IoT solution specific logic that runs on the IoT Gateway. For IoT systems that have Actuators which act on physical entities, a significant capability of the app logic is the provision of ***control logic*** which makes decisions on how the actuators should operate, given input from sensors and data of other kinds, either held locally or held centrally.
- **Analytics -** provides Analytics capability locally rather than in the provider cloud.
- **Agent -** allows management of the IoT Gateway itself and can also enable management of the attached devices by providing a connection to the provider cloud layer's Device Management

service via the device management protocol.

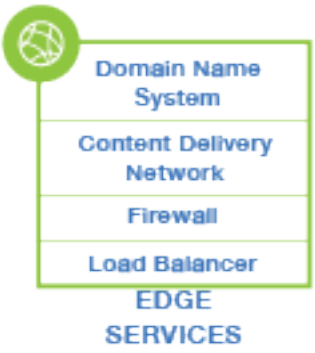
- **Device Data Store -** stores data locally. Devices may generate a large amount of data in real time it may need to be stored locally rather than being transmitted to a central location. Data in the device data store can be used by the application logic and analytics capability in the IoT Gateway.

Public Network - contains the wide area networks (typically the internet), peer cloud systems, the edge services.

Peer Cloud - a 3rd party cloud system that provides services to bring data and capabilities to the IoT platform. Peer clouds for IoT may contribute to the data in the IoT system and may also provide some of the capabilities defined in this IoT architecture. For example an IoT for Insurance solution may use services from partners, such as weather data.

It is likely that for larger IoT systems, such as those involved in Smart Cities, actually involve the combination of a series of smaller IoT systems, each addressing part of the solution – these "systems of systems" involve connections between multiple peer cloud systems, each of which may have IoT devices and associated applications and services. Connecting these individual systems can enable larger more comprehensive solutions.

Edge Services - services needed to allow data to flow safely from the internet into the provider cloud and into the enterprise. Edge services also support end user applications. Edge services include:



- **Domain Name System Server -** resolves the URL for a particular web resource to the TCP-IP address of the system or service that can deliver that resource.
- **Content Delivery Networks (CDN) -** support end user applications by providing geographically distributed systems of servers deployed to minimize the response time for serving resources to geographically distributed users, ensuring that content is highly available and provided to users with minimum latency. Which servers are engaged will depend on server proximity to the user, and where the content is stored or cached.
- **Firewall -** controls communication access to or from a system permitting only traffic meeting a set of policies to proceed and blocking any traffic that does not meet the policies. Firewalls can be implemented as separate dedicated hardware, or as a component in other networking hardware such as a load-balancer or router or as integral software to an operating system.
- **Load Balancers -** provides distribution of network or application traffic across many resources (such as computers, processors, storage, or network links) to maximize throughput, minimize response time, increase capacity and increase reliability of applications. Load balancers can balance loads locally and globally. Load balancers should be highly available without a single point of failure. Load balancers are sometimes integrated as part of the provider cloud analytical system components like stream processing, data integration, and repositories.

Provider Cloud - provides core IoT applications and associated services including storage of device data; analytics; process management for the IoT system; create visualizations of data. Also hosts components for device management including a device registry.

Provider Cloud elements include:

- IoT Transformation and Connectivity
- Application Logic
- Visualization
- Analytics
- Process Management
- Device Data Store
- API Management
- Device Management
- Device Registry
- Device Identity Service
- Transformation and Connectivity

A cloud computing environment provides scalability and elasticity to cope with varying data volume, velocity and related processing requirements. Experimentation and iteration using different cloud service configurations is a good way to evolve the IoT system, without upfront capital investment.

IoT Transformation and Connectivity - enables secure connectivity to and from IoT devices. This component must be able to handle and perhaps transform high volumes of messages and quickly route them to the right components in the IoT solution. The Transformation and Connectivity component includes the following capabilities:

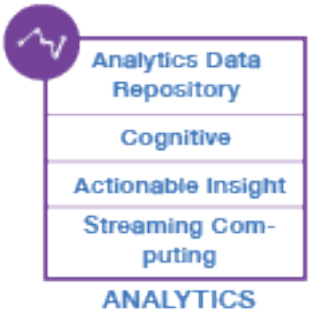
- **Secure Connectivity -** provides the secured connectivity which authenticates and authorizes access to the provider cloud.
- **Scalable Messaging -** provides messaging from and to IoT devices. Scalability of the messaging component is essential to support high data volume applications and applications with highly variable data rates, like weather.
- **Scalable Transformation -** provides transformation of device IoT data before it gets to provider cloud layer, to provide a form more suitable for processing and analysis. This may include decoding messages that are encrypted, translating a compressed formatted message, and/or normalizing messages from varying devices.

