



**Barcelona
Supercomputing
Center**
Centro Nacional de Supercomputación



**UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH**



AI and Predictive Analytics in Data-Center Environments

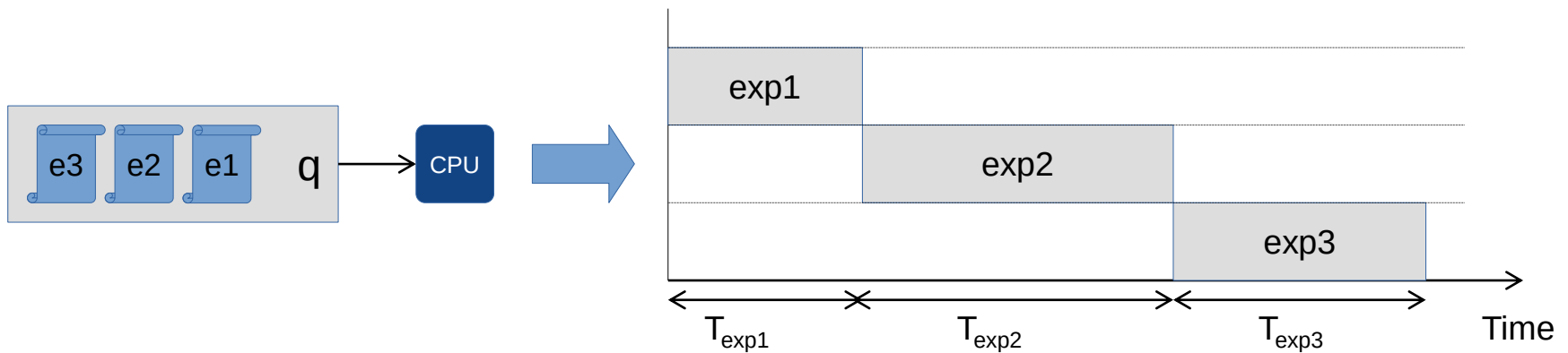
Parallelism and Distributed Executions

Introduction

“The more resources put to work, the earlier we may finish our experiments”

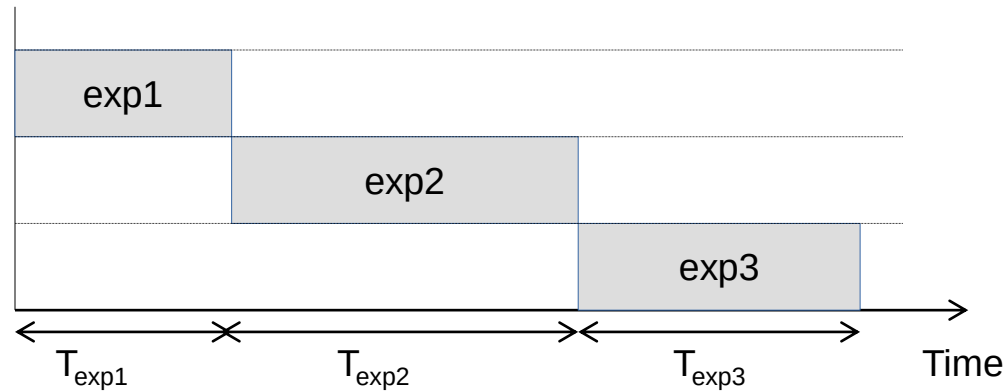
Parallelism

- When we serialize executions:



Parallelism

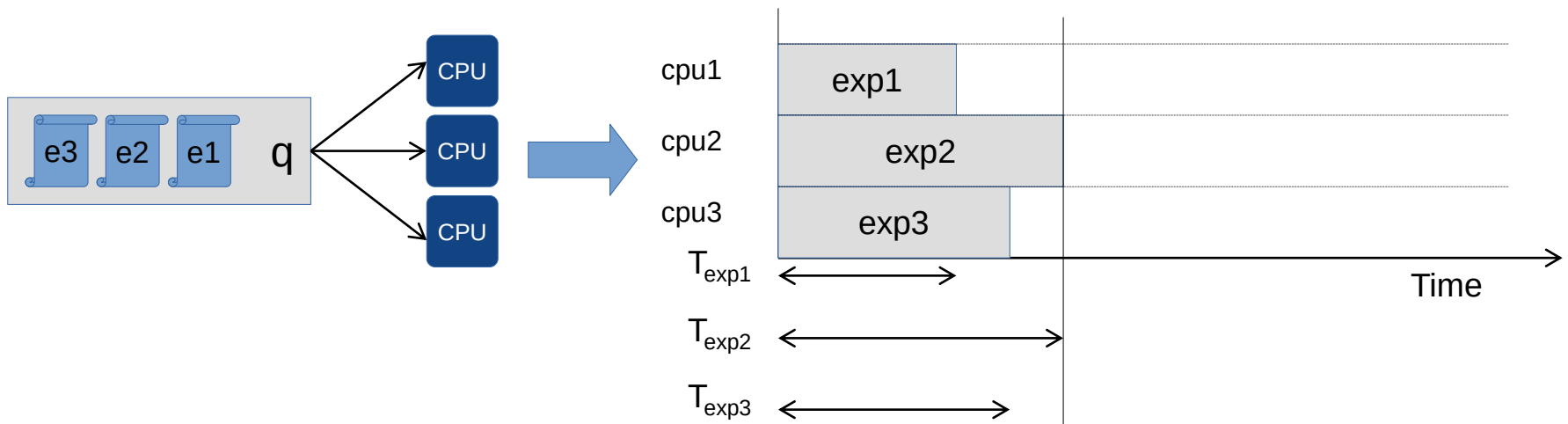
- Required time:



