Solution Brief

Bridging the Infrastructure Gap for the Internet of Things with Object Storage

Table of Contents

The Internet of Thing: Benefits and Challenges

The introduction of the first web browser and the first httpd server in 1990 fueled an evolution of information storage and retrieval. Along with the growth of the World Wide Web, an unprecedented expansion of infrastructure has led to some of the most innovative ideas in the history of humankind, mostly notably Wi-Fi, web 2.0, and mobile technology. For companies, these technologies have enabled never before seen levels of access to customers and their supply chain. We define these elements (devices, machines, mobile apps, websites, partner systems, etc.) as your revenue chain. Every day more of these elements are being enabled to send and receive information via wireless networks. The revenue chain elements, communicating, collecting, and transferring data are what defines your Internet of Things or IoT. To help give perspective, we should explore some key subdivisions of IoT:

- Public IoT

- Consumer IoT

- Commercial IoT

Public Internet of Things

Public IoT can be defined as a network of the sensor enabled elements of public works systems provided by municipalities, state, and federal government. This IoT movement is designed to improve efficiency in the delivery of services and enhance public safety. An example of Public IoT systems is "smart" infrastructure such as smart roads, smart utility meters, and smart electric power grids. Much of this smart infrastructure will make it easier for the governing entity to monitor and maintain the system. As more people begin to live in cities, the need to deliver services to residents quickly is a big driver for these systems.

Part and parcel of the Public IoT is the data generated by governing agencies themselves. Datasets collected by cities, states, and federal governments are a part of the Public IoT as they can be combined with sensor data and other IoT systems. For example, housing and demographic data combined with traffic and environmental information can help pinpoint where additional public transportation is needed more than a wider street. Having this insight can help city planners make those decisions with relevant, contextual and in some cases, near real-time data.

Consumer Internet of Things

Consumer IoT is an extension of web 2.0 that includes the mobile web: phones, devices, apps, and the software infrastructure needed to create a networked, responsive experience. The push to integrate smart devices throughout your personal life from appliances, thermostats, or home entertainment is moving at a rapid pace. At the center of Consumer IoT is your smart device. Your smart phone is the source of computing used by mobile apps to collect data about where, when, and how you use certain products and services. The entities that build mobile apps use the information they collect, as well as information from Public IoT to make their best decision about how to interact with you. A lot of information is crowdsourced from apps to create data products you use. A great example of a crowdsourced application is how map apps use the position of all users to create traffic and other status information. Further analysis and combinations of data can yield revelations not specific to users of the app. This leads us to a third type of IoT called Commercial IoT

Commercial Internet of Things

Commercial IoT can be defined as a company's products and services with sensors designed to send near real-time operational data over a network. Sensors, defined as a device measuring, recording, or responding to a physical property change, capture this data. Data collected and transmitted from these sensors over a network is defined as telemetry. This telemetry can be used to provide a prescriptive approach to service delivery and even predict failures using analytics. Examples of devices outfitted with sensors are engines, manufacturing robots, HVAC systems, and many other everyday finished products. Many of these devices have always had gauges and other instrumentation to indicate status to a trained technician upon inspection but could not capture all of the vital information without taking the products out of service. Sensors enable these products to measure current and other physical properties in real time.

Once captured, this information is transmitted over a network to be collected and further analyzed, providing a monumental shift in the way products are maintained and improved. Technicians can now pinpoint issues and prepare for the next maintenance cycle ahead of time. Product design engineers can analyze the sum total of data to look for defects, overall design changes, and refine maintenance procedure to provide continuous improvements to their products.

A great example is a jet engine manufacturing company looking to capture telemetry streaming from one their products for collection and analysis. Thousands of their products will be in service all around the world moving constantly from place to place.