Application Logic - the core application components, typically coordinating the handling of IoT device data, the execution of other services and supporting end user applications. An Event based programming model with trigger, action and rules is often a good way to write IoT application logic. Application logic can include workflow. Application logic may also include ***control logic***, which determines how to use actuators to affect physical entities, for those IoT systems that have actuators.

Visualization - enables users to explore and interact with data from the data repositories, actionable insight applications, or enterprise applications. Visualization capabilities include End user UI, Admin UI & dashboard as sub components.

- **End User UI** - allows users to communicate and interact with Enterprise applications, analytics results, etc. This also includes internal or customer facing mobile user interfaces.
- **Admin UI** - enables administrators to access metrics, operation data, and various logs.
- **Dashboard** - allows users to view various reports. Admin UI and Dashboard are internal facing user interfaces.

Analytics - Analytics is the discovery and communication of meaningful patterns of information found in IoT data, to describe, to predict, and to improve business performance. It covers the following capabilities for IoT:



- **Analytics Data repository** - supports legacy, new and streaming sources, enterprise applications, enterprise data, cleansed data and reference data, as well as output from streaming analytics. Capabilities include: Exploration & Archive (for storing, exploring and augmenting large data sets using a wide variety of tools); Deep Analytics & Modeling (application of statistical models to yield information from large data sets comprised of both unstructured and weakly-structured elements); Interactive Analysis & Reporting (tools to answer business and operations questions over Internet scale datasets); Data Catalog (results from discovery and IT data curation create a consolidated view of information that is reflected in a catalog). See [2] for more information on Big Data and Analytics Reference Architectures for using cloud computing.
- **Cognitive -** intelligent system that learns at scale, reasons with purpose, analyses to predict, to prescribe, and to discover from massive datasets of interconnected physical, social, enterprise and other entities, and closes the loop with machine-generated advice, assistance and actions, in a manner that self-learns and adapts, for enabling augmented human intelligence through man/ machine collaborations.
- **Actionable Insight -** insights that ultimately drive actions that may be used by business applications from data collected, processed and stored in the data repositories. Capabilities include: Decision Management (analytics-based and operational); Discovery & Exploration (exploration across a variety of sources to provide business users with new visibility into business performance); Predictive Analytics (extracts information from existing datasets to determine the current state, identify patterns and predict future trends); Analysis & Reporting (reports of operational and warehouse data to business stakeholders and regulators where big data typically increases the scope and depth of available data); Content Analytics (enables businesses to gain insight and understanding from their structured and unstructured content); Planning & Forecasting (enables faster and more efficient development of plans, budgets and forecasts by creating, comparing and evaluating business scenarios).
- **Streaming Computing** - accepts and processes in real time large volumes of highly dynamic, time-sensitive continuous data streams from a variety of inputs such as sensor-based monitoring devices, messaging systems and financial market feeds. Capabilities include: Real

Time Analytical Processing (applying analytic processing and decision making to in-motion and transient data with minimal latency) and Data Augmentation (filtering and diverting in-motion data to data warehouses for deeper background analysis).

Process Management - activities of planning, developing, deploying and monitoring the performance of a business process. For IoT systems, real-time process management may provide significant benefits.

Device Data Store - stores data from the IoT devices so that the data can be integrated with processes and applications that are part of the IoT System. Devices may generate a large amount of data in real time calling for the Device Data Store to be elastic and scalable.

API Management - publishes catalogues and updates APIs in a wide variety of deployment environments. This enables developers and end users to rapidly assemble solutions through discovery and reuse of existing data, analytics and services.

Device Management - provides an efficient way to manage and connect devices securely and reliably to the cloud platform. Device management contains device provisioning, remote administration, software updating, remote control of devices, monitoring devices. Device management may communicate with management agents on devices using management protocols as well as communicate with management systems for the IoT solutions.

Device Registry - stores information about devices that the IoT system may read, communicate with, control, provision or manage. Devices may need to be registered before they can connect to and or be managed by the IoT system. IoT deployments may have a large number of devices therefore scalability of the registry is important.

Device Identity Service - ensures that devices are securely identified before being granted access to the IoT systems and applications. In the IoT systems, device identification can help address threats that arise from fake servers or fake devices.

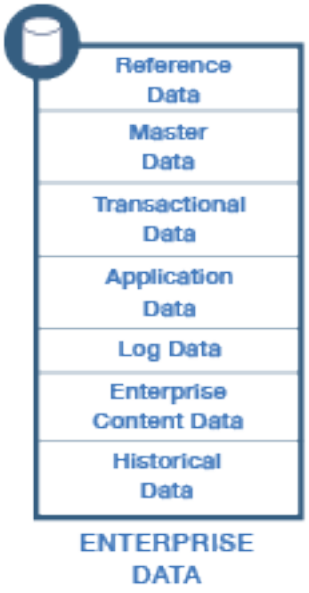
Transformation and Connectivity - enables secure connections to enterprise systems and the ability to filter, aggregate, or modify data or its format as it moves between cloud and IoT systems components and enterprise systems (typically systems of record). Within the IoT reference architecture the transformation and connectivity component sits between the cloud provider and enterprise network. However, in a hybrid cloud model these lines might become blurred. The Transformation and Connectivity component includes the following capabilities:

- **Enterprise Secure Connectivity -** integrates with enterprise data security systems to authenticate and authorize access to enterprise systems.
- **Transformation -** transforms data going to and from enterprise systems.
- **Enterprise Data Connectivity -** enables provider cloud components to connect securely to enterprise data. Examples include VPN and gateway tunnels.

