The OpenFog Consortium Reference Architecture: Executive Summary

This document is the executive summary of the OpenFog Reference Architecture release 1.0, which was published on February 8, 2017. To access the full document, please visit www.openfogconsortium.org/ra.

Fog Computing Overview

Fog computing provides the missing link in the cloud-to-thing continuum. It is a critical architecture for today's connected world as it enables low latency, reliable operation, and removes the requirement for persistent cloud connectivity to address emerging use cases in Internet of Things (IoT), 5G, Artificial Intelligence (AI), Virtual Reality and Tactile Internet applications.

Fog architectures selectively move compute, storage, communication, control, and decision making closer to the network edge where data is being generated in order solve the limitations in current infrastructure to enable mission-critical, data-dense use cases.

Fog computing is:

> ***A horizontal, system-level architecture that distributes computing, storage, control and networking functions closer to the users along a cloud-to-thing continuum.***
>
> *Source: OpenFog Consortium*

Fog computing is an extension of the traditional cloud-based computing model where implementations of the architecture can reside in multiple layers of a network's topology. These extensions to the fog retain all the benefits of cloud computing, such as containerization, virtualization, orchestration, manageability, and efficiency.

The fog computing model moves computation from the cloud closer the edge, and potentially right up to the IoT sensors and actuators. The computational, networking, storage and acceleration elements of this new

model are known as fog nodes. They comprise a fluid system of connectivity and are not completely fixed to the physical edge.

The OpenFog Reference Architecture

The OpenFog Consortium was founded on the principle that an open fog computing architecture is necessary in today's increasingly connected world. Through an independently-run open membership ecosystem of industry, end users and universities, we can apply a broad coalition of knowledge to these technical and market challenges. We believe that proprietary or single vendor fog solutions can limit supplier diversity and ecosystems, resulting in a detrimental impact on market adoption, system cost, quality and innovation.

The OpenFog Reference Architecture (OpenFog RA) is a medium- to high-level view of system architectures for fog nodes and networks. It is the result of a broad collaborative effort of its independently-run open membership ecosystem of industry, technology and university/research leaders. It was created to help business leaders, software developers, silicon architects and system designers create and maintain the hardware, software and system elements necessary for fog computing. It enables fog-cloud and fog-fog interfaces.

OpenFog architectures offer several unique advantages over other approaches, which we term SCALE:

- **S**ecurity: Additional security to ensure safe, trusted transactions
- **C**ognition: awareness of client-centric objectives to enable autonomy
- **A**gility: rapid innovation and affordable scaling under a common infrastructure
- **L**atency: real-time processing and cyber-physical system control
- **E**fficiency: dynamic pooling of local unused resources from participating end-user devices

OpenFog RA Content

The OpenFog RA is a 162-page document that features the following sections:

- A description of the OpenFog Consortium's mission, plans to accelerate fog computing, and an overview of the OpenFog RA itself.

- Selected use cases in early adoption scenarios for fog computing. This list will grow and evolve as the OpenFog RA is refined.
- A description of the eight pillars of the OpenFog architecture. These are the guiding principles for the OpenFog RA.
- An abstract architectural description, providing an in-depth look at the full OpenFog RA.
- A discussion on adherence to the OpenFog architecture with an eye to the objective of the OpenFog RA in driving standardization across the various interfaces.
- Abstract applications of the OpenFog RA to a detailed use case on visual security. This clarifies each aspect of the OpenFog RA and describes what needs to be done for a successful implementation.
- A description of the open areas of fog computing and new opportunities for research.
- A glossary of industry terms.

How Fog Computing Works

Fog computing solves performance challenges in advanced digital deployments in IoT, 5G and artificial intelligence. These include the control of performance, latency and network efficiency. It's important to note that cloud and fog computing are on a mutually beneficial, inter-dependent continuum.

Fog does not replace the cloud; it works with cloud to enable the requirements of selected use cases. Certain functions are naturally more advantageous to carry out in fog nodes, while others are better suited to cloud. The traditional backend cloud will continue to remain an important part of computing systems as fog computing emerges.

