The 8 Pillars of the OpenFog Reference Architecture

Introduction

Fog computing provides the missing link in the cloud-to-thing continuum that is necessary for mission-critical, data-dense use cases. 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, embedded Artificial Intelligence (AI), Virtual Reality and Tactile Internet applications.

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)*

This approach extends the traditional cloud-based computing model where implementations of the architecture can reside in multiple layers of a network's topology. Fog computing retains all the benefits of cloud computing, such as containerization, virtualization, orchestration, manageability, and efficiency.

The unique advantages of fog computing are termed 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

This document describes the 8 foundational architectural pillars, or core principles, of the OpenFog Reference Architecture. For a more indepth description of each pillar plus the architectural approach, use cases, glossary of terms and more, please download the full OpenFog Reference Architecture document at www.openfogconsortium.org/ra.

The OpenFog Reference Architecture

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 the OpenFog Consortium, an independently-run open membership ecosystem of industry, technology and university/research leaders. The intention of the OpenFog RA is to help business leaders, software developers, silicon architects and system designers create and maintain the hardware, software and system elements necessary for fog computing.

The OpenFog RA is driven by a set of 8 core principles called pillars, shown in Figure 1 below. 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, et al) along the cloud-to-thing continuum.

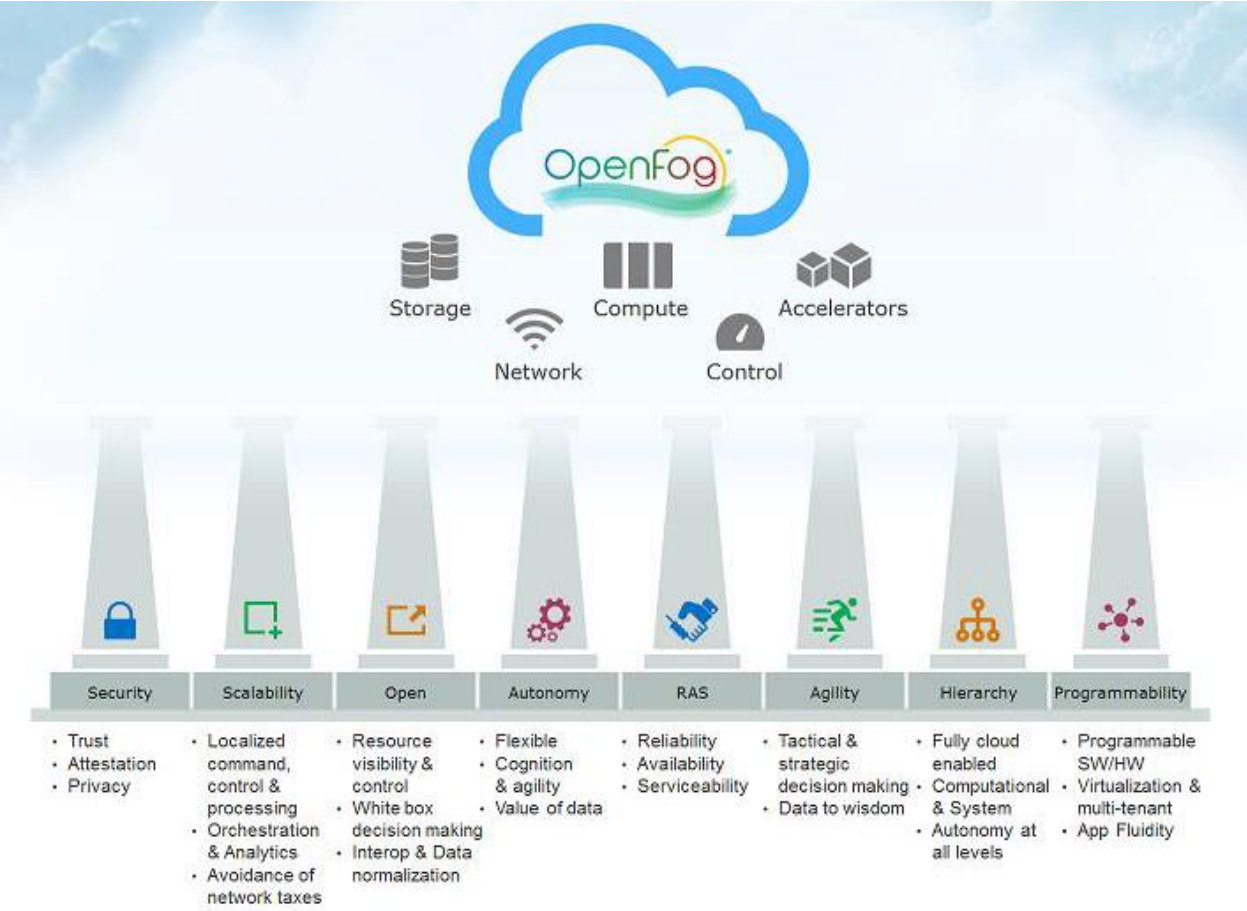

Figure 1: OpenFog Pillars

1. Security Pillar

- **Trust**
- **Attestation**
- **Privacy**

Many IoT, 5G and AI applications supported by the OpenFog RA have privacy critical, mission critical, and even life critical aspects. As such, any security compromise in the fog network can have severe consequences. The OpenFog RA will enable the flexible creation of computing environments that address a broad spectrum of security concerns spanning from IoT devices to cloud and the fog networks in between.

Security in the OpenFog RA is not a one-size-fits-all architecture. Rather, it describes all of the mechanisms that can be applied to make a fog node secure from silicon to software application. Business case, target market, and vertical use case, as well as the location of the node itself, will all create a set of requirements for that node. However, there are certain foundational parts of the architecture, which must be in place in order to build a secure execution environment.

