

# Persistence Arrives on CF

- Paul Warren  
EMC Cloud Platform Team



CLOUD **FOUND**RY

# Cloud Platform Team

- Commit on open source, earn votes, steer CF
- Transform EMC by teaching “the way”

# Persistence on CF

- One of the most popular feature requests

# Persistence on CF

- One of the most popular feature requests



Not logged in [Talk](#) [Contributions](#) [Create account](#) [Log in](#)

[Article](#) [Talk](#) [Read](#) [Edit](#) [View history](#)

## Unstructured data

From Wikipedia, the free encyclopedia

**Unstructured data** (or **unstructured information**) refers to information that either does not have a pre-defined [data model](#) or is not organized in a pre-defined manner. Unstructured information is typically [text](#)-heavy, but may contain data such as dates, numbers, and facts as well. This results in irregularities and [ambiguities](#) that make it difficult to understand using traditional programs as compared to data stored in fielded form in databases or [annotated](#) ([semantically tagged](#)) in documents.

In 1998, [Merrill Lynch](#) cited a rule of thumb that somewhere around 80-90% of all potentially usable business information may originate in unstructured form.<sup>[1]</sup> This rule of thumb is not based on primary or any quantitative research, but nonetheless is accepted by some.<sup>[2]</sup>

IDC and EMC project that data will grow to 40 [zettabytes](#) by 2020, resulting in a 50-fold growth from the beginning of 2010.<sup>[3]</sup> Computer World states that unstructured information might account for more than 70%–80% of all data in organizations.<sup>[4]</sup>

# Persistence on CF

- One of the most popular feature requests



Not logged in [Talk](#) [Contributions](#) [Create account](#) [Log in](#)

Article [Talk](#) [Read](#) [Edit](#) [View history](#)

## Unstructured data

From Wikipedia, the free encyclopedia

**Unstructured data** (or **unstructured information**) refers to information that either does not have a pre-defined [data model](#) or is not organized in a pre-defined manner. Unstructured information is typically [text](#)-heavy, but may contain data such as dates, numbers, and facts as well. This results in irregularities and [ambiguities](#) that make it difficult to understand using traditional programs as compared to data stored in fielded form in databases or [annotated](#) ([semantically tagged](#)) in documents.

In 1998, [Merrill Lynch](#) cited a rule of thumb that somewhere around 80-90% of all potentially usable business information may originate in unstructured form.<sup>[1]</sup> This rule of thumb is not based on primary or any quantitative research, but nonetheless is accepted by some.<sup>[2]</sup>

IDC and EMC project that data will grow to 40 [zettabytes](#) by 2020, resulting in a 50-fold growth from the beginning of 2010.<sup>[3]</sup> Computer World states that unstructured information might account for more than 70%–80% of all data in organizations.<sup>[4]</sup>

t somewhere around 80-90% of all potentially usable business information may originate in unstructured form.

# Persistence on CF

- One of the most popular feature requests



Not logged in [Talk](#) [Contributions](#) [Create account](#) [Log in](#)

Article [Talk](#) [Read](#) [Edit](#) [View history](#)

## Unstructured data

From Wikipedia, the free encyclopedia

**Unstructured data** (or **unstructured information**) refers to information that either does not have a pre-defined [data model](#) or is not organized in a pre-defined manner. Unstructured information is typically [text](#)-heavy, but may contain data such as dates, numbers, and facts as well. This results in irregularities and [ambiguities](#) that make it difficult to understand using traditional programs as compared to data stored in fielded form in databases or [annotated](#) ([semantically tagged](#)) in documents.

In 1998, [Merrill Lynch](#) cited a rule of thumb that somewhere around 80-90% of all potentially usable business information may originate in unstructured form.<sup>[1]</sup> This rule of thumb is not based on primary or any quantitative research, but nonetheless is accepted by some.<sup>[2]</sup>

IDC and EMC project that data will grow to 40 [zettabytes](#) by 2020, resulting in a 50-fold growth from the beginning of 2010.<sup>[3]</sup> Computer World states that unstructured information might account for more than 70%–80% of all data in organizations.<sup>[4]</sup>

t somewhere around 80-90% of all potentially usable business information may originate in unstructured form.<sup>[1]</sup>

l EMC project that data will grow to 40 [zettabytes](#) by 2020, resulting in a 50-fold growth from the beginning of 2010.<sup>[3]</sup>

# However...

- CloudFoundry
  - 12 Factor apps
    - Network attached services only
    - File systems not allowed
    - No ability to lift and shift
- Spring Data
  - No uniform handling for BLOBs