$$\text{TIME} \rightarrow T_{\text{total}} = T_{\text{exp1}} + T_{\text{exp2}} + T_{\text{exp3}}$$

Parallelism

- Applying parallelism
 - When our executions are Independent
 - When we have resources to consume the load
 - e.g.:

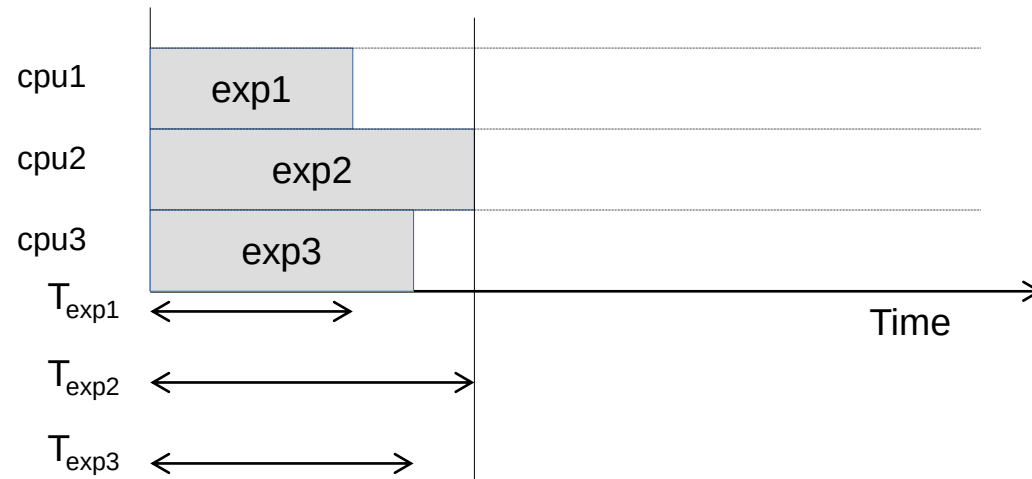


Assumption: each experiment only consumes 1 CPU

Parallelism

- Required Time

*Assumption: each experiment only consumes 1 CPU



$$\text{TIME} \rightarrow T_{\text{total}} = \max(T_{\text{exp1}} + T_{\text{exp2}} + T_{\text{exp3}})$$

Independence

- Parallelism requires independence

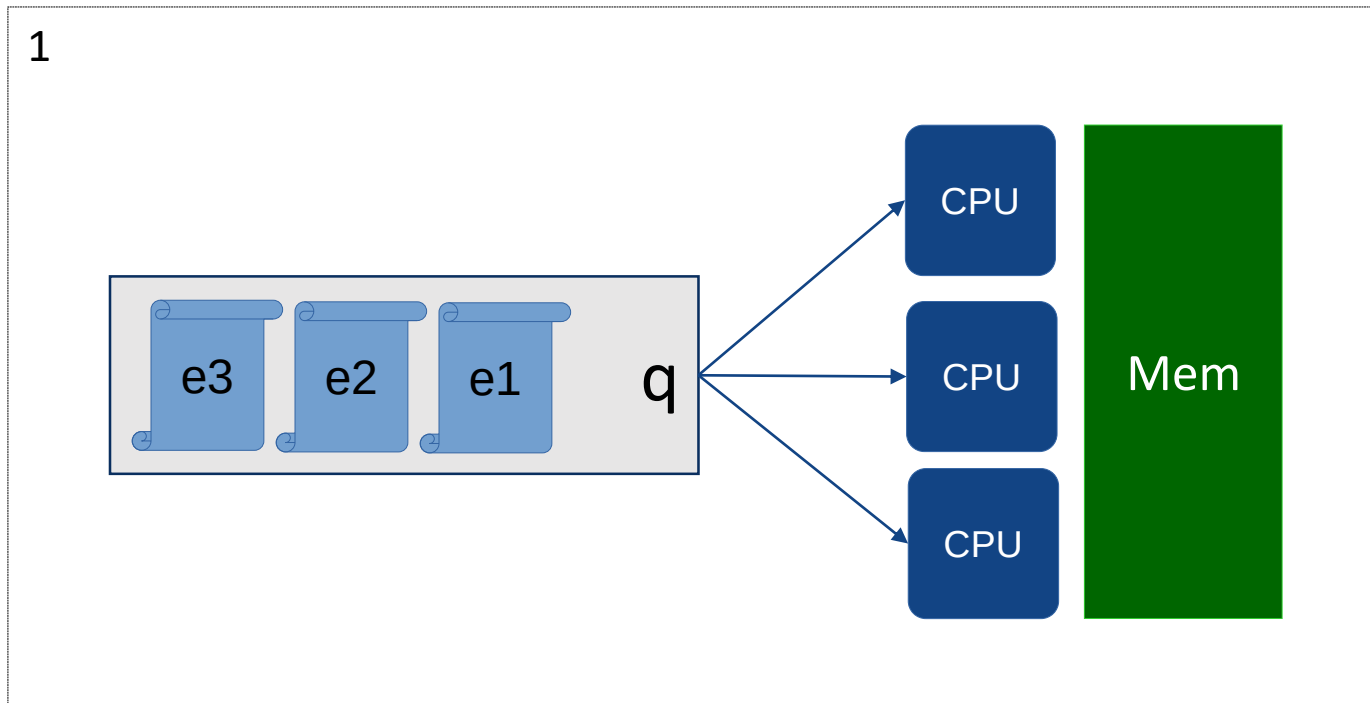
Processes running together must NOT depend on each other results

