

Master of Science HES-SO in Engineering
Av. de Provence 6
CH-1007 Lausanne

Master of Science HES-SO in Engineering

Orientation : Computer science (CS)

Cloud-Native IoT Reference Architecture with Arm SystemReady

Made by
Xavier Clivaz

Under the direction of
Prof. Fabien Vannel
Hepia/inIT/CoRES

Information about this report

Contact Information

Author: Xavier Clivaz
Master Student
HES-SO//Master
Switzerland
Email: xavierclivaz@hotmail.com

Declaration of honor

I, undersigned, Xavier Clivaz, hereby declare that the work submitted is the result of a personal work. I certify that I have not resorted to plagiarism or other forms of fraud. All sources of information used and the author quotes were clearly mentioned.

Place, date: Sion, February 3, 2024

Signature:

A handwritten signature in black ink, appearing to read "Xavier Clivaz".

Acknowledgements

This master's thesis would not have been written over five months without the support of a number of people. It's time to thank all these contributors, whether they are closely involved in the project or are far removed from the digital world.

- First of all, I would like to thank my family and especially my parents, without whom I wouldn't be able to fulfil all my dreams. So I'm always able to fulfil my potential in my daily life. It's the strength of spirit I've inherited that will always enable me to move forward despite the obstacles I encounter.
- Secondly, I would like to thank my supervisor, Mr Fabien Vannel, who was always at my complete disposal. He was able to provide me with invaluable advice throughout the work to achieve the desired result.
- Finally, I would like to thank everyone at [56K.Cloud](#). They are a great team, bringing joy and good humour to the organisation. What's more, they were able to contribute their knowledge when obstacles were encountered.

Abstract

Key words: Arm SystemReady, AWS, Cloud-Native, IoT, reference architecture

Contents

Acknowledgements	v
Abstract	vii
Contents	viii
List of Figures	x
List of Listings	x
1 Introduction	1
1.1 56K.Cloud	1
1.2 Problem	2
1.3 Objectives	3
1.4 Project plan	3
1.5 Research methodology	4
1.6 Structure of this report	5
2 Project methodology	7
2.1 Agile methodology	8
2.2 KanBan methodology	10
2.3 Project management tools	11
2.4 Training	13
3 Analysis	15
3.1 Definitions	16
3.2 Problem formulation	18
3.3 Literature search	19
3.4 History of the Cloud-Native approach	19
3.5 Integrating cloud computing with IoT embedded systems	21
3.6 IoT cloud platforms	25
3.7 Cloud infrastructure tools	27
3.8 Arm SystemReady	31
4 Design	33
4.1 Reference architecture	34
4.2 CI/CD pipeline	35
4.3 Applications	36
5 Implementation	41

Contents

5.1	Reference architecture	42
5.2	CI/CD pipeline	49
5.3	Applications	55
6	Validation	59
6.1	Section 1	60
7	Conclusions	61
7.1	Project summary	61
7.2	Comparison with the initial objectives	61
7.3	Encountered difficulties	61
7.4	Future perspectives	61
A	Plannings	63
A.1	Forward planning	63
A.2	Effective planning	65
	Bibliography	67
	Glossary	73
	Acronyms	75

List of Figures

1.1	56K.Cloud company logo [1]	1
1.2	The problem of linking the hardware to the cloud infrastructure [2]	2
1.3	Overview of objectives	3
1.4	Project plan	4
2.1	Sequential and iterative process [4]	8
2.2	Scrum roles	9
2.3	Scrum life cycle	10
2.4	Example of a virtual KanBan table [10]	11
2.5	Example of version management with two branches	11
2.6	Development and Operations (DevOps) method	12
2.7	AWS Certified Cloud Practitioner	14
3.1	Comparison of different service models [13]	17
3.2	Google trends (01.01.2006 until 22.05.2016) of term Cloud-Native [18]	19
3.3	Traditional server (left) and virtualisation (right)	20
3.4	Model for integrating cloud computing into an embedded system [28]	22
3.5	Cloud Customer Reference Architecture for IoT [30]	23
3.6	Comparison of IoT cloud platform providers (sources from providers)	27
3.7	Overview of AWS CDK [71] deployment [71]	28
3.8	Overview of Terraform deployment [73]	29
3.9	Overview of Pulumi deployment [74]	30
4.1	Reference architecture overview	34
4.2	CI/CD workflows overview	35
4.3	Applications overview	37
4.4	OS update application activity diagram	37
4.5	Certificate rotator application activity diagram	38
4.6	Led and button applications use case	39
5.1	Overview of Pulumi's integration into the reference architecture	42
5.2	Infrastructure deployment workflow	51
5.3	Provisioning setup workflow	52
5.4	OS image building workflow	52
5.5	Applications testing workflow	53
5.6	Applications building workflow	53
5.7	Applications deployment workflow	54
5.8	Infrastructure destruction workflow	55

List of Listings

1 | Introduction

For a long time, embedded systems remained isolated from each other. Now it's time to bring them together with the world of the [cloud](#). Hardware devices have always been limited in terms of resources. Designs have allowed them to evolve according to their use case. However, physical limits have been observed. This is how the [cloud computing](#) came into being, partly in response to these difficulties.

The [cloud](#) provides a huge number of advantages. It frees up hardware resources to optimise the operation of embedded systems. It is likely that these systems will not be able to perform complex calculations. The [cloud computing](#) has the capacity to do this. The storage problem is another constraint. The [cloud computing](#) offers almost unlimited storage. It doesn't stop there. It is capable of offering constant scalability to infrastructures. It is made up of different independent services that can be easily connected together. There is vertical scaling, which means increasing the size of a resource, and horizontal scaling, which means increasing the number of resources.

More and more engineers have decided to take an interest in this area in order to interconnect their embedded systems with the [cloud](#). This practice is an integral part of the [Internet of Things](#). Due to increasingly large projects, not to mention artificial intelligence, they want to use [cloud computing](#) environments in order to no longer be constrained by resources. To make it easier to use a [cloud infrastructure](#), companies have decided to offer [cloud](#) platforms comprising a multitude of services for all types of use ([AWS](#), Microsoft Azure, etc). It's also a way of letting engineers work on their business rather than worrying about maintaining and securing these infrastructures.

1.1 56K.Cloud

[56K.Cloud](#) is a company based in Sion (Valais) since 2018. It offers various services through the [cloud](#). It provides solutions for businesses that want to improve their processes while reducing their costs. It is also a consultancy firm that seeks to promote the [cloud](#) to customers who find it difficult to understand certain concepts in this rapidly expanding digital world. Collaborations with a number of partners are also an asset, enabling us to pool a wide range of skills to provide customers with a complete product. In order to share its vision of the [cloud](#) more widely, it has additional premises in Winterthur (Zurich).



Figure 1.1 [56K.Cloud](#) company logo [1]

The idea of a [cloud](#) solution is to move an on-premise infrastructure to the [cloud](#). It's a way of reducing costs by only paying for what you use (pay-as-you-go), and by having something that can be easily scalable. It's also a good way of increasing the speed of processes by using optimal resources for each application.

Chapter 1. Introduction

In terms of consultancy, [56K.Cloud](#) introduces [Cloud-Native](#) technology to customers who want to find out more. It also covers the subject of [DevOps](#) and containerisation.

1.2 Problem

While [56K.Cloud](#) mainly offers [cloud](#) solutions, this company also works in the [Internet of Things \(IoT\)](#) sector. Recently, it quickly noticed a major problem between embedded systems engineers and those from the world of [cloud](#). More and more engineers are asking for devices to be able to link easily and quickly to a [cloud infrastructure](#). However, there is currently no reference architecture that enables an infrastructure to be deployed and an embedded system to be provisioned directly.

This is a real problem because of the time needed and the difficulty for engineers to link hardware to a [cloud infrastructure](#). Because of the two distinct professions, a person coming from the [cloud](#) will need time to understand and set up an operating system capable of connecting to an infrastructure. And vice-versa, a person coming from the hardware world will need to familiarise themselves with a [cloud infrastructure](#), not to mention setting it up while using the best practices of the [Cloud-Native](#).

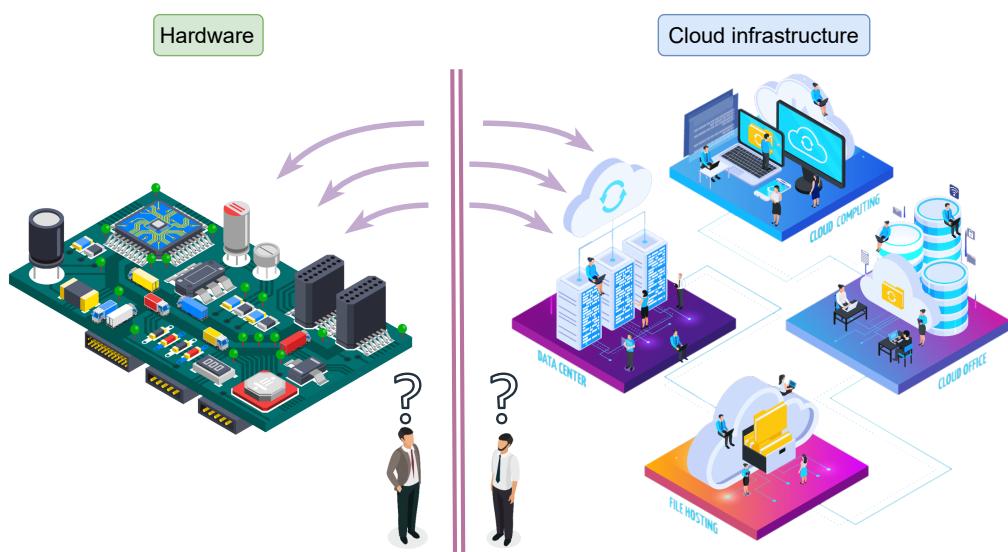


Figure 1.2 The problem of linking the hardware to the [cloud infrastructure](#) [2]

The need to find a solution arose when engineers started asking electronic board manufacturers to design equipment offering services for linking to a [cloud infrastructure](#). Since these manufacturers are not in this discipline, they do not want to spend time developing these types of solutions. It is therefore possible to identify the main players affected. There are the hardware engineers who are unable to meet the new needs of customers. Cloud engineers, who want to do the job for them, have to devote part of their time to introducing embedded concepts. Learning something new can quickly become time-consuming.

Nevertheless, a few solutions are gradually emerging. However, these remain very closed. There is no open source reference architecture capable of satisfying a wide range of products. These solutions therefore remain highly proprietary.

1.3 Objectives

The primary aim of this work is to develop a reference architecture enabling the deployment of a [cloud infrastructure](#) with the [provisioning](#) of a fleet of embedded systems. The idea is to be able to automatically provision devices to the infrastructure when they are first started up. The infrastructure must contain the essential minimum of components to make the architecture as universal as possible. Since [56K.Cloud](#) is a partner of [Arm](#), a processor manufacturer, the embedded systems must be equipped with this. It should be added that [Arm](#) puts certifications on their components such as SystemReady. SystemReady program guarantees that an operating system and the following software layers can function correctly in their processors. The embedded systems used in this project must be certified SystemReady. In addition, the infrastructure must be deployed on the [AWS](#) platform. This is the [cloud](#) provider that [56K.Cloud](#) works with.

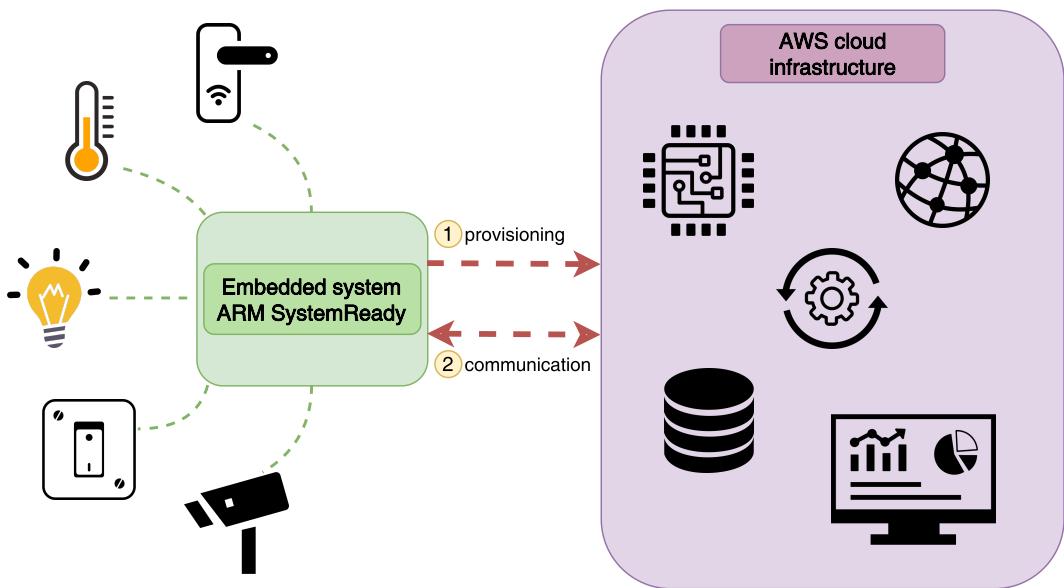


Figure 1.3 Overview of objectives

The aim of this project is to be open source so that a community can form around it. It should enable embedded systems and [cloud](#) engineers to use this reference architecture easily. They need to be able to focus on their final product. The best practices of [Cloud-Native](#) must be used for the development of this project. In order to guarantee a sound architecture, the use of [Continuous Integration \(CI\)](#) and [Continuous Delivery \(CD\)](#) tools are essential. In addition, to ensure compatibility, the architecture must be functional on different embedded systems with a [Arm](#) processor and certified SystemReady.

Secondly, once the reference architecture is functional, a demonstration project must be based on it. This is a proof of concept. The general idea is to deploy a [cloud infrastructure](#) on [AWS](#) and have various [cloud](#) services automatically set up to interact with the embedded system. Data will then transit between these two worlds and it must be viewable from an interface.

1.4 Project plan

Figure 1.4 shows the project plan in the form of [work packages](#).

Chapter 1. Introduction

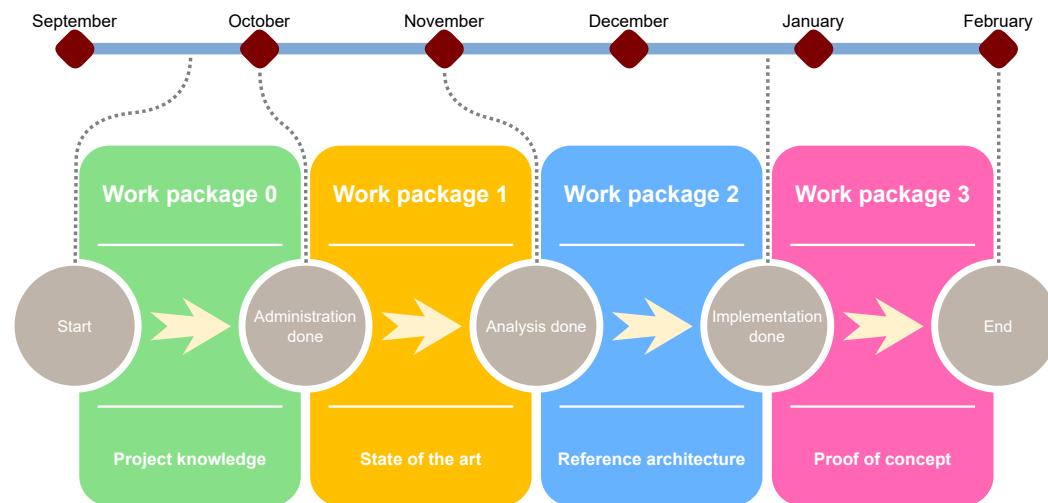


Figure 1.4 Project plan

The first [work package](#) contains all the administration. It involves learning about the project, being aware of the issues and clearly defining the specifications. This takes around two weeks.

About a month is spent researching existing solutions that are closest to the project. This involves drawing up a state of the art by researching scientific articles and other documentation.

The third [work package](#) concerns the reference architecture. Almost two months are devoted to this. This implementation part will be the final product, which will be delivered as open source on a shared repository.

In order to validate this architecture, a proof of concept will be carried out. This is the final [work package](#). It will take just over a month. Tests will be carried out to prove that it works properly.

1.5 Research methodology

The research methodology used in this thesis was first to draw up a state of the art on existing solutions that are as close as possible to this project. An overview of the reference architecture was then drawn up with a view to implementing it more easily. A choice of development tools was also made. At the same time, the implementation was designed. Finally, the results were validated by means of a proof of concept.

The agile methodology, known as [Scrum](#), was used throughout the project. It allows for iterative project management. Weekly meetings were arranged with the professor in charge of the project and, optionally, with some of the company's staff. To enable the project to progress efficiently, the [KanBan](#) method was used. This consists of virtual cards, each referring to a task. These cards are categorised in a table to determine the progress of each task. It's also a way of dividing the project into smaller parts so that time can be better estimated over the course of the project.

1.6 Structure of this report

The [2](#) chapter ([Project methodology](#)) contains a number of project methodologies for effective project management. The tools used are also described.

The [3](#) chapter ([Analysis](#)) contains several definitions and the state of the art. These are definitions of important terms that make up the project. The state of the art concerns the existing solutions that are closest to this project.

The [4](#) chapter ([Design](#)) contains an overview of the implementation as well as ...
To be completed

The [5](#) chapter ([Implementation](#)) contains ... To be completed

The [6](#) chapter ([Validation](#)) contains the analysis and validation of the results of the implementation.

The [7](#) chapter ([Conclusions](#)) concludes the thesis by outlining the current state of the research, the problems encountered and future steps.

2 | Project methodology

Project methodology consists of describing the means used to manage a project. In this chapter, the concept of agility is described in order to explain how the iterative process works. In addition, the [KanBan](#) methodology is described in detail to explain the time aspect of the tasks to be carried out. Finally, the various tools used are discussed.

Contents

2.1	Agile methodology	8
2.1.1	Scrum roles	9
2.1.2	Scrum process	9
2.2	KanBan methodology	10
2.3	Project management tools	11
2.3.1	Version management	11
2.3.2	Continuous Intergration (CI) and Continuous Delivery (CD)	12
2.3.3	Work packages	13
2.4	Training	13
2.4.1	AWS Certified Cloud Practitioner	13

2.1 Agile methodology

Agile methodology is the main project management method used in this work. It enables the project to be managed from the specifications to the final product. The idea for this method was conceived in the 70s or even before [3]. The idea was to change the development process in software engineering by using iterative techniques. Traditional methods worked with sequential techniques, often called Waterfall. Figure 2.1 shows the difference between these two approaches.

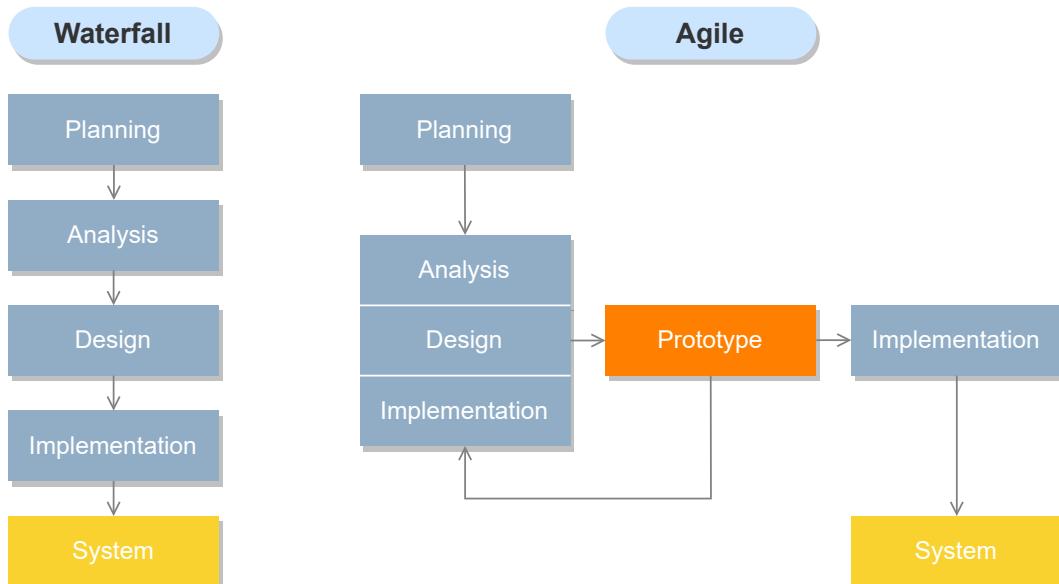


Figure 2.1 Sequential and iterative process [4]

The Waterfall methodology is considered cumbersome. There are deliverables after each phase. Approval is required before moving on to the next phase. It is difficult to go backwards (by moving up the phases). Professors explained this very well during a Masters course given at the HES-SO [4]. They added that this method is best used in very large projects where it is difficult to split teams into small groups to work iteratively.

