DBRepo - Database Repository

What is the core value being generated?	Team	Status
We offer researchers to create a database in a secure, maintained and available environment that allows them to deposit their data from the beginning of a project. This proposal aims to upscale the current deployment via Docker Compose to a deployment where a cluster system is used for the orchestration and load balancing of the containers. Full disclosure: this proposal overlaps with my work as project assistant.	**Project owner / Deputy owner:** Martin Weise **Team members:** Josef Taha Lukas Mahler Tobias Grantner	ACTIVE

Problem space

Why are we doing this?

Problem statement

The current deployment of DBRepo, a repository for databases hosted at TU Wien uses Docker Compose for the deployment and orchestration of ~15 core services, while researcher databases inside DBRepo are deployed with the Docker Daemon. While this setup works well for 4-5 databases, the envisioned capacity of thousands of databases, hosted at TU Wien will not be deployed according to current best practices.

Impact of this problem

Without a technical solution to this problem, the impact of operating ~15 core services and more than 5 databases already is a challenge for most commodity hardware. Even if not used, the databases reserve significant amounts of RAM for eventual queries issued to them, the CPU load is unnecessary when a database is not used.

Who is the audience

- Friendly researchers at ZFDM (test-audience)
- Researchers at TUW (customer/target audience)

How do we judge success?

The project is successful when:

- Each core service is deployed using a cluster software
- Each researcher database is deployed from the container/database service using a cluster software
- Each researcher database can be put to "sleep" where it does not require significant amounts of RAM and is "woken up" when a query is issued to it, while being
- Transparent to the user from the current deployment using Docker Compose

Minimal viable product/service ("MVP")

What needs to be true in order for a prototype to be ready for release?

The MVP must contain:

1. One node for each core service of DBRepo
2. Container Service that creates this node instead of Docker Containers

The MVP may contain:

1. Admin state capabilities where the container service can start/stop nodes

This will be sufficient, for TUW researchers at ZFDM.

What crucial factors are we missing?	Blocking: • Currently we do not have such a cluster software Assumptions: • The knowledge of operating Kubernetes pods can be acquired through documentation • The JUnit test cases for the container service can be re-written to cover the Kubernetes deployment Gaps: • None identified yet

Continued Feedback

What is the key question we would ask to understand if we are on the right track?	**Is each core service and researcher database managed by Kubernetes in a way that allows load balancing for thousands of databases?**
Who are the alpha testers that we can use for validating our assumptions?	Center for Research Data Management / SBA Research • Tomasz Miksa • Barbara Sanchez DS-IFS • Andreas Rauber