# Opportunities

- Add volume support to CloudFoundry
- Add BLOB support to Spring Data



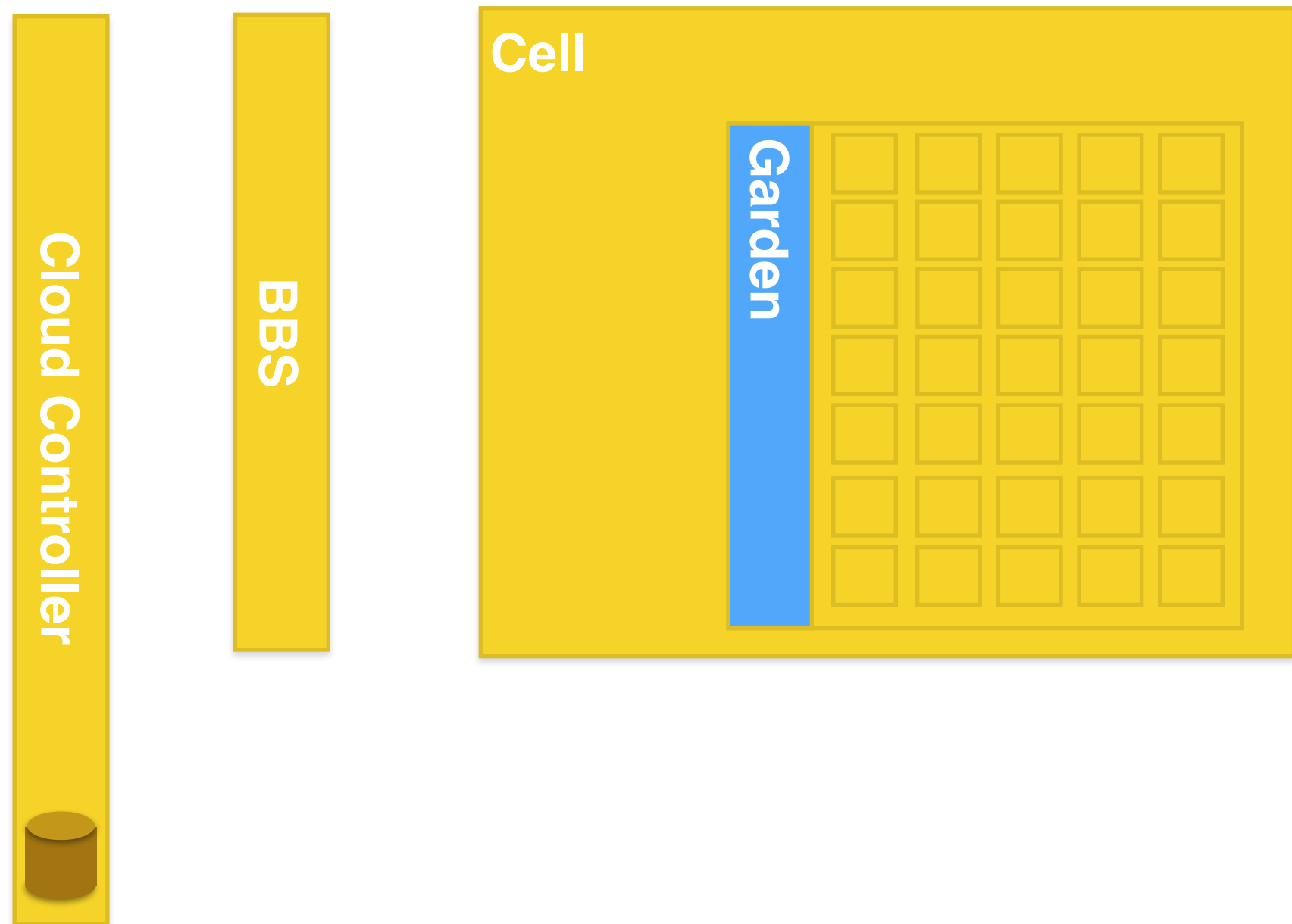
# Spring Content



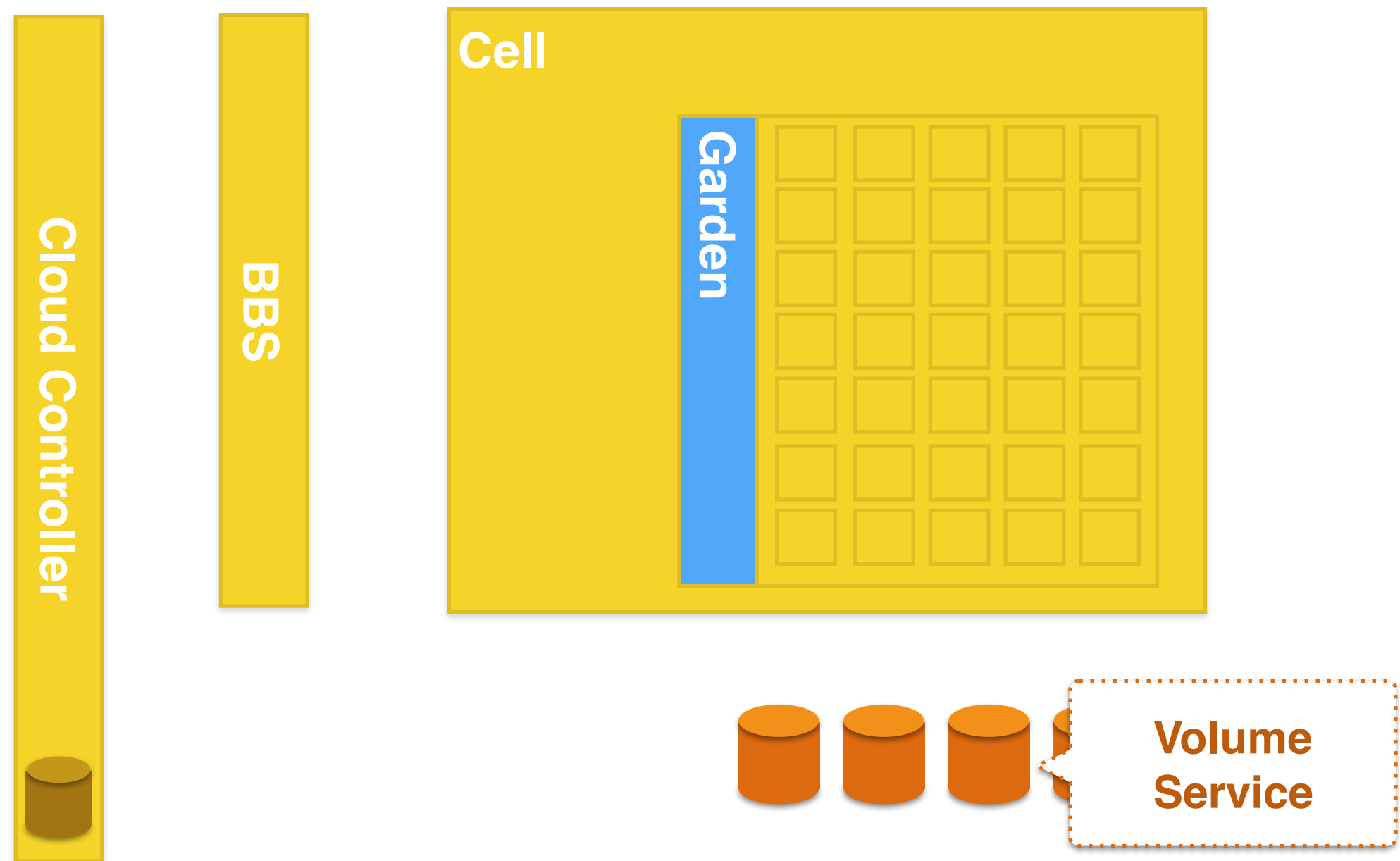
- Not an official Spring project
- Extensions for Spring-Data that add primitives for handling BLOBs
  - Same interface-first approach for unstructured data
- Spring Boot support
- Spring Data REST support
- Extensible Rendition Service
- Supports JPA, Mongo, S3 and Filesystem
- Separating content from metadata is a key characteristic of all CMS

