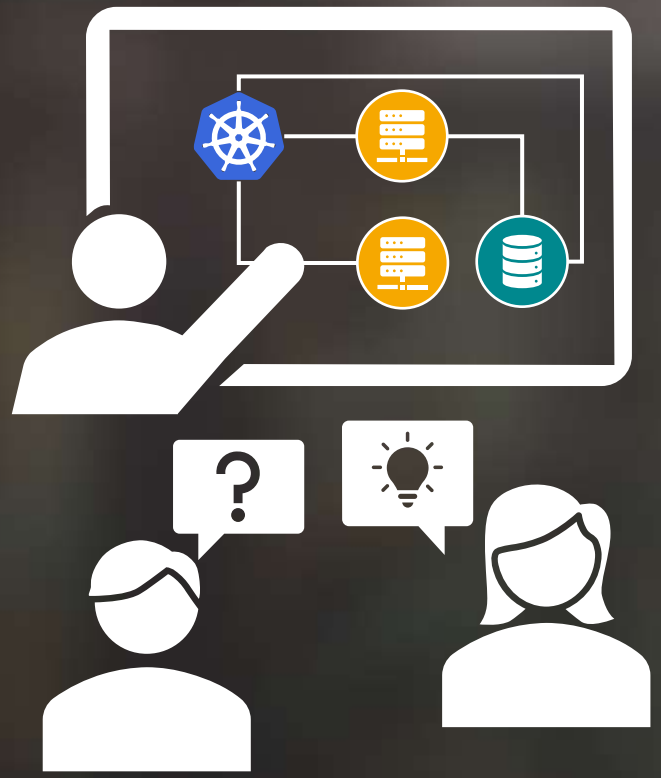


What we do

What a CFO might think we do

CFO = ~~Chief Finance Officer~~ Cloud Funding Officer 😉

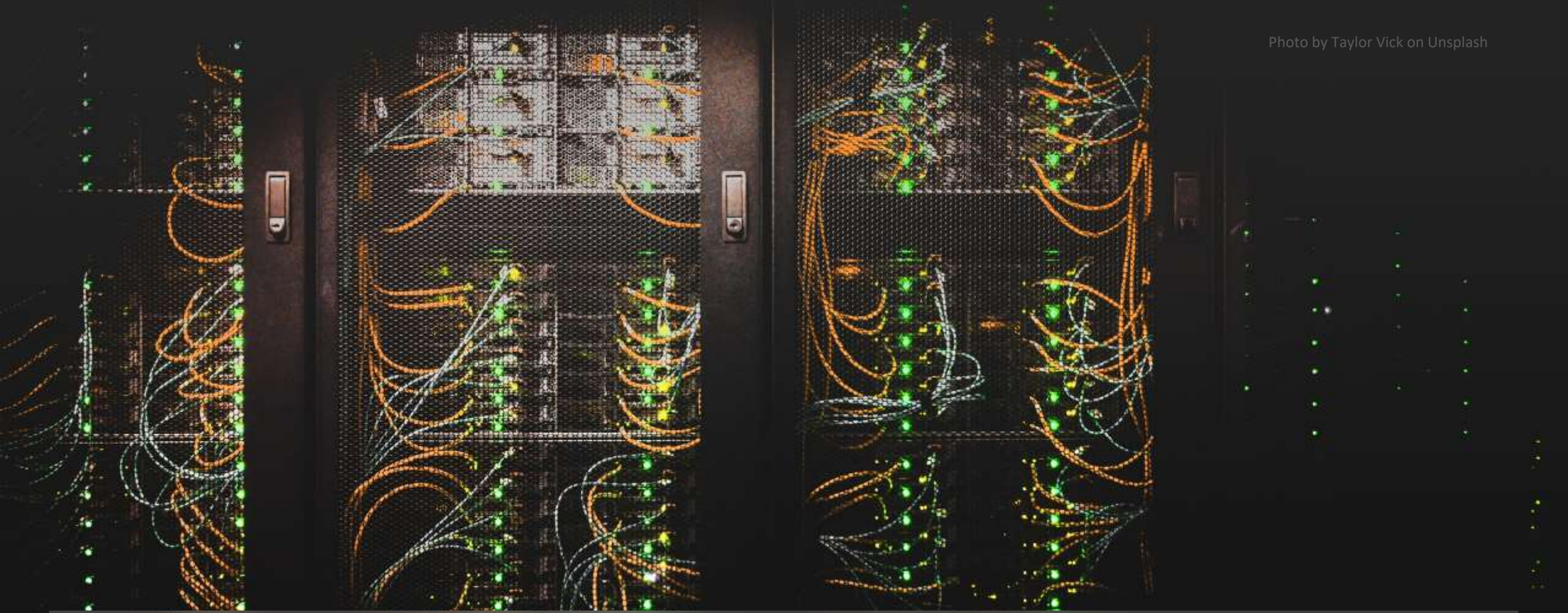


Photo by Taylor Vick on Unsplash

Cloud Cost Estimation & Optimization

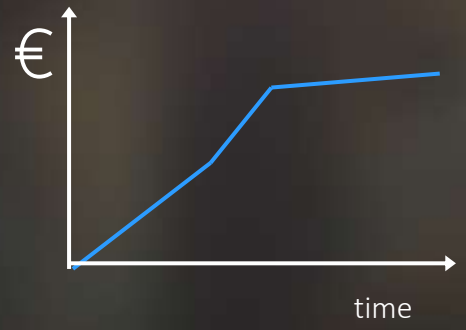
GI Arbeitskreis Micro Services & DevOps | 13th/14th September 2022

