









Cloud-based horizontal scaling of queue-driven

- Presented by Job Verkaik and Mattijn van Hoek (HKV)









Summary

- We can help so you can focus on your core business.
- Make use of **horizontal scalability** and **elasticity** in the cloud.
- Scale FSS up, but especially scale down; Exhaling clouds. In Production.
- Create synergy between your database, Admin Interface API and K8s.
- Reduce spending, reduce energy.
- We **love** to cooperate.



What we want

Can we **control** the number of Forecasting Shell Services (**FSS**) within a **cloud-based** Delft-FEWS setup?

Especially, can we scale down resources when not needed?

Reflection on cloud services & auto scaling

What we did

Benefits & Limitations

Way Forward

Reflection on Cloud Services & Auto-Scaling

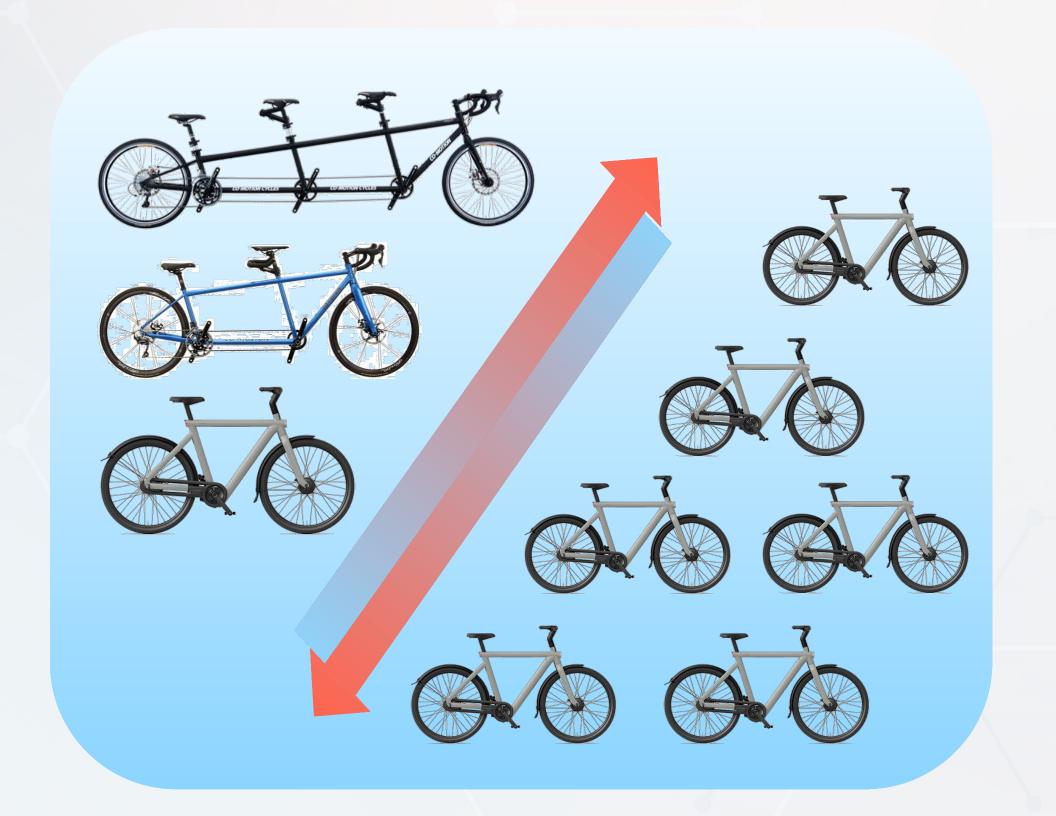


From on-premise to cloud

- Lift and shift. Still static
- Reduce Maintenance Burden
 - Reduce Organizational Costs
- Reduce Computational

Imbalance

Reflection on Cloud Services & Auto-Scaling



Auto-scaling?

- Vertical Scaling (\$ size)
- Horizontal Scaling (⇔instances)
- Scale FSS by CPU, memory or as workload. Cloud thrashing.
- Scale FSS by tasks in FEWS queue

Flowchart What we did Timer - Remove workload. If idle - Add node*. True NOT pendin FSS - Add start Remove FSS end stop workload group from group tasks FEWS-MC False * Using clever bookkeeping