# App Walkthrough

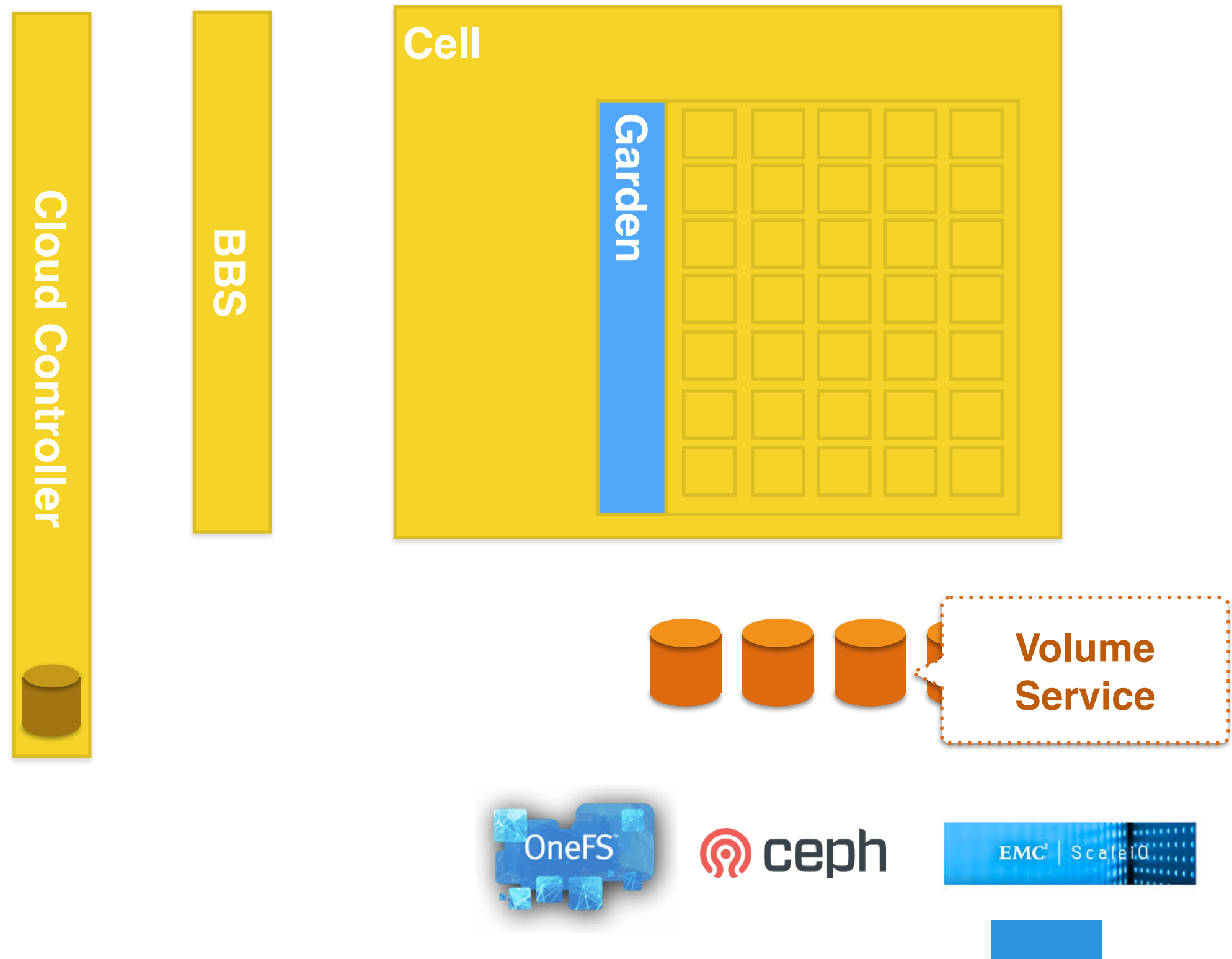
# Architecture



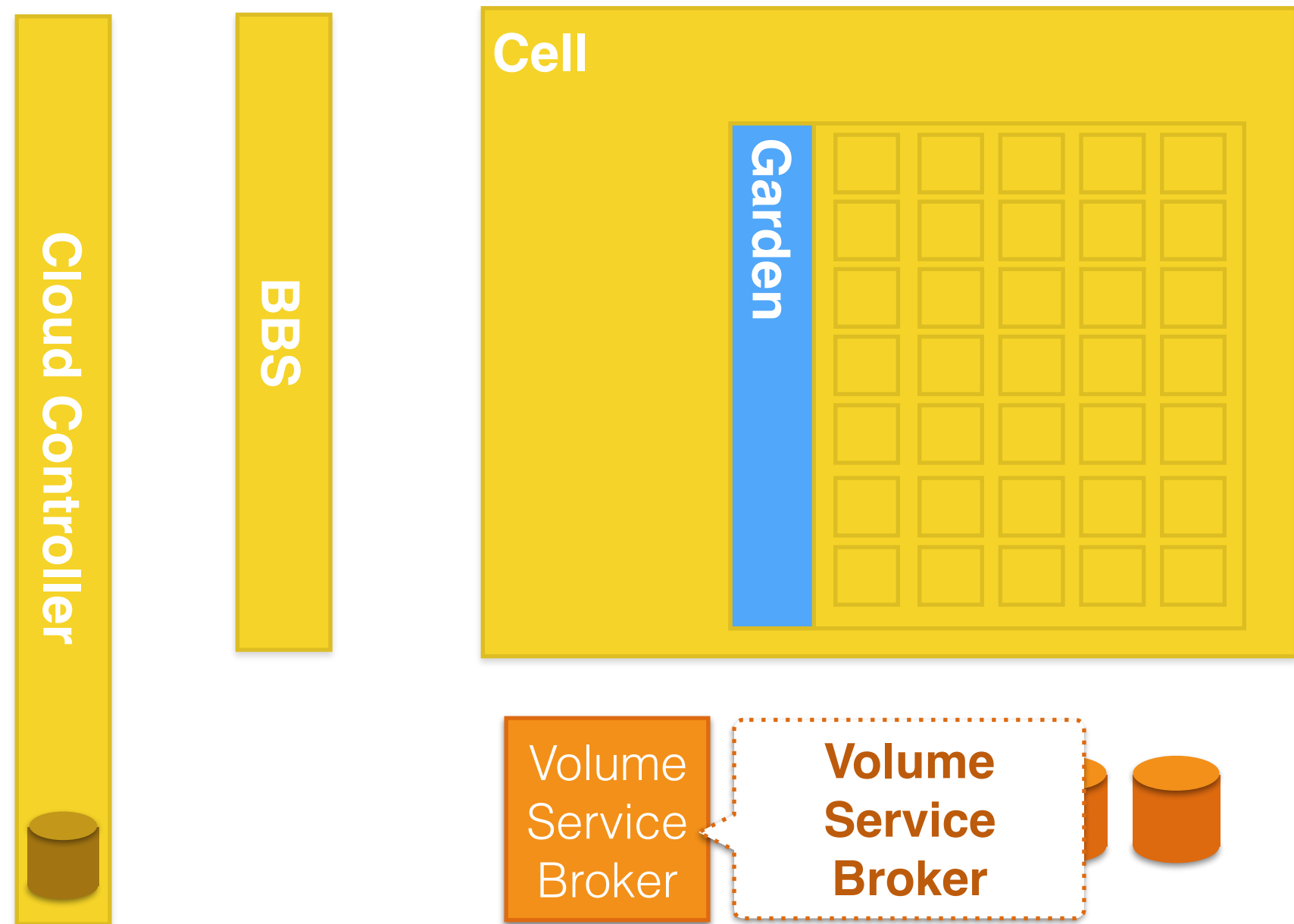
# Architecture



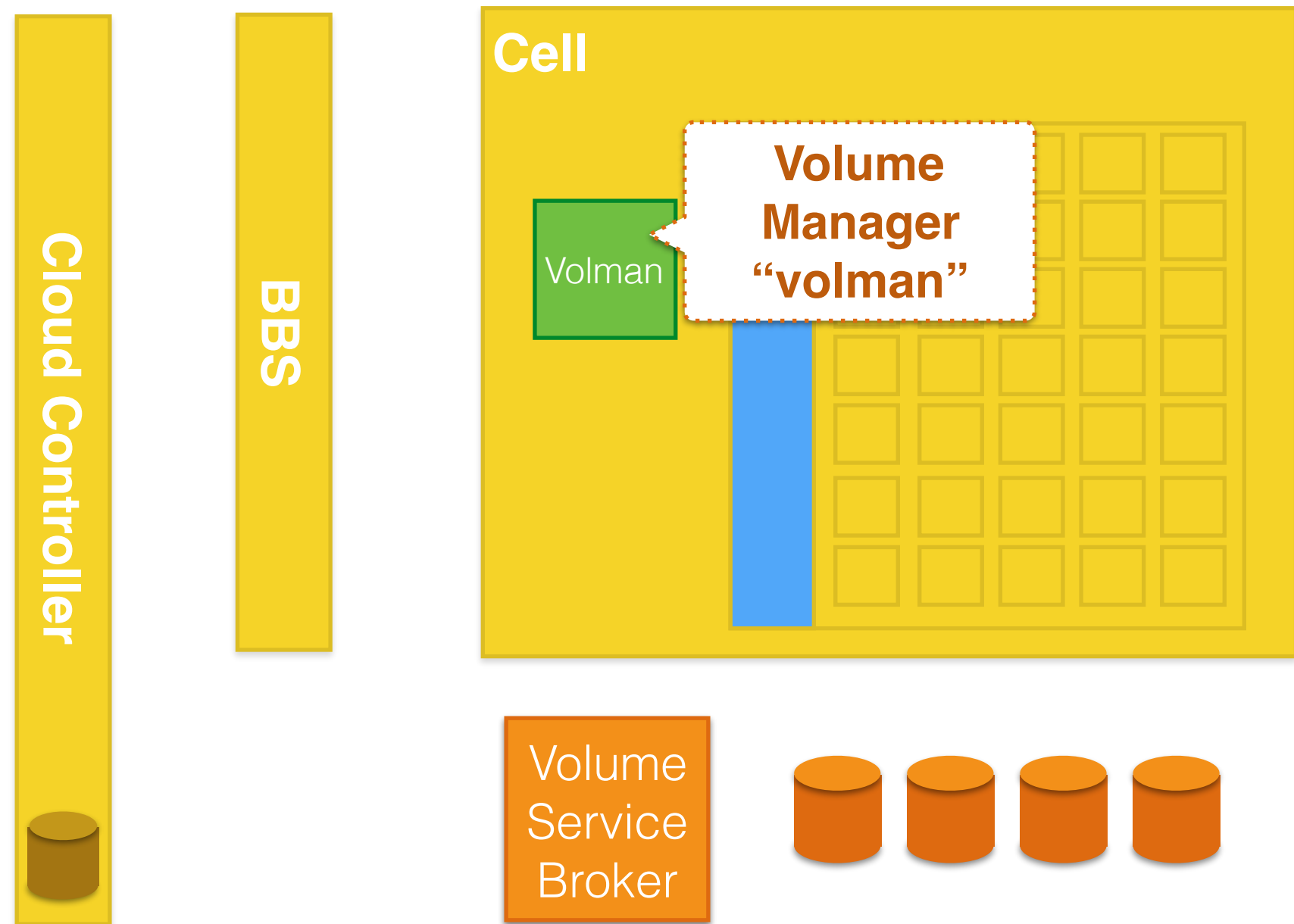
# Architecture



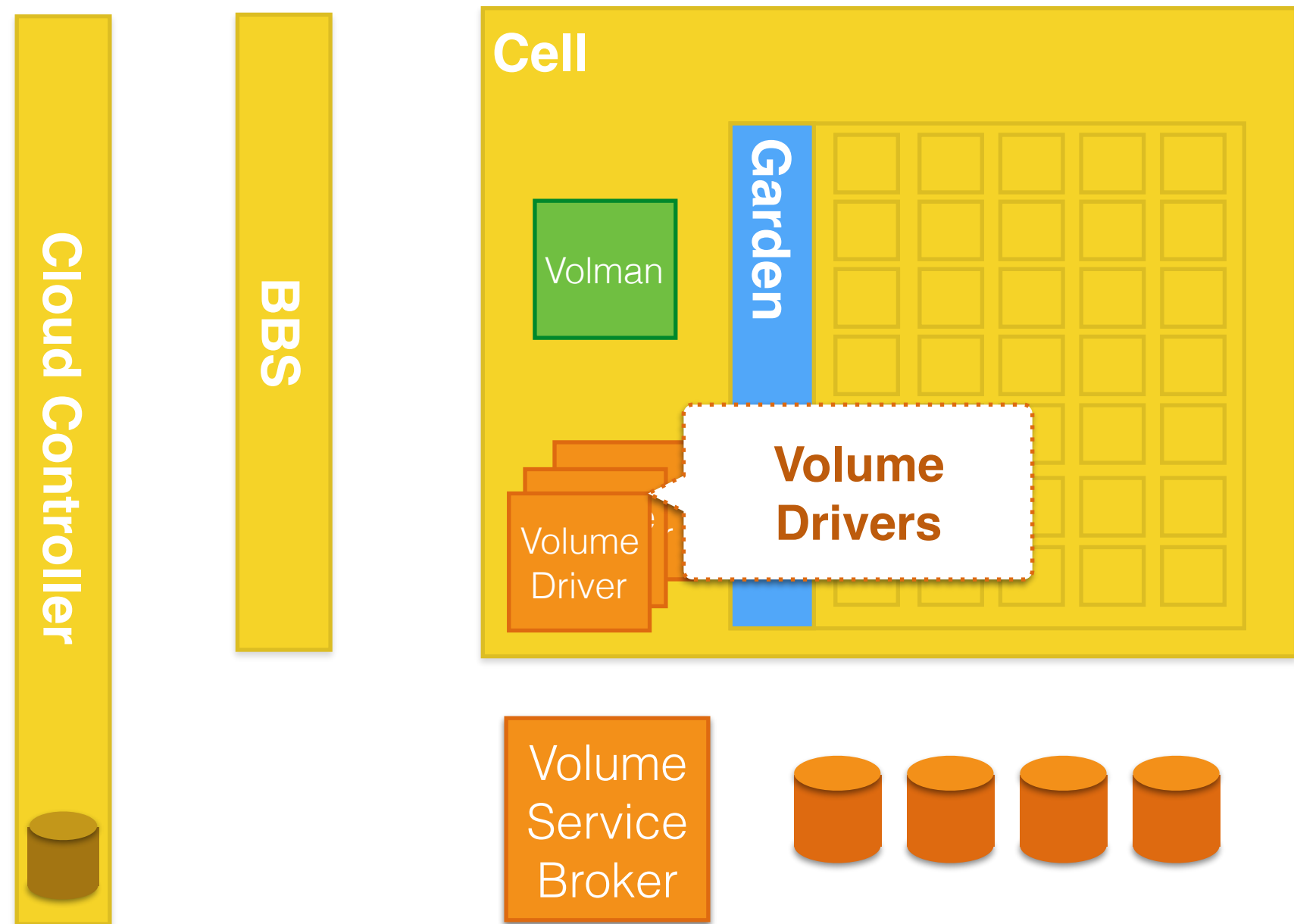