Security implementations have many different descriptions and attributes such as privacy, anonymity, integrity, trust, attestation, verification, and measurement. These are key attributes for the OpenFog RA. Achieving the foundational elements for security requires an approach to discover, attest, and verify all smart and connected "things" before trust can be established.

Conformance to the OpenFog RA requirements will ensure that an OpenFog deployment will be built on a secure end-to-end compute environment. This includes the OpenFog node security, OpenFog network security, and OpenFog management and orchestration security. This will allow architects and designers to focus on the high-value security and privacy problems specific to the types of devices used in their application.

In many applications, particularly for brownfield deployments, or for tiny devices and sensors with little to no security capability, an OpenFog node may act as a device's first point of secure entry into an OpenFog compute hierarchy and the cloud.

The security pillar of the OpenFog RA starts with a clear definition of base building blocks. All fog nodes must employ a hardware-based immutable root of trust. The Hardware Root of Trust is a trusted hardware component which receives control at power-on. It then extends the chain of trust to other hardware, firmware, and software components. The Root of Trust should then be attestable by software agents running within and throughout the infrastructure. Because of the proximity to the edge, nodes in fog networks often act as the first node of access control and encryption.

This means they must provide contextual integrity and isolation, and control aggregation of privacy-sensitive data before it leaves the edge.

As more complex topologies are created in FaaS implementations, the attestation continues as a chain of trust from the fog node, to other fog nodes, and to the cloud. Since fog nodes may also be dynamically instantiated or torn down, hardware and software resources should be attestable. Components that are not attestable should not be fully allowed to participate in the fog node or may not be deemed to have fully trustworthy data.

2. Scalability Pillar

- **Localized command, control & Processing**
- **Orchestration & analytics**
- **Avoidance of network taxes**

The scalability pillar addresses the dynamic technical and business needs behind fog deployments. Elastic scalability cuts across all fog computing applications and verticals. The hierarchical properties of fog and its location at logical edges of networks add additional scaling opportunities:

- Individual fog nodes can scale internally, through the addition of hardware or software.
- Fog networks can scale up and out through the addition of fog nodes to assist heavily loaded nodes, either on the same level of the fog hierarchy or in adjacent levels.
- A network of fog nodes can be scaled up or down in a demand-driven elastic environment.
- Storage, network connectivity, and analytics services can scale with the fog infrastructure.