Parallelism

Different ways to parallelize experiments and jobs

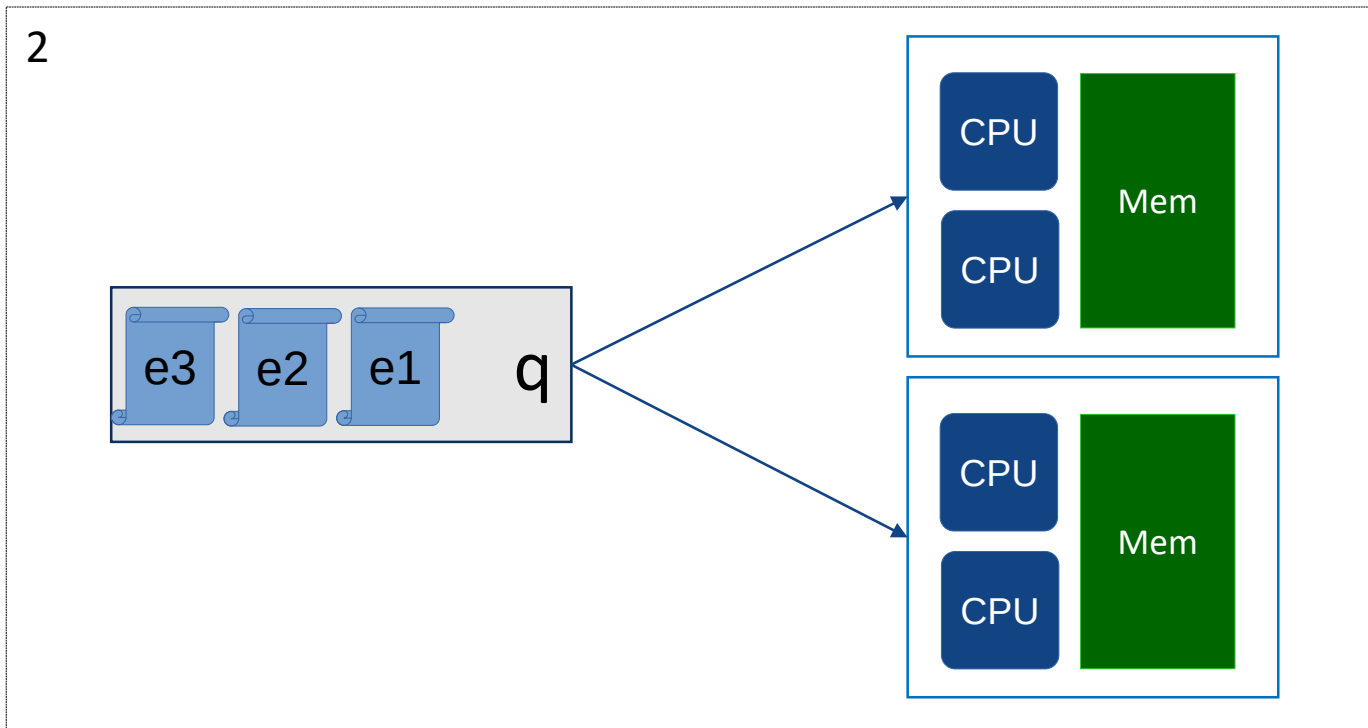
Parallelism

- Parallelism can be applied in many ways
 - “Many experiments & many available resources”