# Architecture



# Architecture

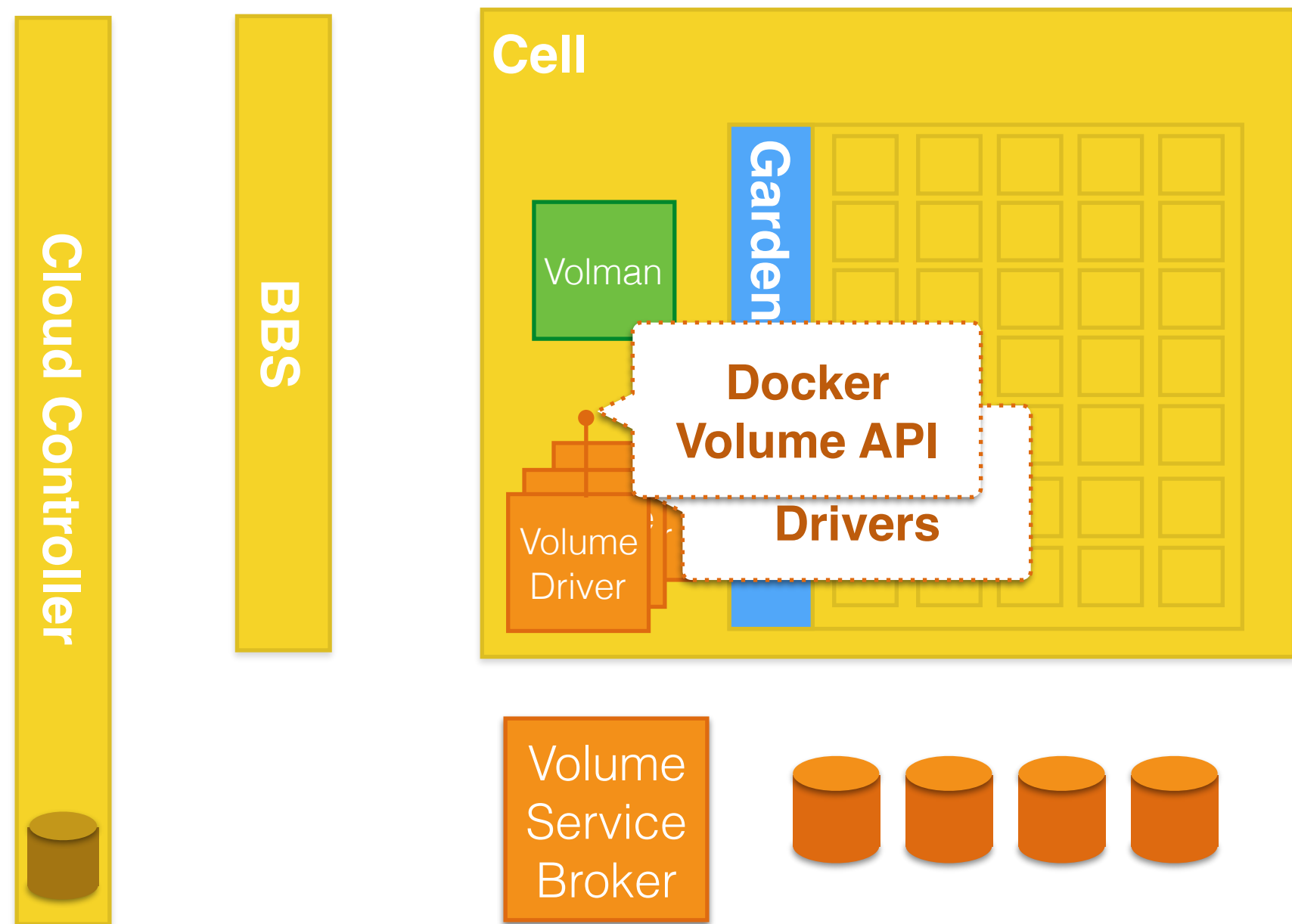


# Architecture



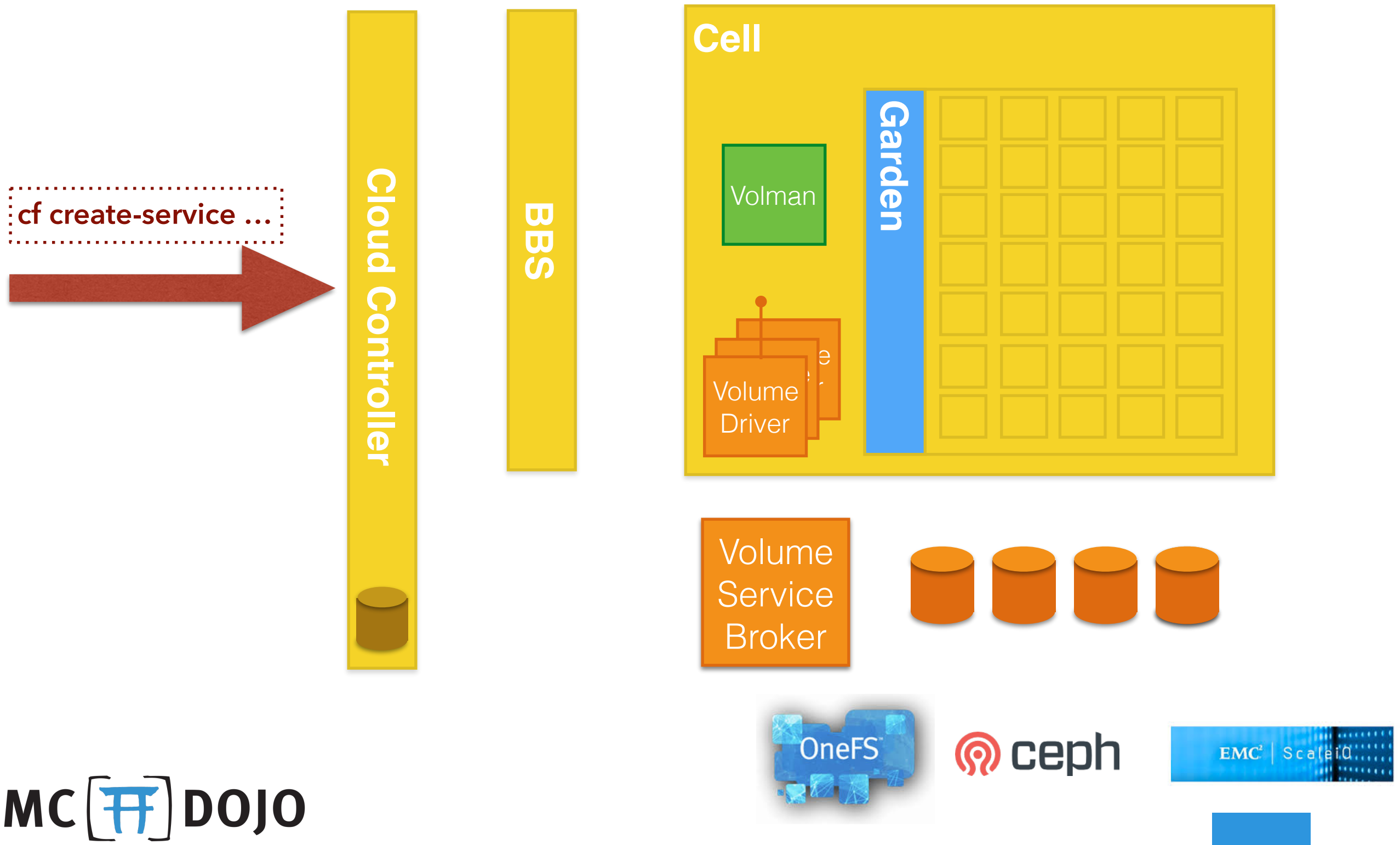


# Architecture

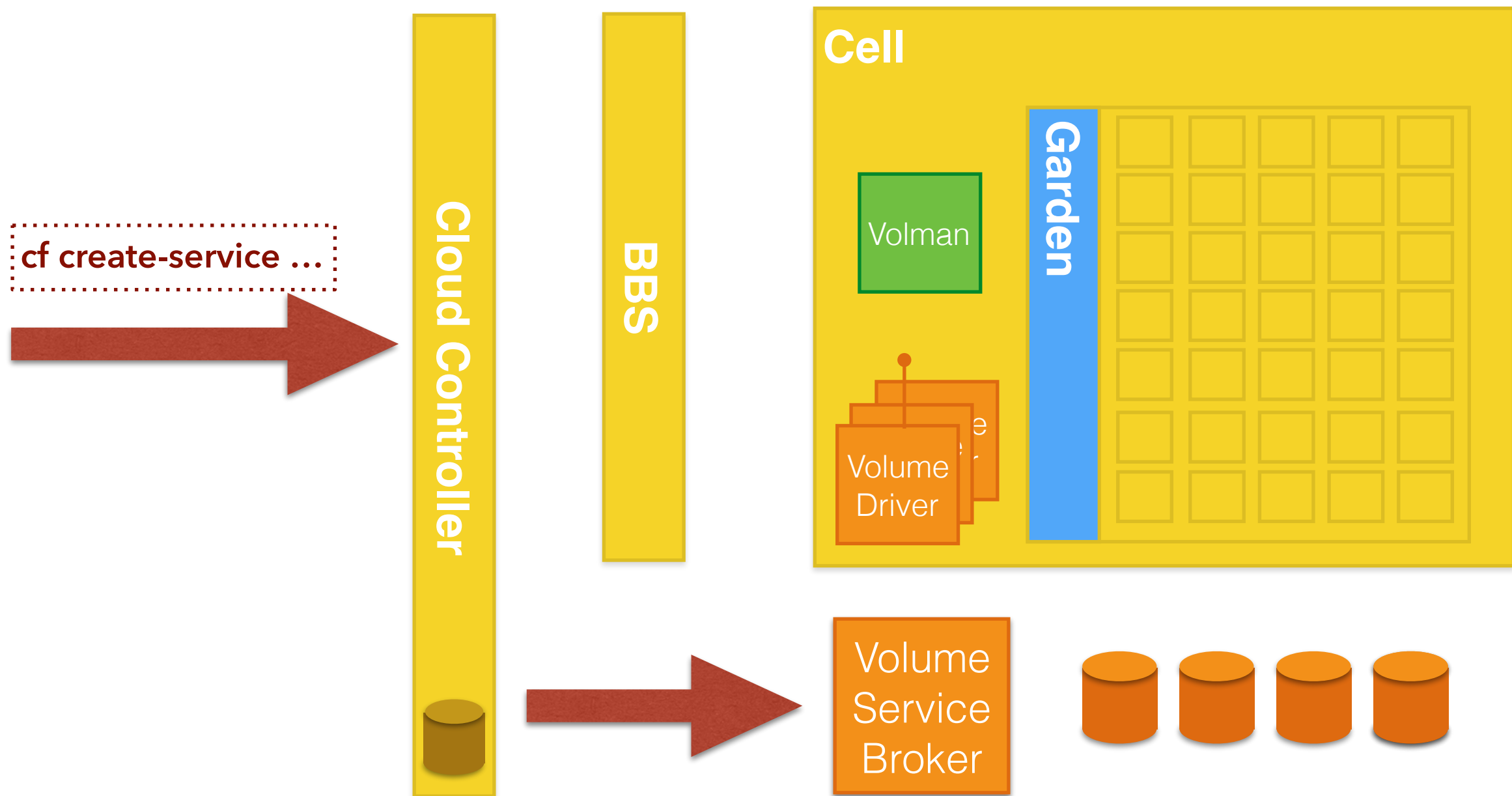


`cf create service`

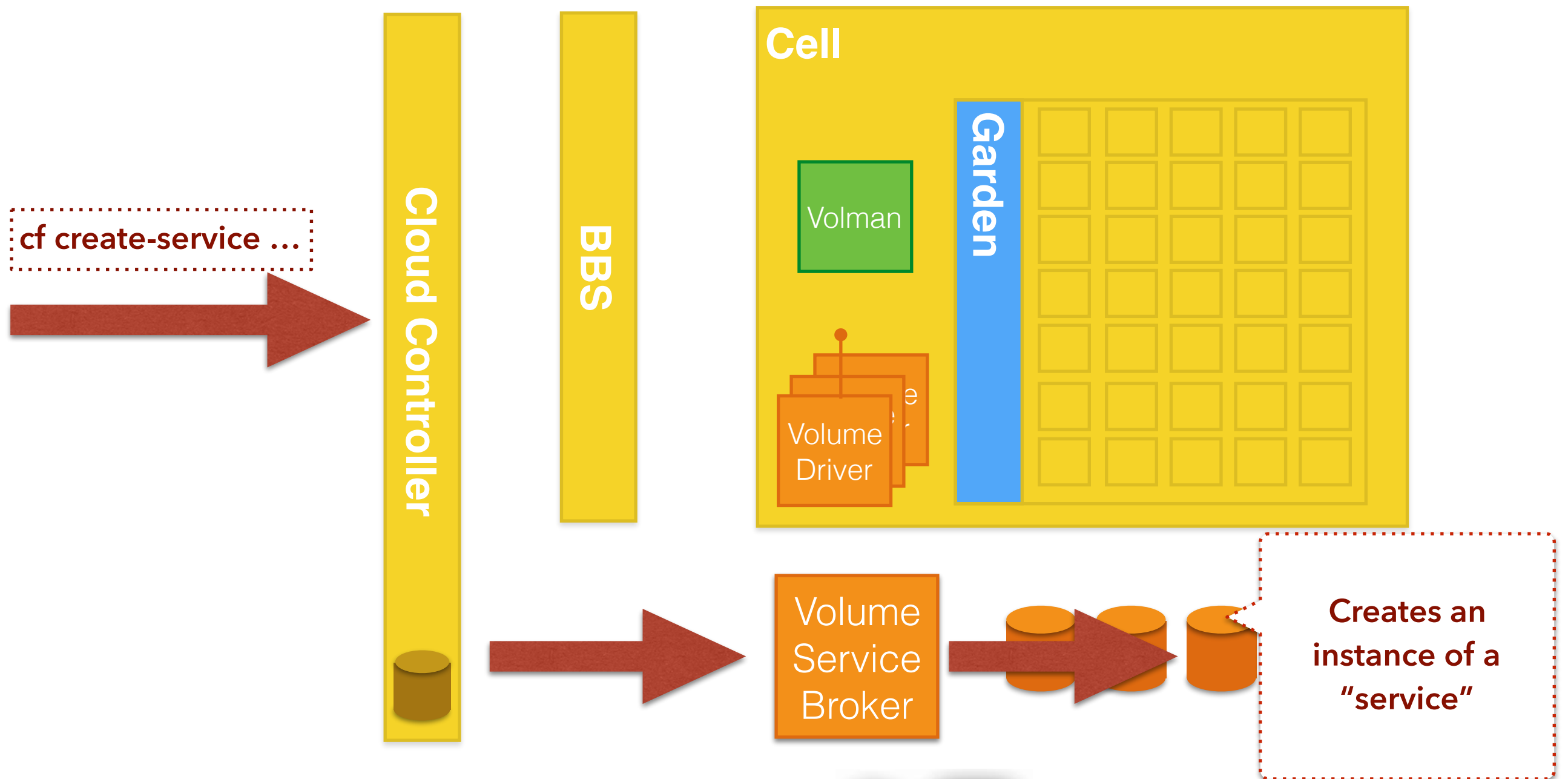
# Architecture



# Architecture

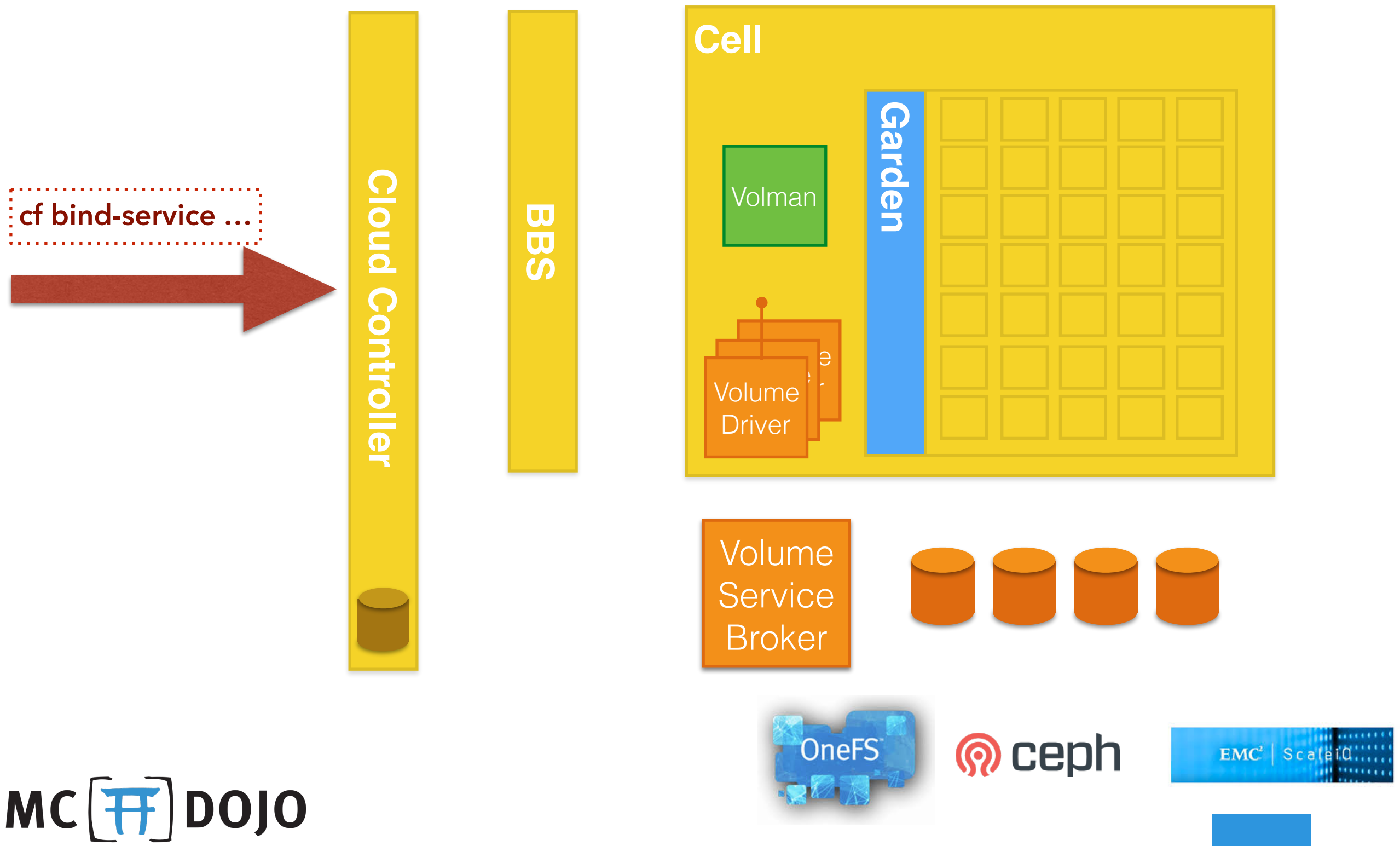