Enterprise Network - host a number of business specific enterprise applications that deliver critical business solutions along with supporting elements including enterprise data. Typically, enterprise applications have sources of data that are extracted and integrated with services provided by the cloud provider. Analysis is performed in the cloud computing environment, with output consumed by the enterprise applications.

Systems of record data have *generally* matured over time and are highly trusted. They remain a primary element in reporting and predictive analytics solutions. Systems of record data include transactional data about or from business interactions that adhere to a sequence of related processes (financial or logistical). This data can come from reference data, master data repositories, and application data used by or produced by enterprise applications functionally or operationally. Typically the data has been improved or augmented to add value and drive insight. Enterprise data may in turn be input into the analysis process through data integration or directly to the data repositories as appropriate.

Enterprise Data - includes metadata about the data as well as systems of record for enterprise applications. Enterprise data may flow directly to data integration or the data repositories providing a feedback loop in the analytical system for IoT. IoT systems may store raw, analyzed, or processed data in appropriate Enterprise Data elements. Enterprise Data includes:



- **Reference Data -** Provide context about collected data.
- **Master Data Repositories** - These repositories can be updated with the output of analytics, to assist with subsequent data transformation, enrichment and correlation. They can support analytics and feed other analytics models when those models execute.
- **Transactional Data** - Data about or from business interactions that adhere to a sequence or related processes (financial or logistical). This data can come from Reference Data, Master Data Repositories, and Distributed Data Storage.
- **Application Data -** Data used by or produced by enterprise applications functionally or operationally. Typically the data has been improved or augmented to add value and drive insight.
- **Log Data -** Data aggregated from log files for enterprise applications, systems, infrastructure, security, governance, etc.
- **Enterprise Content Data -** Data to support any enterprise applications.
- **Historical Data -** Data from past analytics and enterprise applications and systems.

Enterprise User Directory - stores user information to support authentication, authorization, or profile data. The security services and edge services use this to control access to the enterprise network, enterprise services, or enterprise specific cloud provider services.

Enterprise Applications - Enterprise applications consume cloud provider data and analytics to produce results that address business goals and objectives. Enterprise applications can be updated from enterprise data or from IoT applications or they can provide input and content for enterprise data and

IoT applications. Applications might include:

- **Customer Experience -** Customer-facing systems can be a primary system of engagement that drives new business and helps service existing clients at lower cost.
- **New Business Models -** Alternative business models that focus on low cost, fast response and great interactions are all examples of opportunities driven by cloud solutions.
- **Financial Performance -** Financial applications can be made more efficient as data is consolidated and reported faster and more easily.
- **Risk Analytics -** which can be used to evaluate threats to the business, such as fraud or hacking. Elastic resource management means more processing power is available in times of heightened threat.
- **IT Economics –** used to streamline IT operations are streamlined as capital expenditures are reduced while performance and features are improved by cloud deployments.
- **Operations and Fraud -** Cloud solutions can provide faster access to more data allowing for more accurate analytics that flag suspicious activity and offer remediation in a timely manner.

Security and Privacy

Security and Privacy in IoT deployments must address both information technology (IT) security as well as operations technology (OT) security elements. Furthermore, the level of attention to security and the topic areas to address varies depending upon the application environment, business pattern, and risk assessment. A risk assessment will take into account multiple threats and attacks along with an estimate of the potential costs associated with such attacks.

In addition to security considerations, the connecting of IT systems with physical systems also brings with it the need to consider the impact to safety that the IoT system may have. IoT systems must be designed, deployed, and managed such that they can always bring the system to a safe operating state, even when disconnected from communications with other systems that are part of the deployment. Indeed, disconnecting from communications may be part of the security measures put in place to help secure the IoT deployment.

There are several areas of security to consider:

- Identity and Access Management
- Data Protection
- Security Monitoring, Analysis, and Response
- System, Application, and Solution Lifecycle Management

Each of these areas is briefly discussed below.

Identity and Access Management

As with any computing system, there must be strong identification of all participating entities – users, systems, applications, and, in the case of IoT, devices and the IoT gateways through which those devices communicate with the rest of the system. Device identity and management necessarily involves multiple entities, starting with chip and device manufacturers, including IoT platform providers, and also including enterprise users and operators of the devices. In IoT solutions it is often the case that multiple of these entities will continue to communicate and address the IoT devices throughout their operational lifetime.