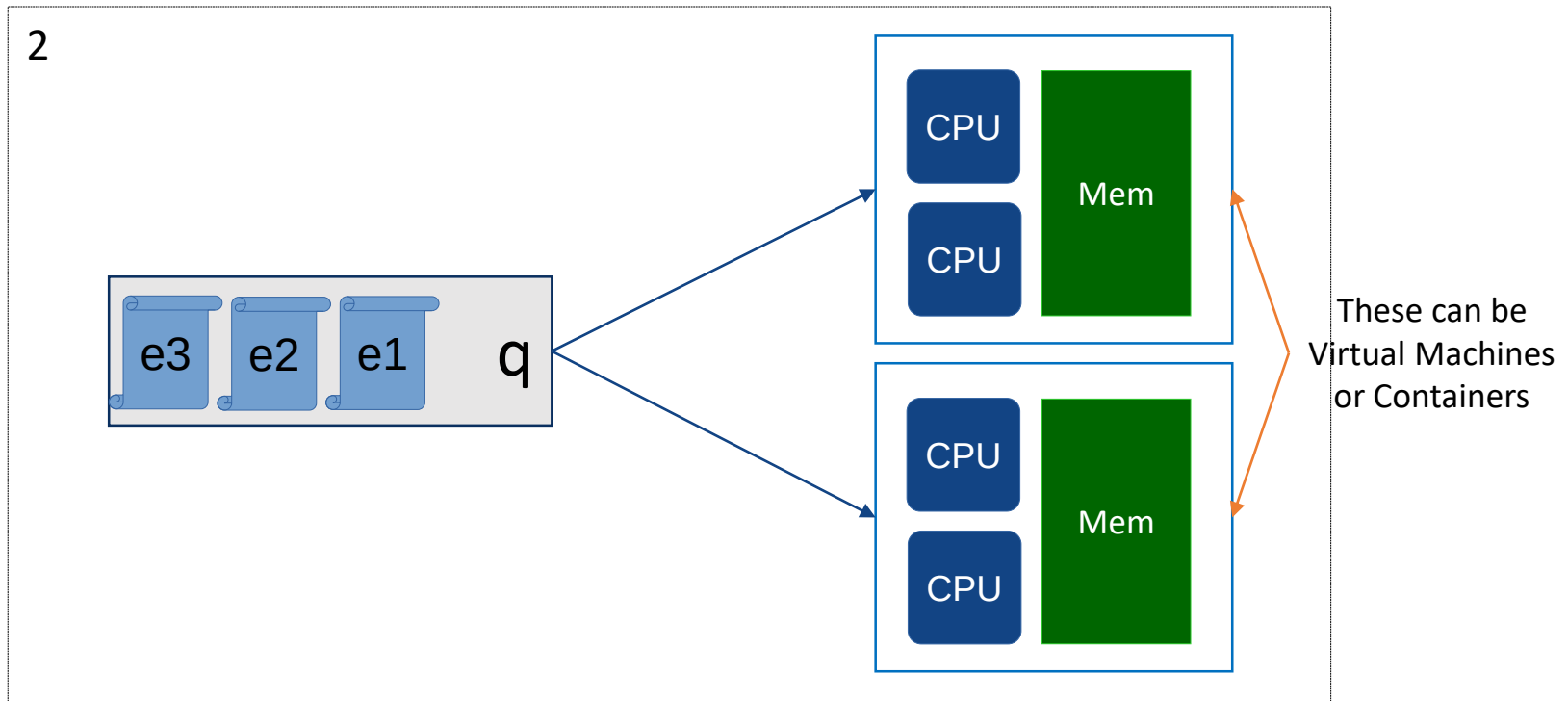
Parallelism

- Parallelism can be applied in many ways
 - “Many experiments & available groups of resources”