plewnia@swc.rwth-aachen.de Christian_Plewnia2 christianplewnia Christian_Plewnia

Software Construction | RWTH AACHEN UNIVERSITY

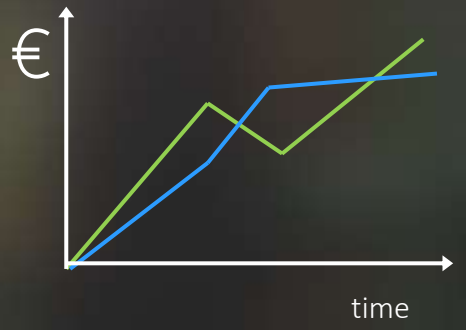
1

Plan



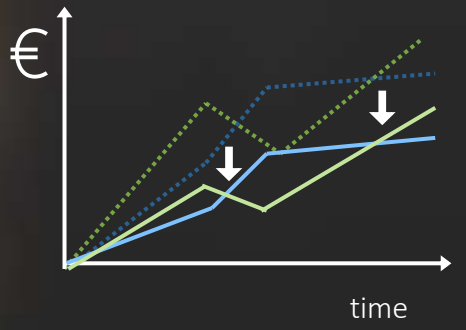
2

Compare



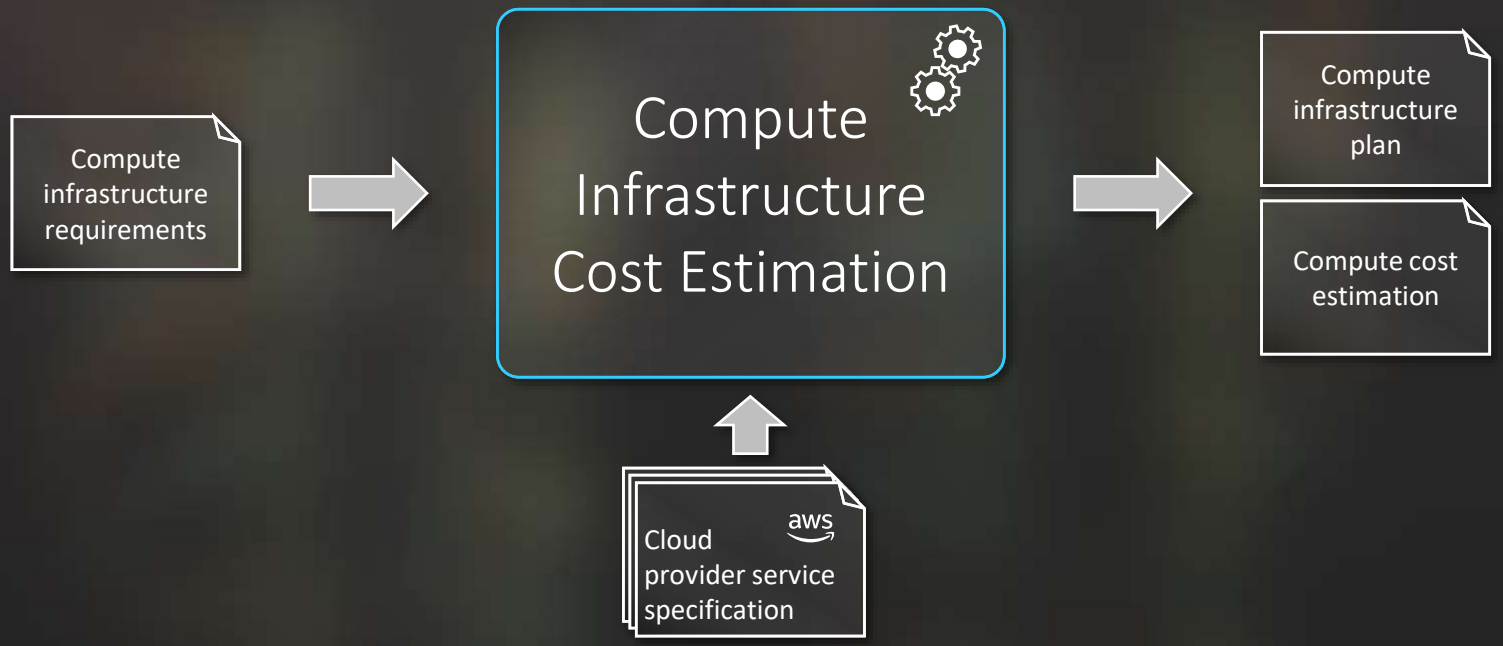
3

Optimize




1
Plan Cloud Compute Infrastructure Costs

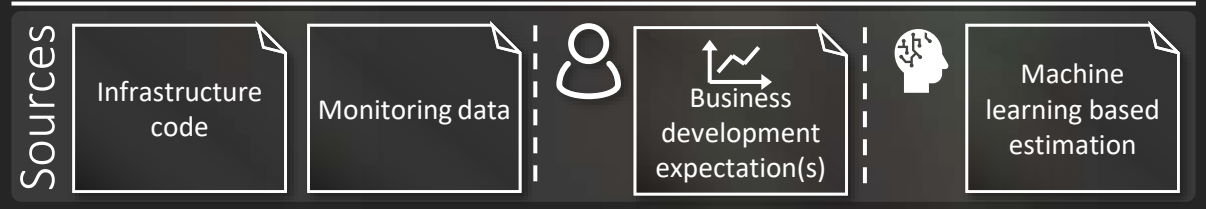
How does the planning work?



Specifications

- What compute infrastructure is required?
  - What types of compute infrastructure are needed? (virtual machines, storage, load balancers)
  - What properties must the compute infrastructure have?
    - Technical properties (e.g., for VMs the CPUs, RAM, GPUs; for storage: size and access performance)
    - Non-technical properties: hosting location, certification
  - How much network traffic is expected to occur?
- How are the requirements expected to develop over time?

Compute infrastructure requirements



- What cloud services are offered?
  - Concrete services (e.g. virtual machines „AWS EC2 t3.large“)
  - Technical and non-technical properties of the services
- What are the pricing models for the offered services?