Data Protection

Data in the device, in flight throughout the public network, provider cloud, and enterprise network, as well as at rest in a variety of locations and formats must be protected from inappropriate access and use. Multiple methods can be utilized, and indeed, in many cases, multiple methods are applied simultaneously to provide different levels of protection of data against different types of threats or isolation from different entities supporting the system. Communications link protection may be used in addition to individual data field level encryption and/or signing done at/in the device in order to provide both end-to-end and point-to-point communications protection. Data at rest in different formats may be encrypted at the field, database, and even whole disk/media level to protect against leakage and improper usage. Increased data collection also results in a need to consider potential privacy implications, requiring additional attention to data segregation, redaction, and special handling requirements.

It is important to consider whether the data involved in an IoT system includes personally identifiable information (PII) – which implies legal and regulatory obligations, sometimes termed "privacy." In some cases, devices may be directly associated with individuals, or individuals may be the physical entities that are the subject of sensor data. If it is possible to associate the device or the data with an individual, then the data is likely to be PII. It is important to recognise that with enough of this observed information, the aggregate data could be used to identify the person it relates to, even where individual elements of data do not. As an example, a domestic electricity meter generates readings that can be related to the individuals who live in the premises concerned – and thus the meter readings should be treated as PII. PII is usually the subject of laws and regulations and the IoT system must be designed to give appropriate protection to these types of data. Protections may involve where and how data can be stored, the identified owner of the information, and what data usage restrictions need to be enforced. Data protection considerations can have a wide range of implications. As just one example, it may be the case that data collected by the device must be stored in the same vicinity where it is collected, either on the device or on an IoT gateway that is close to the device – it cannot be transmitted to a central location such as the provider cloud.

Security Monitoring, Analysis, and Response

Every system must have monitoring of the environment built in so that active attacks as well as anomalous behavior is detected and acted upon. Because of the scale of IoT systems, both in the number of devices as well as the amount of information being processed, there is a requirement for automated response to known attacks as well as automatic detection of suspicious behavior. Response to attacks and suspicious behavior may include temporary isolation, quarantine, or removal of parts of the IoT system as well as having formal incident response processes for addressing vulnerabilities that are discovered after the systems have been put into service. Like IT security, there is a need for disclosure of vulnerabilities such that all affected parties can implement appropriate mitigations, changes, and updates in a timely manner. Note that attacks could come in a wide variety of different forms. As just one example, an attack might come in the form of injection of fake, erroneous, or erratic sensor data into the IoT system in an attempt to steer automated decision-making parts of the system to act in a desired (by the attacker) manner. Such attacks must also be expected, planned for, and responded to.

System, Application, and Solution Lifecycle Management

Lifecycle management of the IoT system is complex, multi-faceted, and has relationships with identity management, device management, the supply chain, application and software development, through to system operations and change management of deployed and in-service systems. Attention to security in

all of these areas is required in order to prevent a variety of attacks ranging from malicious code insertion to inappropriate firmware/software deployment, to effective cryptographic key management. Code, key material, and even physical components must be verified as they flow from procurement and creation through to their installation into the devices, IoT gateways, and systems that make up the IoT system. The IoT system should also provide the capability to update individual components in a secure way, both to address vulnerabilities and also to address functional enhancements over the lifetime of the system. Similar lifecycle considerations apply to privacy and data protection – from "privacy by design" approaches during the creation and assembly of the IoT system through to data deletion when components are decommissioned.

IoT Governance

As described in the IoT Security section, there are many challenges in securing an Internet of Things solution. Oversight and procedures must be used to ensure that when new vulnerabilities and threats are discovered, there is a means and mechanism for addressing these threats in IoT systems.

An important difference in IoT systems from traditional IT systems is that exploits and failures have the potential to cause serious harm to humans, property, and the environment. Physical devices and equipment is usually in-service for much longer periods of time than typical computing systems such as servers, PCs, tablets, and other mobile devices. In addition, this equipment is often installed into locations where change/replacement is not possible, at least not without great cost, inconvenience, or both. This suggests that IoT systems must be designed and deployed with change/update/modification in mind along with strong governance of these systems to ensure that such change is done appropriately, safely, reliably, and securely. Indeed, IoT system change is likely to be needed long after device warranty periods have expired, as it is well known that physical systems are often used for very long periods of time.

Strong governance procedures are needed to determine and enforce the appropriate in-service lifespan for devices and to plan smooth, non-disruptive and secure changeover as new systems are introduced into the system.

The Provider Cloud components may also be subject to change over time – for example, the analytics components and their associated software may undergo regular enhancements to improve their performance and reliability. Appropriate governance must be in place to ensure that changes to these components are understood ahead of time and that the changes do not have an adverse impact on the overall IoT system.