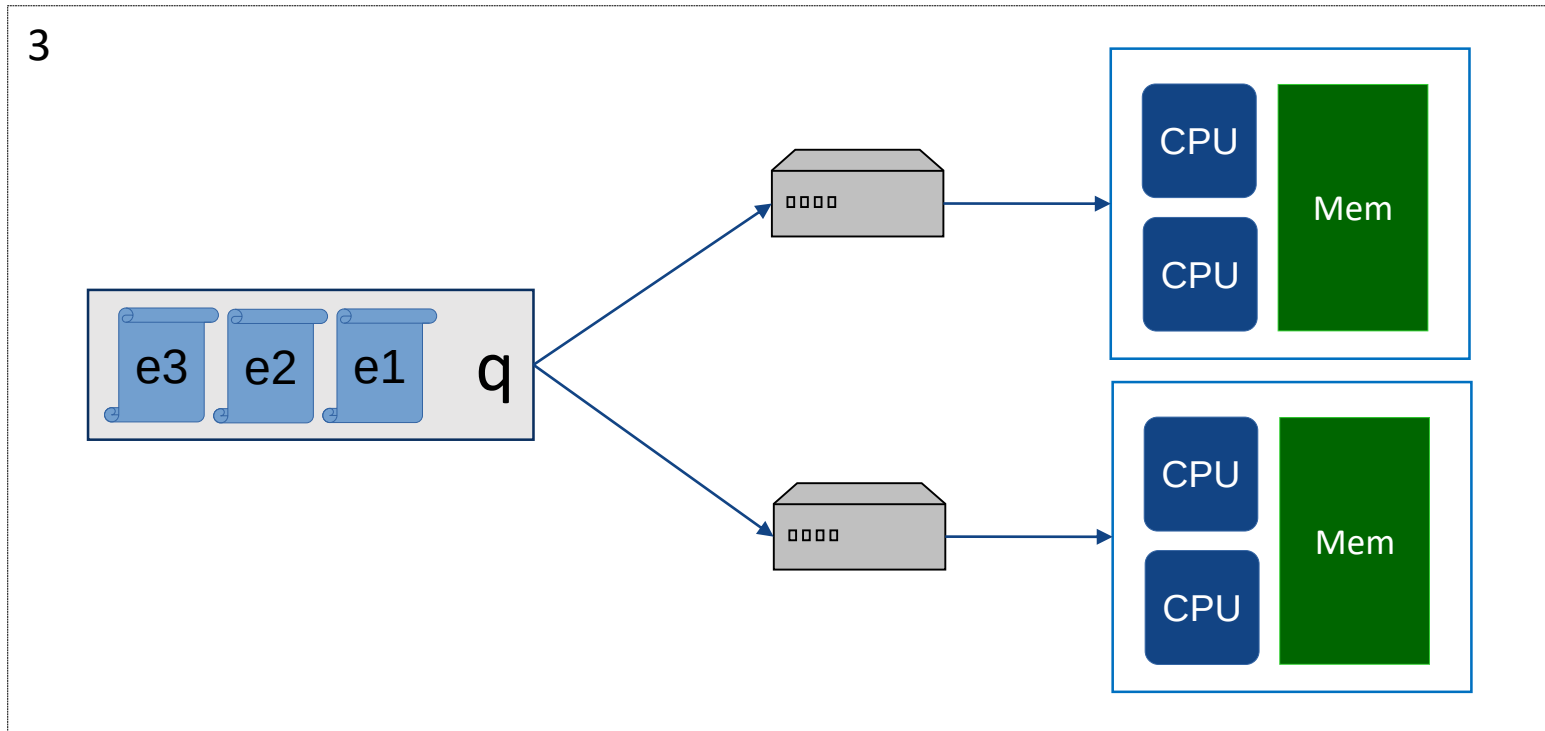
Parallelism

- Parallelism can be applied in many ways
 - “Many experiments & available groups of resources”



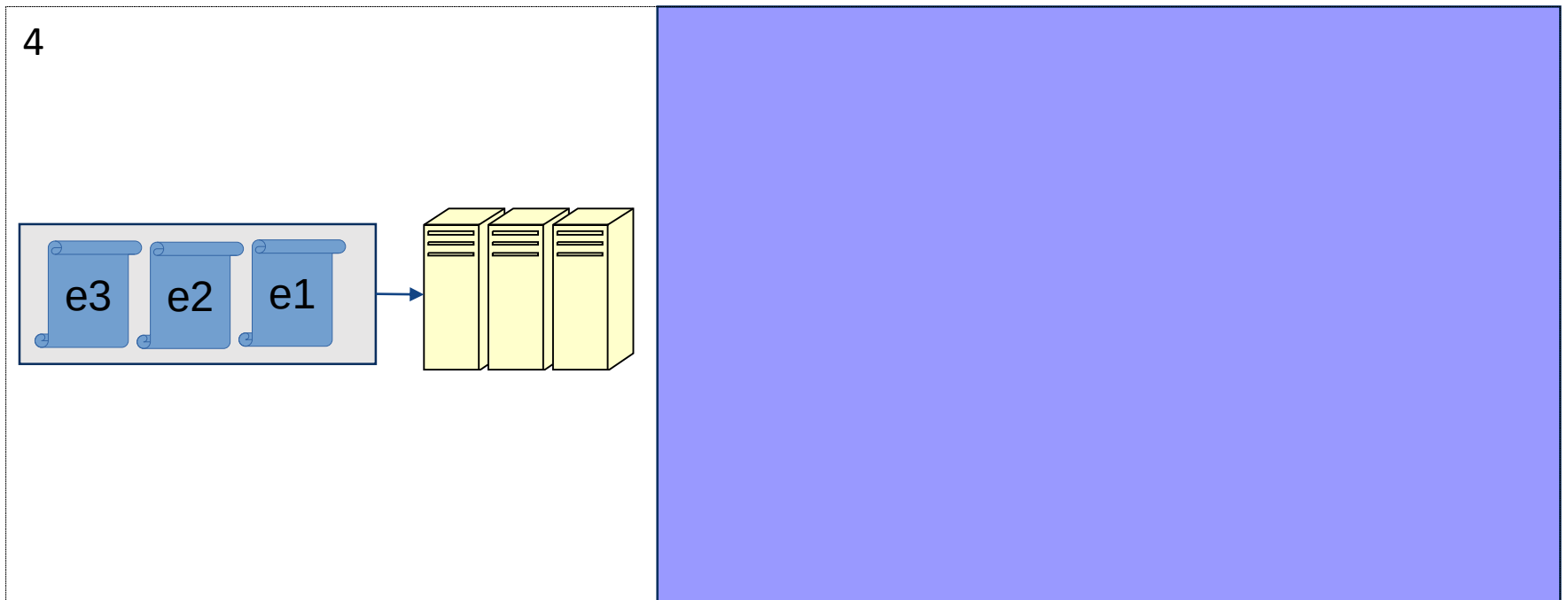
Parallelism

- Parallelism can be applied in many ways
 - “Many experiments & many machines”