Cloud provider service specification

Specifications

```yaml
profile: "deployment"
profileVersion: 211120
configuration:
  timeSpan: "1 Month"
virtualMachine:
  InformationSystemCluster:
    requirements:
      RAM: 377 GB
      vCPU: 72
      Storage: 800 GB
    properties:
      provider: "AZURE"
      location: "germany-north"
      commitment: "1YearReserved"
    workload:
      baselineInstances: 8
      spikes:
       - peak:
           daily:
             additionalInstances: 3
             duration:
               start: "16:00:00"
               end: "20:00:00"
       - prePeak1:
           daily:
             additionalInstances: 3
             duration:
               start: "5:00:00"
               end: "23:30:00"
       - prePeak2:
           daily:
             additionalInstances: 5
             duration:
               start: "6:00:00"
               end: "22:30:00"
       - prePeak3:
           daily:
             additionalInstances: 4
             duration:
               start: "8:00:00"
               end: "21:30:00"
       - prePeak4:
           daily:
             additionalInstances: 1
             duration:
               start: "10:00:00"
               end: "20:30:00"

    network:
      inbound:
        - vm+FileServer:
            # 73 Gb/Day
            volume: "73 Gb"
            timeSpan: "1 Day"
```

Compute infrastructure requirements

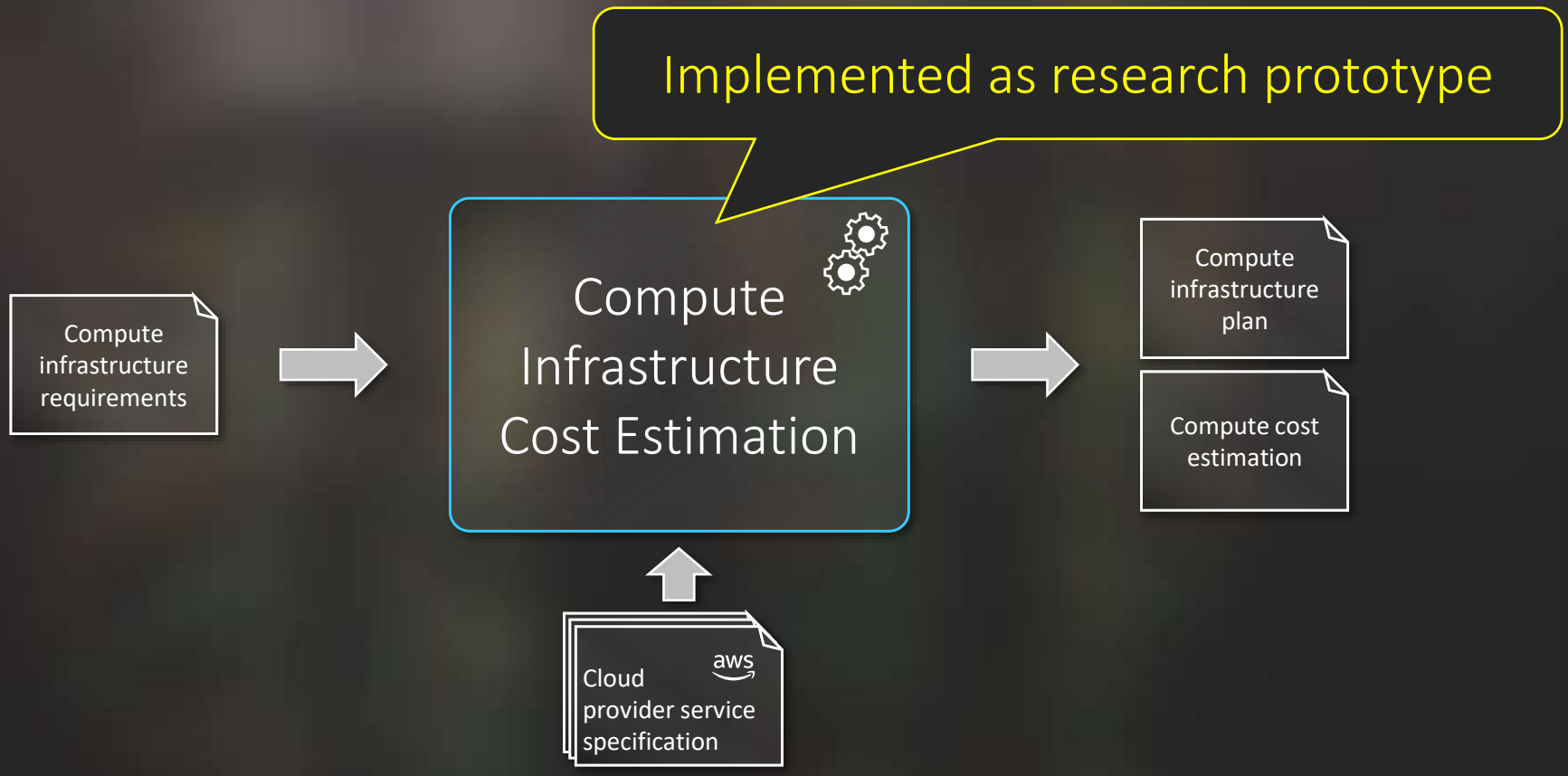
small snippet of the specification

```yaml
AZURE:
  blockStorage:
    germany-north:
      E1:
        capacity:
          options:
          - 4 Gib
        fixedRate:
          cost: 0.39
          currency: USD
          timeSpan: 1 Month
          unit: 4 Gb
        properties:
          provider: AZURE
          storageType: SSD
      E10:
        capacity:
          options:
          - 128 Gib
        fixedRate:
          cost: 12.48
          currency: USD
          timeSpan: 1 Month
          unit: 128 Gib
        properties:
          provider: AZURE
          storageType: SSD
      E15:
        capacity:
          options:
          - 256 Gib
        fixedRate:
          cost: 24.96
          currency: USD
          timeSpan: 1 Month
          unit: 256 Gib
        properties:
          provider: AZURE
          storageType: SSD
```