These same teachers [4] described the agile methodology as follows:

"AGILE" is about values and principles, not practices, but many practices support them. It's not about doing agile, it's about being agile.

Several agile process frameworks were created before the 2000s. Examples include Scrum, XP, RUP, etc. Between the 11th and the 13th of February 2001, 17 people from these different frameworks met to find an alternative to cumbersome, documentation-driven software development processes. They created the agile software development manifesto [5] :

*Individuals and interactions over processes and tools
Working software over comprehensive documentation
Customer collaboration over contract negotiation
Responding to change over following a plan*

The central principle of agile is to have very rapid iterations to build the first business values. The customer is at the centre of the process thanks to frequent communication with the development team [4]. The process framework chosen for this project is **Scrum**. It is very well suited to small projects with small teams. It enables the first software prototypes to be delivered quickly. It adapts easily to changes, unlike a sequential method. It is also based on experience. The idea is to learn continuously through iterations. This is an important point in terms of learning from new developments in the project. This is one of the most popular agile methods. This approach was launched in 1995 by Ken Schwaber and Jeff Sutherland [6].

2.1.1 Scrum roles

Figure 2.2 shows the different roles in a **Scrum** team.

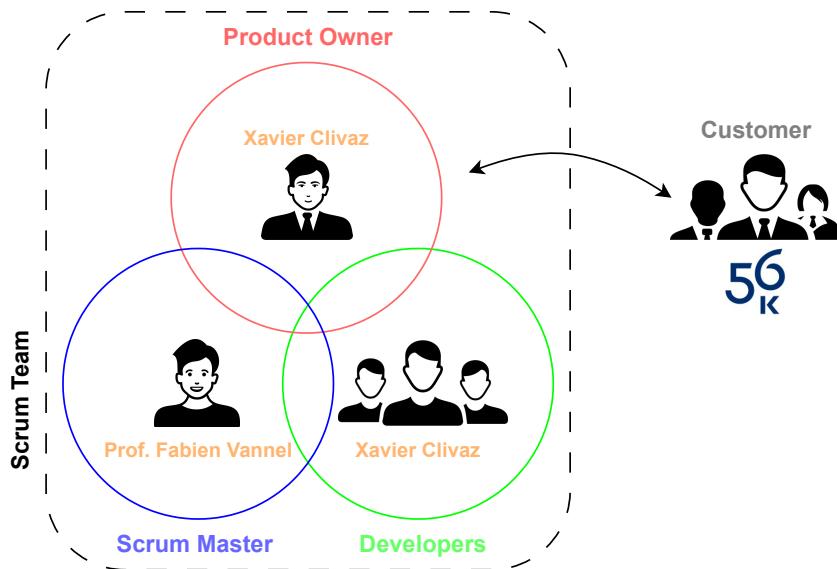


Figure 2.2 **Scrum** roles

A **Scrum** guide has been written by Jeff Sutherland and Ken Schwaber to explain the rules [7]. First of all, there is the **Product Owner (PO)**. This is the customer's spokesperson. It is he who defines the product's functionalities. They are recorded in a Product Backlog in the form of tasks. He must manage these tasks by prioritising some of them. He is also responsible for the value and return on investment (**ROI**) of the product. The development team is generally a small one (around seven people [4]). It is capable of self-management. There are no specific roles or positions. Developers have to deal with tasks that are chosen from the Sprint Backlog. Finally, the **Scrum Master** is required to be of service to the team. He accompanies the team of developers and prevents obstacles from getting in the way of progress. It is the **Scrum** Master who establishes the practices and rules of **Scrum**.

2.1.2 Scrum process

The **Scrum** process is relatively simple. It follows the theory described in the official **Scrum** guide [7]. The life cycle is shown in figure 2.3.

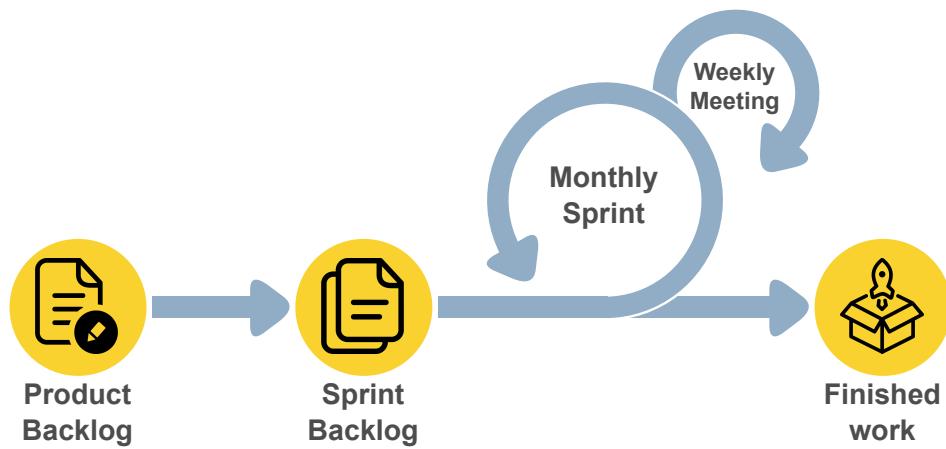


Figure 2.3 *Scrum* life cycle

The process is made up of multiple elements. It begins on the left of the diagram with the Product Backlog. This is a backlog that is filled by the PO. Since the Product Backlog is generally filled with many functionalities divided into tasks, only one part needs to be selected to be placed in the Sprint Backlog. A Sprint generally corresponds to a period of one month during which the tasks in the Sprint Backlog must be completed. During a Sprint, team meetings are organised every day. These are called Daily Meetings and usually last 15 minutes. A Daily Meeting ensures that, at least once a day, the entire team is available to get support on any problems encountered. In this project, only weekly meetings are held, depending on the distance separating the Scrum Master and the team of developers. Before each Sprint, a Sprint Planning is carried out to determine the tasks from the Product Backlog to be put into the Sprint Backlog. At the end of a Sprint, an inspection is made of how the Sprint went, with the aim of improving quality and efficiency. This is called a Sprint Retrospective. Finally, it's important to remember that at the end of each Sprint, one stage of the final product is completed. Sometimes this is called an increment.

2.2 KanBan methodology

The KanBan methodology is very well integrated into Scrum. Its aim is to manage the workflow as efficiently as possible. According to the KanBan guide produced by Scrum.org [8], several fundamental metrics should be taken into account for the flow:

- **Work In Progress (WIP)** : the number of tasks/items started and not completed.
- **Cycle Time** : the time elapsed between the start and end of a task.
- **Work Item Age** : the time elapsed between the start of the task and now.
- **Throughput** : the number of tasks completed per unit of time.

The average cycle time can be predicted using Little's Law [9]:

$$\text{average cycle time} = \frac{\text{average WIP}}{\text{average throughput}} \quad (2.1)$$

2.3 Project management tools

It explains that the more tasks there are in progress, the longer it will take to complete them. A cycle time in [KanBan](#) can be correlated to a Sprint in the [Scrum](#) methodology. Using this methodology, it is therefore possible to define the right number of tasks for each Sprint Planning. Note that the more you divide a feature into smaller tasks, the easier it will be to predict the cycle time.

[KanBan](#) does more than just limit the [WIP](#). It provides a transparent display of the workflow. Tasks can be represented in a table, as shown in figure 2.4. In this case, the table is virtual and can be viewed from anywhere by the whole [Scrum](#) team.

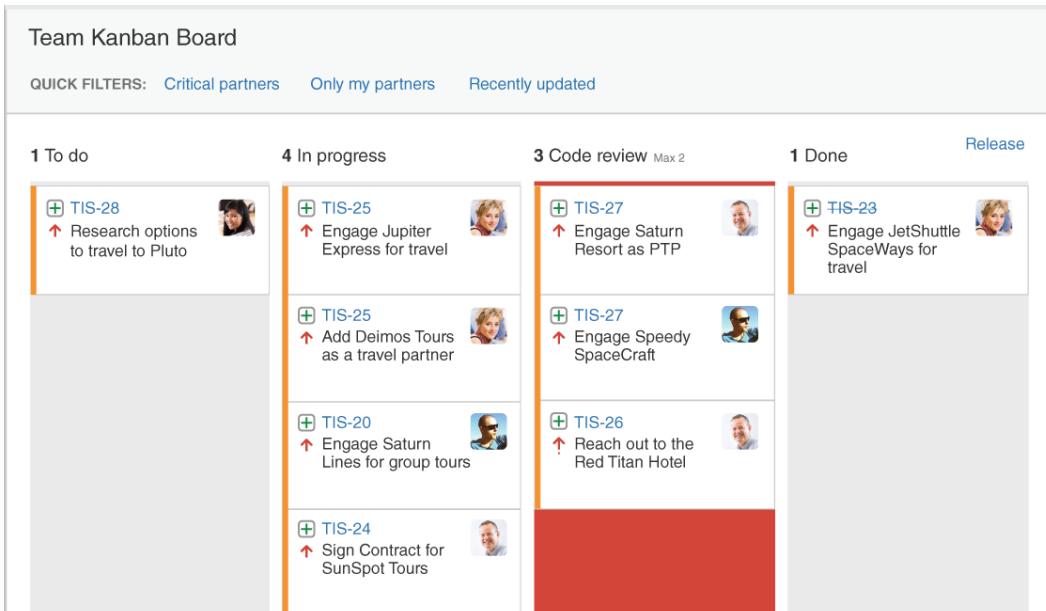


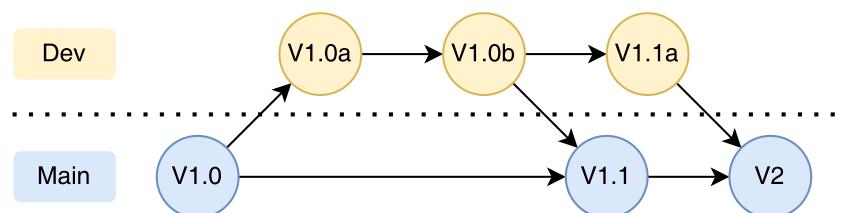
Figure 2.4 Example of a virtual [KanBan](#) table [10]

Each label in the table corresponds to a task. A task can contain a priority level and is often assigned to a member of the [Scrum](#) team. All this is defined during Sprint Planning. The labels are categorised in columns to show their progress.

2.3 Project management tools

2.3.1 Version management

Version management, embodied by tools such as Git, remains an essential part of the software development ecosystem. This practice offers a structured approach to tracking and maintaining changes to a project's code. One of the major benefits lies in its ability to keep a complete history of changes, facilitating collaboration between team members working on the same project.



Chapter 2. Project methodology

Figure 2.5 Example of version management with two branches

An example of a visual representation of a project is shown in figure 2.5, comprising several distinct versions and branches. Branches allow developers to work on alternative versions of the source code, facilitating parallel development, feature isolation and efficient change management without affecting the main branch.

When part of the development is considered complete, it becomes common practice to create a new version. This version freezes the current state of the project at a specific point in time, providing a clear and stable reference. This ability to create versions means that it is possible to return to previous states of the project if problems arise, whether identified by members of the development team or by external stakeholders, such as customers.

In short, version management, embodied by Git, transcends its simple function of tracking changes to become a central pillar of collaborative working, risk management and preserving the integrity of a software project. Its widespread adoption across the industry is testament to its continuing importance in delivering robust, scalable software development projects.

2.3.2 Continuous Intergration (CI) and Continuous Delivery (CD)

CI/CD (Continuous Intergration/Continuous Delivery) integrates synergistically with version management, and in particular with the DevOps framework, as illustrated in figure 2.6. The aim of the DevOps method is to automate software development processes, thereby encouraging collaboration between development teams (grey area) and operational teams (pink area). Two major components of CI/CD have emerged from this approach, namely Continuous Intergration (CI) and Continuous Delivery (CD), which have become fundamental pillars for developers.

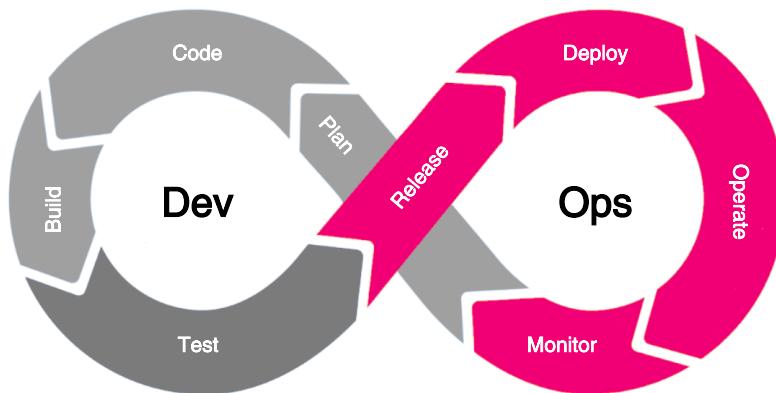


Figure 2.6 DevOps method

CI involves automating various processes with each new software release. These processes typically include building, testing and deploying the software design, as shown in figure 2.6. The main motivation for inventing the CI was to reduce the time needed between launching a project and bringing it to market, a concept known as time-to-market. This automation takes advantage of the constant evolution of machines, contrasting with

the physical and moral limits of human beings. The benefits of CI also extend to team collaboration, simplified debugging, reduced costs, improved software quality, and much more.

CD goes hand in hand with the CI, aimed at automating the deployment of software to production environments. It eliminates the tedious manual steps associated with deployment, ensuring fast, reliable and frequent delivery of new features and software updates.

CI/CD represents a modern and essential approach to the software development lifecycle, offering efficiency gains, improved team collaboration, accelerated time-to-market and an overall improvement in the quality of software products.

2.3.3 Work packages

Work packages are specific, delimited elements of a project which are assigned to a team or an individual for completion. Each work package represents a well-defined task or set of tasks, often associated with concrete deliverables. They facilitate the management and monitoring of the project by allowing a clear division of responsibilities and helping to assess progress. In short, work packages are manageable units that contribute to the achievement of an overall project.

2.4 Training

2.4.1 AWS Certified Cloud Practitioner

AWS (Amazon Web Services) offers a number of certifications designed to validate the skills and knowledge of IT professionals working on the AWS platform. These certifications are designed to cover a variety of areas, from cloud architecture and operations management to security, databases, application development and more.

These certifications are widely recognised in the industry and can help professionals demonstrate their skills and progress in their cloud careers. It is important to note that the AWS certification programme is evolving, and new certifications may be added as new services and features are introduced to the AWS platform.

The certification achieved as part of this project was as follows : **AWS Certified Cloud Practitioner**. This is the first certification to give a general overview of a cloud platform to professionals with no previous experience in this field. AWS defines it as follows :

The AWS Certified Cloud Practitioner validates foundational, high-level understanding of AWS Cloud, services, and terminology. This is a good starting point on the AWS Certification journey for individuals with no prior IT or cloud experience switching to a cloud career or for line-of-business employees looking for foundational cloud literacy. [1]

Online lessons were taken at the start of the project and, within a month, an official exam was successfully passed.

Chapter 2. Project methodology



Figure 2.7 AWS Certified Cloud Practitioner

3 | Analysis

The analysis begins by explaining the fundamental concepts that are being worked on throughout the project. Next, a state of the art is provided on the various tools, methods and technologies used in [cloud computing](#) and [IoT](#).

Contents

3.1	Definitions	16
3.1.1	Reference architecture	16
3.1.2	Cloud computing	16
3.1.3	Cloud-Native	17
3.1.4	Internet of Things	18
3.2	Problem formulation	18
3.3	Literature search	19
3.4	History of the Cloud-Native approach	19
3.4.1	The virtualisation	20
3.4.2	The hybrid	20
3.4.3	All to the cloud	20
3.4.4	The Cloud-Native approach	21
3.5	Integrating cloud computing with IoT embedded systems	21
3.5.1	Overview	22
3.5.2	Reference architecture	22
3.5.3	Existing problems and solutions linked to integration	23
3.6	IoT cloud platforms	25
3.7	Cloud infrastructure tools	27
3.7.1	Integrated IaC tools	28
3.7.2	Universal IaC tools	29
3.7.3	Comparison	30
3.8	Arm SystemReady	31
3.8.1	SystemReady certifications	32

3.1 Definitions

Before going into more detail on the subject of this work, it is important to look at the definitions of the various terms frequently used.

3.1.1 Reference architecture

A reference architecture is a solution model in a specific domain. It must be built so that an architecture can be established on its foundations, to make the task of software developers easier.

The aim is to generalise a solution that shows how it works from an overview. It must be possible to observe the relationships and their interactions between the multiple components of the application based on the reference architecture. There are different layers of abstraction depending on the field of application. A high level of abstraction means that it will be possible to understand the solution through more abstract elements such as the general components that will consolidate the application, for example. A low level of abstraction means that the solution will be more precise and certainly more specific to a use case. It will be possible to find detailed relationships between the services contained in the general components.

3.1.2 Cloud computing

The National Institute of Standards and Technology (NIST) [12] has standardised the definition of [cloud computing](#) as follows:

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. [12]

It is added in the article [12] that this model of [cloud](#) consists of three service models and four deployment models. The service models are as follows:

- **[Infrastructure as a Service \(IaaS\)](#)** : The capability offered to the consumer consists of providing processing, storage, network and other fundamental computing resources where the consumer can deploy and run arbitrary software, which may include operating systems and applications. The consumer manages the operating systems, storage and applications.
- **[Platform as a Service \(PaaS\)](#)** : The capacity provided to the consumer consists of deploying on the [cloud infrastructure](#) applications created or acquired by the consumer using programming languages, libraries, services and tools supported by the provider. The consumer manages only the applications and storage.
- **[Software as a Service \(SaaS\)](#)** : The consumer has the option of using the supplier's applications running on a [cloud infrastructure](#). The consumer manages nothing of the infrastructure apart from any configuration of the applications.

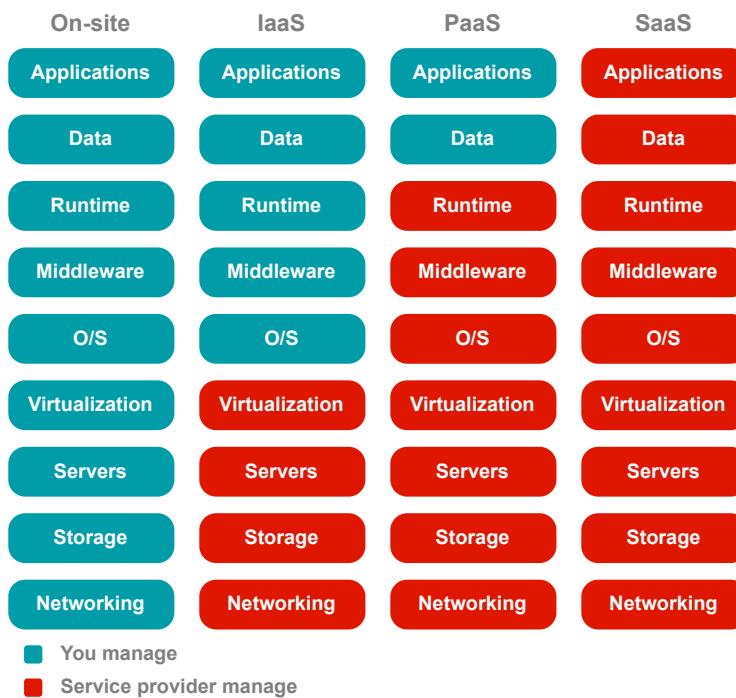


Figure 3.1 Comparison of different service models [13]

The deployment models are as follows [12] :

- **Private cloud** : The **cloud infrastructure** is made available for exclusive use by a single organisation comprising multiple consumers. It may be owned, managed and operated by the organisation, by a third party or by a combination of both, and it may exist on the organisation's premises or off-site.
- **Community cloud** : The **cloud infrastructure** is reserved for the exclusive use of a specific community of consumers from organisations with common concerns. It may be owned, managed and operated by one or more organisations within the community, by a third party or by a combination of such organisations, and it may exist on or off premises.
- **Public cloud** : The **cloud infrastructure** is made available to the general public for open use. It may be owned, managed and operated by a business, educational institution or government organisation, or a combination of these. It is located on the provider's premises.
- **Hybrid cloud** : The **cloud infrastructure** is a composition of two or more distinct infrastructures (private, community or public) which remain unique entities, but which are linked by a standardised or proprietary technology that allows the portability of data and applications.

3.1.3 Cloud-Native

Cloud-Native is an approach to software development which aims to design, implement and manage applications in the **cloud**. **Cloud computing** environments will be necessary for the proper execution of workloads.

Chapter 3. Analysis

The infrastructure can be deployed in private, public or hybrid [clouds](#). This approach will enable the development of modern, easily scalable applications. The flexibility aspect is also emphasised thanks to techniques for decoupling the multiple services of an application. Another strength of the [cloud](#) is its reliability and sustainability. Since the number of resources is incalculable and they are distributed around the world, redundancy guarantees security in the event of an incident, with almost instantaneous migration of an application's execution. The customer is therefore very often spared any unforeseen events. It is also possible to provide updates in real time and on a recurring basis.