Because of the variability in the use cases for fog computing, the OpenFog RA enables elastic scaling of modest deployments through large mission critical deployments based on demand. This scalability is essential for fog computing implementations to adapt to workload, system cost, performance, and other changing business needs. Scalability may involve many dimensions in fog networks:

- **Scalable performance** enables growth of fog capabilities in response to application performance demands (e.g., reducing latency between sensor reading and resulting actuator responses).
- **Scalable capacity** allows fog networks to change in size as more applications, endpoints, "things," users, or objects are added or removed from the network. You can add capacity to individual fog nodes by adding hardware like processors, storage devices, or network interfaces. You can also add capacity through software and various pay-as-you-grow licensing. The converse is also true.
- **Scalable reliability** permits the inclusion of optional redundant fog capabilities to manage faults or overloads. Redundant fog nodes also ensure a large deployment's integrity and reliability at scale, which is part of the reliability, availability, and serviceability (RAS) pillar. There are hardware and software aspects to scalable reliability. The scalability mechanisms supporting the reliability of fog networks must themselves be highly reliable. Availability (which is a scaling measure closely related to reliability) scales through similar methods.
- **Scalable security** is often achieved through the addition of modules (hardware and software) to a basic fog node as its security needs become more stringent. Capabilities like scalable distribution, rights access, crypto processing capacity, and autonomous security features contribute to scalable security.
- **Scalable hardware** involves the ability to modify the configuration of the internal elements of fog nodes, as well as the numbers of and relationships between fog nodes in networks.
    - Processors scale from modest single core CPUs to specialized accelerator chips with thousands of cores or millions of gates.
    - Networking scales from a single wireless or wire line interface to large arrays of wireless, wire line, and fiber interfaces with aggregate capacities of many Gb/s.
    - Storage can scale from a simple flash chip to large arrays of flash / rotating disks and network attached file systems.

- **Scalable software** is also important and includes applications, infrastructure, and management. Orchestration, analytics and composability and modularity are all important aspects of OpenFog scalability.

3. Openness Pillar

- **Resource visibility & control**
- **White box decision making**
- **Interoperability & data normalization**

Openness is essential for the success of a ubiquitous fog computing ecosystem for IoT 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. The openness pillar importance is highlighted in our desire for fully interoperable systems, supported by a vibrant supplier ecosystem.

Openness as a foundational principle enables fog nodes to exist anywhere in a network and span networks. This openness enables pooling by discovery, which means that new software-defined fog nodes can be dynamically created to solve a business mission. The security pillar shares a common theme and requirements in openness characteristics

- **Composability** supports portability and fluidity of apps and services at instantiation. Additional emphasis of composability is visible in the programmability pillar.
- **Interoperability** leads to secure discovery of compute, network, and storage and enables fluidity and portability during execution. The marketplace has clearly articulated its desire for a

vibrant supplier ecosystem, with reasonable expectations that elements from one supplier can be freely substituted for elements from another supplier. This will be addressed through testbeds, FogFests (plug fest), standardization, and open implementations.

- **Open communication** enables features like pooling of resources near the edge of the network to collect idle processing power, storage capacity, sensing ability, and wireless connectivity. For example, a compute-intensive application developed in fog architecture can leverage hundreds of gigabytes sitting idle on nearby laptops, systems, and set-top boxes in a household every evening, or among the passengers of a public transit system. The open discovery of these nearby compute resources is critical. Doing the functional work nearest the edge avoids additional network taxes when moving up the stack towards the cloud. We define network taxes as the cost of transmission.
- **Location transparency** of any node instance to ensure that nodes can be deployed anywhere in the hierarchy. Location transparency provides an alternative to network operator control. This means that any IoT device, such as a smart watch, does need its own carrier-owned data plan. Each thing or software entity can observe its local conditions and make decisions on which network to join. Each endpoint in a fog network can optimize its path to the computational, networking and storage resources it needs (no matter if those resources are in the hierarchical layers of the fog, or in the cloud).

4. Autonomy Pillar

- **Flexible**
- **Cognition & agility**
- **Value of data**

The autonomy pillar enables fog nodes to continue to deliver the designed functionality in the face of the external service failures. In this architecture, autonomy is supported throughout the hierarchy. Decision making is made at all levels of a deployment's hierarchy including near the device or higher order layers. Centralized decision-making in the cloud is no longer the only option. Autonomy at the network edge means intelligence from local devices and peer data can be used to fulfill the business' mission at the point where it makes the most business sense.