Cloud provider service specification

small snippet of the specification

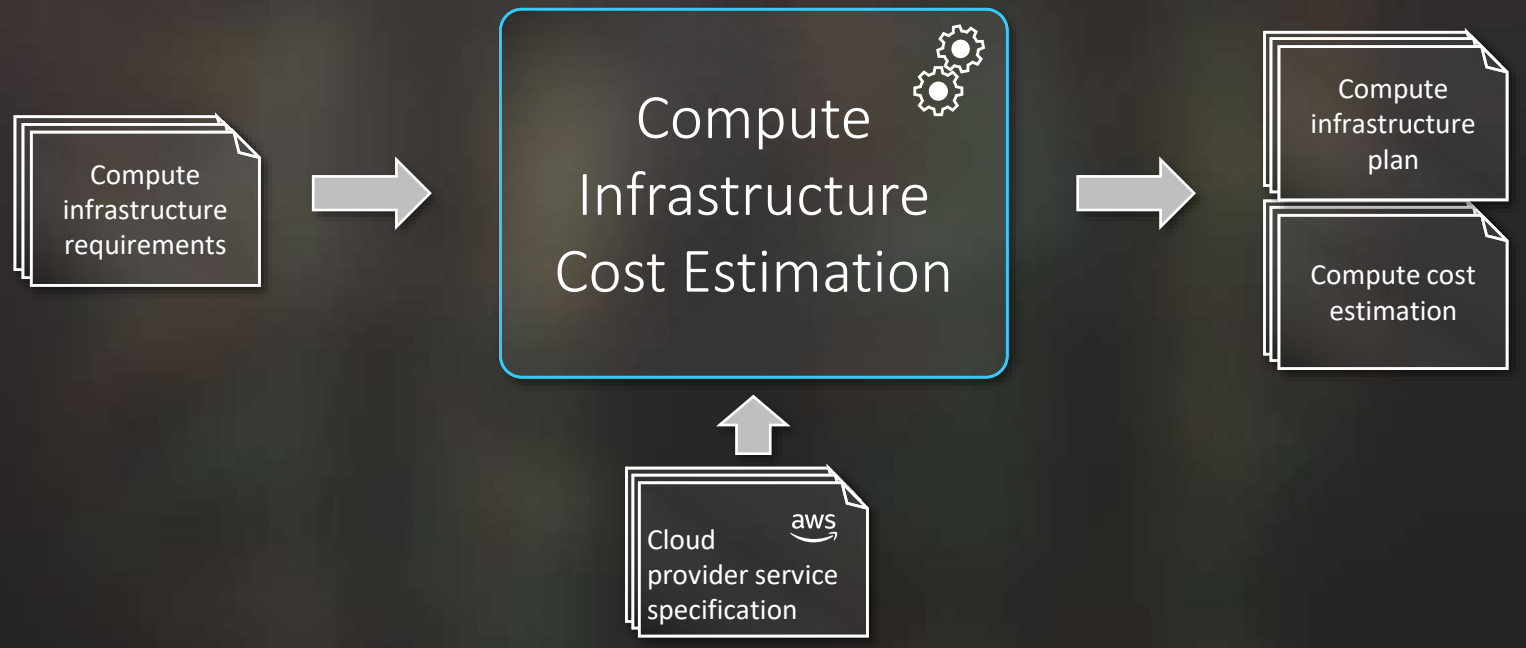
How does the planning work?



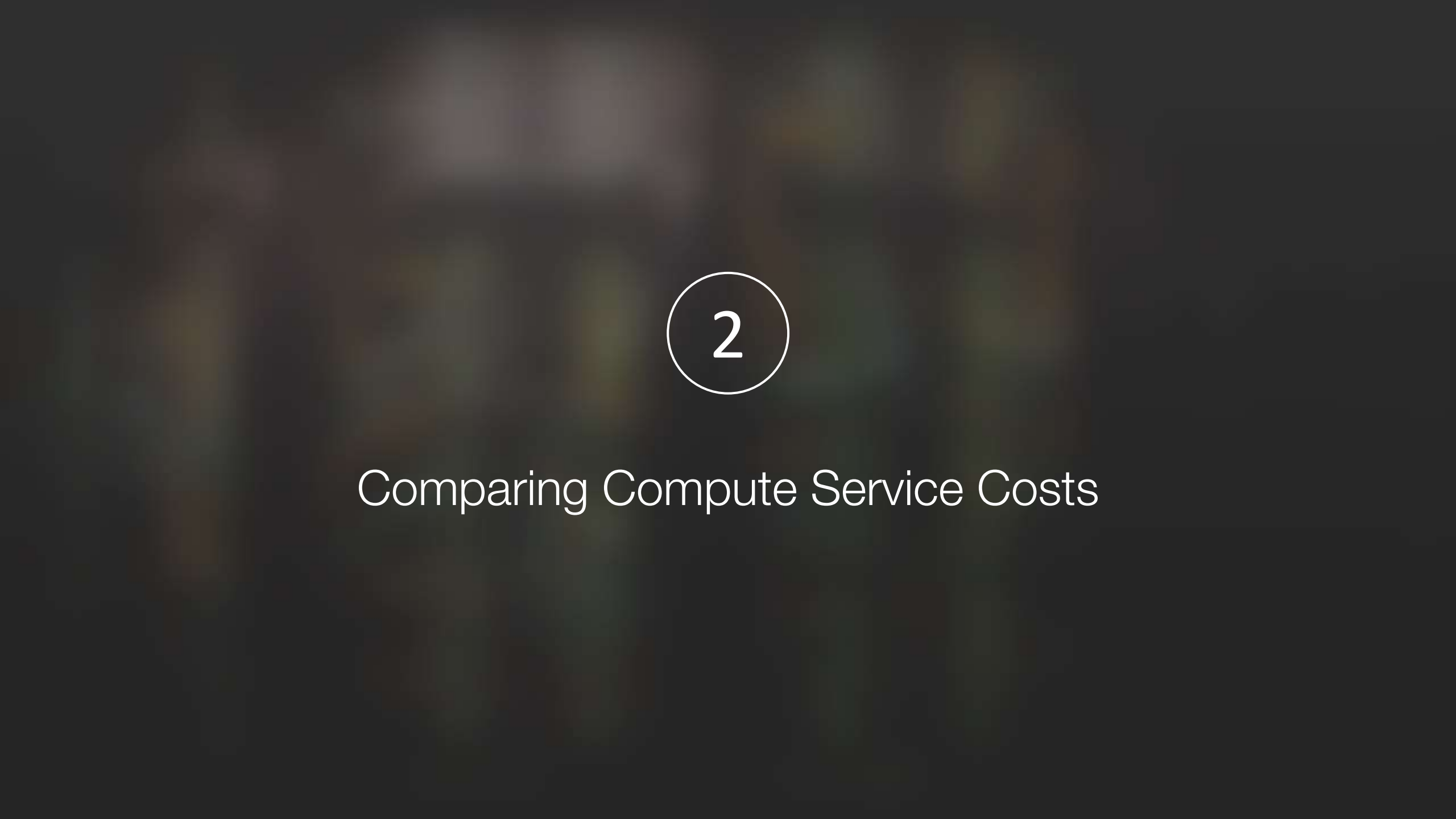
Problem of estimating cost

Problem of estimating required compute infrastructure

From planning to comparing



Compare different „scenarios“


2
Comparing Compute Service Costs

What is there to compare?