# Architecture

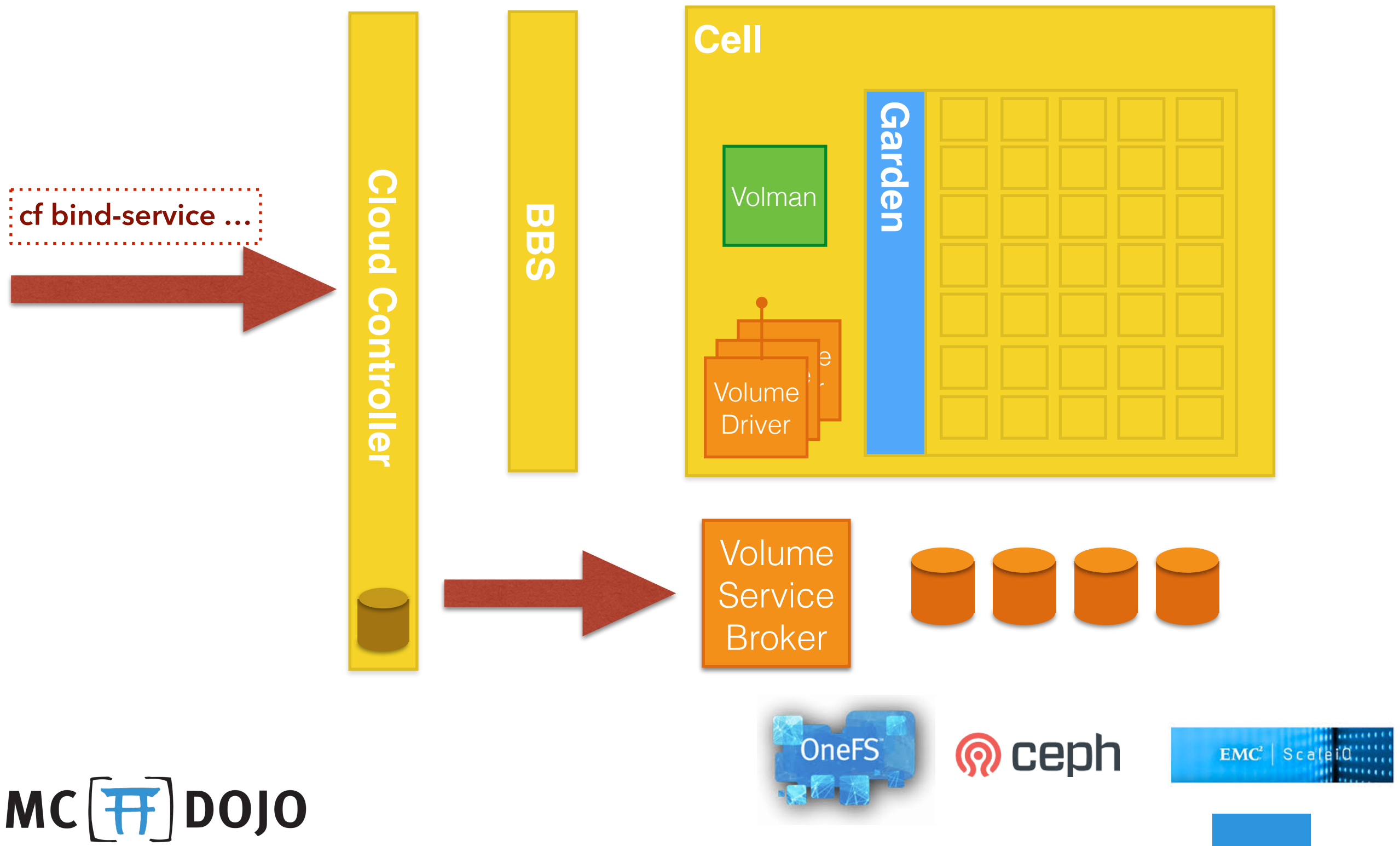


`cf bind-service`

# Architecture

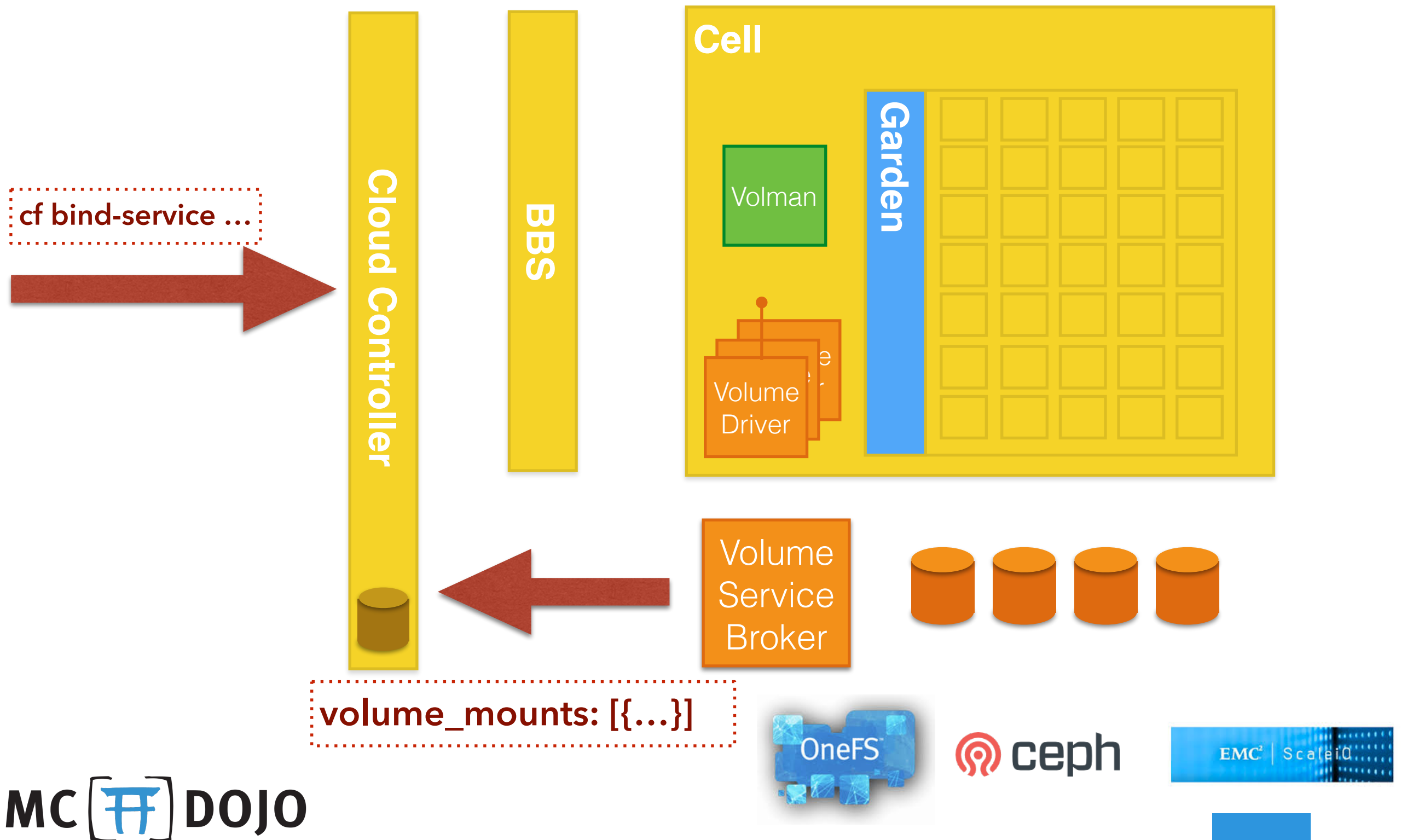


# Architecture

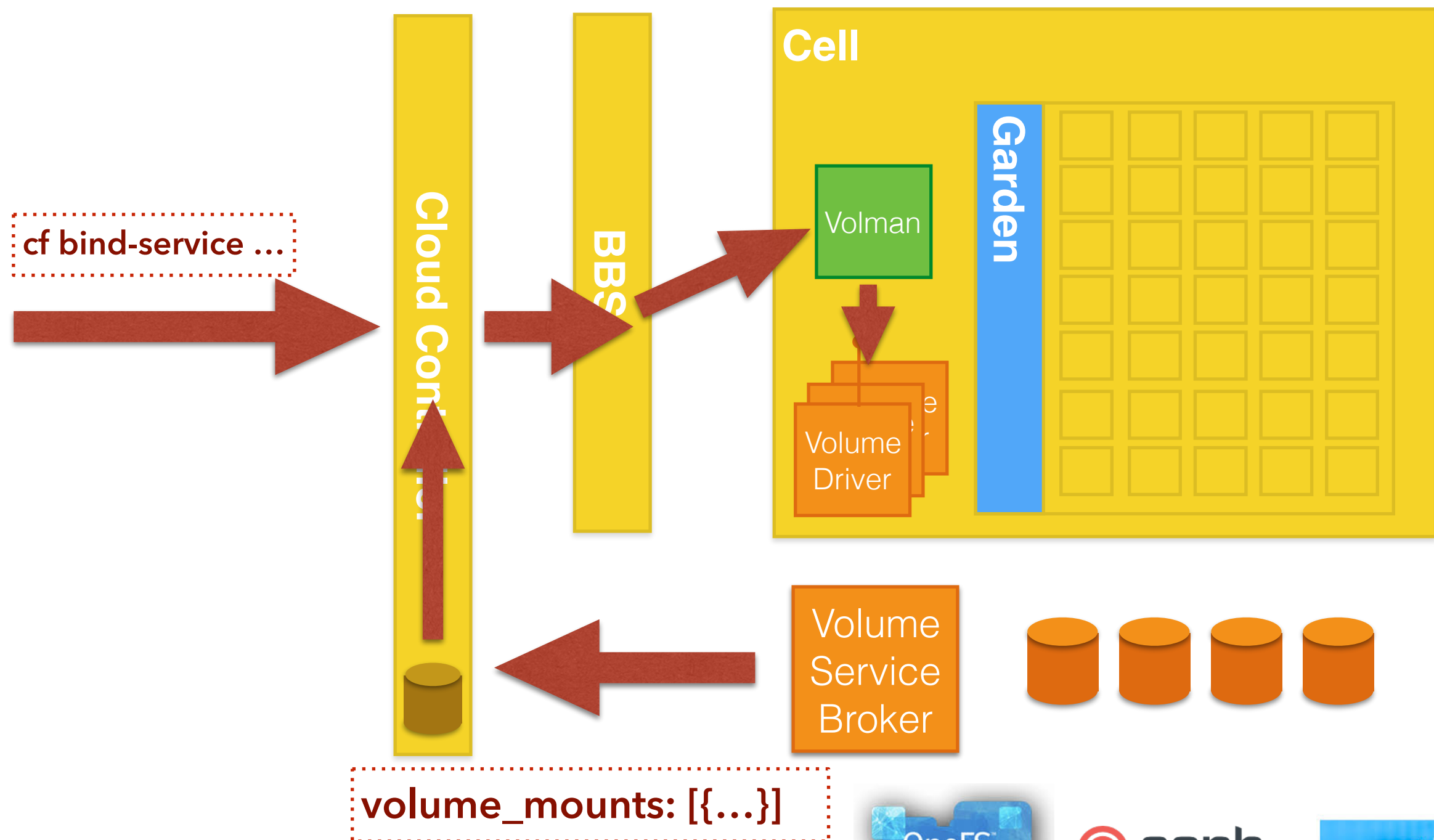




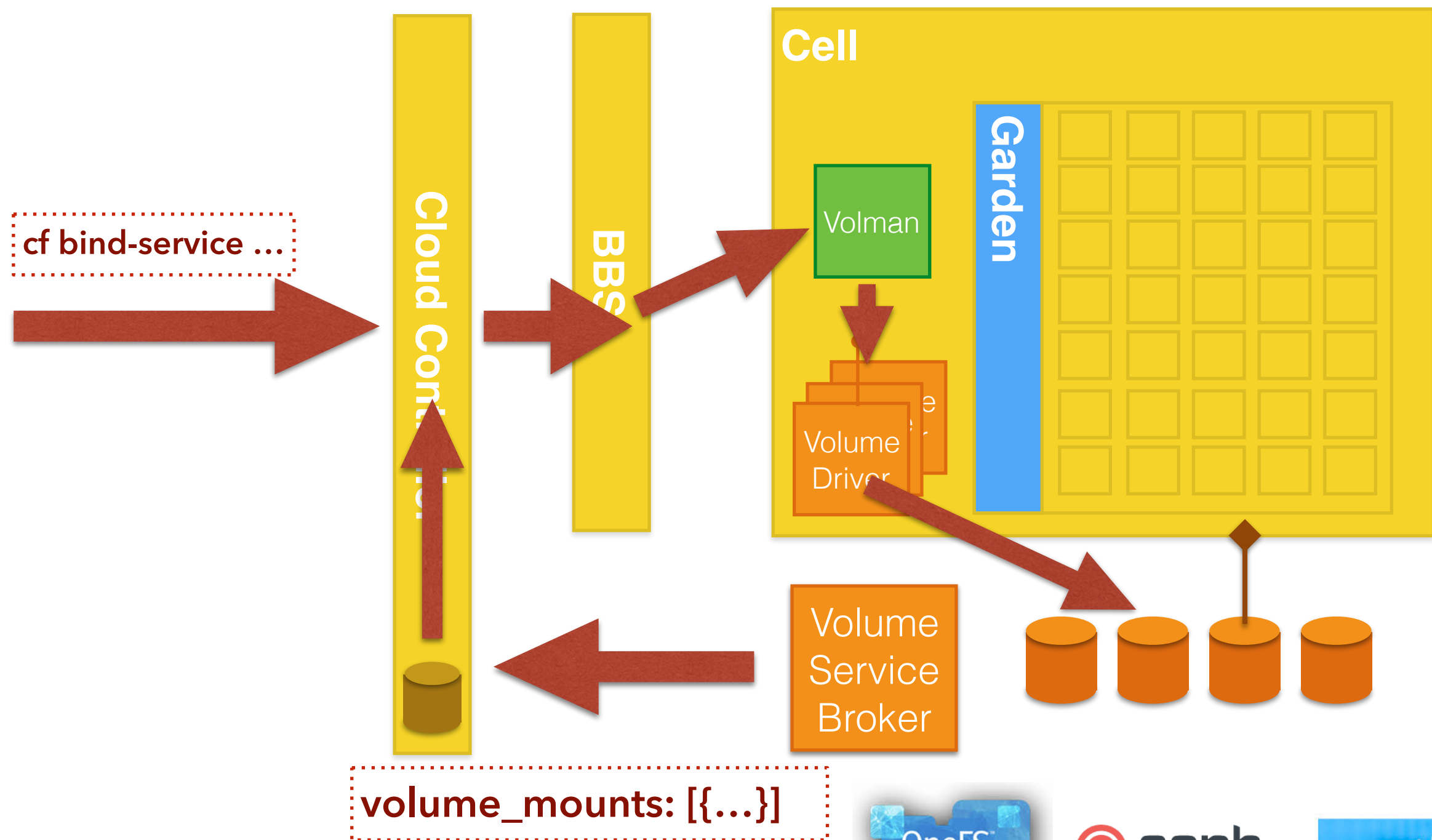