To illustrate how fog computing works, consider an oil pipeline with pressure and flow sensors and control valves. One could transport its sensor readings to the cloud (i.e. using expensive satellite links), analyze the readings in cloud servers to detect abnormal conditions, and send commands back to adjust the positon of the valves.

However, the bandwidth to transport the sensor and actuator data to and from the cloud could cost thousands of dollars per month; those connections could be susceptible to hackers; it may take several hundred milliseconds to react to an abnormal sensor reading, during which time a major leak could spill significant oil; and if the connection to the cloud is down or the cloud is overloaded, control is lost.

In that same scenario, if a hierarchy of local fog nodes is placed near the pipeline, they can connect to sensors and actuators with inexpensive local networking facilities. Fog nodes can add extra security controls, lessening the hacker threat. Fog nodes can react to abnormal conditions in milliseconds, quickly closing valves to greatly reduce the severity of spills.

This example illustrates the advantages of local control in the fog nodes to produce a more robust control system. Moving most of the decision-making functions of this control system to the fog, and only contacting the cloud occasionally to report status or receive commands, creates a superior control system.

Platform as a service (PaaS) is a category of cloud computing services that provides a platform allowing customers to develop, run, and manage web applications without the complexity of building and maintaining the infrastructure typically associated with developing and launching an application.

The OpenFog RA defines the required infrastructure to enable building Fog as a Service (FaaS) to address certain classes of business challenges. FaaS includes Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS), and many service constructs specific to fog. Infrastructure and architecture building blocks show how FaaS may be enabled. This topic is expanded upon in the reference architecture.

The OpenFog RA describes a generic fog platform that is designed to be applicable to any vertical market or application. This architecture is applicable across many different markets including, but not limited to, transportation, agriculture, smart-cities, smart–buildings, healthcare, hospitality, energy and financial services. It provides business value for IoT applications that require real-time decision making, low latency, improved security, and are network-constrained.