- Different compute infrastructure requirements (e.g., for a best case & a worst case estimate)
- Different cloud providers
- Different data centers / regions of a cloud provider
- Different types of contracts (on-demand vs. various commitment options)

Isn‘t comparing cloud providers easy?

IN THE CURRENT WORLD

- Define compute infrastructure needs
- Find cloud services that match my computing needs
  Example:
  - What AWS EC2 instance type to use?
  - What type of storage to use?
- Calculate costs
  - Study pricing models documentation
  - Create a rough estimate using cost calculators

Repeat this for each relevant cloud provider… manually 😢

IN A BETTER WORLD…

- Define compute infrastructure needs
- Describe these needs once
- Choose cloud providers to consider in comparison
- Be done with it

A Case Study

What would migrating to the cloud cost?

Hi, management just asked me whether we should move to the public cloud X.

Okay...

They want to know the monthly costs in the cloud compared to our data center.

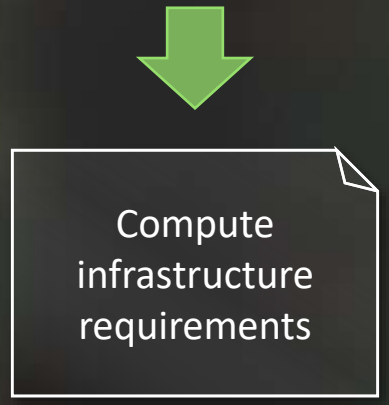
Can you get me these numbers?

I will look into it.

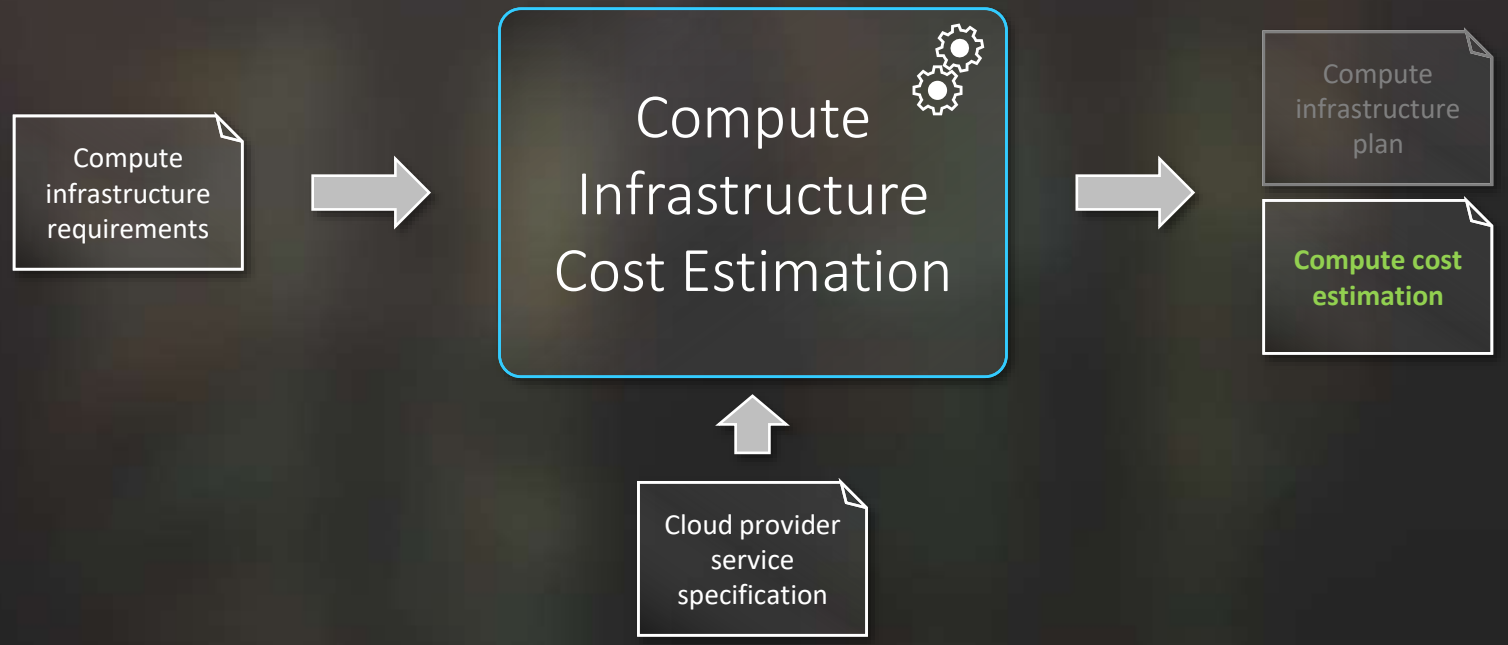
Thanks ☺

Case Study: Setup

- Information system that responds to queries by end users and other systems
- Existing operative setup
  - Cluster with 32 compute servers (approx. 64 cores, 380 GB RAM, and 320 Gb storage)
  - Load balancer distributing the requests
  - Additional smaller infrastructure for supporting software services
  - About 75 GB outgoing traffic per day



Estimating the costs…



If we move our compute infrastructure as it is to the cloud, the costs would be **6.5 times** of what we pay today.