# Architecture



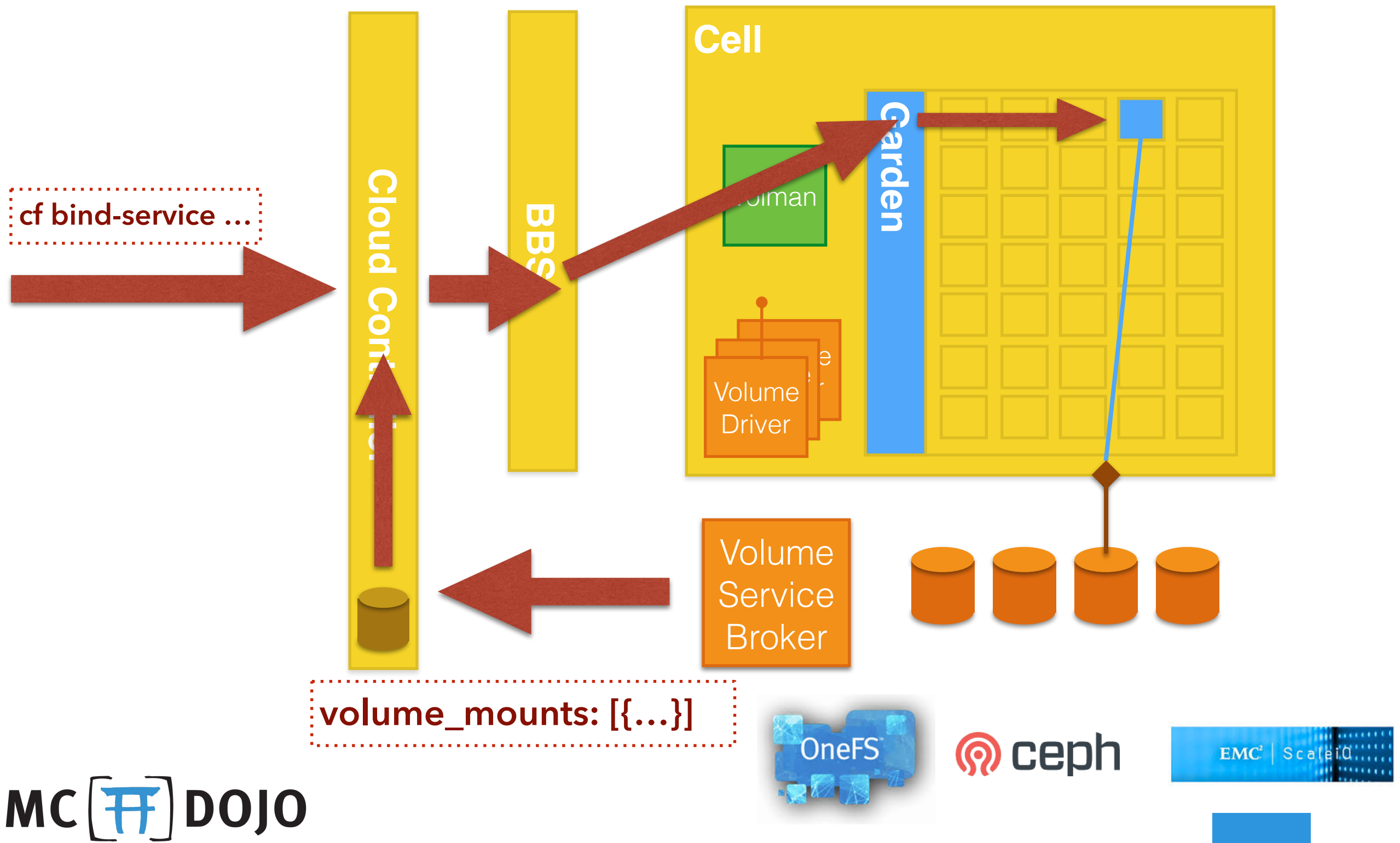
# Architecture



# Architecture



# Architecture



# Futures

- Persistence on CloudFoundry
  - Block devices
    - Requires “Consistent Scheduling” capability
    - ``cf push mysql``
- Spring Content
  - Separate Spring Data BLOB project
  - Make JPA do more sane things!
  - Eventing around getting/setting content
  - Continue to add CM like features like retention services

# Thanks.

- Questions?

[paul.warren@emc.com](mailto:paul.warren@emc.com)

 #persi

[github.com/cloudfoundry-incubator/local-volume-release](https://github.com/cloudfoundry-incubator/local-volume-release)

[github.com/emc-dojo/cf-persist-service-broker](https://github.com/emc-dojo/cf-persist-service-broker)

[github.com/emc-dojo/spring-content](https://github.com/emc-dojo/spring-content)