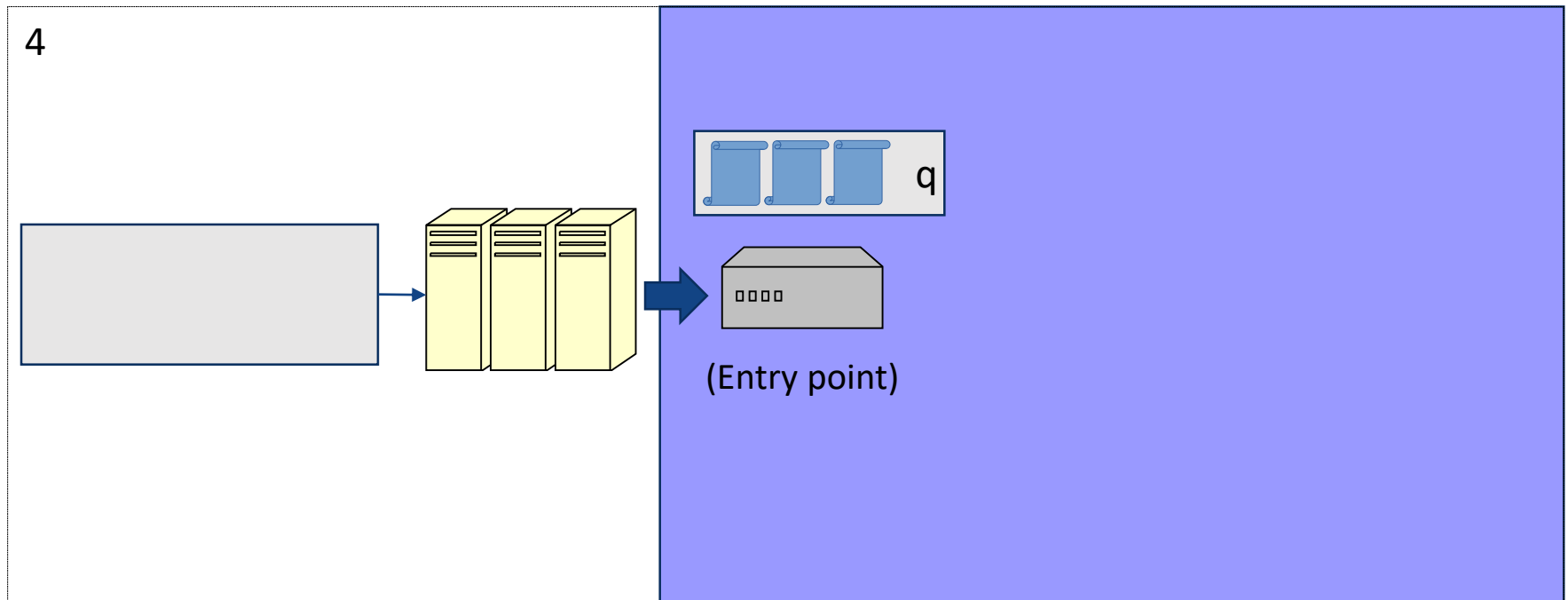
Parallelism

- “Many experiments sent to a Cluster / Managed System”