Pillars of OpenFog Reference Architecture

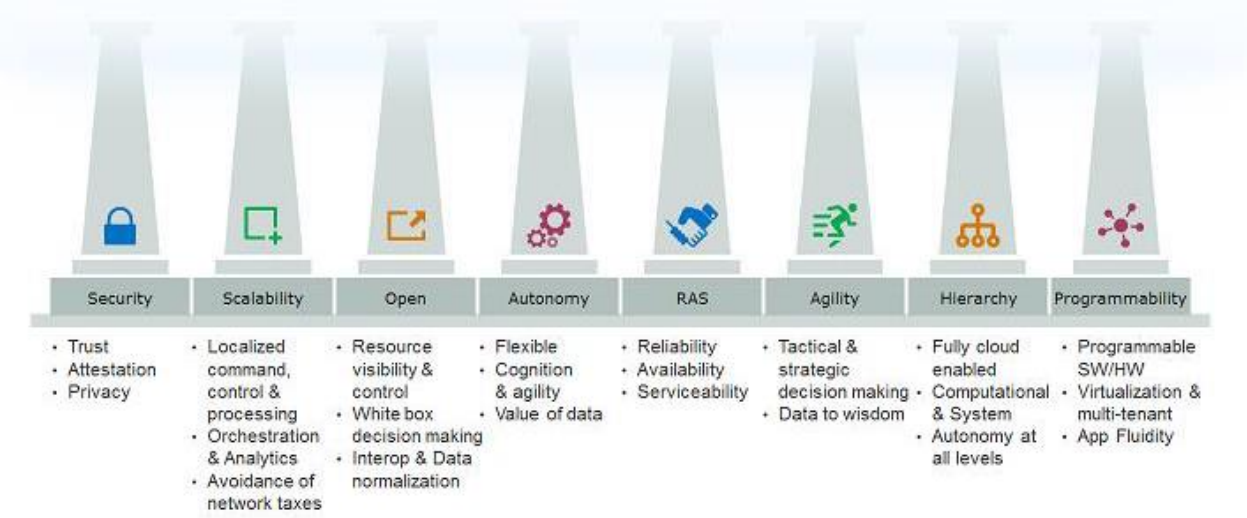

Figure: OpenFog Pillars

The OpenFog RA is driven by a set of core principles called pillars. These pillars form the principals, approach and intention that guided the definition of the reference architecture. They represent the key attributes that a system needs to embody the OpenFog definition of a horizontal, system-level architecture that provides the distribution of computing, storage, control, and networking functions closer to the data source (users, things, etc.) along the cloud-to-thing continuum.

The eight pillars are Security, Scalability, Open, Autonomy, RAS (Reliability, Availability, Serviceability), Agility, Hierarchy, and Programmability. Elements of each are depicted in the above diagram.

Architecture Description

The OpenFog RA description is a composite of perspectives and multiple stakeholder views used to satisfy a given fog computing deployment or scenario. Before going into the lower level details of the view, it is important to first look at the composite architecture description, depicted on the following page:

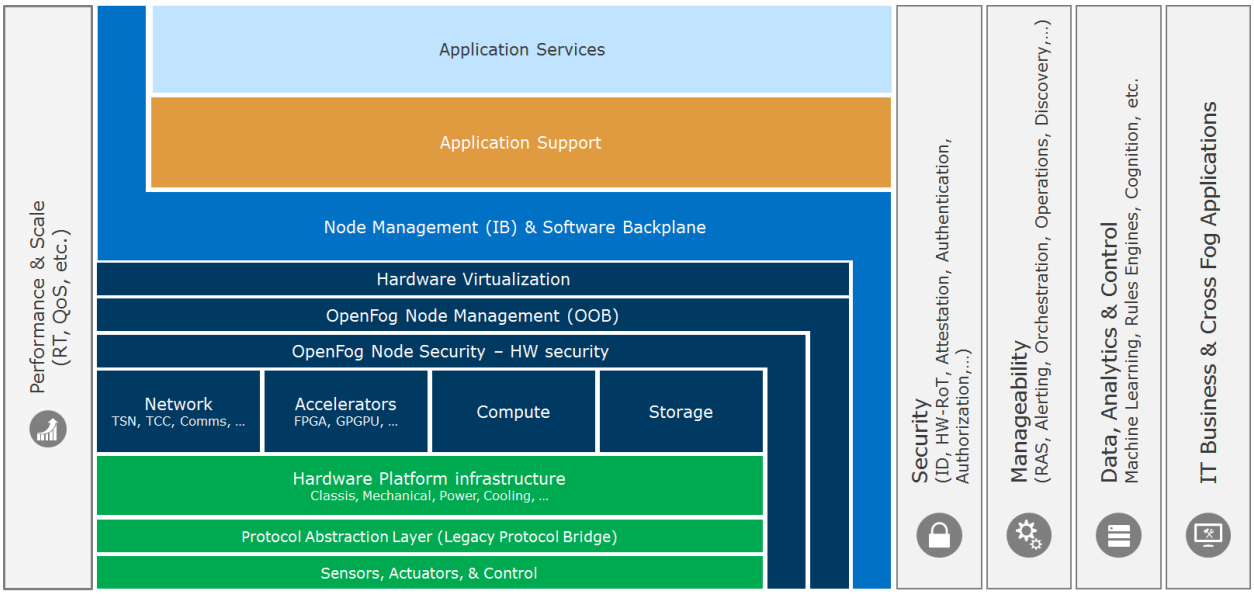

Figure: Architecture Description with Perspectives

The abstract architecture includes perspectives, shown in grey vertical bars on the sides of the architectural description. The perspectives include:

- **Performance:** Low latency is one of the driving reasons to adopt fog architectures. There are multiple requirements and design considerations across multiple stakeholders to ensure this is satisfied. This includes time critical computing, time sensitive networking, network time protocols, etc. It is a cross cutting concern because it has system and deployment scenario impacts.
- **Security:** End-to-end security is critical to the success of all fog computing deployment scenarios. If the underlying silicon is secure, but the upper layer software has security issues (and vice versa) the solution is not secure. Data integrity is a special aspect of security for devices that currently lack adequate security. This includes intentional and unintentional corruption.
- **Manageability:** Managing all aspects of fog deployments, which include RAS, DevOps, etc., is a critical aspect across all layers of a fog computing hierarchy.
- **Data Analytics and Control:** The ability for fog nodes to be autonomous requires localized data analytics coupled with control. The actuation/control needs to occur at the correct tier or location in the hierarchy as dictated by the given scenario. It is not always at the physical edge, but may be at a higher tier.

- **IT Business and Cross Fog Applications:** In a multi-vendor ecosystem applications need the ability to migrate and properly operate at any level of a fog deployment's hierarchy. Applications should be able to span all levels of a deployment to maximize their value.

There are three identified viewpoints in the Architecture description diagram: Software, System, and Node.

- **Software view:** is represented in the top three layers shown in the architecture description, and include Application Services, Application Support, and Node Management (IB) and Software Backplane.
- **System view:** is represented in the middle layers shown in the architecture description, which include Hardware Virtualization down through the Hardware Platform Infrastructure.
- **Node view:** is represented in the bottom two layers, which includes the Protocol Abstraction Layer and Sensors, Actuators, and Control.

Note: The fog platform coupled with the fog software creates the complete fog node. A solution is defined as one or more fog nodes in a given market segment or scenario. The core aspects of a fog node can also be viewed as compute, storage, network, accelerators and control.

However, high-level architectures, including the OpenFog RA, are intended to help engineers, architects, and business leaders understand their specific requirements and how fog nodes can be applied to a given scenario. The goal of the OpenFog Consortium is to increase the number of market segments (use cases) for fog computing and their corresponding business value. The consortium will create testbeds to adapt the high-level architecture to these market segments. These testbeds will also provide opportunities for FogFests (plug fests) to help drive component-level interoperability and accelerate the time to market.

An End-to-End Deployment Use Case

The following example describes an end-to-end use case for Airport Visual Security with outcomes for cloud, the edge and fog.

Airport visual security, called surveillance, illustrates the complex, data-intensive demands required for real-time information collection, sharing, analysis, and action. First, let's look at the passenger's journey:

- Leaves from home and drives to the airport
- Parks in the long-term parking garage
- Takes bags to airport security checkpoint
- Bags are scanned and checked in
- Checks in through security and proceeds to boarding gate
- Upon arrival, retrieves bags
- Proceeds to rental car agency; leaves airport

This travel scenario is without incident. But when one or more threats are entered into this scenario, the visual security requirements become infinitely more complicated. For example:

- The vehicle entering the airport is stolen
- The passenger's name is on a no-fly list
- The passenger leaves his luggage unattended someplace in the airport
- The passenger's luggage doesn't arrive with the flight
- The luggage is scanned and loaded on the plane, but it is not picked up by the correct passenger.
- An imposter steals or switches a boarding pass with another passenger and gets on someone else's flight.
- The passenger takes someone else's luggage at the arrival terminal

Catching these possible threats requires an extensive network of surveillance cameras across the outbound and inbound airports, involving several thousand cameras. Approximately one terabyte of data per camera per day must be transmitted to security personnel or forwarded to local machines for scanning and analysis.

In addition, law enforcement will need data originating from multiple systems about the suspect passenger's trip, from the point of origination to arrival. Finally, all of the video and data must be integrated with a real-time threat assessment and remediation system.