The main benefits are high efficiency, reduced costs and high availability. Applications will be able to exploit resources that are optimal for their use case. The term pay-as-you-go is widely used, as the cost will depend solely on the use made of the resources and not on anything else, such as their maintenance, hardware security, etc. Availability is high because of the astronomical number of resources made available by the various [cloud](#) environment providers.

3.1.4 Internet of Things

[Internet of Things \(IoT\)](#) refers to the interconnection between physical objects and the internet. Objects can range from light bulbs to medical devices and much more. The main areas concerned are home automation and medical technology.

This involves linking different objects and applications to move into an automated world. As this becomes more and more widespread, the services needed to make it work need a huge amount of resources. This has been made possible by the [cloud](#). Everything must be accessible from anywhere in the world, quickly and securely. Several communication technologies are possible to guarantee excellent accessibility and reliability.

3.2 Problem formulation

The general question for analysis is: "Which technologies should be used to develop a reference architecture".

To help answer this question, it is important to consider the following aspects:

- Technology
- Environment
- Tools
- Certification

Since the development uses the [Cloud-Native](#) approach, its history can be described first. The integration of the [IoT](#) world into [cloud computing](#) must then be observed through various scientific works and research. A study of [cloud](#) service providers must be undertaken to differentiate the [iot](#) services offered by each one. As the reference architecture is based on an [Infrastructure as Code \(IaC\)](#), research into [IaC](#) tools is carried out. Finally, it is worth finding out whether SystemReady certification exists for electronic boards using an [Arm](#) processor architecture.

3.3 Literature search

Literature search has focused on academic search engines. The following search engines were used :

- [Google Scholar](#)
- [IEEE Xplore](#)
- [ScienceDirect](#)

The sources were mainly conference papers. Some information was found on websites. The following keywords were used to find sources that met the requirements of this analysis:

- integration
- embedded systems
- IoT
- Cloud-Native
- cloud
- cloud computing
- cloud providers
- cloud platforms
- Infrastructure as Code
- IaC tools
- Arm SystemReady

3.4 History of the [Cloud-Native](#) approach

[Cloud-Native](#) is a term that has been around for several years. Figure 3.2 shows the evolution of this term since 2006. The boom took place around 2016. It is undoubtedly due to the birth of Docker (2013) [14] and Kubernetes (2014) [15]. In 2015, the Cloud Native Computing Foundation [16] was created with the aim of making the [Cloud-Native](#) approach ubiquitous [17].

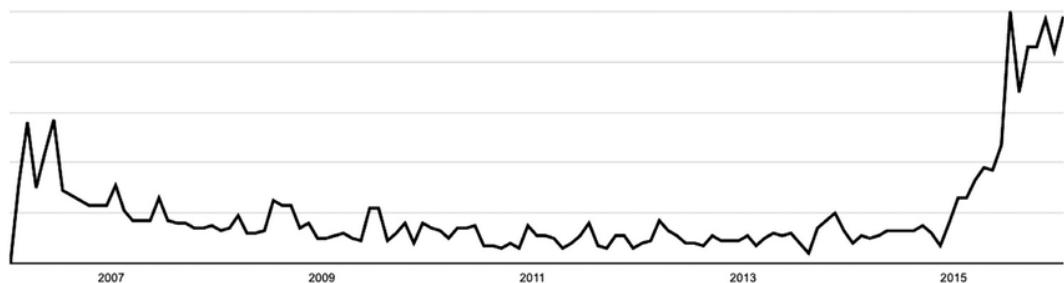


Figure 3.2 Google trends (01.01.2006 until 22.05.2016) of term [Cloud-Native](#) [18]

Before all this, there were only on-site data centres. Generally speaking, each company had its own servers running in its own data centre. This meant that servers were always set up for a specific application. [19]

3.4.1 The virtualisation

The first change came in the 2000s with the virtualisation of servers, although this has been around since the 1960s. When a new application had to be developed, new physical servers had to be bought. With virtualisation, it is no longer necessary to buy new ones. It is possible to run several applications on the same server using virtualisation, which limits the hardware resources for each application (figure 3.3). [19]

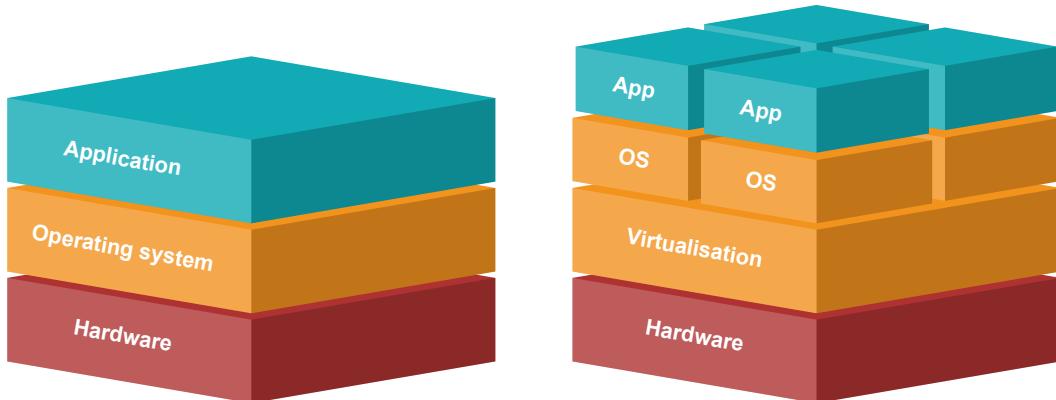


Figure 3.3 Traditional server (left) and virtualisation (right)

3.4.2 The hybrid

Despite this, a minimum of one server had to be purchased for an application to work. There were also resource limits for a certain number of applications. Unfortunately, not everyone could afford to buy new servers. In 2006, [AWS](#) launched three web services, [Elastic Compute Cloud \(EC2\)](#), Simple Storage Service (S3) and Simple Queue Service (SQS), to allow organisations and individuals to use Amazon's IT infrastructure on an as-needed basis at basic prices [20]. This initiative stems from a restructuring of the Amazon platform for better scalability and this has made it possible to sell virtual servers as a service [21]. By extension, the myth of making money from servers not in use for the majority of the year has been debunked by Benjamin Black, co-founder of the [EC2](#) [21]. This is where the world of [cloud](#) really began. Kratzke and Quint confirm this in their scientific article [18]. The purpose of the [EC2](#) service, which is still used today, is to offer a virtual machine in the [cloud](#). This is a [cloud computing](#) environment for running workloads. It is now possible to migrate your infrastructure from the database to the [cloud](#). The corresponding expression is "Lift and Shift" [22]. These were undoubtedly the first so-called [Cloud-Native](#) approaches [19]. However, this term only appeared in papers for the first time in 2012 [18]. These were two conference papers proposing solutions for [Cloud-Native](#) applications [23, 24]. Organisations have learned a lot about only paying for what you use. From there, there's a move to hybrid where companies are still using their on-premises servers alongside servers in the [cloud](#). [19]

3.4.3 All to the [cloud](#)

A new barrier has now been crossed. It's no longer a question of doing hybrid, but of transferring the entire infrastructure to the [cloud](#). This was made possible when [cloud](#) service providers such as [AWS](#), Microsoft Azure, Google Cloud and many others developed several services for all types of use. The reason there are different services is quite simply to decouple application functionality as much as possible. This is known

3.5 Integrating cloud computing with IoT embedded systems

as a microservices architecture. The risk of an entire application being interrupted in the event of a problem is much less likely. However, some organisations still have applications that are several decades old and have a monolithic architecture. It is not necessarily possible to decouple functionalities. That's why they work on a hybrid basis. [19]

3.4.4 The Cloud-Native approach

If we ask several engineers today about the definition of the **Cloud-Native** approach, we will find varying interpretations. There are currently three main schools of thought on this subject [19]. The first group considers that an approach is **Cloud-Native** when all the workloads run in the **cloud**. A minority within this group argue that you can be partially **Cloud-Native**, i.e. have one complete application deployed in the **cloud** while maintaining another application locally. However, others feel that this is still a hybrid approach. A second group says that to be truly **Cloud-Native**, you need to fully exploit the capabilities offered by the **cloud**. This means not only using the basic services of a **cloud** provider, but also making full use of the advanced features on offer, such as serverless functions. Some even consider that a company has to be born in the **cloud** to be truly **Cloud-Native**, a phenomenon that is increasingly being observed, with companies launching their first applications directly in the **cloud** without ever having used on-premises servers. This is the case for the company **56K.Cloud** **To be confirmed**. [19]

In 2018, a brief history of **cloud** application architectures described the term **Cloud-Native** as follows :

Cloud infrastructures (IaaS) and platforms (PaaS) are built to be elastic. Elasticity is understood as the degree to which a system adapts to workload changes by provisioning and de-provisioning resources automatically. Without this, cloud computing is very often not reasonable from an economic point of view. Over time, system engineers learned to understand this elasticity options of modern cloud environments better. Eventually, systems were designed for such elastic cloud infrastructures, which increased the utilization rates of underlying computing infrastructures via new deployment and design approaches like containers, microservices or serverless architectures. This design intention is often expressed using the term Cloud-Native. [25]

3.5 Integrating cloud computing with IoT embedded systems

With technology advancing at a rapid pace, today's world appreciates help in making everyday life better [26]. Embedded systems that have become increasingly connected now offer this service. However, since they need to attend any location for any domain and task, they need to be small for better integration. As a result, resources are limited for computing operations, data storage, data processing and so on. The built-in security and low energy consumption that these devices must also ensure are not enough to allow them to move forward. This is why **cloud computing** needs to be integrated to fill the gaps left by **IoT** [26]. Cloud of Things is one of the paradigms given to the combination of these two technologies in 2014 [27].

3.5.1 Overview

A model for integrating [cloud computing](#) into an embedded system is shown in figure 3.4. [28]

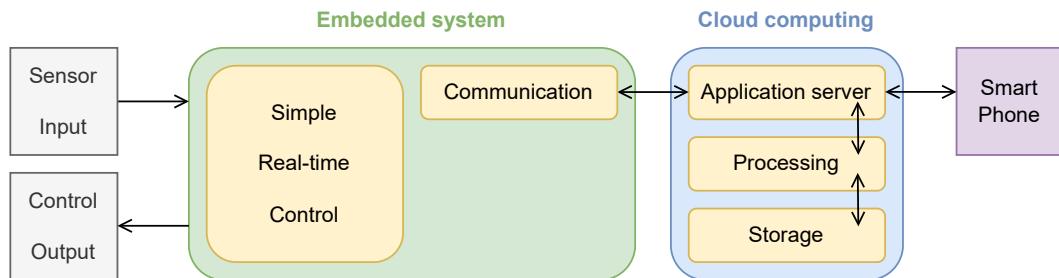


Figure 3.4 Model for integrating [cloud computing](#) into an embedded system [28]

The main idea is to use only what is necessary on the [IoT](#) device. In other words, all computing and storage functions should be moved to the [cloud](#). The device should only be linked to the input and output peripherals using a simple controller and establish communication with the [cloud](#). For the integration to work, the embedded system must have an internet connection. [Cloud computing](#) would contain the core of the application as well as various services to perform analysis, processing and calculation operations, and store data in real time. From there, it would also be possible to view the data from a smartphone. This approach is a solution that was thought up by Furuichi and Yamada in 2014 [29]. It offers a number of advantages, such as reduced energy consumption by eliminating large workloads and reduced size by freeing up electronic components. The [cloud](#) also makes it easy to scale up if necessary. The application can then be highly scalable. [28]

Furuichi and Yamada wanted to use a project to prove that their suggested approach worked. They developed a traffic jam detection system that recognises car licence plate numbers along with their location and time. For a plate number to be recognised from a raw image, image processing is carried out using machine learning. They found that this application, split between an [IoT](#) device and a [cloud computing](#) environment, scored much better than the same application managed entirely on a laptop or embedded system. The evaluation looked at cost, battery life, performance, scalability and reliability. [29]

3.5.2 Reference architecture

A reference architecture has been developed to support [IoT](#) objects in [cloud computing](#) (figure 3.5). The Standards Development Organization has decided to make this a standard by launching the Cloud Standards Customer Council programme to drive forward the adoption of [cloud computing](#). In this architecture, various aspects are taken into account: scalability, security, reliability and protection of privacy. [30]

3.5 Integrating cloud computing with IoT embedded systems

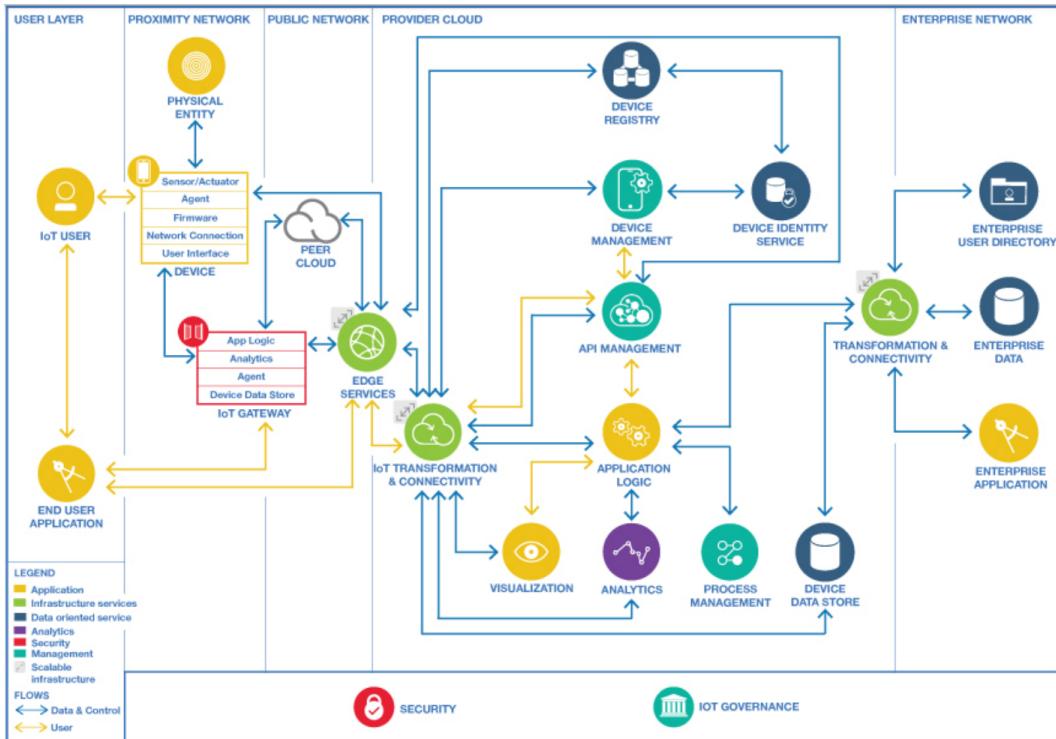


Figure 3.5 Cloud Customer Reference Architecture for IoT [30]

The architecture proposed in figure 3.5 is made up of components and their relationships. They are separated under a three-tier architecture model, called 3-Tier (presentation, logic, data). There is edge-tier, the platform tier and the enterprise tier.

The edge-tier contains the "Proximity Network" and "Public Network" parts. It represents the place where data is collected. It is collected from an IoT device in a proximity network. The data will either be sent to an IoT gateway or directly to the cloud provider.

The cloud provider tier manages the entire logical part of the application. It includes the collection, processing and analysis of data flows. Each device is also managed in this tier, as is its identity, for example.

The enterprise tier contains the "Enterprise Network" part. It includes the enterprise data and the application. Data from cloud computing can be stored in enterprise data.

3.5.3 Existing problems and solutions linked to integration

Integrating IoT and cloud computing poses a number of challenges. They are to be found both on the hardware side and in the cloud. [27]

There are many communication protocols. There are protocols for every type of use. On the IoT side, for example, there are Zigbee [31], Bluetooth and **Bluetooth Low Energy (BLE)** [32], LoRaWAN (long range) [33], WiFi, Thread [34] and the latest Matter protocol [35]. In an IoT network, there can be a multitude of sensors linked to a gateway. Interoperability is a problem in this case. Since 2022, version 1.0 of the Matter protocol has been published to address this problem. However, few devices are yet compatible. In research projects, Zigbee is very popular [26, 27, 36, 37]. Two main technologies are used

Chapter 3. Analysis

for communication between the IoT gateway and the cloud: **Message Queuing Telemetry Transport (MQTT)** [38] and **Constrained Application Protocol (CoAP)** [39]. Several projects have used CoAP for resource reasons [40, 41, 42]. It is a protocol dedicated to constrained devices with small amounts of memory [39], using UDP. It is also suitable for devices that need to bind to web services as it uses the **REpresentational State Transfer (REST)** model. REST is an architectural style for web services [40]. It would have been possible to use the **Service-Oriented Architecture (SOA)** style, but it is less suited to this use case [40]. Service-based architecture styles allow a high degree of decoupling from the application. This is an advantage in the event of a problem with a service, so as not to completely freeze the application. Finally, MQTT is specially designed for IoT [38]. It uses the TCP transport protocol. Cloud provider AWS adopts the MQTT protocol in its IoT sector [43]. A communication between an Arm microcontroller and AWS proved that this protocol works [42]. In fact, the data passing through is often in **JavaScript Object Notation (JSON)** format [28, 37, 44]. This is a lightweight data representation syntax for storing and exchanging textual information [37]. There is also the **HyperText Transfer Protocol (HTTP)** communication protocol. It would be suitable for updating IoT devices. It allows more bytes to be sent per packet, which increases the update speed. Furthermore, in industry, HTTP is generally less constrained by firewalls than MQTT or CoAP [45].

A lot of energy is consumed in IoT devices. The reason for this is the increasing amount of data being transferred to the cloud. Energy needs to be managed efficiently, for example by setting a sleep mode. If this is not possible, natural energy could be used to provide power, such as solar, wind or vibration. [27]

Resource allocation is another debate. Each IoT object may have a different function. They do not necessarily require the same resources in the cloud. What's more, unexpected events could occur requiring a greater workload. The requirements could not necessarily be predictable. Even if they were, the cost of increasing unused resources for most of the time would be a loss. Cloud service providers offer a number of solutions to this problem. For example, AWS offers a service that balances the load in the cloud computing environment [46]. The combinatorial auction approach is popular for resource allocation in the cloud [47]. It must satisfy **Quality of Service (QoS)** constraints and maximise the cloud provider's profit. Provider and user profit, resource utilisation and quality of service are the dominant performance factors. Research has succeeded in optimising this approach [47]. Nevertheless, as soon as a new IoT node is added, a sample of data could be sent to the cloud to set up the minimum resources required. [27]

Identity management in an internet network is very important to ensure that each device is unique. Given the number of IoT devices today, IPv6 addressing would be the optimal solution. It would be sufficient to support this type of network. However, deployment remains problematic. The coexistence of IPv6 with IPv4 is not so easy [27]. There are three possible techniques for cooperation between the two [48]. The IoT device may support both types of addressing. Alternatively, IPv6 packets must be encapsulated in IPv4 packets, or a **Network Address Translation (NAT)** must be available that translates both types of packet.

The discovery of new IoT devices to be integrated into the cloud needs to be managed. At any time, a device can be part of a service and leave it. It is also necessary to monitor its status once it is connected and keep it up to date. AWS has developed a solution to this which must be installed on IoT devices. From there, the devices can be easily integrated. In addition, it is possible to perform Over-The-Air (OTA) updates [49] and monitor the state of devices [50]. Azure also provides these different services [51, 52]. An update method was designed in this area in 2022 [45]. It is based on the standard update architecture for IoT [53], proposed by the Internet Engineering Task Force (IETF).

Cloud computing environments must respond correctly to data flows, whether large or small. The urgency with which certain data is transmitted is also an integral part of this QoS. QoS is assessed in terms of bandwidth, delay, packet loss rate and delay in the transmission of data packets (jitter). [27]

Another point to consider is where the data is stored. Depending on the volume of data, it is best to store it in the physical location closest to the user. Sensitive data must also be managed in accordance with each country's data protection laws. Fortunately, cloud providers offer the option of choosing the region where data is stored. [27]

Security seems to be an issue in integration. In 2016, there was a lack of trust in cloud service providers and a lack of awareness of Service-Level Agreements (SLAs) [54]. A Service-Level Agreement is a document defining the provision of cloud services and the responsibilities of the provider and the customer. 56K.Cloud claims some negligence with developers. However, when data is transferred to the cloud, the location of databases is not always transparent. Distributing data over several locations, while ensuring high availability, also increases the chances of sensitive data being leaked. A distributed system of this kind is exposed to various potential attacks, such as SQL injections, cross-site scripting attacks and many others. Significant vulnerabilities can also be exposed, including session hijacking and virtual machine evasion. In addition, the computing power constraints associated with connected objects limit the application of public key cryptography to all layers of the system. Two years later, in 2018, in an attempt to partially address this problem, two security models have been proposed using two encryption algorithms (AES and RSA) [36]. They can be used to integrate IoT and cloud computing. However, they remain sub-optimal and could be the subject of future research.