The Complete Picture

Figure 3 provides a more detailed view of components, subcomponents and relationships for a cloud-based IoT solution architecture.

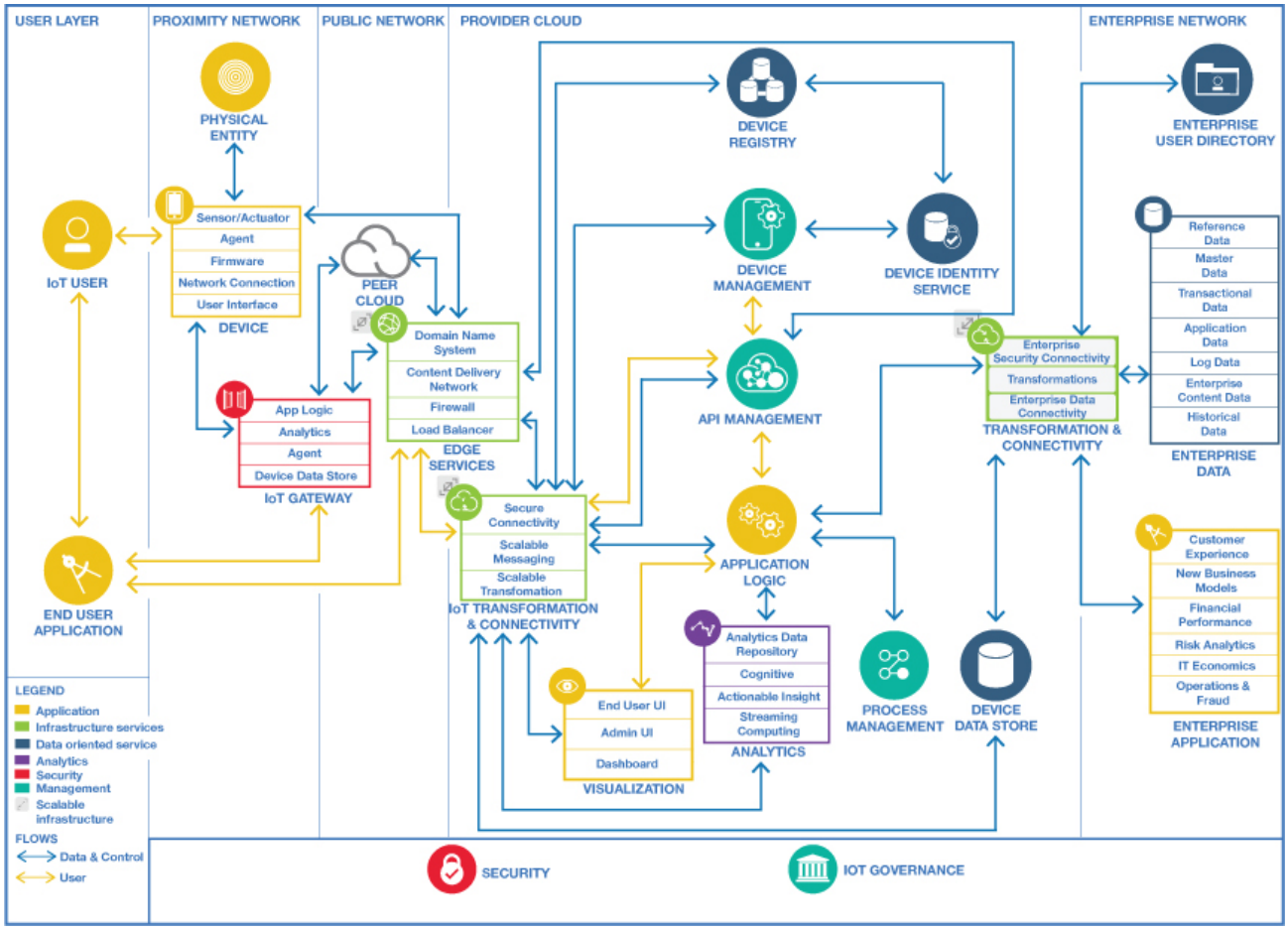


Figure 3: Detailed Components Diagram

Runtime Flow

Figure 4 illustrates the flow of a connected insurance service use case for IoT.

