



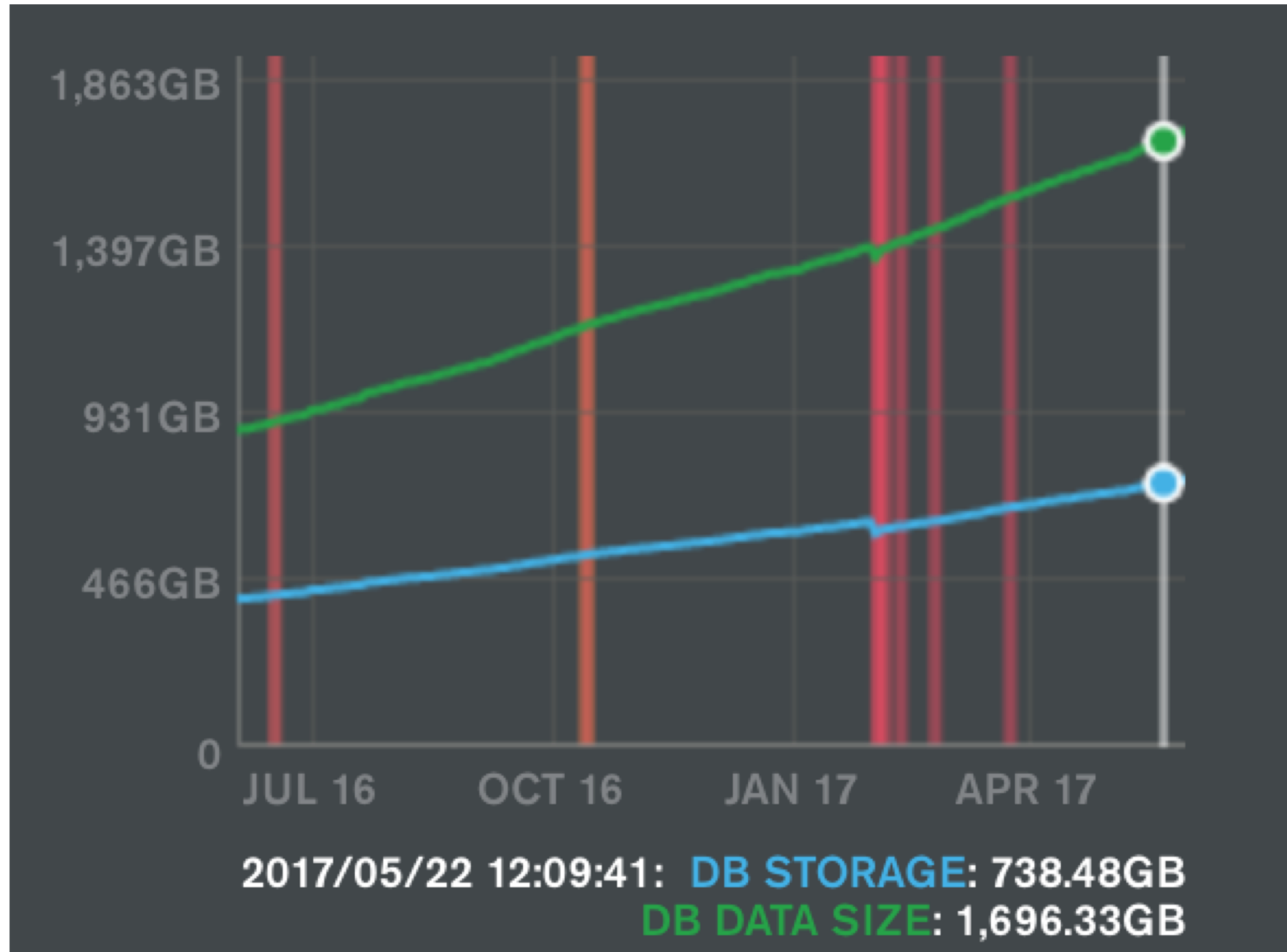
Reduce Open edX Mongo Disk Usage

Modulestore Structure Pruning

Presented by:

Elton Carr, II (Microsoft)

David Ormsbee (edX)



Setting: Open edX Ops Hangout

We noticed that course content uses a lot of space on MongoDB and it is costing us a lot...