The OpenFog RA supports autonomy for a wide range of functions. It does not rely upon centralized entity for operation (e.g., a backend cloud). Some of the typical areas for autonomy at the edge include:

- **Autonomy of discovery** to enable resource discovery and registration. For example, an IoT device coming online in the field would typically "phone home" first to let the backend cloud know it is alive and its associated functions are available. But when an uplink network to the cloud is unavailable, it can stop the device from going live. An autonomous fog node can potentially act as a proxy for the device registration, which then allows the device to come online without the backend cloud.
- **Autonomy of orchestration and management** (O&M) automates the process of bringing services online and managing them through the operational lifecycle and decommissioning. Autonomy of O&M entails a number of actions including: instantiation of services; provisioning the environment around the services, such as routing of data flows; and keeping track of the health and status of the resources. All these actions should be as automated as possible through programmability and policies. The architecture includes an autonomous and scalable O&M function that is set up to handle any surge of demand for resources, without real-time reliance on the cloud or the need for significant human labor.

- **Autonomy of security** enables devices and services to come online, authenticate themselves against a minimal set of fog security services, and perform their designed functions. In addition, these security services can store records for future audits. With autonomy, these actions can be performed where they are needed, when they are needed, and regardless of connectivity to the cloud. Fog nodes can autonomously react to evolving security threats, such as updating virus screening algorithms, determination of denial-of-service (DoS) attacks, etc. without administrator involvement.
- **Autonomy of operation** supports localized decision making by IoT systems. Sensors provide data, which is the basis for autonomous actions at the edge. If the cloud or a single place in the system's hierarchy is the only location where decisions can be made, this violates the ability to ensure reliability and as such, the architecture ensure operational autonomy.
- **Cost savings** is a key motivator for autonomy. Connectivity today costs money. The more data that is sent through the network, the higher the costs are for businesses due to network taxes. This drives the need for more processing at the edge of the network, with just-in-need and just-in-time data sent to the cloud as required for additional business insights. For example, when an oil rig generates 30,000 data points a second, not all of the data must be sent through an expensive satellite link. Local and fog domain analytics and pre-processing can autonomously filter out the unimportant data points and extract the more mission critical ones to be delivered to the next layer in the hierarchy.

5. RAS Pillar

- **Reliability**
- **Availability**
- **Serviceability**

Reliability, availability, and serviceability (RAS) is resident throughout successful system architectures and, as such, takes on great importance in the OpenFog RA. Hardware, software, and operations are the three main areas of the RAS pillar.

A **reliable** deployment will continue to deliver designed functionality under normal and adverse operating conditions. The reliability of the RAS pillar includes but is not limited to the following properties:

- Ensuring reliable operation of the underlying hardware upon which the software is operating, enabling reliable and resilient software and a reliable fog network, which is generally measured in uptime.
- Safeguarding the availability and integrity of data and compute on edge gateways using enhanced hardware, software, and network designs.
- Autonomous predictive and adaptive self-managing capabilities when required by the health of the system to initiate self-healing routines for hardware and software and upgrade new firmware/application and security patches.
- Increasing customer satisfaction by simplifying support and device self-optimization and healing.
- Initiating requests for preventative maintenance, including new hardware and software patches, network re-routing, etc.
- Testing and validation of system components, including device drivers and diagnostic tools under a variety of environmental conditions.
- Providing alarms, reports, logs, etc.
- Validation of system platforms and architectures through interoperability certification test suites.

Availability ensures continuous management and orchestration, which is usually measured in uptime. The availability of the RAS pillar includes but not limited by the following properties:

- Secure access at all levels of a fog hierarchy for orchestration, manageability, and control, which includes upgradeability, diagnostics and secure firmware modification.
- Fault isolation, fault syndrome detection, and machine learning to help improve Mean Time To Repair (MTTR) of a failed system to achieve higher availability.
- Concept of cloud based back-end support with availability of interfaces throughout the system.
  - Secure remote access from a plurality of devices (not just a single console).
  - Redundant/duplicate device (peer-to-peer) IoT platform.
  - Mesh access capabilities of end-point sensor/peering.
  - Remote boot capabilities of the platform.
  - Support for redundant configurations for persistent productivity.