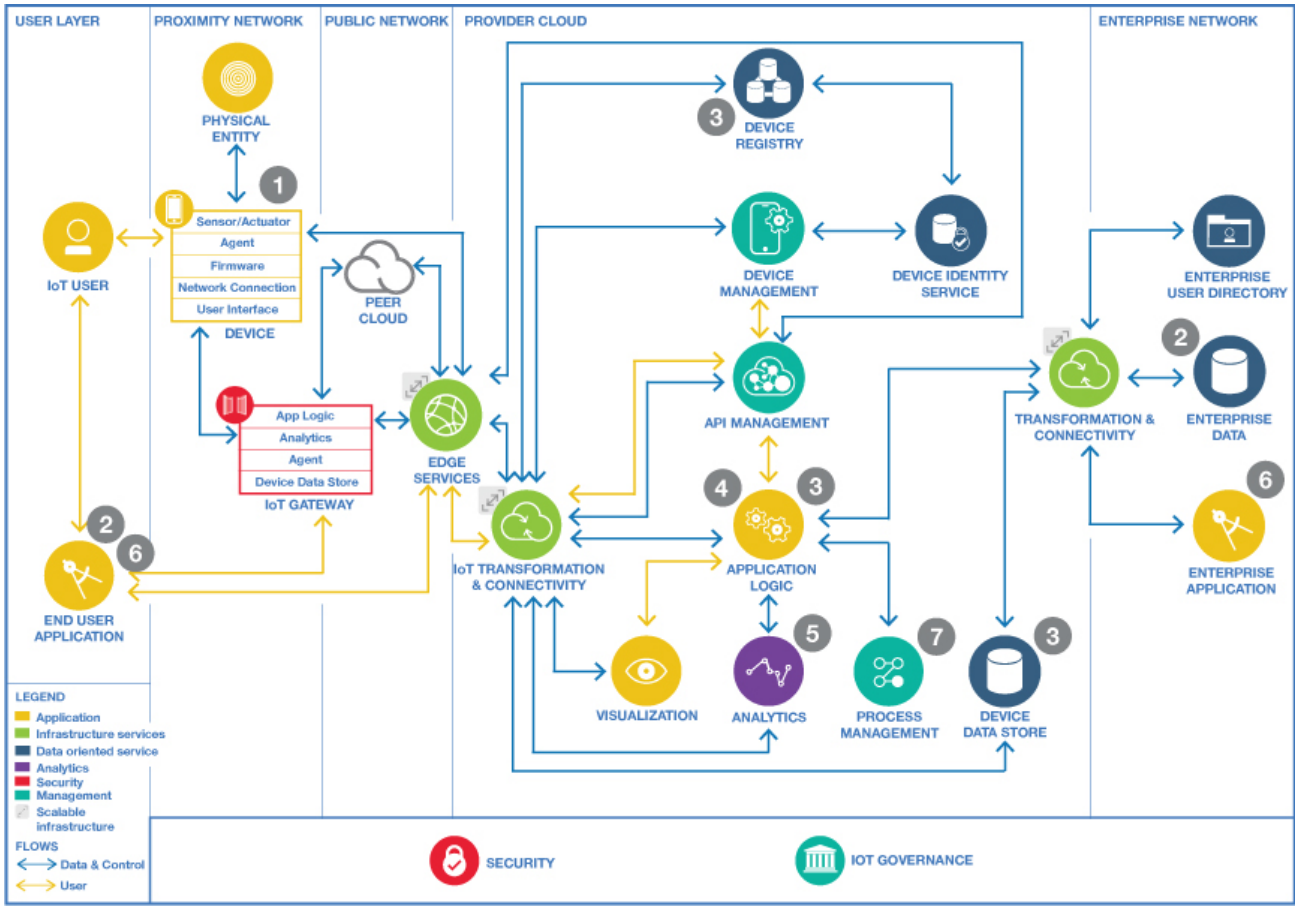


Figure 4: Flow for Insurance Scenario for IoT

In this example, Smart Homes with connected devices/sensors provide insurance companies the ability to improve the service to their policyholders while gaining insight into risks in the home. This allows the policyholder to receive notification of potential danger to the home and engage with the insurer in a more proactive manner. By connecting home ecosystem partners, insurers and other services such as weather information the connected insurance service leverages key components of the IoT Reference Architecture. As an example there is the use of leak detection sensors and valves that provide the policyholder with monitoring of water leaks and protection from resulting damage. The sensors are purchased from multiple sources and installed in the home that includes connecting them to the device makers cloud services. Then the policyholder authorizes the insurance cloud service to connect to the device makers cloud service granting access to the device data. The device maker is responsible for the lifecycle of the devices and the insurance company benefits from access to the data from these devices and provides an improved experience to its policyholders. Basic information flow includes:

1. Sensors and actuators are deployed in the home and attached to the device makers cloud service. As an example the sensors can include water leak detection, water flow, temperature and the actuators can include automatic water shutoff valves.