Unnecessary data communication is a current trend. IoT devices generate all kinds of data. It would be interesting to have an IoT gateway that manages traffic by letting data circulate only when it is necessary and only that which is necessary [27]. This would save energy and make better use of network and cloud resources. An architecture for this approach has been presented [55].

3.6 IoT cloud platforms

A number of projects have created their own cloud platform [56]. The University of Glasgow decided to create its own cloud computing environment based on a Raspberry Pi [57], called PiCloud [58]. It emulates every layer of a cloud stack, from resource virtualisation to network behaviour. A fleet of 56 Raspberry Pi devices was able to

Chapter 3. Analysis

consolidate this [cloud](#) platform. This project is ideal for education, thanks to its low cost. Another study, based on the previous one, carried out the same style of [cloud](#) platform, but with 300 Raspberry Pi devices. Called the Bolzano Raspberry Pi [59], this [cloud computing](#) environment incorporates several [Network Attached Storage \(NAS\)](#) as storage units. However, these are traditional [cloud infrastructures](#).

[Cloud infrastructures](#) have been adapted for the use of [IoT](#). What differentiates them from traditional infrastructures is the processing of data generated by events in real time [60]. There are now a large number of public [cloud](#) service providers for [IoT](#). There are also a number of open source [IoT](#) solution providers. The main ones today are [AWS](#), Microsoft Azure and Google Cloud [61].

[AWS](#) [62] provides a [cloud](#) platform with over 200 services. Resources are distributed across 32 regions worldwide. It claims to be the most widely adopted platform in the world. Products include: compute, storage, databases, analytics, networking, mobile, development tools, management tools, [IoT](#), security and enterprise applications. It offers its services on demand and payment is on a pay-as-you-go basis. Through a survey published in 2019 on the various platforms [63], a solution dedicated to [IoT](#), called [AWS IoT](#), was summarised. This service aims to collect, store and analyse data from devices. Amazon FreeRTOS and [AWS IoT Greengrass](#) are made available by [AWS IoT](#), facilitating the development of embedded applications on these devices. For embedded systems, both [Arm](#) and x86 architectures with Linux are supported [64]. For device management, there is [AWS IoT Core](#), which facilitates device connectivity with [cloud](#) services. This can be complemented by [AWS IoT Things Graph](#) to visualise the data emitted by devices. In terms of data processing, the platform provides [AWS IoT Analytics](#), offering developers a convenient way to analyse the data generated. The results of this process can be routed to other devices or systems via the [AWS IoT Events](#) service [63] service. The aim of a thesis project was to define an [IoT](#) architecture for connecting [Arm](#) microcontrollers to [AWS](#) [42]. An example of a smart home application was made on this architecture. A machine learning service was used in the [cloud computing](#) to make predictions from data sent from the [IoT](#) devices. The company [56K.Cloud](#) is also carrying out a few projects linking [Arm](#) devices with a [cloud infrastructure AWS](#). It typically uses the [AWS IoT Greengrass Core](#) solution to run applications on embedded systems.

Microsoft Azure [65] is another public [cloud](#) provider. Launched in 2008, it offers more than 200 [cloud](#) products and services. Its data centres are located in more than 60 regions around the world. The company operates in the healthcare, finance, public sector, manufacturing and retail sectors. It says it invests a billion dollars a year in security to protect customer data. As far as [IoT](#) is concerned, it offers a solution called [Azure IoT](#). This solution includes several services. [Azure IoT Hub](#) is a service that establishes connections, administers and develops the ability to manage billions of [IoT](#) devices, from the peripheral to the [cloud](#). [Azure IoT Central](#) provides a user interface and [APIs](#) for connecting and managing large-scale embedded systems. Data is visualised using [Azure Time Series](#). [Azure Sphere](#) offers security to protect [IoT](#) devices. Physical device spaces can be replicated using the [Azure Digital Twins](#) service. System logic can be managed by [Azure IoT Edge](#). For the development of integrated applications, there is [Azure RTOS](#). [65]

Google Cloud [66] is one of the world's leading **cloud** providers. It currently distributes its resources in 39 regions across the globe. Launched in 2008, the provider is trying to keep up with the competition by offering around a hundred services for all types of business. However, Google has announced that it will be pulling out of the **IoT** business in 2023. The Google Cloud spokesperson said that the needs of their customers could be better served by specialist **IoT** partners [67].

There are other **cloud** providers offering **IoT** services. Some of these are mentioned in figure 3.6, where a comparison is made.

	<i>IoT platform</i>	<i>Open source</i>	<i>Protocols</i>	<i>Data store</i>	<i>Push notification</i>	<i>Trigger</i>	<i>Visualisation</i>
AWS IoT	No	MQTT, HTTPS	Yes	Yes	Yes	Yes	Yes
Azure IoT	No	MQTT, AMQP, HTTPS	Yes	Yes	Yes	Yes	Yes
IBM Watson IoT Platform	No	MQTT, JMS	Yes	Yes	Yes	Yes	Yes
Kaa Enterprise IoT Platform	Yes	MQTT, HTTPS	Yes	Yes	Yes	Yes	Yes
ThingsBoard	Yes	MQTT, CoAP, HTTP, LwM2M, SNMP	Yes	Yes	Yes	Yes	Yes

Figure 3.6 Comparison of **IoT cloud** platform providers (sources from providers)

Looking at the back end of platforms, they would have an advantage in using containerisation over traditional virtual machines [68]. It is seen as a lightweight virtualisation solution, with greater flexibility. In addition, containers are proving particularly suitable for solving the platform issues commonly associated with **PaaS** services in the **cloud**. This includes aspects such as application packaging and coordination [68]. Docker is the most popular container solution [14]. As for container orchestration, there's Kubernetes, an open source system [15].

3.7 Cloud infrastructure tools

When you want to host applications in the **cloud**, you need to deploy a **cloud infrastructure**. It is perfectly possible to build the infrastructure manually. All you have to do is visit a **cloud** provider's web page and select the resources you need from the services on offer. However, this infrastructure could be deployed several times. Fortunately, there are tools that can automatically deploy **cloud infrastructures**, called **Infrastructure as Code (IaC)**. The automatic deployment of **cloud infrastructures** corresponds to **provisioning**. These tools use templates containing a description of the infrastructure based on executable code or a configuration file. They save time on deployment and ensure that the implementation policy behind the infrastructure is always the same (security, rules, etc.). In addition to **provisioning**, **IaC** tools are capable of updating and deleting an infrastructure. There are universal **provisioning** tools for different **cloud** providers and there are integrated tools for each provider. [69]

3.7.1 Integrated IaC tools

Having mentioned the AWS and Microsoft Azure providers in section 3.6, each of them has its own IaC tool reserved for use with its platform. Google Cloud is not mentioned in this section due to the cessation of IoT activities.

At AWS, there are two IaC tools. AWS CloudFormation [70] is the first. AWS Cloud Development Kit (CDK) [71] is the second for defining and building infrastructure in the AWS cloud environment. AWS CDK gives users the flexibility to define infrastructure in distinct programming languages in the form of an application, with imperative syntax. It uses AWS CloudFormation as the deployment engine and allows users to define the infrastructure using programming idioms to model the system design. The CDK application deployment process includes the construction of the defined elements, preparation, validation and synthesis of the deployment artefacts. The CDK application is then transferred to the AWS CloudFormation service for real deployment. An overview of the deployment is shown in figure 3.7. The basic elements of CDK applications are called "constructs", and they represent cloud components containing services. Constructs can be nested to create a hierarchy of dependencies called a "construct tree". Ultimately, this hierarchy defines how constructs are synthesized into resources in the final AWS CloudFormation model. In a CDK application, it is possible to define several stacks. These are unique deployment units, each containing a hierarchy of constructs. [69]

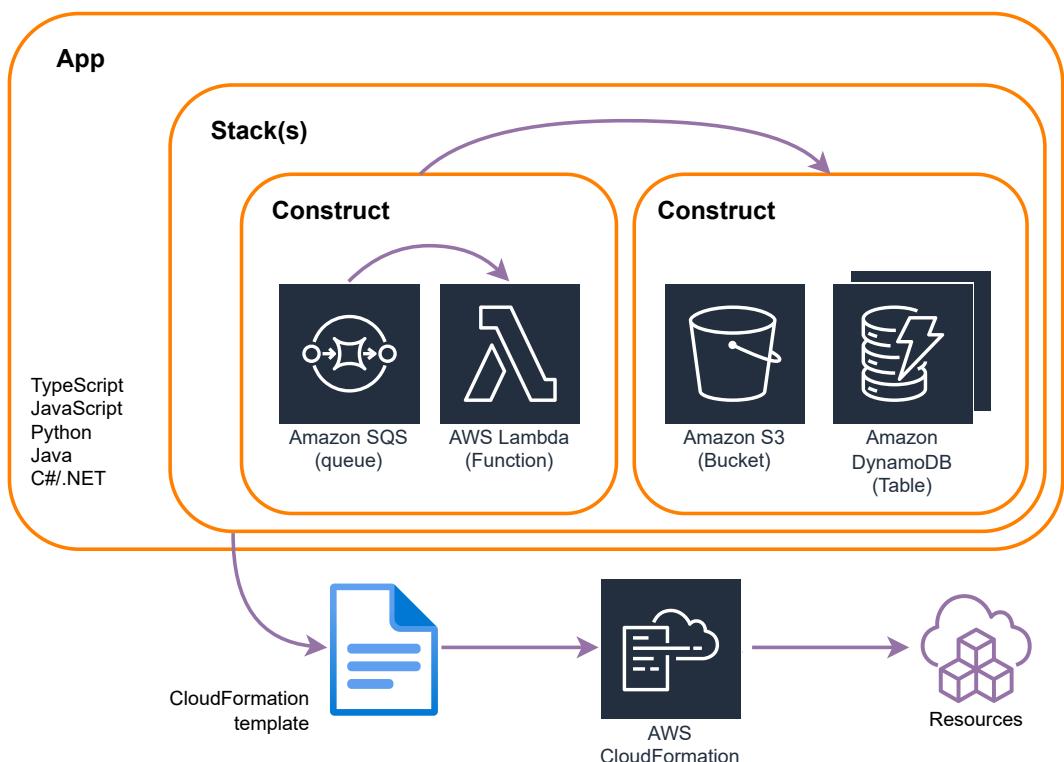


Figure 3.7 Overview of AWS CDK [71] deployment [71]

Microsoft Azure offers the [Azure Resource Manager \(ARM\)](#) deployment and management service [72]. All actions performed on resources by the user go through the [ARM](#) manager. It authenticates and authorises them. At the deployment level, there are [ARM](#) models, which are [JSON](#) files defining the infrastructure and configuration of a [cloud](#) environment. [JSON](#) uses declarative syntax. Before deployment, the [ARM API](#) checks the models for possible errors, guaranteeing successful deployment. [ARM](#) models ensure an identical architecture for deployments in different environments, such as development, testing and production, for example. The fewer dependencies there are, the faster [ARM](#) can deploy infrastructure resources in parallel. [ARM](#) models can be segmented into modular files for easy reuse. A group of resources deployed in the Azure platform can be extracted in the form of an [ARM](#) model. [69]

3.7.2 Universal IaC tools

There are several [IaC](#) tools compatible with different [cloud](#) platforms such as [AWS](#), Microsoft Azure, and many others. The [56K.Cloud](#) company mainly uses Terraform [73] and Pulumi [74].

Terraform is a tool created in 2014 by HashiCorp, written in the Go programming language [75]. It mainly provisions [Infrastructure as a Service](#), but can also deploy [Platform as a Service](#) and [Software as a Service](#). Terraform works using two inputs: configuration and state. The desired resources in the [cloud infrastructure](#) are described in a configuration file. This file works with the Terraform language using declarative syntax. The state corresponds to the current state of the infrastructure, or in other words the stack. The state is managed by the Terraform [API](#) and is stored locally by default. The deployment engine compares the desired infrastructure with the current state and determines which resources need to be created, updated or deleted. Terraform uses provider plugins to interact with [cloud](#) providers. It offers five basic commands for different stages: "init", "validate", "plan", "apply" and "destroy", enabling the infrastructure to be managed efficiently and easily [73]. The "init" command prepares the project directory. The "validate" command checks that the configuration is valid. The "plan" command displays the changes required by the current configuration. The "apply" command creates or updates the [cloud infrastructure](#). Finally, the "destroy" command destroys the [cloud infrastructure](#). An overview of the deployment can be seen in figure 3.8. [69]

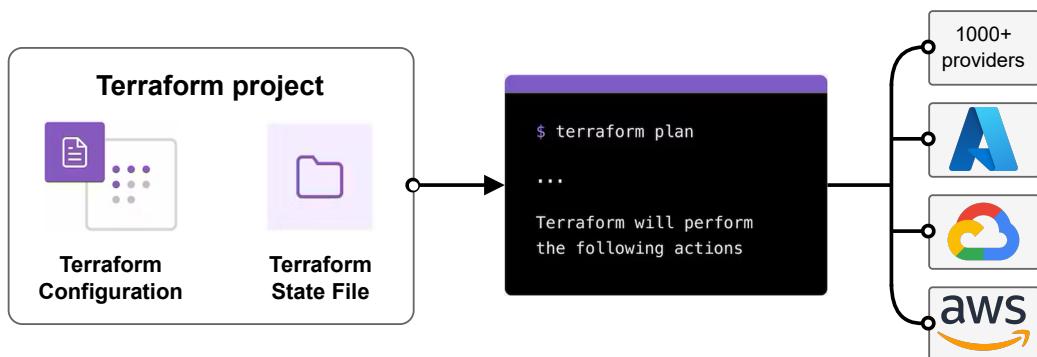


Figure 3.8 Overview of Terraform deployment [73]

Pulumi is an [IaC](#) tool similar to Terraform, but open source. It is written in the Go language [76]. Version 1.0 of Pulumi was released in 2019. This tool can be used with several programming languages, such as TypeScript, JavaScript, Python, Go and .NET. It therefore offers developers the possibility of creating [cloud infrastructures](#) using standard technologies, unlike Terraform. Deployment follows the Terraform model, where a Pulumi application is run to schedule the desired resources in the [cloud infrastructure](#). The deployment engine compares these desired resources with the current state of the stack and determines the necessary actions. The state of the last infrastructure deployment is saved in the Pulumi [Cloud](#) platform by default. With the Pulumi tool, we're talking about applications, because it uses imperative syntax. Applications describe how the [cloud infrastructure](#) should be composed by allocating resource objects whose properties correspond to the desired state. Resources are deployed in parallel if possible, but some resources have dependencies between them. Deployment is managed by the Pulumi [CLI](#). An overview of deployment is shown in figure 3.9. [69]

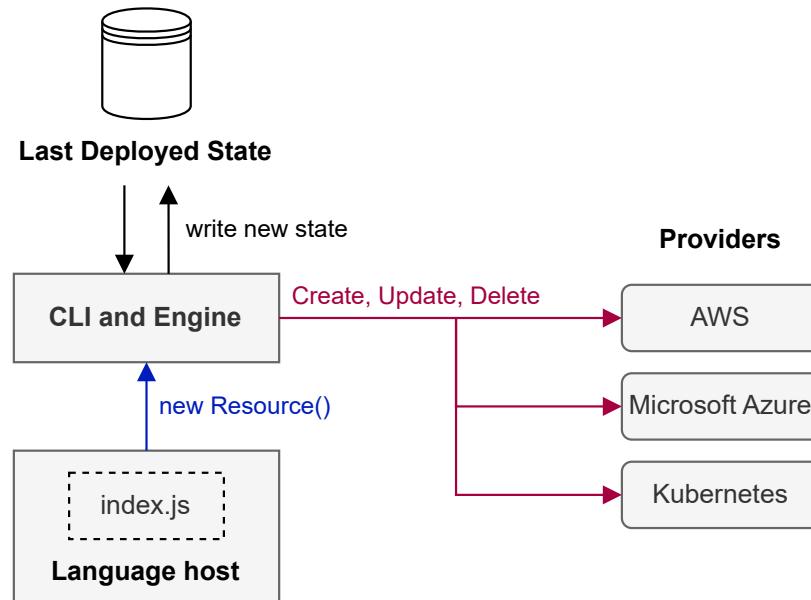


Figure 3.9 Overview of Pulumi deployment [74]

3.7.3 Comparison

Without even using the [IaC](#) tools, it is possible to identify a few differences. The tools integrated with the [cloud](#) providers cannot be used outside their platform. The other two tools, Terraform and Pulumi, can be provisioned by different providers. One notable point concerns the syntax. Terraform and Microsoft Azure use descriptive syntaxes. These limit the variety of programming languages. Terraform has its own language which you need to familiarise yourself with the first time. [AWS](#) and Pulumi use imperative syntax, which leaves the choice of programming languages open.

Focusing on the practical side, other disparities were noted. These stem from a project that was carried out with the aim of comparing the different [IaC](#) tools mentioned in this section 3.7 [69]. Terraform has several advantages, including clear documentation, an understandable configuration language, real-time execution plans and the ability to structure the infrastructure by dividing scripts into separate files. Because [cloud](#) providers

have different services, it is impossible to reuse the same code with each of them. It is possible to maintain the overall architecture of the infrastructure if the entities are clearly separated in the scripts. On the Pulumi side, resources can be separated into classes or functions. Common code between providers is therefore limited to abstractions of classes and functions. Pulumi has developed a tool capable of converting Terraform scripts into Pulumi scripts. It differs from Terraform by using Pulumi [Cloud](#) for infrastructure state management by default instead of a self-managed [API](#). This is an advantage of not needing to store state files locally and manage them. Terraform and Pulumi generally use [APIs](#) from [cloud](#) providers to create resources. One case stands out with Pulumi, when it works with [AWS](#), it uses the [AWS SDK](#) service directly, which will create the resources. Concerning the [Azure Resource Manager](#) tool, it is less efficient than [AWS CDK](#) because of the complexity of the configuration files. [AWS CDK](#) offers better readability by being able to divide up the code as required. Despite all this, it should be noted that these four [IaC](#) tools were able to provision the project in question. [69]

3.8 Arm SystemReady

Manufacturer [Arm](#) has decided to introduce a certification programme called SystemReady [77]. In 2020, as part of [Arm's](#) Cassini project, SystemReady was introduced to address the compatibility concerns of users planning to move from x86 architectures to systems based on [Arm's](#) processors [78]. [Arm](#) defines its programme as follows:

Arm SystemReady is a compliance certification program based on a set of hardware and firmware standards: [Base System Architecture \(BSA\)](#) and [Base Boot Requirements \(BBR\)](#) specifications, plus a selection of supplements. This ensures that subsequent layers of software also 'just work'. The compliance certification program tests and certifies that systems meet the SystemReady standards, giving confidence that operating systems [OS](#) and subsequent layers of software just work. [77]

The [Arm BSA](#) specification establishes a hardware basis for system software, including [OSs](#), hypervisors and firmware, based on the 64-bit [Arm](#) architecture. This specification ensures that users can install, boot and run generic [OSs](#) and hypervisors [79]. The [BBR](#) specification complements [BSA](#) by defining the basic firmware requirements necessary to support any [OS](#) or hypervisor compatible with the [BSA](#) specification [80].

A number of advantages were mentioned in relation to the program [77]:

- Compliance allows [OSs](#) and workloads to operate seamlessly on different [Arm](#) platforms.
- Standardisation of a range of different devices and systems provides a stable basis and a choice of systems for all sectors.
- Compliance with standards inspires confidence in software compatibility, so developers can concentrate on innovation and adding value to their products.
- Compliance simplifies and accelerates time-to-market for partners.
- Compliance makes it easy to identify [Arm-based](#) devices thanks to the "System-Ready certified" stamp.

- Compliance makes it easy to deploy and maintain standard firmware interfaces, reducing maintenance costs.
- Compliance reduces the costs associated with adopting a new platform by eliminating custom firmware engineering.

3.8.1 SystemReady certifications

To be able to adapt to different types of devices and markets, [Arm](#) has introduced four certifications. The first is SystemReady SR. It ensures the smooth operation of servers or workstations on a [Arm](#) chip. It is specially designed for Windows, Linux, VMware and [Berkeley Software Distribution \(BSD\)](#) environments. It targets generic out-of-the-box [OSs](#) and also supports legacy [OSs](#) on new devices. [77]

SystemReady ES is designed to meet the needs of Windows, Linux, VMware and [BSD](#) ecosystems based on [Arm](#) embedded systems. It also targets generic out-of-the-box [OSs](#) and supports legacy [OSs](#) on new devices. [77]

SystemReady IR ensures that Linux and [BSD](#) work perfectly on [Arm](#) embedded systems. It is ideally suited to the [IoT](#) sector. It mainly targets the Linux environment, but also custom images (Yocto, OpenWRT, buildroot) and pre-built images (Debian, Fedora, SUSE). [77]

SystemReady LS specifies the correct operation of Linux [OSs](#) on [Arm](#) chips designed for servers. This certificate is mainly for hyperscalers. Hyperscalers are large [cloud](#) service providers. [77]

At present, a multitude of hardware design companies are partnering with [Arm](#) to follow the SystemReady standards. Several embedded systems are already certified and on the market.