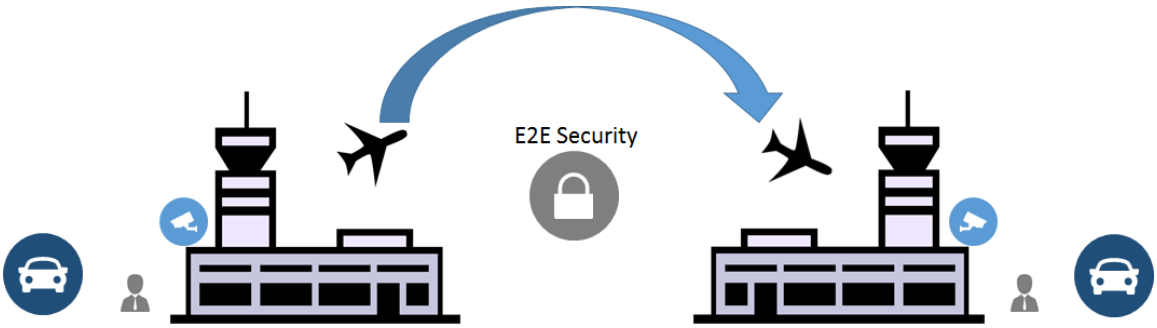

Figure: Airport scenario

Cloud and Edge Approaches. In an edge-to-cloud design, every camera (edge device) in the airport transmits directly to the cloud for processing, as well as the other relevant data collected from the passenger's travel records.

	Advantages	Disadvantages
Edge-to-Cloud Approach	• Store shared data in a common location • Historical analytics for threat prevention planning	• Latency (inability to process images and alert authorities with millisecond turnaround) • High cost of data transfer • Reliance on always available cloud
Edge-only Approach	• Low latency	• Limitations in sharing data and information across systems within the airport. • Limitations with sharing data between airports in near real time

While there are advantages to both approaches, the disadvantages can lead the systems susceptible to incidents.

Fog Computing Approach. The OpenFog pillars are present throughout the airport visual security in an end-to-end architecture. In this deployment scenario, some of the OpenFog RA pillars are emphasized:

- **Security**: The airport visual security scenario is a physically distributed fog deployment. Thus, physical possession is in scope for the security analysis. Transportation and storage of data must also be secure as much of the data which may contain personally identifiable information.
- **Scalability**: The OpenFog RA must adapt with the business needs as it relates to system cost and performance. When a new airport terminal, gate, or additional sensors and equipment are added, the solution must scale and not require a completely new deployment.
- **Open**: Openness is essential for the success of a ubiquitous fog computing ecosystem for IoT or 5G platforms and applications. Proprietary or single vendor solutions can result in limited supplier diversity, which can have a negative impact on system cost, quality and innovation.
- **RAS**: The various aspects of the solution must be reliable, available, and serviceable which includes orchestration of existing or new resources. As new object recognition models are trained for visual

analytics, these inference engine models should be updated on near edge devices without impacting availability of the solution.

- **Programmability**: Visual analytics is utilized to facilitate this scenario, requiring programming at the hardware level. For example, accelerators such as FPGAs could perform inference on images.

Adherence to OpenFog RA

The OpenFog Consortium intends to partner with standards development organizations and provide detailed requirements to facilitate a deeper level of interoperability. This will take time, as establishing new standards is a lengthy process. Prior to finalization of these detailed standards, the Consortium is laying the groundwork for component level interoperability and certification. Testbeds will prove the validity of the OpenFog RA through adherence to the architectural principles.

Next Steps

The OpenFog RA is the first step in creating industry standards for fog computing. It represents an industry commitment toward cooperative, open and inter-operative fog systems to accelerate advanced deployments in smart cities, smart energy, smart transportation, smart healthcare, smart manufacturing and more. Its eight pillars describe requirements to every part of the fog supply chain: component manufacturers, system vendors, software providers, application developers.

Looking forward, the OpenFog Consortium will publish additional details and guidance on this architecture, specify APIs for key interfaces, and work with standards organizations such as IEEE on recommended standards. The OpenFog technical community is working on a suite of follow-on specifications, testbeds which prove the architecture, and new use cases to enable component-level interoperability. Eventually, this work will lead to certification of industry elements and systems, based on compliance to the OpenFog RA.

For more information, please contact info@openfogconsortium.org.