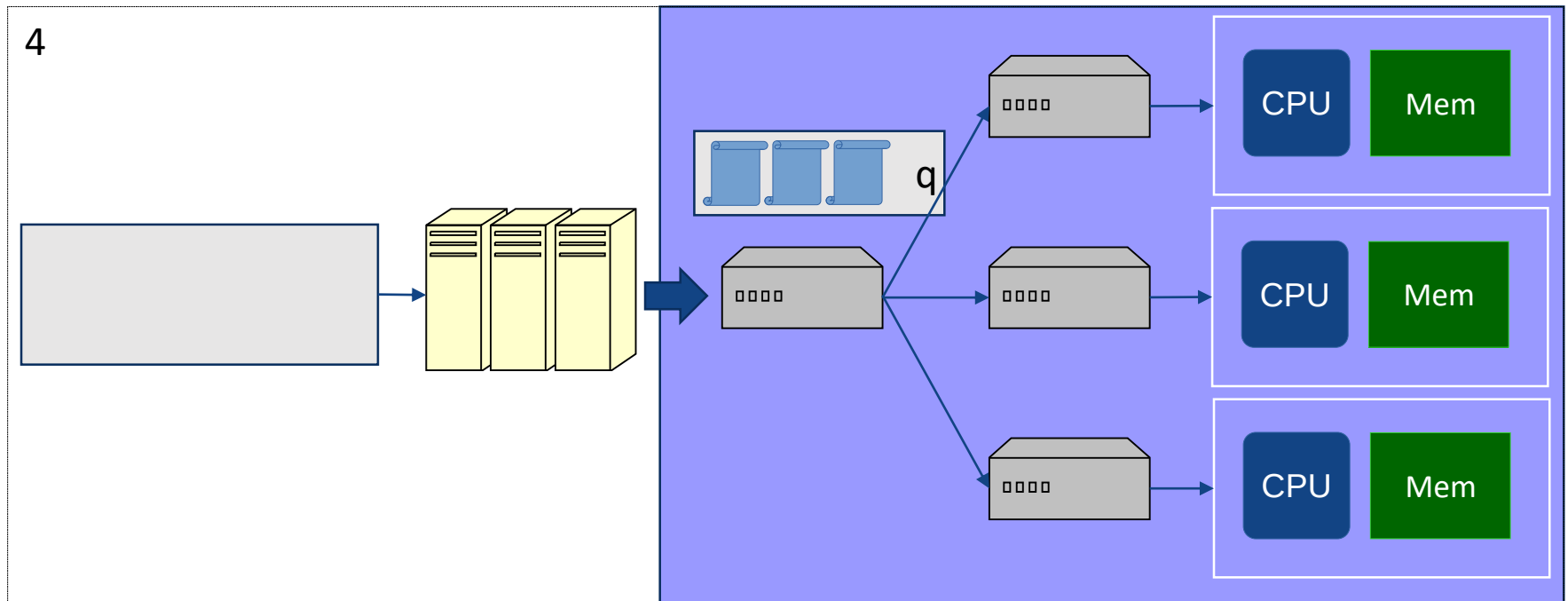
Parallelism

- “Many experiments sent to a Cluster / Managed System”



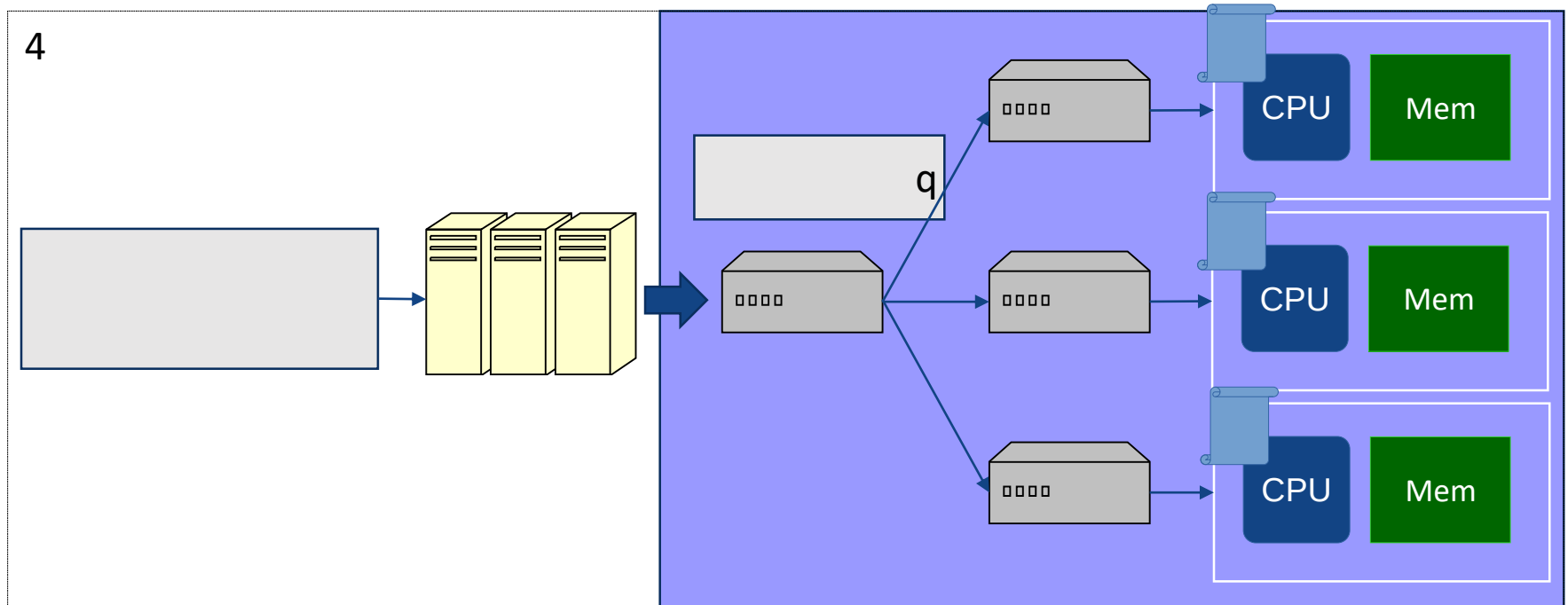
Parallelism

- “Many experiments sent to a Cluster / Managed System”