4 | Design

The design shows an overview of the different parts of the reference architecture. It describes the idea behind the structure of the [cloud infrastructure](#) with the integration of several embedded systems. The general operation of each of the applications included in this architecture is also described.

Contents

4.1	Reference architecture	34
4.1.1	Cloud infrastructure	34
4.1.2	Embedded systems integration	35
4.2	CI/CD pipeline	35
4.3	Applications	36
4.3.1	OS update	37
4.3.2	Certificate rotation	38
4.3.3	Led application	39
4.3.4	Button application	39
4.3.5	Corrupt application	39

4.1 Reference architecture

The reference architecture is divided into two parts: the first concerns embedded systems based on [Arm](#) architectures and certified [Arm SystemReady](#) with a Linux operating system, while the second relates to the [cloud](#) and is hosted by the [AWS](#) provider. An overview is clearly shown in figure 4.1.

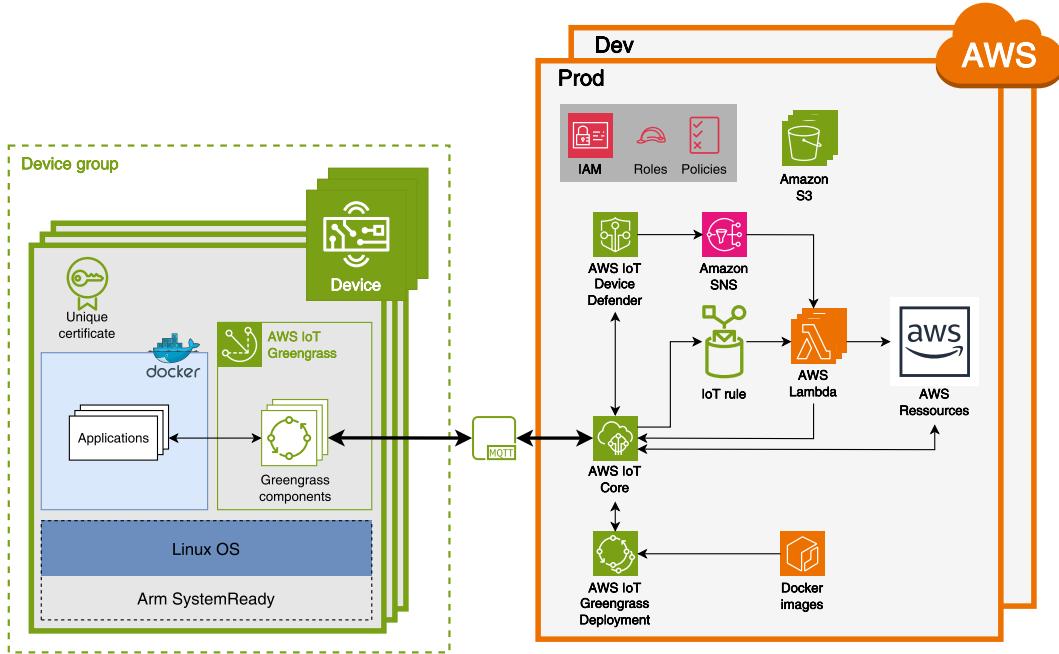


Figure 4.1 Reference architecture overview

4.1.1 Cloud infrastructure

In the [cloud](#), two distinct infrastructures are presented. The first infrastructure is dedicated to development, enabling all kinds of operations to be tested. The second infrastructure is the production infrastructure, encompassing the final product intended for customers. Each infrastructure uses different [AWS](#) services.

The central service managing the [IoT](#) is [AWS IoT](#) Core, which exchanges data with all connected devices. Its preferred communication protocol is [MQTT](#). Provisioned devices are registered in a specific group within this service. Only authorised devices are provisioned. [AWS IoT](#) Greengrass Deployment handles the deployment of new Greengrass components or updates to a group of devices. A component, in the context of [AWS IoT](#) Greengrass, is a unit of deployment and execution that encapsulates code, dependencies and resources, making it easy to deploy, manage and update [IoT](#) applications. In this architecture, Docker components are used, encapsulating Docker containers to offer greater flexibility in software deployment. Docker images are stored in the Amazon [Elastic Container Registry \(ECR\)](#). Another service, [AWS IoT](#) Device Defender, checks daily whether a certificate is about to expire, as each device has a unique certificate enabling a secure connection. If a certificate expires within 30 days, it is replaced by a new one. Other Lambda functions, triggered by rules, can enrich the architecture, as can other [AWS](#) resources, depending on the needs of each developer. Storage spaces such as Amazon S3 are used to save the state of the infrastructure, configuration files

and OS images. The configuration files contain a list of the serial numbers of all the IoT devices authorised to be provisioned. If a device is not listed and attempts to be provisioned, access will be denied. If a device is already provisioned and its serial number is removed from the list, its access will be withdrawn and it will be deleted from the cloud. The IAM service is there to assign roles and policies, limiting authorised actions on resources to the strict minimum.

4.1.2 Embedded systems integration

Embedded systems must be based on Arm architecture and be Arm SystemReady certified. The devices are flashed with a custom Linux OS image. On first boot, two software programs are installed : Docker Engine for running Docker containers, and AWS IoT Greengrass Core for provisioning, communicating with AWS IoT Core and managing Greengrass components. This software provisions the device by exchanging the claim certificate with a unique certificate, securing the connection of the device, which is then registered in the device group. Deployment of Greengrass components from the cloud to devices is automated when a new device is associated or a new version of a component becomes available. Applications are contained in Docker containers to ensure optimum portability.

4.2 CI/CD pipeline

The pipeline is made up of several workflows, each dedicated to a specific function. An overview of the various stages in the pipeline is shown in figure 4.2. The pipeline uses CI/CD methods to automate the deployment of the cloud infrastructure, as well as preparing for the integration of embedded systems.

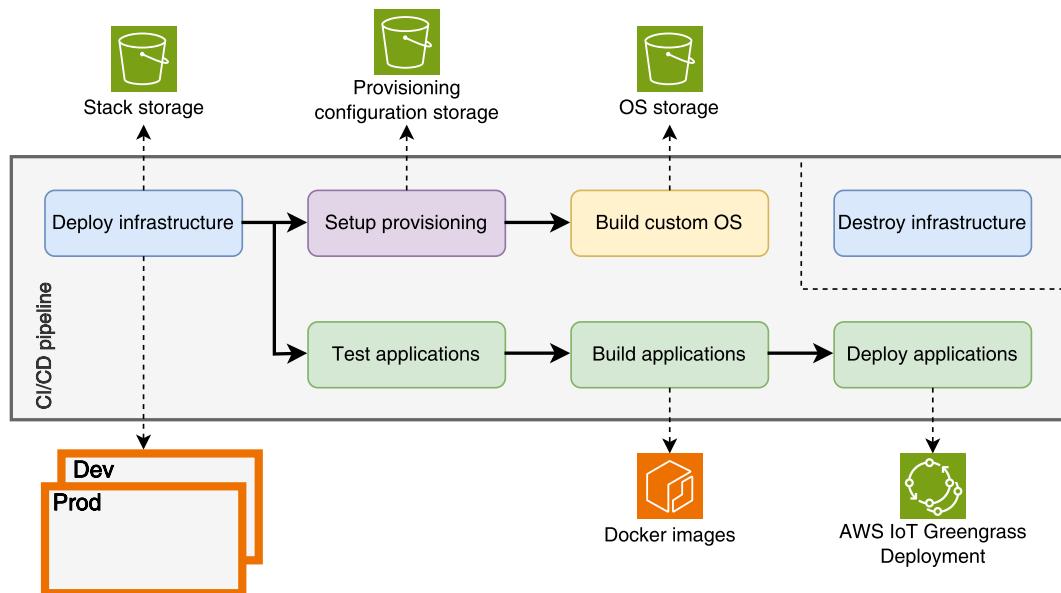


Figure 4.2 CI/CD workflows overview

The first job is to deploy the AWS cloud infrastructure (*Deploy infrastructure*). It can enable the deployment of development or production infrastructure. Since it uses a tool called IaC, the state of the infrastructure must be stored. This flow also enables resources to be updated.

If the entire infrastructure is deployed without error, the provisioning configuration workflow (*Setup provisioning*) takes care of preparing configuration files for *provisioning* in a storage space. It also manages device control. It checks the list of authorised devices to identify whether a previously provisioned device has been removed from the list. If so, it will be removed from the *cloud*. It may also be the case that authorised devices have already been provisioned and need to be connected to the new infrastructure. This operation is also checked and carried out.

If the previous step has been successful, the next step (*Build custom OS*) is to prepare and create the *OS* image with all the files and scripts needed to automatically provision the devices when they start up for the first time. The image must be saved in a storage space that allows it to be downloaded for flashing the devices' SD cards.

In parallel with the previous two workflows, the application workflows are launched. The first (*Test applications*) performs various tests on the applications. It includes security and unit tests.

If the tests have not failed, the next job (*Build applications*) is to build the Greengrass components from the applications. A Docker image is first created for each of the applications, coded in a freely chosen programming language. These images are then linked to their respective components. Finally, the components must be saved in the *AWS IoT* Greengrass service to be ready for deployment.

The last flow (*Deploy applications*) takes care of deploying applications to the device group. It uses the Greengrass Deployment service.

After all these tasks, there is still the infrastructure dismantling workflow (*Destroy infrastructure*). It is independent and can be launched manually. It is available in the event that a problem occurs and the infrastructure needs to be redeployed.

4.3 Applications

Several applications have been developed. Some are essential to the reference architecture, while others are demonstration applications. An overview of these is shown in figure 4.3.

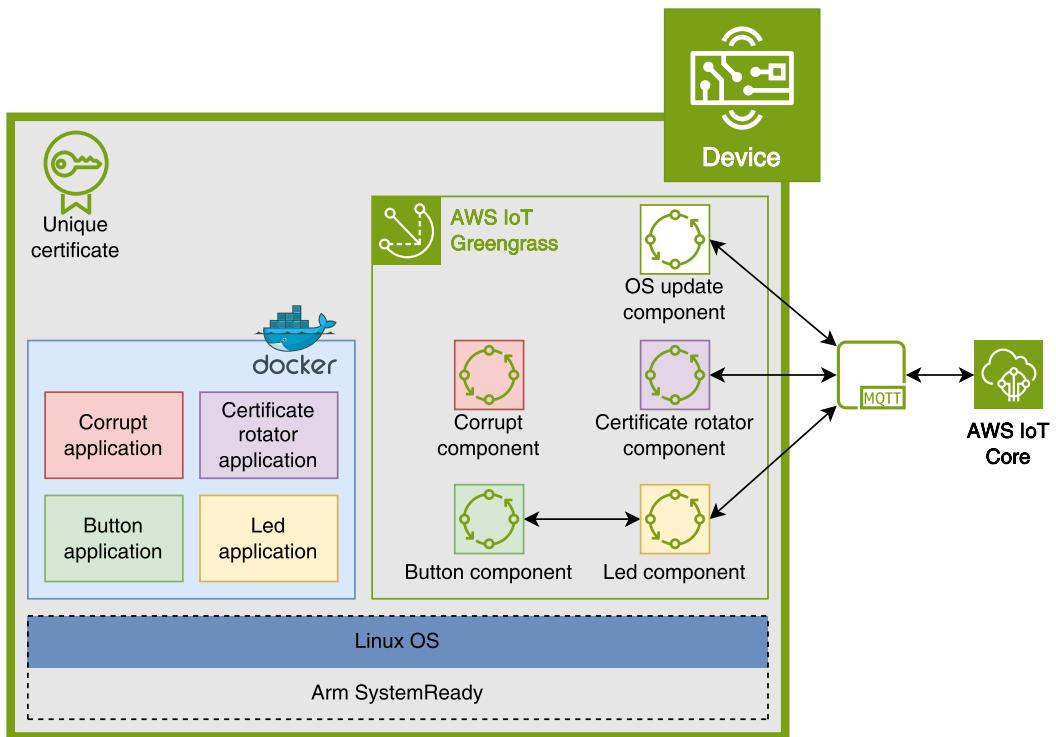


Figure 4.3 Applications overview

4.3.1 OS update

The application used to update **Operating System** features is essential to the reference architecture. Keeping Linux up to date is essential for security, stability, performance and compatibility. Security updates correct vulnerabilities, minimising the risk of exploitation by malicious software. Bug fixes and performance optimisations improve system stability and efficiency. Updating also ensures compliance with security requirements and adaptation to new features and standards. By keeping a Linux **OS** up to date, users can ensure that their system runs at optimum performance while keeping pace with technological developments and industry requirements.

This application does not run in a Docker container, but directly in the Greengrass component. The operation of the application must be kept simple. An activity diagram in figure 4.4 shows this.

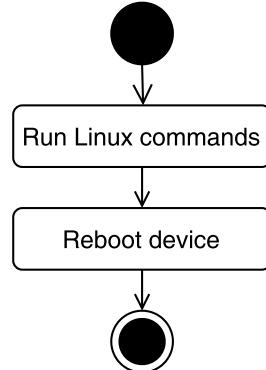


Figure 4.4 *OS update application activity diagram*

When the component starts up the first time the device is launched or after it has been updated, it will execute the Linux commands that have been written. To enable a complete update, the device will restart automatically. The component executes the Linux commands only once for each new version of the component.

4.3.2 Certificate rotation

On embedded systems communicating with AWS, unique certificates are used to establish secure connections via the protocols. Certificates have expiry dates for security reasons. The use of certificates with an expiry date makes it possible to limit the time during which a certificate is considered valid. This helps to reduce the risks associated with compromised private keys and ensures that certificates are regularly renewed, incorporating the latest security standards. It is therefore crucial to update certificates on embedded systems before their expiry date. If a certificate expires, secure connections with cloud services may be interrupted, leading to communication problems. To avoid this, this application must allow certificates to be managed automatically. It works with the AWS IoT Device Defender service, which alerts the application when it is time to change a certificate.

This application runs in a Docker container. An activity diagram illustrates its behaviour in figure 4.5.

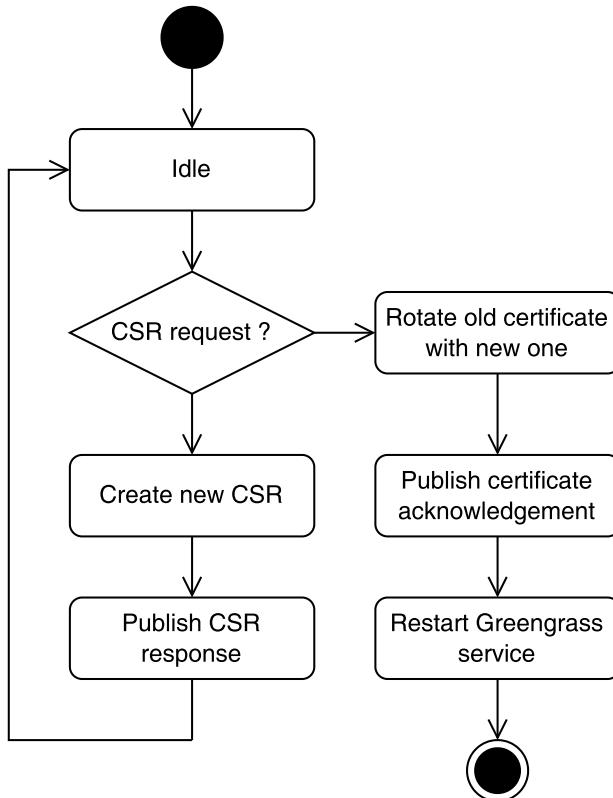


Figure 4.5 *Certificate rotator application activity diagram*

The general idea is that it is alerted by the IoT Device Defender service when a certificate change is required. The application then creates a new private key and makes a certificate creation request ([Certificate Signing Request \(CSR\)](#)) to the Amazon root certificate authority. A signed certificate is returned. This will replace the old certificate and the old private key with the new one. In order for the AWS IoT Greengrass Core service to take the new certificate into account during future communications, this service must be restarted on the device.

4.3.3 Led application

This application is not essential to the reference architecture. It is used as a demonstration for MQTT communication with AWS IoT. It is an application in a Docker container that blinks an led on an embedded system. The blinking frequency can be adjusted from an application associated with AWS IoT using the MQTT protocol. Blinking can also be enabled or disabled from the button application. Two pieces of data, the blink frequency and the blink state, are transmitted to AWS IoT so that they can be viewed on an interface. This is shown in figure 4.6.

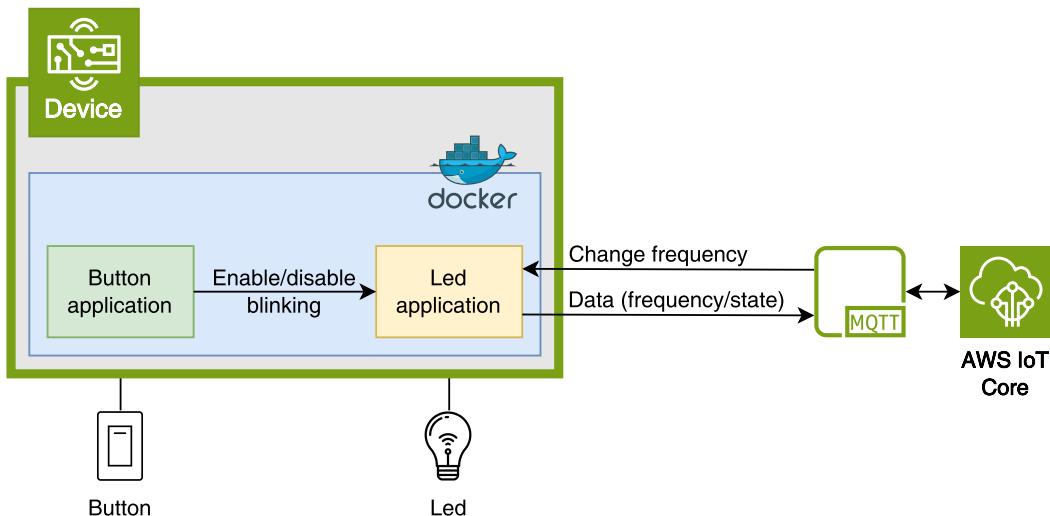


Figure 4.6 Led and button applications use case

4.3.4 Button application

This application is connected to the led application while retaining its independence. It is designed as a demonstration of local MQTT communication. Also running in a Docker container, it simply interacts with the led application to activate or deactivate the blinking of the led. The user presses a button built into the device to reverse the blinking state. A high-level view is shown in figure 4.6.

4.3.5 Corrupt application

A corrupted application is also developed to examine the behaviour of the device with AWS services. The approach is to run a simple Docker application by deliberately providing an incorrect environment for launching the container.

5 | Implementation

The implementation sets out the concrete realisation of the reference architecture, detailing the tools used to automate the deployment of the [cloud infrastructure](#) and the integration of the embedded systems. Particular attention is paid to the in-depth description of the automated pipeline. The various applications that accompany this architecture are also highlighted.

Contents

5.1	Reference architecture	42
5.1.1	Cloud infrastructure	42
5.1.2	Embedded systems integration	43
5.1.3	Security	48
5.2	CI/CD pipeline	49
5.2.1	GitHub Actions	49
5.2.2	Access to AWS resources	50
5.2.3	Infrastructure deployment	50
5.2.4	Provisioning setup	51
5.2.5	OS image building	52
5.2.6	Applications testing	52
5.2.7	Applications building	53
5.2.8	Applications deployment	54
5.2.9	Infrastructure destruction	54
5.3	Applications	55
5.3.1	Greengrass components	55
5.3.2	Containerization	57

5.1 Reference architecture

5.1.1 Cloud infrastructure

An [Infrastructure as Code \(IaC\)](#) tool was selected to orchestrate the efficient deployment of the [cloud infrastructure](#) on [AWS](#), and Pulumi was chosen because of its choice of programming language, in this case Python. The use of the same programming language for the implementation of applications and the description of [AWS](#) resources provides consistency within the project.

Integrating Pulumi is simple. The resources required are described in Python files. The deployment is managed by the Pulumi [CLI](#), which records the state of the infrastructure in an Amazon S3 compartment. This compartment is located in the same [AWS](#) account as the infrastructure, using Amazon S3 to store the state rather than the default Pulumi [Cloud](#) platform, thus eliminating external dependencies. Figure 5.1 shows an overview of this part of the implementation.