Servicing a fog deployment ensures correct operation. Serviceability of the RAS pillar includes but is not limited by the following properties:

- Highly automated installation, upgrade, and repair to efficiently deploy fog computing at scale.
- Hardware or software can either autonomously heal or be serviced by the various manufacturers.
- Ease of use to accommodate maintenance.
- Serviceability of the system:
  - Hardware, software, applications, networking, and data.
  - Ease of access/swap-out of the hardware (component interoperability).
  - Ease of secure upgradeability of software, BIOS, and applications locally or remotely and in real time.
  - Replication of system configuration over cloud on replaced/swap-out systems.
- Support for redundant configurations for persistent productivity.

RAS is especially important for OpenFog RA deployments in harsh environmental conditions and remote locations. This is why aspects from RAS are found throughout the architecture.

6. Agility Pillar

- **Tactical & strategic decision making**
- **Data to wisdom**

The agility pillar addresses business operational decisions for an OpenFog RA deployment. It is not possible for humans alone to analyze the data generated at the scale predicted by IoT as the basis for rapid, sound business and operational decisions. The agility pillar focuses on transforming this volume of data into actionable insights. Agility also deals with the highly dynamic nature of fog deployments and the need to respond quickly to change.

Data generation by sensors and systems in an OpenFog RA deployment are turbulent, bursty, and are often created in huge volumes. Most importantly, data may not have context, which is created only when the data is collated, aggregated, and analyzed. The analysis of data can be executed at the cloud level, but this subjects the data to increasing levels of latency. The ideal approach is to make operational decisions as soon as data can be turned into a meaningful context.

The architecture enables the creation of context close to the data generation where it makes the most sense for a given scenario. More strategic, system-wide decisions and policy management can be made further up the layers in the fog hierarchy. This avoids network dependencies we termed as "network taxes" as described in other OpenFog RA pillars.

The OpenFog RA allows system developers to optimize the placement of their applications as decision making components.

7. Hierarchy Pillar

- **Fully cloud enabled**
- **Computational & system**
- **Autonomy at all levels**

Computational and system hierarchy is not required for all OpenFog architectures but it is still expressed in most deployments. The OpenFog architecture is complementary to traditional cloud architectures due in part to the OpenFog hierarchy pillar.

OpenFog RA computing resources can be seen as a logical hierarchy based on the functional requirements of an end-to-end IoT system. Depending on the scale and nature of the scenario being addressed, the hierarchy may be a network of smart and connected partitioned systems arranged in physical or logical layers, or it may collapse into a single physical system (scalability pillar).

Using building automation from smart cities as an example, a company that manages a single office complex may have the entire fog deployment located locally. A large commercial property management company may have distributed fog deployments at local and regional levels feeding information to centralized systems and services. Each fog node is autonomous (autonomy pillar) to ensure uninterrupted operations of the facility it manages.

The figure below shows a logical view of the IoT system from a computational perspective. Each layer in the hierarchy addresses a specific concern of the IoT system.