Yeah... that's because we keep every edit version of every course and library (structure docs).

... Let's fix that.

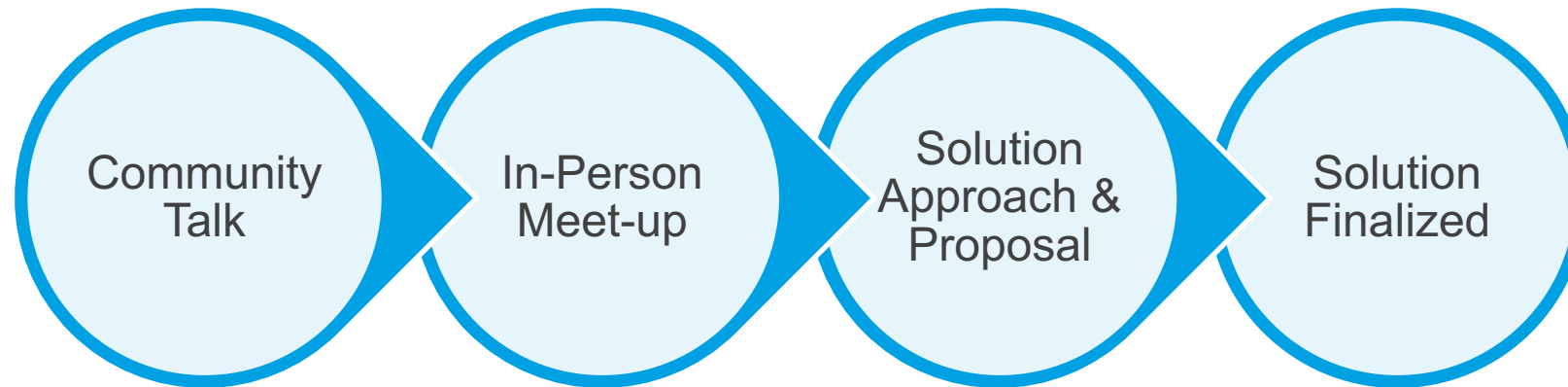


Collaborating From Ideation to Solution

Microsoft	edX
Elton Carr, II Manikarthik Konakalla	Kevin Falcone Joseph Mulloy David Ormsbee



<https://github.com/edx/tubular/pull/185>



- Open edX Ops Hangout
- [PLAT-697](#)



- Pull Request [#184](#)

- Refactored for performance & memory
- Pull Request [#185](#)

```
python ./structures.py make_plan plan.json
```

edX = 5M structures in 2 hours

Microsoft = 136K structures in 15 minutes

```
python ./structures.py prune plan.json
```

edX = 5M structures in 8 hours

Microsoft = 136K structures in 1 hour

Real Usage and Source

<https://github.com/edx/jenkins-job-dsl/blob/master/devops/resources/mongo-pruner.sh>

<https://github.com/edx/tubular/blob/master/tubular/scripts/structures.py>

Operational Safety

- First prune is the hardest
- Can run without downtime
 - Test and tune `--batch-size` and `--delay`
 - Pruning from old plan files is safe
- Sync new MongoDB instances for immediate gains
- `make_plan` command is read-only
- `prune` command can be interrupted/restarted
 - Monitor MongoDB memory usage

MongoDB Disk Usage Impact

Microsoft

- Logical: 96 GB → 16 GB (-83%)
- Compressed: **10.3 GB → 5.9 GB (-43%)**

edX.org

- Logical: 2,126 GB → 380 GB (-82%)
- Compressed: **866 GB → 272 GB (-69%)**

No Better Time Than the Present



Use ModuleStore Structure Pruning to save space



Our community is strengthened when we collaborate



Engaging early is an absolute plus