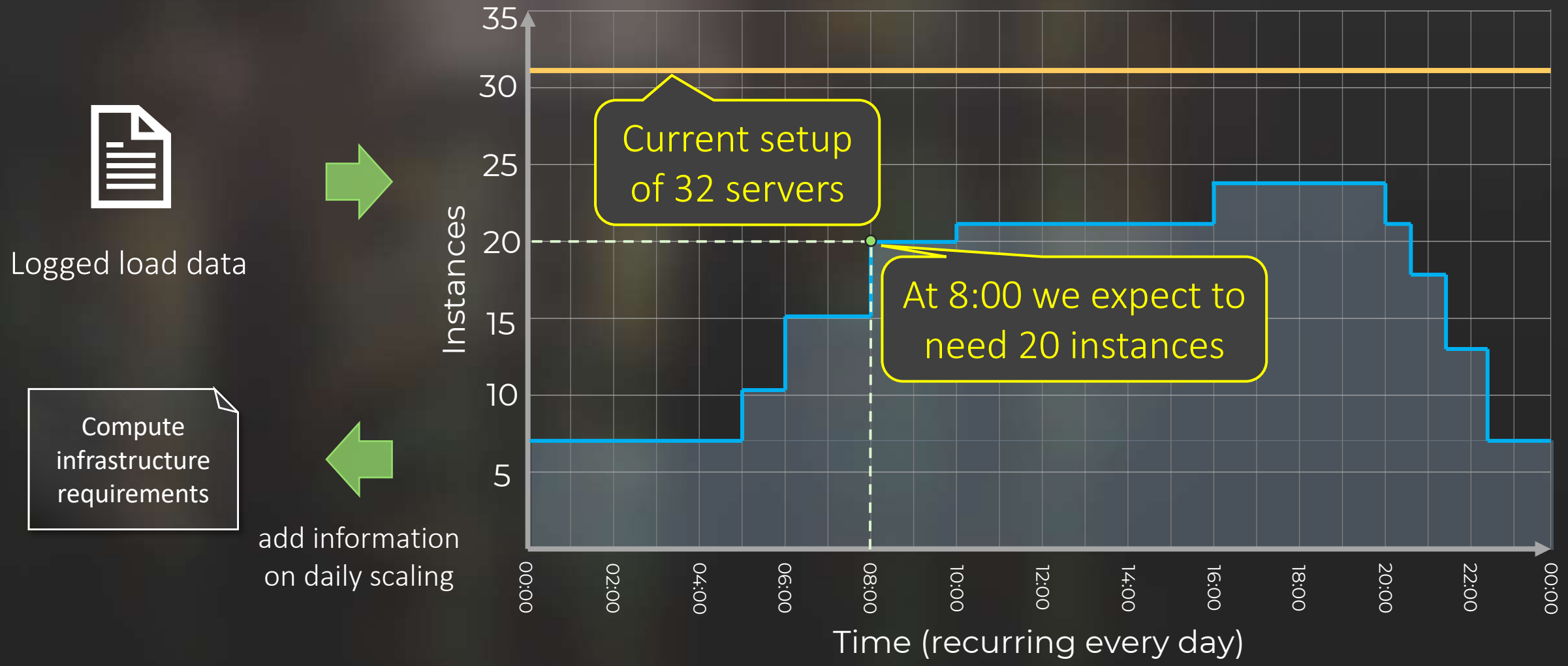
It will cost that much?! 😬

But isn‘t this comparison unfair? Cloud is all about making use of elasticity 🤨

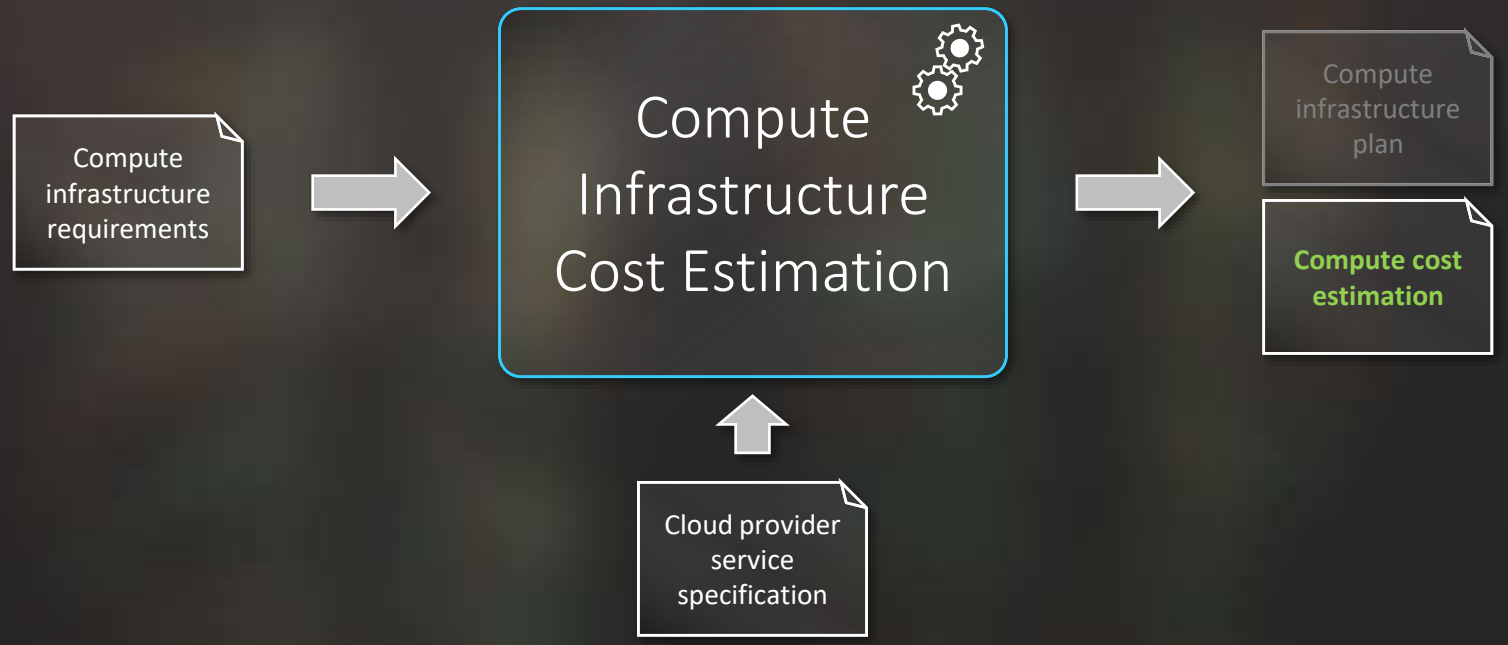
True. Most of our traffic occurs during the day with a peak in the evening.

I will look into this. 🙂

Scenario 2: Elasticity with 100% on-demand instances



Estimating the costs again…



If we **make use of elasticity**, the costs would be **3.5 times** of what we pay today.

That is more than I expected.

But we can reduce the costs by committing to **reserved virtual machine instances**.