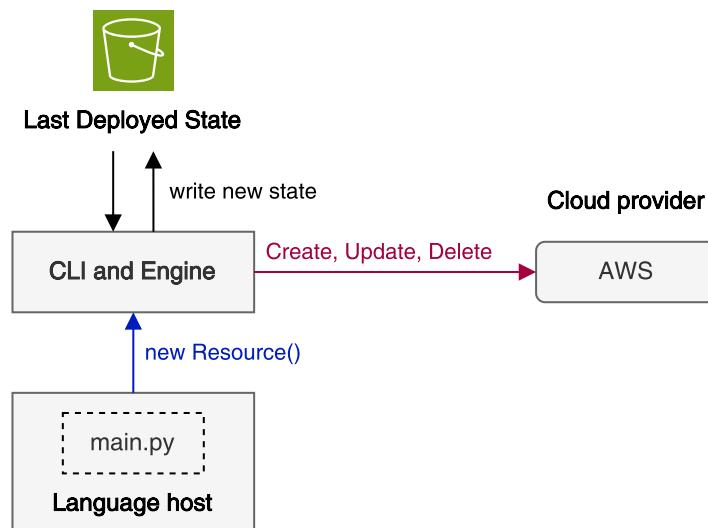


Figure 5.1 Overview of Pulumi's integration into the reference architecture

This [cloud infrastructure](#) part is implemented in a specific folder as follows :

```

cloud-infrastructure
├── Pulumi.yaml
├── Pulumi.dev.yaml
└── Pulumi.prod.yaml
└── main.py
└── iam.py
└── requirements.txt

```

The main configuration file for the Pulumi tool, `Pulumi.yaml`, includes crucial information such as the name of the stack and the programming language used, in this case Python.

This implementation aims to facilitate the deployment of two distinct infrastructures, namely the development environment and the production environment. Therefore, two additional configuration files, `Pulumi.dev.yaml` and `Pulumi.prod.yaml`, are present to

differentiate these two environments. *Pulumi.dev.yaml* contains parameters such as the AWS development account ID and the deployment region, while *Pulumi.prod.yaml* includes the same parameters with production-specific values. The identifiers of the AWS accounts must be different.

The Python files *main.py* and *iam.py* detail the description of the resources to be deployed. *iam.py* focuses on IAM roles and policies, while *main.py* mainly covers resources related to IoT and other resources.

Finally, the *requirements.txt* file lists the Pulumi libraries to be installed, including the main Pulumi library, which is essential for all IaC projects.

In terms of the Pulumi libraries used, AWS Native, a new preview, uses the AWS Cloud Control API to manage and provision AWS resources, generally aligning with the latest AWS features as they are released. The resources available in this library are based on those defined in the AWS CloudFormation registry. In addition, AWS Classic is another library that is used to fill in resources that are not yet available in AWS Native. This library uses the AWS SDK to manage and provision these resources.

5.1.2 Embedded systems integration

5.1.2.1 AWS IoT Greengrass

Integrating embedded systems with the AWS ecosystem is crucial to reaping the full benefits of cloud computing. To this end, AWS IoT Greengrass was chosen, a powerful solution that facilitates the execution of local processing on IoT devices, while enabling seamless interaction with AWS cloud services. AWS defined this solution as follows :

AWS IoT Greengrass is an open source Internet of Things (IoT) edge runtime and cloud service that helps you build, deploy and manage IoT applications on your devices. You can use AWS IoT Greengrass to build software that enables your devices to act locally on the data that they generate, run predictions based on machine learning models, and filter and aggregate device data. AWS IoT Greengrass enables your devices to collect and analyze data closer to where that data is generated, react autonomously to local events, and communicate securely with other devices on the local network. Greengrass devices can also communicate securely with AWS IoT Core and export IoT data to the AWS Cloud. You can use AWS IoT Greengrass to build edge applications using pre-built software modules, called components, that can connect your edge devices to AWS services or third-party services. You can also use AWS IoT Greengrass to package and run your software using Lambda functions, Docker containers, native operating system processes, or custom runtimes of your choice. [81]

In this implementation, AWS IoT Greengrass Core software is deployed on embedded systems, acting as an intelligent hub that manages interactions with the AWS cloud. It adapts perfectly to Linux OS. Greengrass components, encapsulating the necessary code, dependencies and resources, are then deployed automatically using mechanisms such as AWS IoT Greengrass Deployment.

Chapter 5. Implementation

A component must be installed. *Greengrass nucleus* (`aws.greengrass.Nucleus`) is a mandatory component and the minimum requirement for running [AWS IoT Greengrass Core](#) software on a device. It is configurable to customize and update the [AWS IoT Greengrass Core](#) software [Over-The-Air](#).

5.1.2.2 Communication

In this project, components running on an embedded system use the [AWS IoT Greengrass Core InterProcess Communication \(IPC\)](#) library, available in the [AWS IoT Device SDK](#), to exchange data with the [AWS IoT Greengrass nucleus](#) and other Greengrass components. The [IPC](#) interface supports the [Message Queuing Telemetry Transport](#) communication protocol. [MQTT](#) offers lightweight, asynchronous communication.

Publish/subscribe messaging allows messages to be sent and received in topics. Components can publish messages in topics to communicate with other components, and those who have subscribed to the topic can act on messages received. In this case, the communication is local to the [IoT](#) device.

The [IPC](#) interface also facilitates the sending and receiving of [MQTT](#) messages between [AWS IoT Greengrass](#) and [AWS IoT Core](#). Components can publish messages to [AWS IoT Core](#) and subscribe to topics to react to [MQTT](#) messages from other sources.

5.1.2.3 OTA update

All devices with [AWS IoT Greengrass Core](#) software can obtain [OTA](#) updates via [MQTT](#).

[AWS IoT Greengrass Core](#) software update

This functionality is built into the [AWS IoT Greengrass Core](#) software. It is made possible by the *Greengrass nucleus* component and other optional components. The following information is taken from the [AWS documentation](#) [49].

A few prerequisites must be met before an update can be carried out. Firstly, the Greengrass device must be connected to the [cloud AWS](#) in order to receive the deployment. It must be properly configured with certificates and authentication keys to interact with both the [AWS IoT Core](#) and [AWS IoT Greengrass](#). Finally, the [AWS IoT Greengrass Core](#) software must be configured and run as a system service.

There are a few things to bear in mind when upgrading. The Greengrass nucleus stops. This stops all the other components present. When the nucleus component is shut down, the device's connection to [AWS](#) is lost.

The following table summarises the behaviour of Greengrass updates :

5.1 Reference architecture

Action	Deployment configuration	Nucleus update behavior
Add new devices to a thing group targeted by an existing deployment without revising the deployment.	The deployment does not directly include Greengrass nucleus. The deployment directly includes at least one AWS -provided component, or includes a custom component that depends on an AWS -provided component or on the Greengrass nucleus.	On new devices, installs the latest patch version of nucleus that meets all component dependency requirements. On existing devices, does not update the installed version of the nucleus.
	The deployment directly includes a specific version of the Greengrass nucleus.	On new devices, installs the specified nucleus version. On existing devices, does not update the installed version of the nucleus.
Create a new deployment or revise an existing deployment.	The deployment does not directly include Greengrass nucleus. The deployment directly includes at least one AWS -provided component, or includes a custom component that depends on an AWS -provided component or on the Greengrass nucleus.	On all targeted devices, installs the latest patch version of the nucleus that meets all component dependency requirements, including on any new devices that you add to the targeted thing group.
	The deployment directly includes a specific version of the Greengrass nucleus.	On all targeted devices, installs the specified nucleus version, including any new devices that you add to the targeted thing group.

In this reference architecture, the version of *Greengrass nucleus* is not specified. However, there is one component provided by [AWS](#) (`aws.greengrass.Cli`) and custom components which depend on several components provided by [AWS](#).

Greengrass components update

For custom or public [AWS](#) components, the [AWS IoT](#) Greengrass Deployment service enables OTA updates.

Chapter 5. Implementation

The prerequisites for an update are the same as for the software update. When an update is deployed, only components with a new version are updated. These components will then be stopped and restarted automatically after being updated. If a deployment error occurs, a rollback is performed and the components resume their course with the old version.

The following table explains two main actions:

Action	Components behaviour
Add new devices to a group of devices targeted by an existing deployment without modifying the deployment.	All the components and dependencies present in the deployment are automatically deployed with their latest version on the new devices. On existing devices, nothing happens.
Create a new version of one or more Greengrass components.	Only components with a new version will be deployed on all devices in the group of devices targeted for deployment.

5.1.2.4 Fleet provisioning

With [AWS IoT fleet provisioning](#), it is possible to securely set up [AWS IoT](#) to generate and distribute X.509 device certificates and private keys to each device when they first connect to [AWS IoT](#) [82]. These client certificates, signed by the Amazon Root Certificate Authority, are provided by [AWS IoT](#). It is very flexible to define specific device groups, device types and permissions for Greengrass Core devices to be provisioned using fleet provisioning. A [provisioning](#) template is used to describe the [provisioning](#) process for each device, detailing the device, policy and certificate resources to be created during [provisioning](#). This template is created from Pulumi.

[AWS IoT](#) Greengrass offers a dedicated [AWS IoT](#) fleet provisioning plugin (`aws.greengrass.FleetProvisioningByClaim`), which facilitates the installation of [AWS IoT](#) Greengrass Core software using the resources created by [AWS IoT](#) fleet provisioning. This plugin uses claim-based [provisioning](#), where devices use a [provisioning](#) claim certificate and private key to obtain a unique X.509 device certificate, along with a private key, enabling regular operations. The claim certificate and private key are integrated as soon as the Linux [OS](#) image is created, enabling devices to be activated later when they are connected. The same claim certificate and private key can be used for multiple devices.

To deploy [AWS IoT](#) Greengrass Core software with [AWS IoT](#) fleet provisioning, configuration of resources in the [AWS](#) account is required. These resources include a provisioning template, a claim certificate and a token exchange IAM role. Once these elements have been created from Pulumi, apart from the certificate from the [AWS CLI](#), they can be reused to provision several embedded systems within the same fleet. The token exchange IAM role authorises calls to [AWS](#) services from the [IoT](#) device.

In this architecture, devices are registered by their serial number to ensure that each one has a unique name.

5.1.2.5 OS image

To enable the devices to be provisioned to integrate automatically on their first boot, a Linux **OS** image is created specifically for this purpose. The base image used for this architecture is *Raspberry Pi OS Lite* based on the Debian distribution. A directory dedicated to the creation of the image is present as follows:

```
os-image
  └── raspios-lite.json
  └── set_os.sh
  └── provision.sh
  └── config.yaml
```

Image creation is managed by Packer [83], an open source tool for automating image creation. Packer follows a process defined in a configuration file **JSON** (*raspios-lite.json*).

It works as follows :

1. **Packer configuration** (*raspios-lite.json*) : The Packer configuration file describes all the steps required to create the image. It specifies the constructors (in this case, the Arm constructor), the actions to be performed, the base image sources and other parameters.
2. **Running Packer** : In this project, Packer runs in a Docker environment. When the container is launched, the tool starts the build process following the specified configuration.
3. **Packer builder ARM plugin** : A plugin [84] to the Packer tool is used to build images specific to the Arm architecture. It uses the Arm build model to create a custom image based on the specification.
4. **Actions** : The actions in the configuration file are carried out to configure the image. This may involve installing software, configuring system parameters, etc. In this case, some system parameters are configured using the *set_os.sh* script. The **provisioning** script *provision.sh* used to install the necessary software is saved in the image and configured to run the first time the operating system is booted. Finally, the claim certificate and its private key are saved in the image.
5. **Build complete** : Once all the build and configuration steps have been successfully completed, Packer creates a custom image in the specified format (.img in this case).

The **provisioning** file *provision.sh* is a bash script. When it first runs on a device, it first retrieves the device's serial number, so that the device is registered in **AWS IoT** under that name. The script then installs the Docker Engine software, paving the way for future deployment of Greengrass components in containers. After this step, **AWS IoT** Greengrass Core is installed with the plugin specially dedicated to fleet **provisioning**. Finally, the **provisioning** command is executed using the *config.yaml* configuration file, which contains various parameters such as the device serial number, as well as the paths to locate the claim certificate and private key.

5.1.3 Security

At AWS, security is a top priority [85]. Security is a shared responsibility between AWS and the developer :

- **Security of the cloud** : AWS is responsible for protecting the infrastructure that runs AWS services in the AWS cloud.
- **Security in the cloud** : The developer's responsibility is determined by the AWS service that he uses.

When using AWS IoT Greengrass, the developer is also responsible for securing these devices, the connection to the local network and the private keys.

5.1.3.1 Authentication

Greengrass devices use X.509 certificates for authentication. X.509 certificates are digital certificates that adopt the X.509 public key infrastructure standard to bind a public key to the identity specified in the certificate. These X.509 certificates are issued by a trusted entity known as a **Certificate Authority (CA)**. The CA holds one or more specific certificates, called CA certificates, which it uses to issue X.509 certificates. Only this CA has the privilege of accessing the certificates. In this process, the authority is Amazon Root Certificate Authority. When provisioning a device, the fleet provisioning plugin requests a unique client certificate from this authority. [85]

Only devices authorised to be provisioned will receive a certificate. In this implementation, a list allowing the serial numbers of authorised devices to be inserted has been created in a dedicated file (*allowlist.txt*). When the provisioning plugin attempts to provision a device, it will first trigger a Lambda function which will check the list to see if the device is authorised. If so, it will receive its unique certificate. If not, it will not receive it and will not be provisioned.

Greengrass devices store certificates in the default Greengrass root folder (/v2/greengrass/).

5.1.3.2 Authorization

Greengrass devices rely on AWS IoT policies for authorization. These AWS IoT policies determine the scope of operations permitted for AWS IoT devices. They specify, in detail, permissions or denials of access to AWS IoT Core and AWS IoT Greengrass data plane operations, including actions such as publishing MQTT messages. In this architecture, an AWS IoT policy is created directly from Pulumi. [85]

5.1.3.3 Data protection

Greengrass devices frequently collect data for transmission to AWS services for further processing. AWS IoT Greengrass uses encryption to protect data in transit (via the internet or local network) and at rest (stored in the AWS cloud). [85]

AWS IoT Greengrass offers two distinct modes of communication for data in transit :

- **Data in transit over the internet** : Communication over the internet between a device and AWS IoT is encrypted using the Transport Layer Security (TLS) protocol. TLS is a security protocol that guarantees the confidentiality and integrity of data when it is exchanged over a network. All data sent to the AWS cloud therefore uses the MQTT protocol based on TLS.
- **Data on the device** : Communication between components on the Greengrass device is not encrypted, as the data does not leave the device.

When it comes to data at rest, AWS IoT Greengrass adopts different approaches :

- **Data at rest in the AWS cloud** : Customer data stored in the AWS cloud is encrypted by the AWS IoT Greengrass service using AWS KMS keys under its management.
- **Data at rest on the Greengrass device** : Data at rest on the device is protected by Unix file permissions and full disk encryption (if enabled, not enabled in this implementation). The user is responsible for securing the file system and the device.

5.2 CI/CD pipeline

A CI/CD process derived from the DevOps methodology has been put in place. It automates several workflows by integrating all the tools and software already described. The entire reference architecture has been developed in a public GitHub repository [86].

5.2.1 GitHub Actions

GitHub Actions is the Continuous Integration (CI) and Continuous Delivery (CD) service integrated directly into the GitHub platform and used in this project. It automates various aspects of the software development lifecycle, from building code to deploying it.

GitHub Actions works like this :

1. **Definition of workflows** : These describe the specific steps to be carried out during particular events, such as code validations, new version releases, etc. Several workflows can be run sequentially or in parallel.
2. **Triggering workflows** : These are triggered automatically in response to specific events defined by the developers. For example, a workflow can be triggered every time a new branch is created.
3. **Execution of steps** : Each workflow is made up of steps (jobs) that are executed sequentially or in parallel. These steps can include predefined actions, custom scripts, tests, deployments, etc.
4. **Runtime environments** : GitHub Actions provides runtime environments managed by GitHub. In this context, the Ubuntu-based Linux environment is used.
5. **Notifications and reports** : The results of each stage of the workflow are recorded and presented in the GitHub interface. Developers can receive notifications of successful or unsuccessful execution of the workflow.
6. **Integration with other services** : GitHub Actions can be integrated with other services and tools, such as cloud deployment services, notification services, secret management tools, etc.

7. **Actions ecosystem** : GitHub Actions offers a rich ecosystem of pre-built actions that developers can use in their workflows. These actions are reusable units of functionality (tests, deployments, notifications, etc.).

All workflows are implemented in a specific folder as follows :

```
.github/workflows
  ├── deploy-infra.yml
  ├── setup-provisioning.yml
  ├── build-image.yml
  ├── test-apps.yml
  ├── build-apps.yml
  ├── deploy-apps.yml
  └── destroy-infra.yml
```

5.2.2 Access to AWS resources

GitHub Actions integrates seamlessly with the [AWS cloud](#). It uses the [OpenID Connect \(OIDC\)](#) credentials protocol to access resources when deploying or updating the [cloud infrastructure](#). [OIDC](#) allows workflows to interact with [AWS cloud](#) service provider resources without having to permanently store credentials as GitHub secrets. Before deploying the infrastructure, the [AWS](#) account configuration is adjusted to recognise the GitHub [OIDC](#) as a trusted federated identity. In addition, this identification is based on IAM roles. This limits the possible actions on resources to the minimum required. Everything must be configured before the [CI/CD](#) process is launched.

To enable this connection to be used, [AWS](#) provides a pre-built action which is integrated into the flows ([aws-actions/configure-aws-credentials](#)).

5.2.3 Infrastructure deployment

This initial flow is triggered each time a new modification is pushed into Git. Before deploying the [cloud infrastructure](#), the decision was taken to generate a claim certificate and store it in a dedicated S3 bucket. This operation is carried out only once, ensuring that the same claim certificate is always preserved, even if the infrastructure is destroyed and redeployed at a later date. This ensures that devices ready to be provisioned can always do so using the same claim certificate.

A second S3 bucket is specifically created to record the state of the infrastructure. Pulumi connects to it, orchestrating the deployment of the configured infrastructure, whether for the development or production environment.

Successful completion of this first stage simultaneously triggers two other processes: provisioning configuration and testing of the developed applications. If deployment fails, the entire [CI/CD](#) process is interrupted.

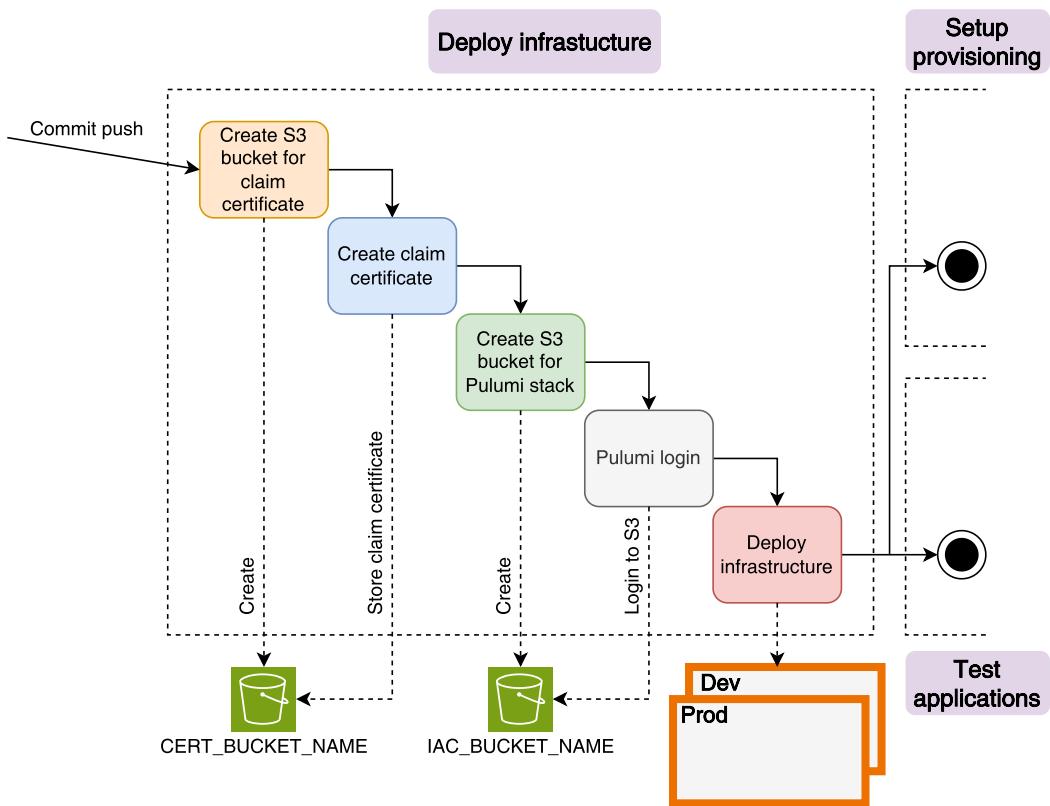


Figure 5.2 Infrastructure deployment workflow

5.2.4 Provisioning setup

This flow manages the preparation of configuration files for provisioning embedded systems, as well as checking devices that have already been provisioned. Two tasks are performed simultaneously. The first checks the serial numbers of devices in a dedicated list. If previously provisioned devices are no longer in the list, they are removed from the [AWS cloud](#). Authorised devices are reassigned to the new infrastructure and their unique certificate is reassociated with the [AWS IoT](#) policy for actions in the [cloud](#).

