MSc. in Computing
Practicum Approval Form

Section 1: Student Details

Project Title:	The applicability of container technologies for building a Secure Production ready PaaS
Student ID:	16212630
Student name:	Ruben Vasconcelos
Student email	ruben.vasconcelos3@mail.dcu.ie
Chosen major:	Software Engineering
Supervisor	Geoff Hamilton
Date of Submission	

Section 2: About your Practicum

Please answer all questions below. Please pay special attention to the word counts in all cases.

What is the topic of your proposed practicum? (100 words)

The aim of this practicum is to research the challenges involved in building a container platform that is capable of running production software in a reliable manner.

In this practicum, I will be exploring the following areas directly related to building a PaaS (Platform as a Service) for deploying containers, also known as CaaS (Containers as a service):

- Different layers of security and security requirements.
  - Such as networking requirements
  - And Container/Image security scanning
- Logging monitoring
- Availability
  - Failure recovery
  - How to implement a high availability platform
- User management
- Container Orchestration
- Continuous integration
  - For applications and the platform itself

I am aware that some companies might have very specific business needs, trying to create a platform for every use case would be nearly impossible. However, keeping this in mind and the Docker principle of “batteries included but removable” - Solomon Hykes, I will try to come up with an architecture suitable for general applications and research projects that can easily be extended or downgraded to cater for different needs.

Please provide details of the papers you have read on this topic (details of 5 papers expected).

1. Containerization and the PaaS cloud – Claus Pahl, Irish Centre for Cloud Computing and Commerce, IEEE Cloud Computing (Volume: 2, Issue: 3, May-June 2015)

2. Container-based orchestration in cloud: state of the art and challenges – Andrea Tosatto, Pietro Ruiu and Antonio Attanasio from the Istituto Superiore Mario Boella Torino, Italy: 8-10 July 2015 IEEE: 15399224

3. Continuous integration: improving software quality and reducing risk - Book by Andrew Glover, Paul M. Duvall, and Steve Matyas, 2007.

4. Virtualization vs Containerization to Support PaaS - Rajdeep Dua VMWare Bangalore, A Reddy Raja IIIT Hyderabad, Dharmesh Kakadia IIIT Hyderabad, 2014 IEEE International Conference on Cloud Engineering

5. Security Challenges for the Public Cloud - Kui Ren, Cong Wang, and Qian Wang • Illinois Institute of Technology IEEE Internet Computing (Volume: 16, Issue: 1, Jan.-Feb. 2012)

How does your proposal relate to existing work on this topic described in these papers? (200 words)

It’s a relatively new area, there are some papers that touch on the basic principle of containers, container security, and possible applicabilities. Although, very few actually dive into the challenges associated with the implementation of the platforms required for running production containers.

Using docker it’s very easy to spin up a container for development purposes, but running containerized applications in production is a totally different matter, as a lot of different things have to be taken into account such as:

1: Selecting a Cloud Provider

2: Image Management

3: Container Orchestration

4: Service discovery

5: Continuous Integration

6: Log Management & Monitoring of applications and the platform itself.

7: Database Management

8: Secret management

A large number of companies are currently working on developing their own container platforms including the company I work for.

This practicum aims to help fill that gap and provide the interested parties with a tool and the knowledge for standing up such platforms in fast and reliable manner.

What are the research questions that you will attempt to answer? (200 words)

What would be a viable PaaS Container architecture?

How are all the components put together?

What are the security concerns?

How will you explore these questions? (Please address the following points. Note that three or four sentences on each will suffice.)

- What software and programming environment will you use?

Environment: Linux on the cloud.

I will be using a number of DevOps tools for building, configuring and maintaining the platform such as Chef, Terraform or Cloudformation, Jenkins .....

Other tools like Docker and container technologies for orchestrating/managing containers such as Swarm or Kubernetes.

- What coding/development will you do?

I will be developing a prototype for the automated implementation of the proposed architecture.

- What data will be used for your investigations? Not applicable.

- Is this data currently available, if not, where will it come from? Not applicable.

- What experiments do you expect to run?

Architecture's fault tolerance, where one or multiple components fail unexpectedly.

Service and infrastructure scalability, performance and security.

Cost of running the platform over a period of time.

Comparison between the different technologies available for implementing the architecture's components.

- What output do you expect to gather?

In relation to the Architecture's fault tolerance, I expect to gather information on how well the proposed Platform is able to recover on the event, that some component fails in order to identify potential denial of service attack targets and high-risk components.

In relation to scalability, I expect to gather data on how much load can be exhorted on the system at a given time and how easily can the underlying infrastructure be scaled up or down. This will also allow me to gather Information about different cluster architectures and how different orchestration tools behave, allowing me to deduce a suggested ratio of cluster manager to workers.

I believe that research papers and online resources on the topic often, overlook the importance of cost management which is obviously a very important part of any successful business. I expect to gather some cost information that directly correlates to how the system scales. By mixing spot instances with dedicated ones I hope to identify a cluster architecture that maximizes cost without compromising reliability.

- How will the results be evaluated?

Costs results will be evaluated by directly mapping them to the number of cluster nodes executing at a given time. Keeping in mind that the use of spot instances will require an extra code wrapper, which may increase the platforms complexity and introduce new potential points of failure.

Architecture's fault results will be used to subgroup the different components into high risk and low risk. Depending on how the platform responds to the induced failure of different components.

Scalability and performance results will be evaluated by examining system load metrics at different scales, which could cause different components to fail. The results may then be used to help identify a viable auto-scaling policy.