FEWS Admin Interface API KubeCTL/Azure API

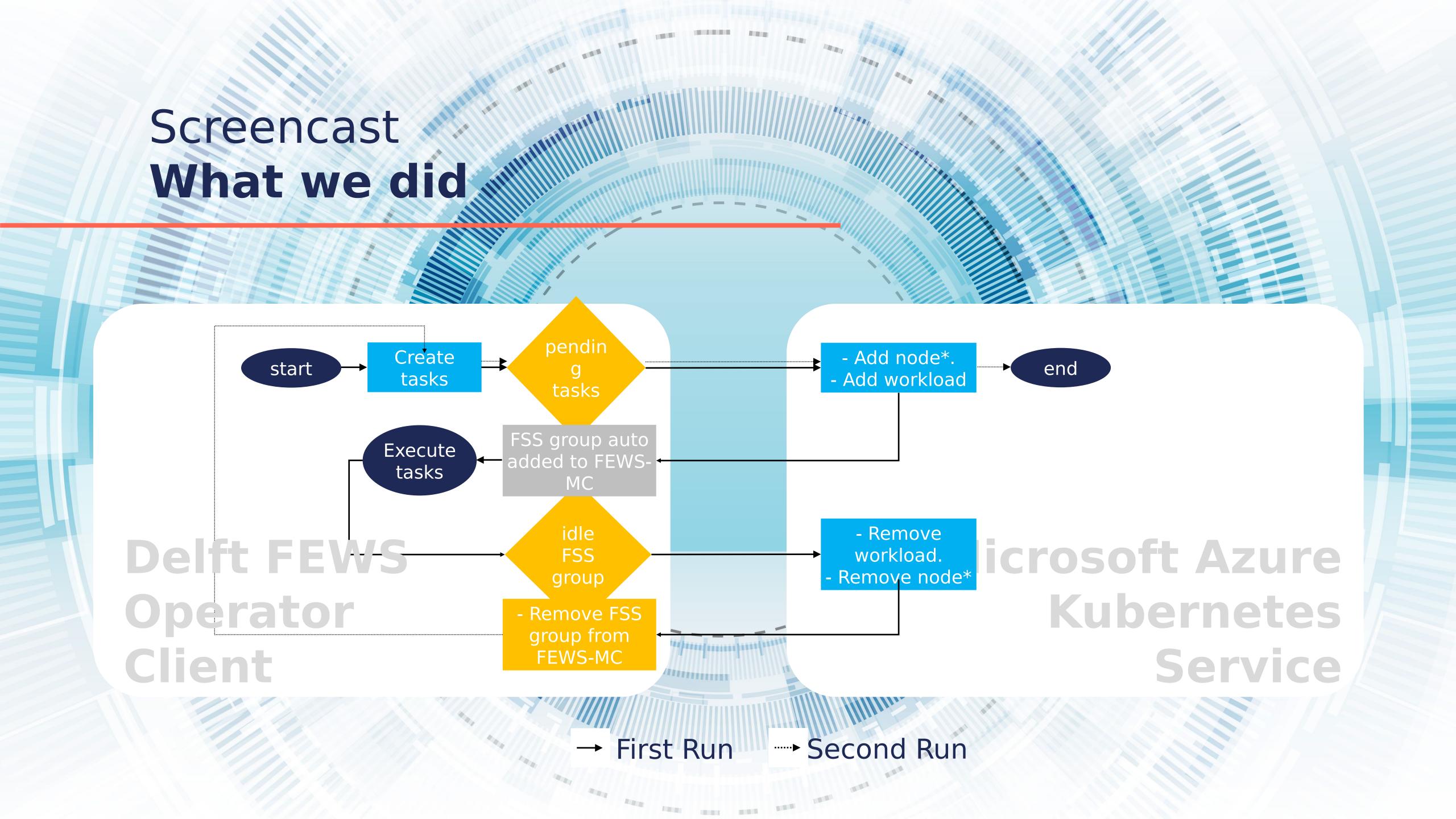
Screencast What we did

Left panel will show

- Delft FEWS
Operator
Client

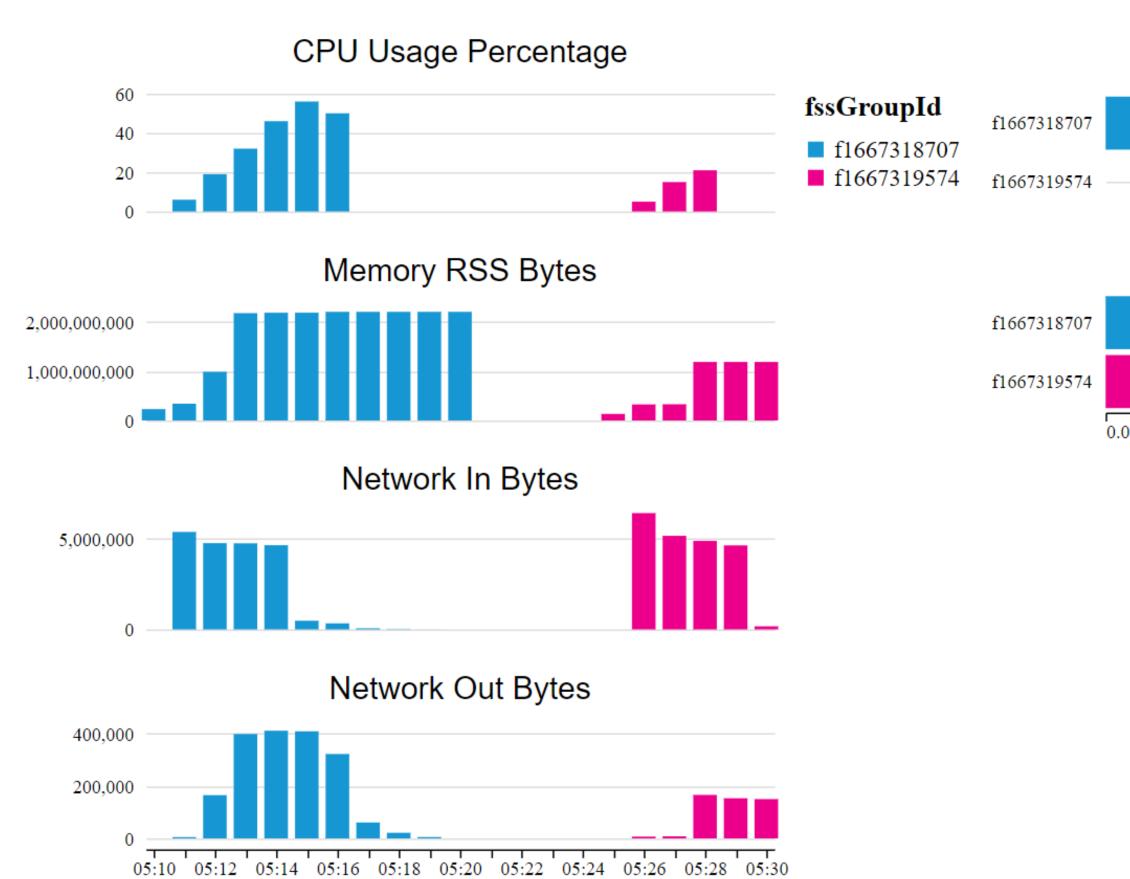
Right panel will show

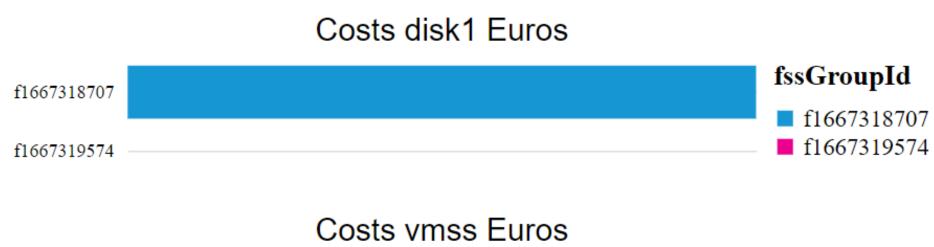
- Few seconds Delft FEWS MC
 - Microsoft
 Azure
 Kubernetes
 Service
- Few seconds Delft FEWS MC