The second task, related to configuration, retrieves the [IoT](#) endpoints for future communications between the devices and [AWS IoT](#). Next, the parameters are saved in a configuration file required for provisioning. This file is stored, along with the list of authorised devices, in the same S3 bucket as the claim certificate.

Only if this flow is successful will the creation of the Linux image begin.

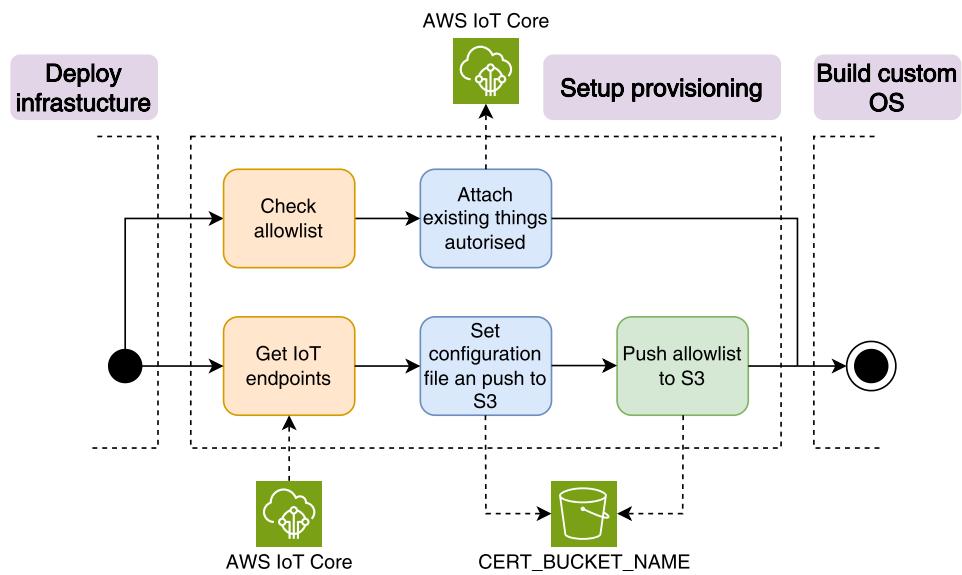


Figure 5.3 *Provisioning* setup workflow

5.2.5 OS image building

This process begins by retrieving the configuration file and the claim certificate, along with its associated private key, which have previously been saved in the S3 compartment. The image is then created using the Packer tool and its plugin. The image is created in a Docker container preconfigured with the necessary environment. Once the image has been generated, it is saved in its corresponding S3 compartment.

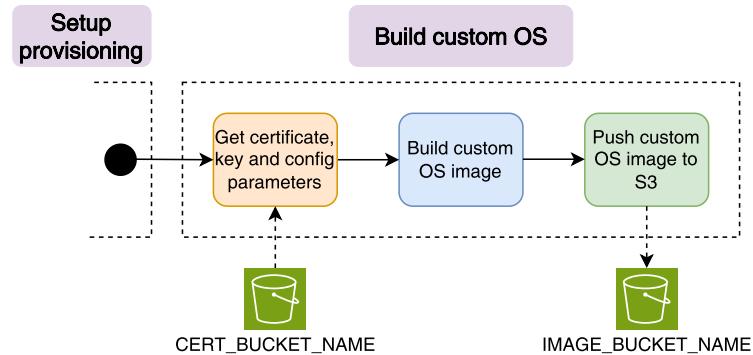


Figure 5.4 *OS* image building workflow

5.2.6 Applications testing

Simultaneously with the two previous workflows, this stage aims to test all the applications developed. It runs three tests concurrently. The first tests the linting of each application, helping to improve the quality of the code. The second performs static security tests for each application. Finally, unit tests are carried out using test code developed in parallel with the applications. The tools used in this task will depend on the programming language used.

If all the tests have been passed, the applications can be built.

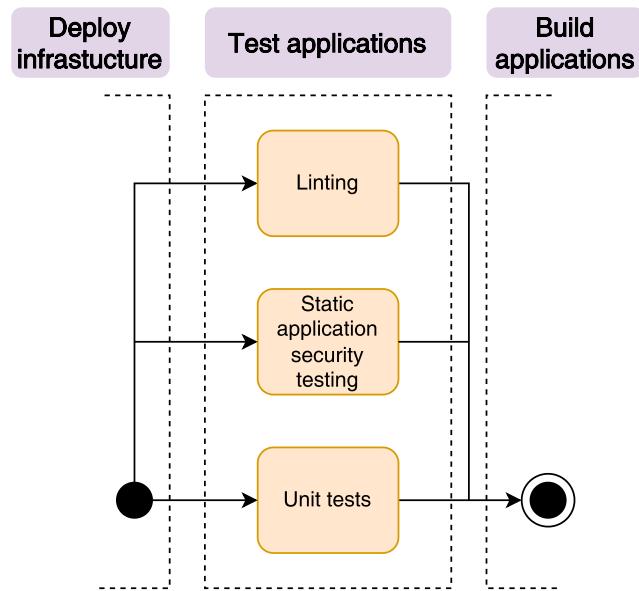


Figure 5.5 Applications testing workflow

5.2.7 Applications building

This process is divided into two stages. The first involves containerising applications using Docker, with this operation applied only to new applications or those that have undergone modifications. The resulting images are then stored in the dedicated Amazon ECR service.

The second part of the process focuses on creating a Greengrass component for each application. When a component is updated, a new version is created. However, only applications that have undergone modifications will see a new version of their component generated. It is also possible for a component's parameters to change, causing it to be updated. Each time a new version of a component is released, it is published in AWS IoT Greengrass, orchestrated by the AWS Greengrass Development Kit (GDK).

Once this workflow has been successfully completed, the Greengrass components are deployed.

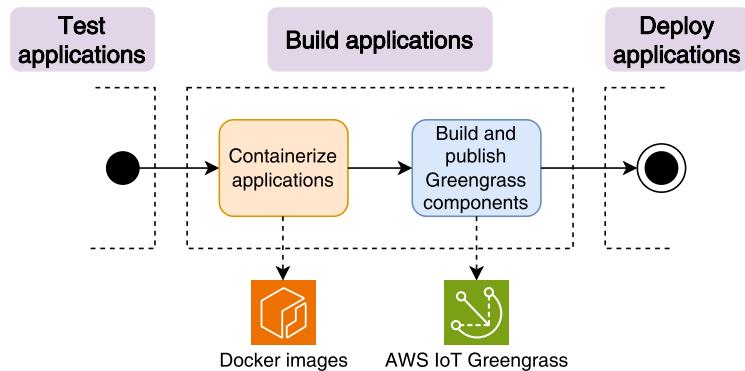


Figure 5.6 Applications building workflow

5.2.8 Applications deployment

This workflow supports the deployment of Greengrass components on a group of devices. It does this using the [AWS CLI](#). It doesn't matter whether the components have been updated or not, deployment takes place either way. [AWS IoT](#) Greengrass Deployment is smart enough to see if a new version of a component has been released. If this is the case, only the component with the new version will be deployed in the device group.

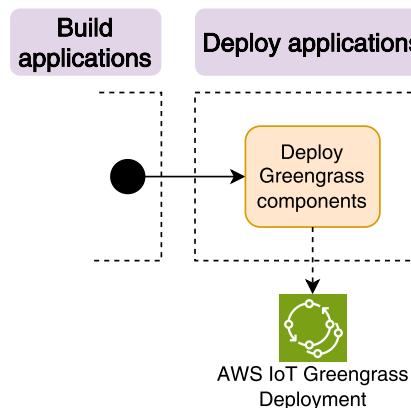


Figure 5.7 Applications deployment workflow

5.2.9 Infrastructure destruction

This last workflow is optional and has the task of dismantling the [AWS cloud infrastructure](#). To enable Pulumi to delete the S3 bucket containing the [OS](#) image, it must first be emptied. The [OS](#) image is then deleted. The bucket containing the configuration file and the list of authorised devices is emptied, with the exception of the claim certificate and its private key. Embedded systems integrated into the [AWS cloud](#) retain their unique certificate in [AWS](#), but lose their [AWS IoT](#) policy, dissociating them from any action. Finally, Pulumi connects to the S3 bucket containing the stack to completely remove the infrastructure. Once all the operations have been completed, the S3 bucket containing the stack is also deleted.

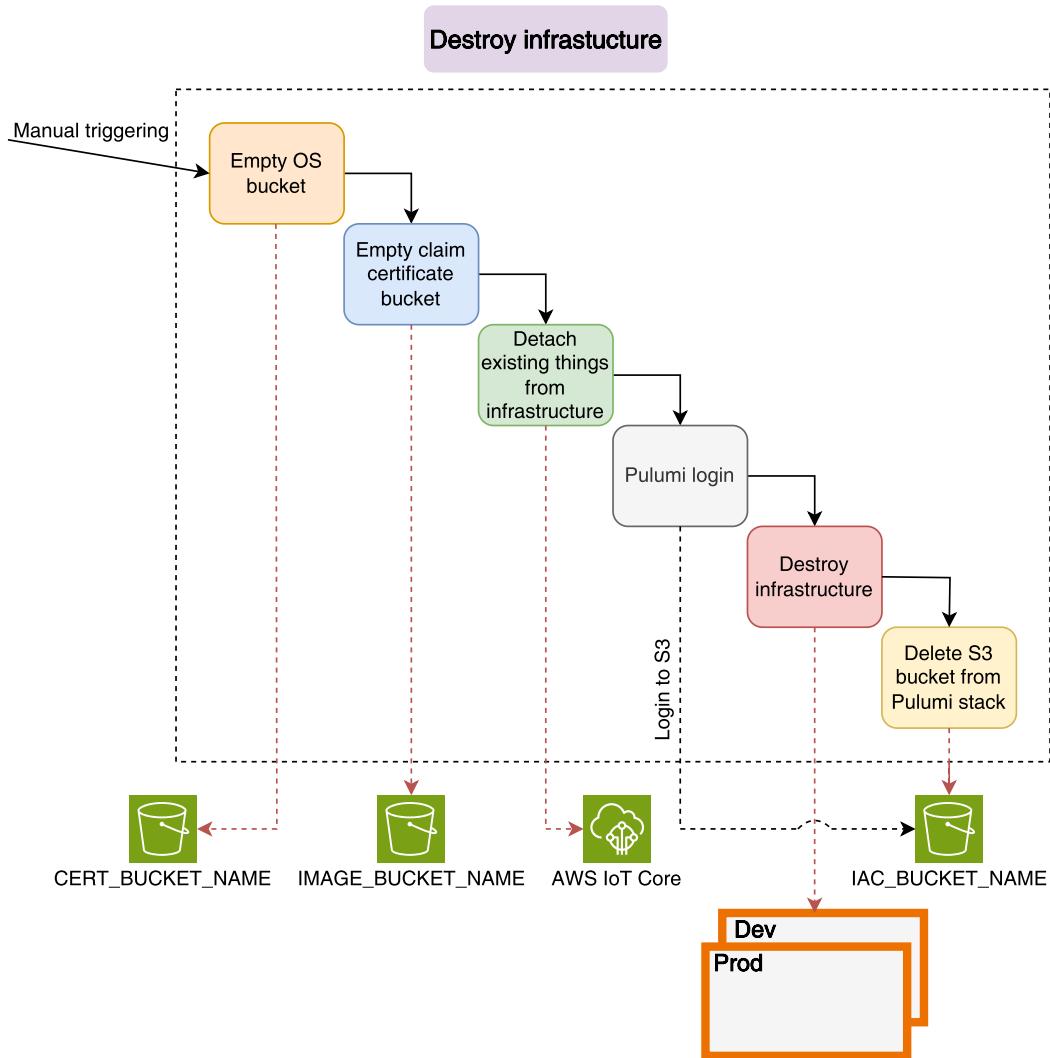


Figure 5.8 Infrastructure destruction workflow

5.3 Applications

5.3.1 Greengrass components

Greengrass components are software modules that are deployed on devices containing the [AWS IoT Greengrass Core](#) program. Components can be applications, runtime installers, libraries, or any code that is executable on a device. Components may depend on other components. For example, one component installs Python and other components depend on it to run Python applications. When components are deployed on a fleet of devices, [AWS IoT Greengrass](#) deploys only the software modules that the devices need. [87]

Here are some key points about Greengrass components :

- **Greengrass component definition** : A component can be defined as a combination of code (for example, a Python application), software dependencies and system resources. These may include libraries, configuration files, executables, etc.

Chapter 5. Implementation

- **Component management** : [AWS IoT Greengrass](#) enables centralised management of components. Components can be created, published and managed in [AWS IoT Greengrass](#) and deployed to groups of Greengrass devices.
- **Types of components** : Two types of component are used
 - Nucleus (`aws.greengrass.nucleus`): The Greengrass nucleus is the mandatory installed component that provides the minimum functionality for the operation of the [AWS IoT Greengrass](#) Core software.
 - Generic (`aws.greengrass.generic`): The Greengrass nucleus executes lifecycle scripts for a generic component. This type is the default for custom components.
- **Deployment** : Components are deployed on groups of Greengrass devices to run device-level applications. Deployments are orchestrated and managed from the [AWS IoT Greengrass](#) Deployment service.
- **Versioning** : Components can be versioned, enabling different versions of the software to be managed and updates to be rolled out in a controlled manner.

In this project, a component is defined in a folder as follows :

```
component_1
└── gdk-config.json
└── recipe.yaml
```

The `gdk-config.json` file is specific to the [AWS GDK](#) tool, used to generate and publish a component to [AWS IoT Greengrass](#). It only includes parameters such as the component name, for example.

The `recipe.yaml` file describes the component, its required dependencies, its lifecycle and the location of the executable. An example of this file is shown below.

```
RecipeFormatVersion: "2020-01-25"
ComponentName: COMPONENT_NAME
ComponentVersion: COMPONENT_VERSION
ComponentDescription: COMPONENT_DESCRIPTION
ComponentPublisher: AUTHOR
ComponentConfiguration:
  DefaultConfiguration:
    var1: 1
    var2: "frequency"
    accessControl:
      aws.greengrass.ipc.mqttproxy:
        com.app:mqttproxy:1:
          policyDescription: "Allows access to subscribe to input
            ↳ topics from AWS IoT."
          operations:
            - "aws.greengrass#SubscribeToIoTCore"
        resources:
          - "topic1"
      com.app:mqttproxy:2:
        policyDescription: "Allows access to publish to output
          ↳ topics to AWS IoT."
        operations:
          - "aws.greengrass#PublishToIoTCore"
        resources:
          - "topic2"
ComponentDependencies:
```

```

aws.greengrass.DockerApplicationManager
Manifests:
- Platform:
  os: linux
  architecture: aarch64
Lifecycle:
  run:
    RequiresPrivilege: true
    Script: docker run DOCKER_IMAGE@DIGEST
Artifacts:
- URI:
  ↳ "docker:AWS_ACCOUNT_ID.dkr.ecr.AWS_REGION.amazonaws.com/DOCKER_IMAGE"

```

Starting at the top of the example, information about the component is presented. Next, the configuration (*ComponentConfiguration*) is detailed, allowing the creation of variables and the configuration of access controls (*accessControl*). The latter is used to define authorised operations, typically using the [MQTT](#) protocol, such as the topics to which the component can publish or subscribe. Next, the dependent components are specified (*ComponentDependencies*). Finally, the lifecycle (*Lifecycle*) is detailed, along with the specification of the environment platform. In this example, a Docker application is launched and its image has been retrieved from the Amazon [ECR](#) service via its URI.

5.3.2 Containerization

The Python applications developed in this project are run in Docker containers. The creation of the application image requires the use of a Dockerfile. This specifies the architecture of the embedded system on which the container is to run, as well as the base image used. Finally, the Dockerfile copies the directory containing the application into its environment, then simply executes a bash script. This bash script installs the necessary Python libraries and launches the Python application.

```

docker_app_1
├── app_1
│   ├── start.sh
│   ├── requirements.txt
│   └── main.py
└── Dockerfile

```


6 | Validation

This part presents ...

Contents

6.1 Section 1	60
-------------------------	----

6.1 Section 1

7 | Conclusions

- 7.1 Project summary**
- 7.2 Comparison with the initial objectives**
- 7.3 Encountered difficulties**
- 7.4 Future perspectives**

A | Plannings

A.1 Forward planning

	Task	Duration	Begin	End	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13	Week 14	Week 15	Week 16	Week 17	Week 18	Week 19
1	Project duration	19 weeks	18.09.2023	09.02.2024																			
2	Project knowledge	4 weeks	18.09.2023	15.10.2023																			
3	Administration	1 week	18.09.2023	24.09.2023																			
4	Documentation classification	1 week	18.09.2023	24.09.2023																			
5	Specifications	2 weeks	18.09.2023	01.10.2023																			
6	Planning	3 weeks	25.09.2023	15.10.2023																			
7	State of the art	5 weeks	25.09.2023	29.10.2023																			
8	Formulating the problem	1 week	25.09.2023	01.10.2023																			
9	Literature search	2 weeks	02.10.2023	15.10.2023																			
10	Evaluation documentation	2 weeks	09.10.2023	22.10.2023																			
11	Analysis and interpretation	2 weeks	16.10.2023	29.10.2023																			
12	Presentation	2 weeks	16.10.2023	04.02.2024																			
13	Implementation	18 weeks	18.09.2023	01.10.2023																			
14	Familiarisation with 56KCloud	2 weeks	18.09.2023	01.10.2023																			
15	IaC testing	1 week	30.10.2023	05.11.2023																			
16	GitHub Actions testing	1 week	30.10.2023	05.11.2023																			
17	IaC reference architecture	6 weeks	06.11.2023	17.12.2023																			
18	CI/CD pipeline	1 week	13.11.2023	19.11.2023																			
19	Reference architecture testing	2 weeks	11.12.2023	22.12.2023																			
20	Clean code	1 week	18.12.2023	22.12.2023																			
21	Reference architecture documentation	1 week	18.12.2023	22.12.2023																			
22	Proof of concept	4 weeks	08.01.2024	04.02.2024																			
23	Thesis	19 weeks	18.09.2023	09.02.2024																			
24	Preparation	1 week	18.09.2023	24.09.2023																			
25	Writing	19 weeks	18.09.2023	09.02.2024																			
26	Validation	1 week	05.02.2024	09.02.2024																			

POC | Thesis

Reference architecture

State of the art

Milestones

A.2 Effective planning

Bibliography

- [1] *56K.Cloud*. Let's Start your Cloud Journey. URL: <https://www.56k.cloud> (visited on 09/21/2023).
- [2] *Freepik*. Images by macrovector. URL: <https://www.freepik.com/author/macrovector> (visited on 09/29/2023).
- [3] Noura Abbas, Andrew M. Gravell, and Gary B. Wills. "Historical Roots of Agile Methods: Where Did "Agile Thinking" Come From?" en. In: 2008. ISBN: 978-3-540-68255-4. DOI: [10.1007/978-3-540-68255-4_10](https://doi.org/10.1007/978-3-540-68255-4_10).
- [4] Elena Mugelini and Omar Abou Khaled. *MSE university course - IT Project Management*. PHASE III: execute the IT project - choosing a SDLC methodologies. 2022.
- [5] *Manifesto for Agile Software Development*. URL: <https://agilemanifesto.org/iso/en manifesto.html> (visited on 10/05/2023).
- [6] *Scrum Guides*. Jeff Sutherland. URL: <https://scrumguides.org/jeff.html> (visited on 10/05/2023).
- [7] Ken Schwaber and Jeff Sutherland. *The Scrum Guide*. 6th ed. 2020.
- [8] Scrum.org, Daniel Vacanti, and Yuval Yeret. *The Kanban Guide for Scrum Teams*. 2021.
- [9] *Wikipedia - Little's law*. URL: https://en.wikipedia.org/wiki/Little%27s_law (visited on 10/06/2023).
- [10] *Atlassian - Kanban*. URL: <https://www.atlassian.com/agile/kanban> (visited on 10/06/2023).
- [11] *AWS - AWS Certified Cloud Practitioner*. URL: <https://aws.amazon.com/certification/certified-cloud-practitioner/> (visited on 01/29/2024).
- [12] Peter Mell and Timothy Grance. "The NIST Definition of Cloud Computing". en. In: 2011. DOI: [10.6028/NIST.SP.800-145](https://doi.org/10.6028/NIST.SP.800-145).
- [13] *RedHat - IaaS vs. PaaS vs. SaaS*. URL: <https://www.redhat.com/en/topics/cloud-computing/iaas-vs-paas-vs-saas> (visited on 10/13/2023).
- [14] *Docker*. URL: <https://www.docker.com> (visited on 10/12/2023).
- [15] *Kubernetes*. URL: <https://kubernetes.io> (visited on 10/12/2023).
- [16] *Cloud Native Computing Foundation*. MAKE CLOUD NATIVE UBIQUITOUS. URL: <https://www.cncf.io> (visited on 10/12/2023).
- [17] *Cloud Native Computing Foundation*. CNCF Charter. URL: <https://github.com/cncf/foundation/blob/main/charter.md> (visited on 10/12/2023).
- [18] Nane Kratzke and Peter-Christian Quint. "Understanding cloud-native applications after 10 years of cloud computing - A systematic mapping study". en. In: 2017. DOI: [10.1016/j.jss.2017.01.001](https://doi.org/10.1016/j.jss.2017.01.001).