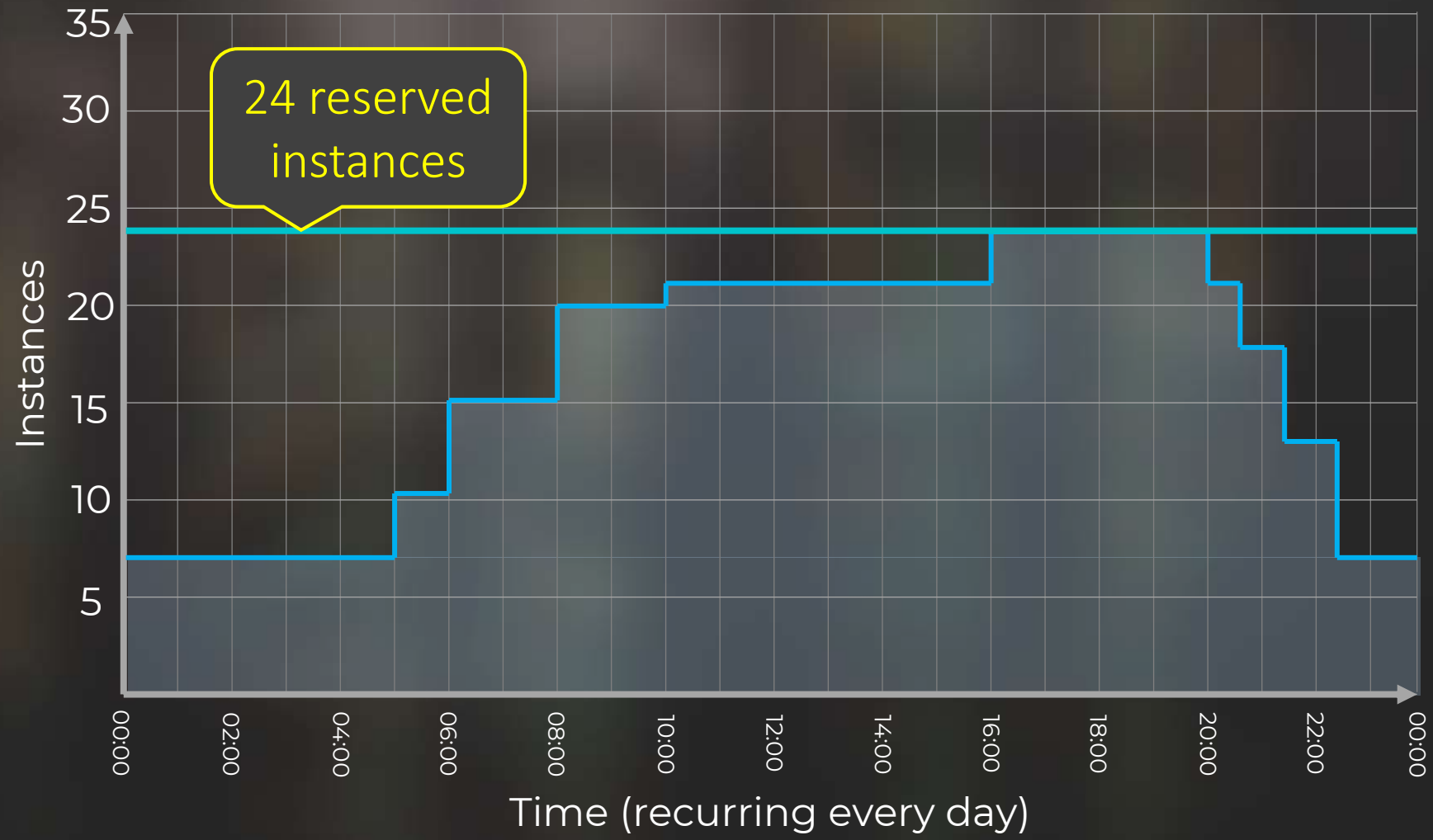
Great, please check this.

There is one problem though. I don‘t know the optimal reservation number. 🤔

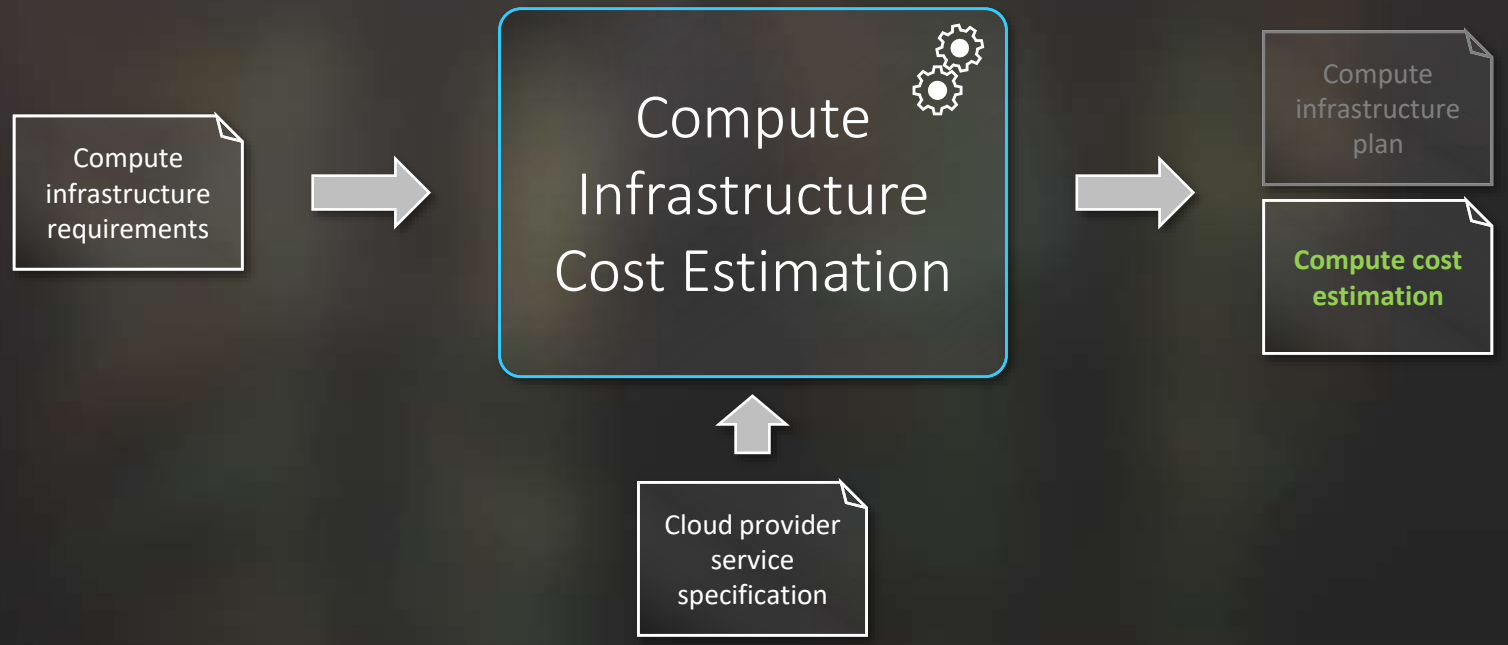
I can quickly get the costs for all servers reserved and find the optimal number later. Okay?