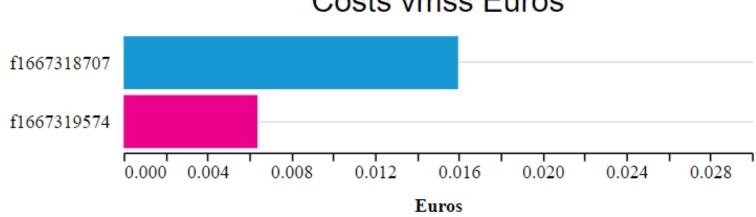


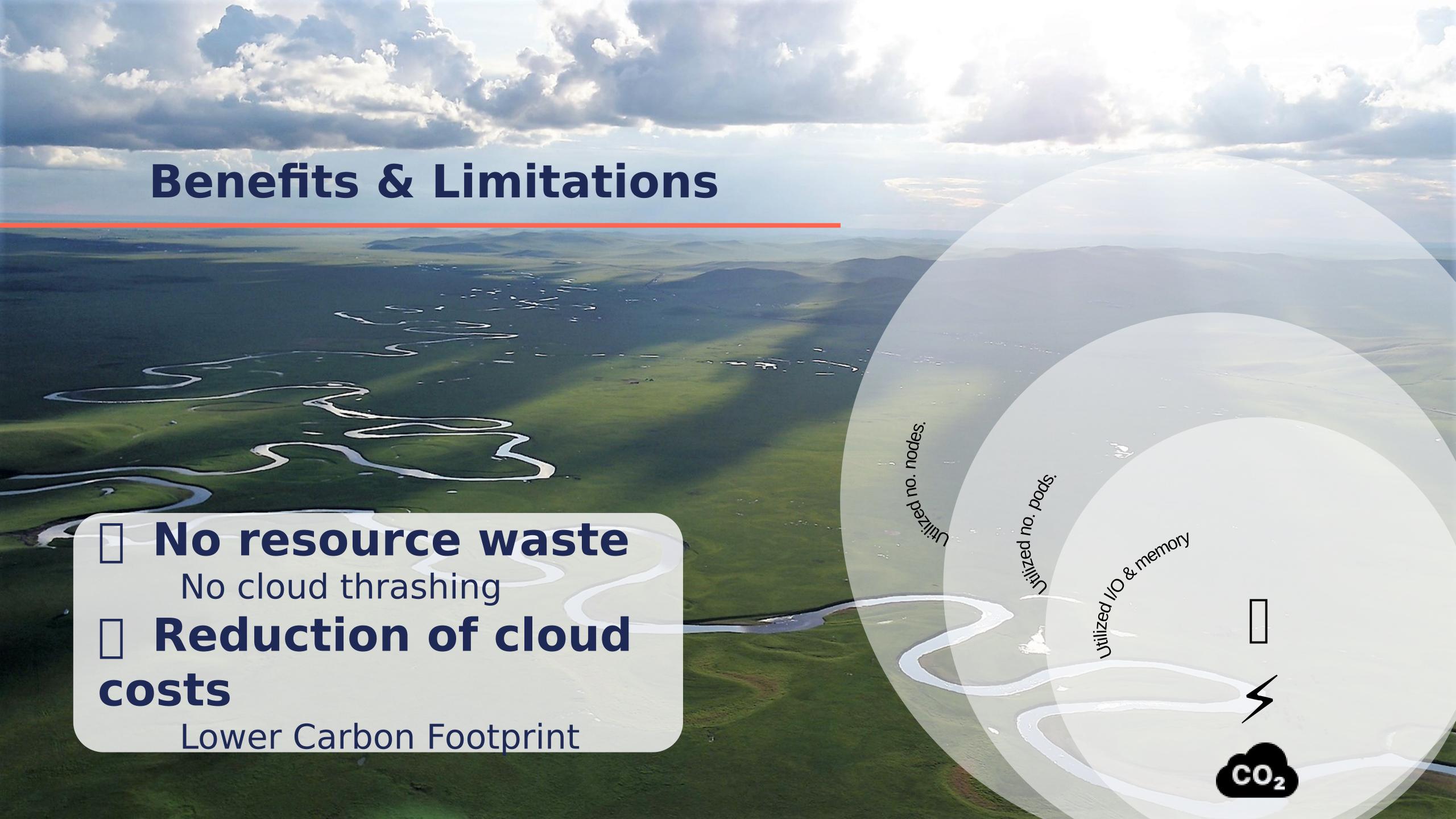


Stats What we did









Benefits & Limitations

Latencies

Time to add FSS: 5 minutes

Time to remove FSS: 3 minutes

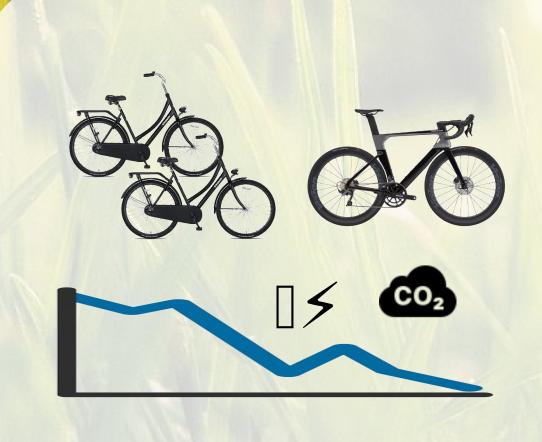
Type Tasks

Suitable for heavy tasks in parallel

- Short tasks can result in interference.
- Once FSS is up, queue is empty



Way Forward





Get ahead of the curve We invite you to discuss

- Reduce Maintenance Burden
- Reduce Organizational Costs
- Reduce Computational Imbalance

Make FEWS WorkFlows Cloud Native

Can it?

Proactive vs Reactive scaling

Wrt latencies listen to queue or

Meet the team

We believe in the synergy between Delft-FEWS systems and Cloud Infrastructure. You too?



Job Verkaik verkaik@hkv.nl



Mattijn van Hoek hoek@hkv.nl



Rolf Waterman



Marc Jansen