2. Homeowner logs into the insurance mobile application and authorizes the insurance service to access the device makers (peer) cloud and their device data. The mobile application sends the authorization token and insurance company identifier to the cloud service.
3. This information is used to map the user, devices and insurance policy within the cloud service. The device cloud service is used because the device makers have already deployed into their own cloud and owns the lifecycle of the device as well as the user experience with the devices.
4. The insurance service receives authorization/device details/insurance id from the insurance mobile application and processes this in several nodes (application logic, device registry and device data store). The devices are registered with the device registry and data mapping is updated in the application logic component.
5. Insurance service application connects to the device maker (peer) cloud using the authorization token and requests the data. The application is setup to pull data on a configured interval. In addition to device data, the application can be configured to access other data sources such as a weather data service for use in analysis.
6. Data from devices and other sources such as the weather service are continually updated and sent to analytics to determine if a potential risk threshold has been exceeded. This data is analyzed to determine if there is a potential for damage to the home (including water damage, freeze potential, etc.). Once it is determined that there is a problem, using the analysis from step 5 notifications are sent to the homeowner and to the insurance company. The homeowner can then take an action to respond to the notification and determine if damage has occurred and the insurance company can initiate a claim process.
7. If damage has occurred then the insurance business process of claims management is initiated. The insurance business processes can be accomplished in the cloud service, their enterprise applications, or their mobile applications. This is dependent on how and where the insurance company decides to perform the business logic.

Cloud architecture makes this type of solution easier to implement and maintain. As demand increases, more resources must be acquired.

Deployment Considerations

Cloud environments offer tremendous flexibility with less concern for how components are physically connected. The need for advanced planning is reduced but still important. This section offers suggestions for better provisioning of data and computing resources.

Initial Criteria

- Scalability & Elasticity
- Data Bandwidth
- Data Sovereignty
- Resilience
- CPU and Computation
- Data Volume
- Security

No single cloud environment optimizes all these criteria. A little advanced planning goes a long way towards ensuring user satisfaction – and it helps keep costs in line with expectations.

Scalability & Elasticity

In IoT architecture the number of sensors can be very large and the associated number of transactions can be even larger. This is further multiplied in cases such as connected cars with other data such as traffic and the weather. IoT transformation and connectivity needs to provide scalable messaging and scalable transformation of data in cloud for these data flows. Elasticity is the ability for a cloud solution to provision and de-provision computing resources on demand as workloads change. Public clouds have a distinct advantage since they generally have larger pools of resources available. You also benefit by only paying for what you use. Private clouds and dedicated hardware can make up some of the difference with higher bandwidth data paths. Setting up auto-scale for the queue is not necessarily a one-time event; adjust as usage is better understood to avoid over or under subscribing.

Data Bandwidth

Public and private clouds need to be optimized for big data. Large cloud data sets requiring fast access benefit from processing components with fast and efficient data access. In many cases, this means moving the processing to the data, or vice versa. Cloud systems can effectively hide the physical location of data and processing. Tuning activities can be carried out continuously with minimal impact on deployed applications.

Data Sovereignty

The physical location in which data is stored may be regulated, with the regulations varying from country to country. This is particularly the case for personally identifiable information (PII) and for sensitive data such as health data and financial records. The European Union has particularly stringent regulations that apply to the PII of European citizens. As a result, any IoT cloud system must take into account data sovereignty rules and store and process data only in those locations permitted by the regulations – this requires that the provider cloud used provides the cloud service customer with control over storage and processing locations.

Resilience

In IoT systems resilience and fault tolerance is very important. IoT systems should not depend on one single component at any point and should tolerate the failure of a single component, such as a single IoT device. Components in the provider cloud can be made resilient through the use of multiple instances of programs and cloud services allied with data replication and redundancy on multiple storage systems. The networks should also be resilient, for example with multiple paths and multiple providers in the public network. There is no silver bullet to make the entire network available all the time but it should be a highly available and resilient. It is important to ensure that the connectivity capabilities can support resilience.

CPU and Computation

The availability of inexpensive commodity processors means that public, private and hybrid cloud server farms are typically highly scalable. Modern development environments using Hadoop, Spark and Jupyter (iPython) take advantage of these massively parallel systems. Streams and high-speed analytics are an emerging area where cloud applications leverage more powerful processor pools to enable real-time, in-motion data solutions. Dedicated hardware allows for faster development and testing prior to migration towards hybrid and public environments.

Data Volume
In IoT systems the data volume can exceed a threshold at which the traditional analytic toolsets and approaches may no longer scale in meeting performance requirements. So careful planning to store data in public cloud or private cloud or traditional data center is very important. Streaming of data in cases of weather or map use for GPS may result in a huge data set for analysis. Also all data loses relevance over time. Data retention requires a little experimentation, unless specifically governed by regulatory or other policies. Public clouds offer the flexibility to store varying amounts of data with no advance provisioning. In-house cloud storage solutions can offer long term storage cost advantages when volume is predicted in advance.