Yes, thank you. ☺

Scenario 3: Elasticity with 100% one-year reserved instances



Estimating the costs again…



Concluding the case study

Scenario	Monthly costs
REFERENCE Operating the setup in own data center	$x \frac{\text{EUR}}{\text{month}}$
SCENARIO 1 Migrate the setup to the cloud with the **infrastructure as it is**	$6.5 \cdot x \frac{\text{EUR}}{\text{month}}$
SCENARIO 2 Migrate the setup to the cloud and **make use of elasticity** by **scaling up and down**	$3.5 \cdot x \frac{\text{EUR}}{\text{month}}$
SCENARIO 3 Migrate the setup to the cloud and **commit to reserved instances**	$2.1 \cdot x \frac{\text{EUR}}{\text{month}}$

Remarks

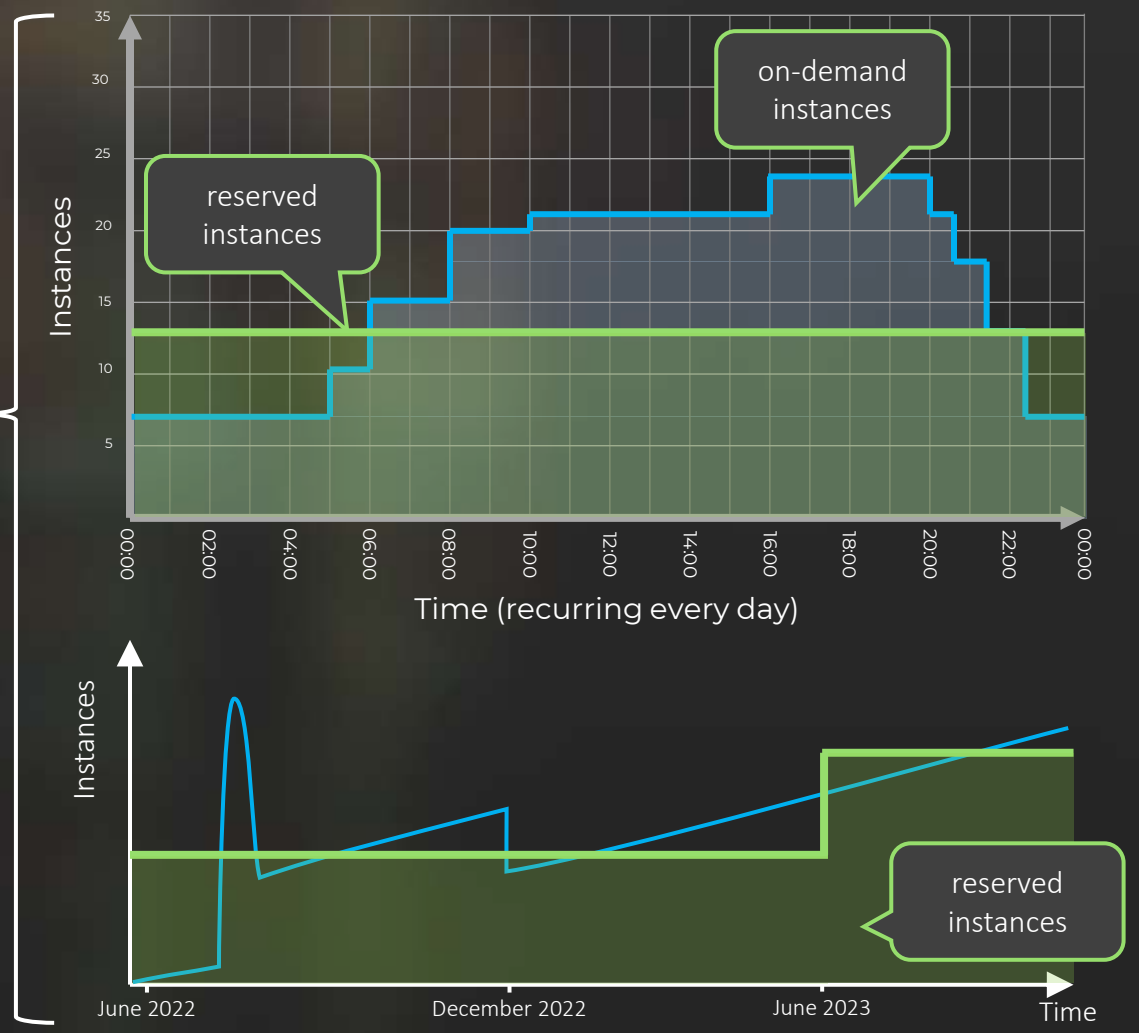
1. Detailed setup & cost information were not shown for confidentiality reasons
2. In this case study, the operating costs in the own data center were **exceptionally low** (to my experience)
3. There is still unused potential for optimizing the cloud costs

~~Compare~~… Optimize!

Optimization

- First „easy“ optimizations
  - Don‘t keep resources longer than necessary
  - Right-size resources
- Implement cost-efficient scaling
- Find cost-efficient commitments to reserved instances
- Optimize cloud services including licenses
  - Cloud services with bring your own license (e.g., Microsoft Azure Hybrid benefit)

I want to address automating this next



4

Conclusion

Take aways

- Estimation of cloud costs manually is laborious
  → Developed a research prototype to support this and to compare estimation scenarios

- Proactively managing cloud costs can make a significant difference
  → as shown by the case study

- Further automated optimizations seem useful
  → future work