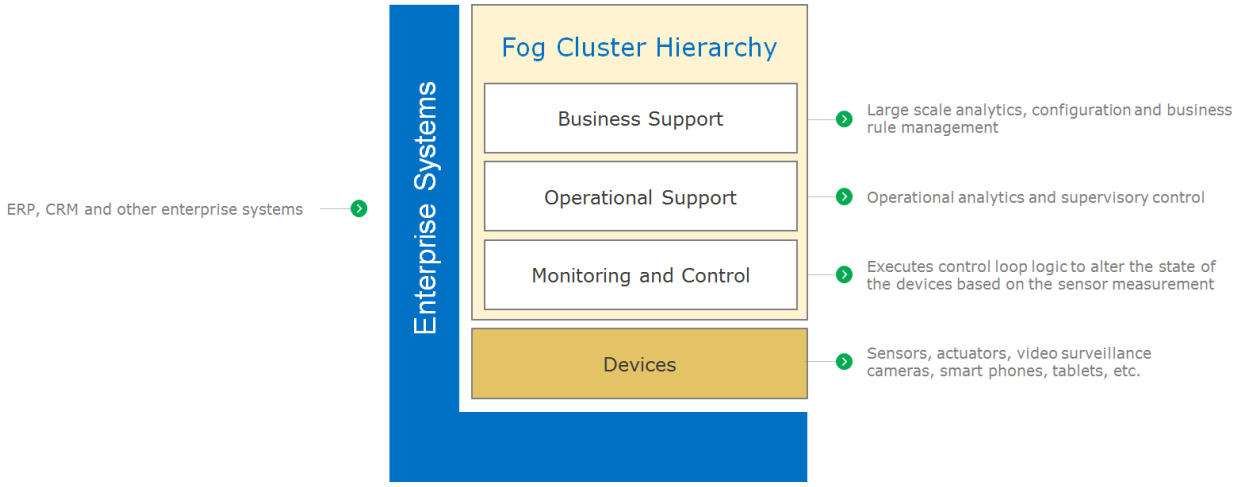

Figure 2: Layered architecture view of an IoT system.

Devices in the hierarchy: Sensors and actuator devices are the physical things and produce telemetry data to be consumed by the monitoring and control layer. This layer analyzes the telemetry and generates actuation commands if the process being monitored deviates from the desired state.

Monitoring and control in the hierarchy: Sensors and actuators are connected to microcontrollers that are programmed to monitor and control the state of the processes. A process state is represented by a set of parameters measured by the sensors and modified by the actuators. The main responsibility of this layer is to execute control logic through stateful inspection of the sensor telemetry. This involves computing alarms and generating events, which may trigger workflows through machine-to-machine or human intervention.

Operational support in the hierarchy: The operational support layer is responsible for analyzing streaming telemetry and storing operationally oriented analytics. The analytics may be presented through interfaces like control room dashboards and mobile applications. The scope of the analytics at this layer is narrow; it focuses on the operational aspects of the physical environment for which the system is responsible. This layer combines drill down historical analytics with streaming analytics for a composite picture of real-time operations with some short-term history.

Surrogacy in the hierarchy: The computation of a complex operation in the fog nodes may be delegated to the hierarchical nodes to leverage adjacent resources. Consider virtual reality tasks associated with a wearable such as smart glasses. Some of the information retrieval and computation tasks may be carried out on the glasses, while an associated element in the hierarchy (e.g., a smartphone) may handle its storage and connectivity requirements. This hierarchical architecture may leverage all of these devices at the same time, with an intelligent division of labor across them.

Business support in the hierarchy: The primary responsibility of this layer is to store and analyze the entire history of the IoT operations that span multiple systems. This is the system of record for IoT operations as governed by the compliance and record retention policies. Petabyte scale analytics will help in mining insights, business planning, comparing the operational efficiency of processes, operational optimization through training machine learning models, etc. Additionally, metadata and reference data management, business rule management, and the operational health of lower layers are the other aspects of this layer. These are also viewed in the agility pillar.

8. Programmability Pillar

- **Programmable software & hardware**
- **Virtualized & multi-tenant**
- **Application fluidity**

The programmability pillar enables highly adaptive deployments including support for programming at the software and hardware layers. This means that re-tasking a fog node or cluster of fog nodes for accommodating operational dynamics, can be completely automated. The re-tasking can be done with the help of the fog nodes inherent programmability interfaces which we describe using general purpose compute or accelerator interfaces. Programmability of a fog node includes the following benefits:

- **Adaptive infrastructure** for diverse IoT deployment scenarios and support changing business needs.
- **Resource efficient deployments** maximizing the resources by using a multitude of features including containerization. This increases the portability of components and is a key design goal enabled by programmability.
- **Multi-tenancy** to accommodate multiple tenants in a logically isolated runtime environment.

- **Economical operations** that results adaptive infrastructure to changing requirements.
- **Enhanced security** to automatically apply patches and respond more quickly to evolving threats.

To review the pillars in more detail, please download the full OpenFog Reference Architecture document at www.OpenFogConsortium.org/RA. If you have questions or comments about these pillars, please contact the OpenFog Consortium at info@OpenFogConsortium.org.

About the OpenFog Consortium

The OpenFog Consortium is a global nonprofit formed to accelerate the adoption of fog computing in order to solve the bandwidth, latency, communications and security challenges associated with IoT, 5G and artificial intelligence. Our work is centered around creating a framework for efficient and reliable networks and intelligent endpoints combined with identifiable, secure, and privacy-friendly information flows in the Cloud-to-Things continuum based on open standard technologies. For more information, please contact us at info@OpenFogConsortium.org.