The company currently estimates 12 petabytes or more of telemetry data will be captured each year. This scenario, visualized in Figure 1 introduces several challenges.
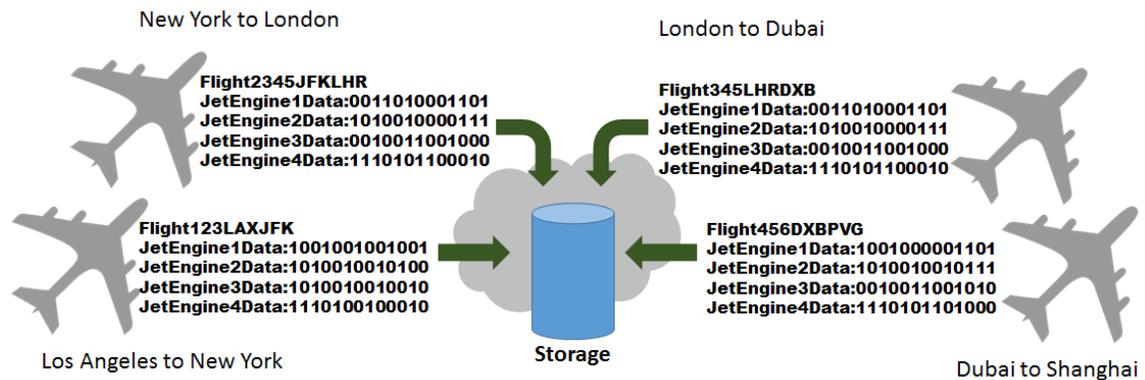

Figure One: Telemetry Collection from a Jet Engine Manufacturer's Internet of Things

The jet engines on the plane generate telemetry in flight. The sensors record the telemetry and send it over a network to be stored for analysis. Support of hundreds of engines on thousands of flights worldwide, running millions of hours in all conditions and generating billions of bytes of telemetry requires an IoT. This data deluge from anytime-anyplace telemetry capture as well as data storage scalability are a challenge for IT architectures. Whether the large jet manufacturer looks to an IT partner or do it themselves, most of today's storage technology does not measure up and introduces an infrastructure gap within IoT.

Figure Two, illustrates the infrastructure gap creation of an Internet of Things and the systems currently running a business can create.
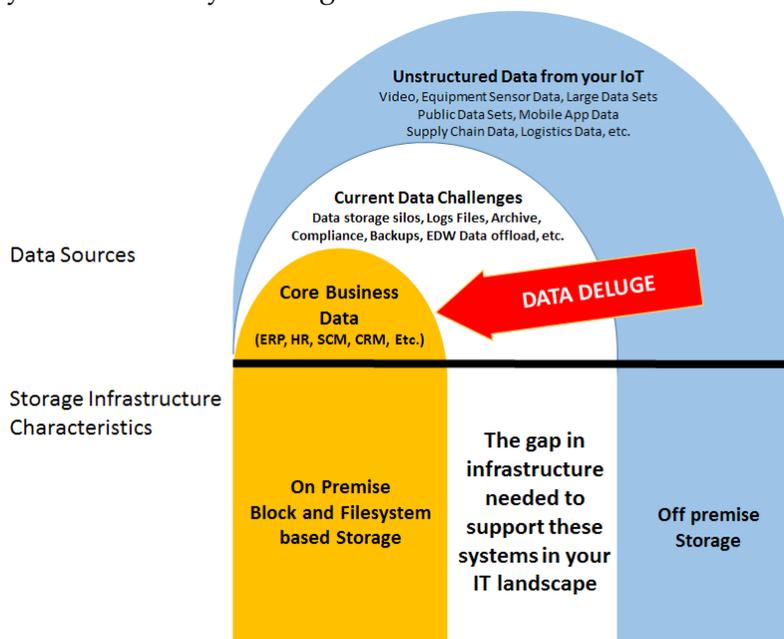

Figure Two: Infrastructure Gap for Data Storage for Internet of Things Telemetry

The characteristics of telemetry are very different from typical IT data. Telemetry is generated once and cannot be regenerated. Telemetry can have different volume, velocity, and variety across each individual product. Sensors capturing telemetry need to securely transmit and directly access storage via a network without a complex setup, minimizing human interaction. Collection and management of certain data provides additional non-IT challenges. Some of these challenges include security concerns, data locality based on regulations, privacy laws, as well as the total cost of ownership (TCO) when scaling past the petabyte range. Storage technology, at the foundation of this challenge, is experiencing a revolution. Object storage is a solution that can address the challenges of the infrastructure gap in your Internet of Things.

In this solution brief, Edison Group explores IoT and the infrastructure gap and evaluates object storage and EMC Elastic Cloud Storage (ECS) as the solution to fill that gap.