Bibliography

- [19] *Rancher - The History of Cloud Native*. URL: https://www.suse.com/c/rancher_blog/the-history-of-cloud-native/ (visited on 10/12/2023).
- [20] Simson Garfinkel. "An Evaluation of Amazon's Grid Computing Services: EC2, S3 and SQS". en. In: 2007.
- [21] Benjamin Black — EC2 Origins. URL: <http://blog.b3k.us/2009/01/25/ec2-origins.html> (visited on 10/31/2023).
- [22] IBM. What is lift and shift? URL: <https://www.ibm.com/topics/lift-and-shift> (visited on 10/12/2023).
- [23] Vasilios Andrikopoulos, Christoph Fehling, and Fran Leymann. "DESIGNING FOR CAP - The Effect of Design Decisions on the CAP Properties of Cloud-native Applications". In: *Proceedings of the 2nd International Conference on Cloud Computing and Services Science - CLOSER*. INSTICC. SciTePress, 2012, pp. 365–374. ISBN: 978-989-8565-05-1. DOI: [10.5220/0003931503650374](https://doi.org/10.5220/0003931503650374).
- [24] Sergio García-Gómez et al. "4CaaSt: Comprehensive Management of Cloud Services through a PaaS". In: 2012, pp. 494–499. DOI: [10.1109/ISPA.2012.72](https://doi.org/10.1109/ISPA.2012.72).
- [25] Nane Kratzke. "A Brief History of Cloud Application Architectures". en. In: Applied Sciences, 2018, p. 1368. DOI: [10.3390/app8081368](https://doi.org/10.3390/app8081368).
- [26] Manuel Díaz, Cristian Martín, and Bartolomé Rubio. "State-of-the-art, challenges, and open issues in the integration of Internet of things and cloud computing". en. In: Journal of Network and Computer Applications, 2016, pp. 99–117. DOI: [10.1016/j.jnca.2016.01.010](https://doi.org/10.1016/j.jnca.2016.01.010).
- [27] Mohammad Aazam et al. "Cloud of Things: Integrating Internet of Things and cloud computing and the issues involved". en. In: 2014, pp. 414–419. DOI: [10.1109/IBCAST.2014.6778179](https://doi.org/10.1109/IBCAST.2014.6778179).
- [28] Nakita Oza. "Integration of Cloud with Embedded Systems for IoT devices". en. In: International Journal of Latest Transactions in Engineering and Science, 2018.
- [29] Tetsuo Furuichi and Kunihiro Yamada. "Next Generation of Embedded System on Cloud Computing". en. In: Procedia Computer Science, 2014, pp. 1605–1614. DOI: [10.1016/j.procs.2014.08.244](https://doi.org/10.1016/j.procs.2014.08.244).
- [30] Cloud Standards Customer Council. "Cloud Customer Architecture for IoT". en. In: (2016).
- [31] Zigbee. The Full-Stack Solution for All Smart Devices. URL: <https://csa-iot.org/all-solutions/zigbee/> (visited on 10/20/2023).
- [32] Bluetooth. URL: <https://www.bluetooth.com> (visited on 10/29/2023).
- [33] LoRa Alliance. URL: <https://lora-alliance.org> (visited on 10/29/2023).
- [34] Thread. URL: <https://www.threadgroup.org> (visited on 10/29/2023).
- [35] Connectivity Standards Alliance - Matter. URL: <https://csa-iot.org/all-solutions/matter/> (visited on 10/29/2023).
- [36] Christos Stergiou et al. "Secure integration of IoT and Cloud Computing". In: *Future Generation Computer Systems* (2018), pp. 964–975. ISSN: 0167-739X. DOI: [10.1016/j.future.2016.11.031](https://doi.org/10.1016/j.future.2016.11.031).

- [37] Moataz Soliman et al. "Smart Home: Integrating Internet of Things with Web Services and Cloud Computing". en. In: *2013 IEEE 5th International Conference on Cloud Computing Technology and Science*. 2013, pp. 317–320. DOI: [10.1109/CloudCom.2013.155](https://doi.org/10.1109/CloudCom.2013.155).
- [38] *MQTT*. The Standard for IoT Messaging. URL: <https://mqtt.org> (visited on 10/29/2023).
- [39] Zach Shelby, Klaus Hartke, and Carsten Bormann. *The Constrained Application Protocol (CoAP)*. RFC 7252. 2014. DOI: [10.17487/RFC7252](https://doi.org/10.17487/RFC7252). URL: <https://datatracker.ietf.org/doc/rfc7252>.
- [40] Jiehan Zhou et al. "CloudThings: A common architecture for integrating the Internet of Things with Cloud Computing". en. In: CSCWD, 2013, pp. 651–657. DOI: [10.1109/CSCWD.2013.6581037](https://doi.org/10.1109/CSCWD.2013.6581037).
- [41] Matthias Kovatsch, Martin Lanter, and Simon Duquennoy. "Actinium: A RESTful runtime container for scriptable Internet of Things applications". In: *2012 3rd IEEE International Conference on the Internet of Things*. 2012, pp. 135–142. DOI: [10.1109/IOT.2012.6402315](https://doi.org/10.1109/IOT.2012.6402315).
- [42] Sergio Martin et al. *Arm Mbed – AWS IoT System Integration*. en. Research Reports or Papers. 2018. URL: <https://eprints.gla.ac.uk/157277/>.
- [43] *AWS IoT Core - MQTT*. URL: <https://docs.aws.amazon.com/iot/latest/developerguide/mqtt.html> (visited on 10/29/2023).
- [44] Euripides G. M. Petrakis et al. "Internet of Things as a Service (iTaaS): Challenges and solutions for management of sensor data on the cloud and the fog". In: *Internet of Things* (2018), pp. 156–174. ISSN: 2542-6605. DOI: [10.1016/j.iot.2018.09.009](https://doi.org/10.1016/j.iot.2018.09.009).
- [45] Maria Júlia Berriel de Sousa et al. "Over-the-air firmware update for IoT devices on the wild". In: *Internet of Things* (2022). ISSN: 2542-6605. DOI: [10.1016/j.iot.2022.100578](https://doi.org/10.1016/j.iot.2022.100578).
- [46] *Amazon Elastic Load Balancer*. URL: <https://aws.amazon.com/en/elasticloadbalancing/> (visited on 10/30/2023).
- [47] Yeongho Choi and Yujin Lim. "Optimization Approach for Resource Allocation on Cloud Computing for IoT". en. In: *International Journal of Distributed Sensor Networks* (2016). ISSN: 1550-1329. DOI: [10.1155/2016/3479247](https://doi.org/10.1155/2016/3479247).
- [48] *Comment établir la coexistence entre IPv4 et IPv6*. URL: <https://community.fs.com/fr/article/how-to-achieve-ipv4-and-ipv6-coexistence-dual-stack-or-mpls-tunnel.html> (visited on 10/30/2023).
- [49] *Update the AWS IoT Greengrass Core software (OTA)*. URL: <https://docs.aws.amazon.com/greengrass/v2/developerguide/update-greengrass-core-v2.html> (visited on 10/30/2023).
- [50] *Check Greengrass core device status*. URL: <https://docs.aws.amazon.com/greengrass/v2/developerguide/device-status.html> (visited on 10/30/2023).
- [51] *What is Device Update for IoT Hub?* URL: <https://learn.microsoft.com/en-us/azure/iot-hub-device-update/understand-device-update> (visited on 10/30/2023).

Bibliography

- [52] *Monitor device connection status*. URL: <https://learn.microsoft.com/en-us/azure/iot-hub/monitor-device-connection-state> (visited on 10/30/2023).
- [53] Brendan Moran et al. *A Firmware Update Architecture for Internet of Things*. 2021. DOI: [10.17487/RFC9019](https://doi.org/10.17487/RFC9019).
- [54] Alessio Botta et al. "Integration of Cloud computing and Internet of Things: A survey". In: *Future Generation Computer Systems* (2016), pp. 684–700. ISSN: 0167-739X. DOI: [10.1016/j.future.2015.09.021](https://doi.org/10.1016/j.future.2015.09.021).
- [55] Mohammad Aazam and Eui-Nam Huh. "Fog Computing and Smart Gateway Based Communication for Cloud of Things". In: *2014 International Conference on Future Internet of Things and Cloud*. 2014, pp. 464–470. DOI: [10.1109/FiCloud.2014.83](https://doi.org/10.1109/FiCloud.2014.83).
- [56] Muhammad Ilyas, Muneeb Ahmad, and Sajid Saleem. "Internet-of-Things-Infrastructure-as-a-Service: The democratization of access to public Internet-of-Things Infrastructure". en. In: *International Journal of Communication Systems* (2020). DOI: [10.1002/dac.4562](https://doi.org/10.1002/dac.4562).
- [57] *Raspberry Pi*. URL: <https://www.raspberrypi.com> (visited on 10/20/2023).
- [58] Fung Po Tso et al. "The Glasgow Raspberry Pi Cloud: A Scale Model for Cloud Computing Infrastructures". In: *2013 IEEE 33rd International Conference on Distributed Computing Systems Workshops*. 2013, pp. 108–112. DOI: [10.1109/ICDCSW.2013.25](https://doi.org/10.1109/ICDCSW.2013.25).
- [59] Pekka Abrahamsson et al. "Affordable and Energy-Efficient Cloud Computing Clusters: The Bolzano Raspberry Pi Cloud Cluster Experiment". en. In: IEEE, 2013, pp. 170–175. ISBN: 978-0-7695-5095-4. DOI: [10.1109/CloudCom.2013.121](https://doi.org/10.1109/CloudCom.2013.121).
- [60] Ranjana Sikarwar, Pradeep Yadav, and Aditya Dubey. "A Survey on IOT enabled cloud platforms". In: *2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT)*. 2020, pp. 120–124. DOI: [10.1109/CSNT48778.2020.9115735](https://doi.org/10.1109/CSNT48778.2020.9115735).
- [61] *Gartner - Cloud Infrastructure and Platform Services Reviews and Ratings*. URL: <https://www.gartner.com/reviews/market/cloud-infrastructure-and-platform-services> (visited on 10/26/2023).
- [62] *Amazon Web Services*. URL: <https://aws.amazon.com/> (visited on 10/23/2023).
- [63] Yohanes Yohanie Fridelin Panduman, Srirusti Sukaridhoto, and Anang Tjahjono. "A Survey of IoT Platform Comparison for Building Cyber-Physical System Architecture". In: *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*. 2019, pp. 238–243. DOI: [10.1109/ISRITI48646.2019.9034650](https://doi.org/10.1109/ISRITI48646.2019.9034650).
- [64] Alexandra A. Kirsanova, Gleb I. Radchenko, and Andrei N. Tchernykh. "Fog computing state of the art: concept and classification of platforms to support distributed computing systems". In: (2021). DOI: [10.48550/arXiv.2106.11726](https://doi.org/10.48550/arXiv.2106.11726).
- [65] *Microsoft Azure*. URL: <https://azure.microsoft.com/> (visited on 10/23/2023).
- [66] *Google Cloud*. URL: <https://cloud.google.com/> (visited on 10/23/2023).

Bibliography

- [67] Akenza AG - *Google IoT Core is Shutting Down*. URL: <https://blog.akenza.io/google-iot-core-shutdown> (visited on 10/25/2023).
- [68] Claus Pahl and Brian Lee. "Containers and Clusters for Edge Cloud Architectures – A Technology Review". In: *2015 3rd International Conference on Future Internet of Things and Cloud*. 2015, pp. 379–386. DOI: [10.1109/FiCloud.2015.35](https://doi.org/10.1109/FiCloud.2015.35).
- [69] Matti Valkeinen. "Cloud Infrastructure Tools For Cloud Applications : Infrastructure management of multiple cloud platforms". en. In: (2022). URL: <https://trepo.tuni.fi/handle/10024/137424>.
- [70] AWS *CloudFormation*. URL: <https://aws.amazon.com/en/cloudformation/> (visited on 10/26/2023).
- [71] AWS *CDK*. URL: <https://aws.amazon.com/en/cdk/> (visited on 10/26/2023).
- [72] Azure *Resource Manager*. URL: <https://azure.microsoft.com/en-gb/get-started/azure-portal/resource-manager> (visited on 10/26/2023).
- [73] Terraform. URL: <https://www.terraform.io> (visited on 10/26/2023).
- [74] Pulumi. URL: <https://www.pulumi.com> (visited on 10/26/2023).
- [75] GitHub - hashicorp/terraform. URL: <https://github.com/hashicorp/terraform/tags> (visited on 10/26/2023).
- [76] GitHub - pulumi/pulumi. URL: <https://github.com/pulumi/pulumi> (visited on 10/26/2023).
- [77] SystemReady Certification Program. URL: <https://www.arm.com/architecture/system-architectures/systemready-certification-program> (visited on 10/27/2023).
- [78] Google, Oracle's Ampere VMs get Arm's SystemReady seal of approval. URL: https://www.theregister.com/2022/11/02/google_oracles_cloud_arm/ (visited on 10/27/2023).
- [79] SystemReady Pre-Silicon Reference Guide BSA integration and compliance. BSA and SBSA overview. URL: <https://developer.arm.com/documentation/102858/0100/BSA-and-SBSA-overview> (visited on 10/27/2023).
- [80] GitHub - ARM-software/bbr-ac. URL: <https://github.com/ARM-software/bbr-ac> (visited on 10/27/2023).
- [81] What is AWS IoT Greengrass? URL: <https://docs.aws.amazon.com/greengrass/v2/developerguide/what-is-iot-greengrass.html> (visited on 01/31/2024).
- [82] Install AWS IoT Greengrass Core software with AWS IoT fleet provisioning. URL: <https://docs.aws.amazon.com/greengrass/v2/developerguide/fleet-provisioning.html> (visited on 02/01/2024).
- [83] Packer by HashiCorp. URL: <https://www.packer.io> (visited on 02/01/2024).
- [84] GitHub - mkaczanowski/packer-builder-arm. URL: <https://github.com/mkaczanowski/packer-builder-arm> (visited on 02/01/2024).
- [85] Security in AWS IoT Greengrass. URL: <https://docs.aws.amazon.com/greengrass/v2/developerguide/security.html> (visited on 02/01/2024).

Bibliography

- [86] *GitHub - 56kcloud/aws-iot-reference-architecture*. URL: <https://github.com/56kcloud/aws-iot-reference-architecture> (visited on 02/01/2024).
- [87] *Greengrass components*. URL: https://docs.aws.amazon.com/en_en/greengrass/v2/developerguide/greengrass-components.html (visited on 02/03/2024).

Glossary

Arm Arm is a British company that designs processors with RISC-type (Reduced Instruction Set Computer) architectures. [viii](#), [3](#), [15](#), [18](#), [19](#), [24](#), [26](#), [31](#), [32](#), [34](#), [35](#)

AWS AWS stands for Amazon Web Services. It is a company that provides a cloud platform with over 200 services. The resources are spread over more than 32 regions around the world. [x](#), [1](#), [3](#), [7](#), [13](#), [14](#), [20](#), [24](#), [25](#), [26](#), [28](#), [29](#), [30](#), [31](#), [34](#), [35](#), [36](#), [38](#), [39](#), [41](#), [42](#), [43](#), [44](#), [45](#), [46](#), [47](#), [48](#), [49](#), [50](#), [51](#), [53](#), [54](#), [55](#), [56](#)

cloud The term cloud refers to a remote infrastructure made up of servers that can be accessed from anywhere in the world via an internet connection. [v](#), [viii](#), [x](#), [1](#), [2](#), [3](#), [7](#), [13](#), [14](#), [15](#), [16](#), [17](#), [18](#), [19](#), [20](#), [21](#), [22](#), [23](#), [24](#), [25](#), [26](#), [27](#), [28](#), [29](#), [30](#), [31](#), [32](#), [34](#), [35](#), [36](#), [38](#), [42](#), [43](#), [44](#), [48](#), [49](#), [50](#), [51](#), [54](#)

cloud infrastructure A cloud infrastructure is the set of computing resources in the cloud that make up an IT environment. This includes servers, storage, network and security. [x](#), [1](#), [2](#), [3](#), [16](#), [17](#), [21](#), [26](#), [27](#), [29](#), [30](#), [33](#), [34](#), [35](#), [41](#), [42](#), [50](#), [54](#)

HES-SO The HES-SO is a Swiss university based in French-speaking Switzerland. It offers a wide range of specialisations in different fields such as technology, health, social work, economics and art. It has several buildings spread across the French-speaking cantons. [8](#)

KanBan KanBan is a project management tool designed to efficiently manage a workflow over a given period. It is widely adopted in agile methodology. [x](#), [4](#), [7](#), [10](#), [11](#)

provisioning Provisioning means the automated creation and installation of an IT infrastructure. Device provisioning involves creating the resources needed to integrate it into an IT infrastructure and integrating it into the existing infrastructure. [x](#), [3](#), [27](#), [35](#), [36](#), [41](#), [46](#), [47](#), [48](#), [51](#), [52](#)

Scrum Scrum is a process framework based on agile methodology for software engineering. It is suitable for small development teams. It ensures good communication between the customer and the team. It enables the first versions of a software product to be delivered quickly.. [x](#), [4](#), [7](#), [8](#), [9](#), [10](#), [11](#)

work package In project management, work packages are commonly used to define a plan. The plan provides an overview of the progress of the project. [3](#), [4](#), [7](#), [13](#)

Acronyms

API Application Programming Interface. 26, 29, 31

ARM Azure Resource Manager. 29, 31

BBR Base Boot Requirements. 31

BLE Bluetooth Low Energy. 23

BSA Base System Architecture. 31

BSD Berkeley Software Distribution. 32

CA Certificate Authority. 48

CD Continuous Delivery. viii, ix, x, 3, 7, 12, 13, 33, 35, 41, 49, 50, 51, 53

CDK Cloud Development Kit. x, 28, 31

CI Continuous Integration. viii, ix, x, 3, 7, 12, 13, 33, 35, 41, 49, 50, 51, 53

CLI Command Line interface. 30, 42, 46, 54

CoAP Constrained Application Protocol. 24

CSR Certificate Signing Request. 39

DevOps Development and Operations. x, 2, 12, 49

EC2 Elastic Compute Cloud. 20

ECR Elastic Container Registry. 34, 53, 57

HTTP HyperText Transfer Protocol. 24

IaaS Infrastructure as a Service. 16, 21, 29

IaC Infrastructure as Code. 15, 18, 19, 27, 28, 29, 30, 31, 35, 42, 43

IoT Internet of Things. viii, x, 1, 2, 15, 18, 19, 21, 22, 23, 24, 25, 26, 27, 28, 32, 34, 35, 36, 38, 39, 43, 44, 45, 46, 47, 48, 49, 51, 53, 54, 55, 56

IPC InterProcess Communication. 44

JSON JavaScript Object Notation. 24, 29, 47

Acronyms

MQTT Message Queuing Telemetry Transport. [24](#), [34](#), [39](#), [44](#), [48](#), [49](#), [57](#)

NAS Network Attached Storage. [26](#)

NAT Network Address Translation. [24](#)

OIDC OpenID Connect. [50](#)

OS Operating System. [x](#), [31](#), [32](#), [33](#), [35](#), [36](#), [37](#), [38](#), [41](#), [43](#), [46](#), [47](#), [52](#), [54](#)

OTA Over-The-Air. [25](#), [44](#)

PaaS Platform as a Service. [16](#), [21](#), [27](#), [29](#)

PO Product Owner. [9](#), [10](#)

QoS Quality of Service. [24](#), [25](#)

REST REpresentational State Transfer. [24](#)

ROI Return On OInvestment. [9](#)

SaaS Software as a Service. [16](#), [29](#)

SDK Software Development Kit. [31](#)

SOA Service-Oriented Architecture. [24](#)

TCP Transmission Control Protocol. [24](#)

TLS Transport Layer Security. [49](#)

UDP User Datagram Protocol. [24](#)

WiFi Wireless Fidelity. [23](#)

WIP Work In Progress. [10](#), [11](#)