Security
As more data about people, financial transactions and operational decisions is collected, refined and stored, the challenges related to information governance and security increase. The data privacy and identity management of devices and individual is very important from the cloud computing point of view. The cloud generally allows for faster deployment of new compliance and monitoring tools that encourage agile policy and compliance frameworks. Cloud data hubs can be a good option by acting as focal points for data assembly and distribution. Tools that monitor activity and data access can actually make cloud systems more secure than standalone systems. Hybrid systems offer unique application governance features: Software can be centrally maintained in a distributed environment with data stored in-house to meet jurisdictional policies.

Optimized Provisioning
Optimized cloud provisioning can help you select the right product family for a given set of usage criteria.

Hybrid Cloud and IoT

An enterprise routinely needs a combination of public cloud, private cloud and on-premises components that when linked, create a *hybrid cloud*. Hybrid cloud computing is a deployment model which involves combining the use of multiple cloud services across different deployment models – in particular, combining the use of public cloud services with private cloud services and existing on-premises enterprise systems. See the CSCC *Practical Guide to Hybrid Cloud Computing* [5] for more information about hybrid cloud.

Businesses implementing hybrid cloud solutions are looking for flexibility and agility in delivering new capabilities.

Some examples that explain the need for hybrid cloud deployment for IoT systems:

IoT for Health - consider the IoT system to support a diabetic patient with a continuous glucose-monitoring device, an insulin pump and activity monitoring devices on his body. The core of the solution is implemented in a cloud provider platform, with the patient's devices registered with the device registry and managed remotely using the device management capability. Processing of the incoming streams of data from the devices is handled with the device data store and custom application logic, along with analytics to identify patterns in the data that might indicate any problems that require special attention.

The patient is registered in the enterprise user directory of the health organization (the registration is used for other purposes as well as the IoT capabilities) and the general health systems are enterprise applications and involve enterprise data such as the patient's health records. Such applications and data are very likely to be too sensitive to place into the provider cloud and need strong access controls. It is likely that the cloud provider components are deployed in a private cloud, either on-premises or off-premises, due to the sensitive personal information involved, with rigorous security controls in place to avoid unauthorized access to the data. The applications directly used by the patient and the patient's relatives / helpers are supported from the provider cloud, typically using apps on smartphones. The applications used by health professionals are a combination of ones running in the provider cloud and ones running in the enterprise network.

This IoT system uses a hybrid cloud deployment, with many of the components supporting the health devices and patient applications running on the provider cloud, while the core health systems and their associated data run in the enterprise network of the health organization.

IoT for Connected Cars - The IoT solution for Connected Cars is a real time event detection and management system designed to detect, analyze and handle events generated by connected cars in a secure manner. Some of the information with historic and maintenance data for car manufacturer will stay in their dedicated private cloud or in their traditional data centers while other generic information and their integration with 3rd party cloud services may stay in public cloud. Connected cars need real-time information about weather, traffic and map data that comes from peer cloud services. For the data privacy and sovereignty requirements, data with personal information about customers may reside in on-premises data centers in specific countries. With use of Hybrid cloud only can we handle these specific needs.

Summary

The Internet of Things is a dynamic and exciting area of IT. Many IoT systems are going to be created over the next few years, covering many and varied use cases in wide variety of domestic, commercial, industrial, health and government contexts.

IoT systems involve many distinct components, each with their own challenges. Aspects of scale, speed, safety, security and privacy are pervasive in IoT systems and need careful attention. The reference architecture described in this white paper provides a sound basis for understanding IoT systems and for addressing the various challenges in a systematic and logical way.

References

[1] Cloud Standards Customer Council 2015, *Cloud Customer Architecture for Big Data and Analytics, Version 1.1*
http://www.cloud-council.org/deliverables/CSCC-Customer-Cloud-Architecture-for-Big-Data-and-Analytics.pdf

[2] Cloud Standards Customer Council 2015, *Cloud Customer Architecture for Web Application Hosting, Version 2.0*
http://www.cloud-council.org/deliverables/CSCC-Customer-Cloud-Architecture-for-Web-Application-Hosting.pdf

[3] Cloud Standards Customer Council 2015, *Cloud Customer Architecture for Mobile*
http://www.cloud-council.org/deliverables/CSCC-Customer-Cloud-Architecture-for-Mobile.pdf

[4] The Industrial Internet Consortium's Industrial Internet Reference Architecture IIRA paper
http://www.iiconsortium.org/IIRA.htm

[5] Cloud Standards Customer Council 2016, *Practical Guide to Hybrid Cloud Computing*
http://www.cloud-council.org/deliverables/CSCC-Practical-Guide-to-Hybrid-Cloud-Computing.pdf

Acknowledgements

Major contributors to this whitepaper are: Glenn Daly (IBM), Mike Edwards (IBM), Tim Hahn (IBM), Gopal Indurkhya (IBM), Heather Kreger (IBM), Eric Libow (IBM), Bob Marcus (ET-Strategies), Bob Martin (MITRE), Peter Niblett (IBM), Alex Tumashov (Schlumberger).