Filling the Infrastructure Gap for the Internet of Things with Object Storage

While not a new concept, object storage is one of the hottest terms in IT today. As a result, startups are being acquired, and "born on the web" companies are innovating heavily in this field. Using a flat namespace, object storage uses a globally unique address and metadata to store data. This method reduces the overhead needed to manage storage systems such as:

- LUN creation, expansion, or migration

- Applying data protection schemes

- Creating, extending, and managing file systems

With object storage, data can be of any size and type. From documents, images, audio and video, there is no need to apply special techniques to store these and other data types. Along with a unique global ID, metadata is embedded with each object. Users and applications can embed additional metadata to further increase and customize ease of identification.

Key Advantages of Object Storage for Internet of Things

Let's explore the advantages of object storage for storing data from the IoT. Traditional storage systems are built to interact with operating systems and people, object storage is built to interact with applications and many different data sources. Block storage is needed to provide a place for the applications to run on the operating systems to live and generate data. The telemetry and other sensor data from the IoT will come from devices connect via TCP/IP which exist outside of the traditional operating system. These telemetry-generating devices need simplified and secured access to their storage infrastructure. Using block storage or traditional network attached storage (NAS) is not an option for this type of workload due to characteristics such as:

- One time generation of data that cannot be reproduced

- Data is generated in multiple geographic locations

- Data is unpredictable in volume, variety and velocity

When data is generated only once, the system capturing the data needs to be resilient but easy to scale. TCP/IP inherent resilience helps with transmission of data with an underlying need to capture multiple copies of data to physically different media. Object

systems, by default and automatically, execute this replication of data with rules to govern locality. Replication rules describe where extra copies of data will be stored to provide resiliency and availability. Locality describes what those locations are within the replication rules. The rules for data replication, which can be user defined, specify locality with granularity starting at the physical disk all the way to the datacenter level. This provides an assurance data will be available and reproducible in the event of a failure of one or more components.

When you set the replication rules to provide datacenter locality of data, this characteristic supports data collection in multiple geographic locations. The replications process will bring eventual consistency to the data stored. Data integrity during the replication process is maintained using hash functions. This feature, for example, enables the jet engine manufacturer to ingest and store data from each of their engine flights regardless of geographic location.

Collecting and keeping this information secure and available could mean life and death for passengers and crew as well as the company itself. Having this "real-time, single generation" dataset available helps improve a company's ability to radically change their maintenance and customer service habits. Instead of maintenance checks based on time, the capture and analysis of this data decouples the dependency on scheduled maintenance to find problems. This sets groundwork for predictive analysis as you combine other datasets using analytics and object based storage systems.

Edison Group believes object storage technology is best equipped to address the infrastructure gap between your business, your revenue chain, and the unstructured data your Internet of Things generates. We recommend that organizations begin to investigate this technology within the next 6-12 months and plan to implement it in their environment within 12-18 months. Now that we understand the challenge associated with data deluge from the Internet of Things and what technology is needed to bridge the infrastructure gap, let's get an overview of how an object storage solution, EMC ECS, can close your storage infrastructure gap for the Internet of Things.

EMC ECS and Object Storage

EMC provides a cloud-scale object storage platform that meets the storage demands of today and beyond through their ECS solution. ECS is a turnkey, on-site solution offering all the advantages of commodity infrastructure with enterprise grade reliability, availability, and serviceability. ECS can efficiently store PBs of data – whether billions of small files and/or large files – in a cost appropriate state-of-art, storage system.

EMC ECS Appliance features include:

- Universal protocol support in a single platform with support for object, file (NFS) and HDFS

- Single management view across multiple types of infrastructures

- Geo-federated, active-active architecture with a single global namespace, enabling the management of a geographically distributed environment as a single logical resource using metadata-driven policies to distribute and protect content

- Multi-tenancy support, detailed metering, and an intuitive self-service portal, as well as billing integration

These features allow customers to extend automation capabilities and deliver improved efficiencies across their storage environments, providing better control of operating expenses as their IoT drives data growth at unprecedented rates — one of the key pain points customers face in the current IT landscape.

To help put this data growth in context, the Digital Universe is growing 40 percent yearly into the next decade.[1] By 2020, it will contain as many digital bits as there are stars in the universe. This vast amount of data makes storing, accessing, and managing all this data difficult, not to mention expensive. The way customers distribute and protect their data at scale today will play a very important role in how successful they are in the future.

---

[1] http://www.emc.com/leadership/programs/digital-universe.htm