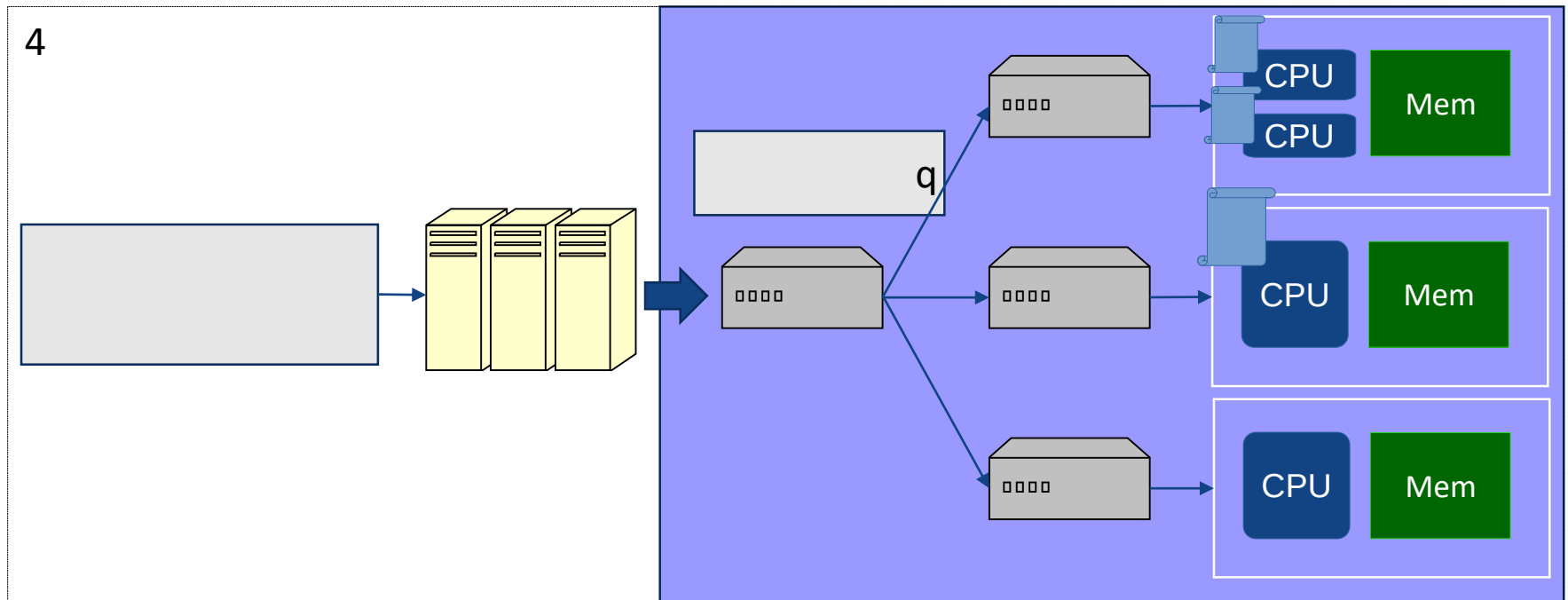
Parallelism

- “Many experiments sent to a Cluster / Managed System”



Parallelism

- “Many experiments sent to a Cluster / Managed System”

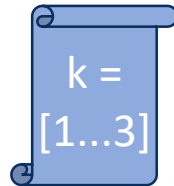


Parallelizing Experiments

Single experiments/jobs can be split

Parallelizing Experiments

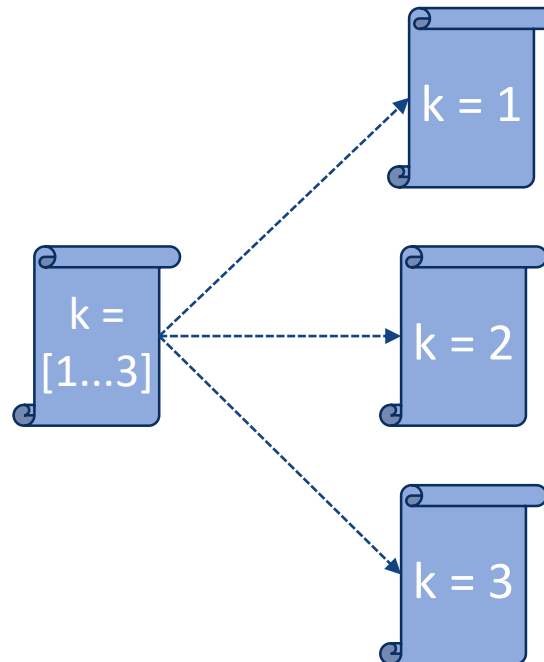
- E.g. Send different experiments to separate machines



“Search the best hyper-parameter ‘k’ for experiment E”

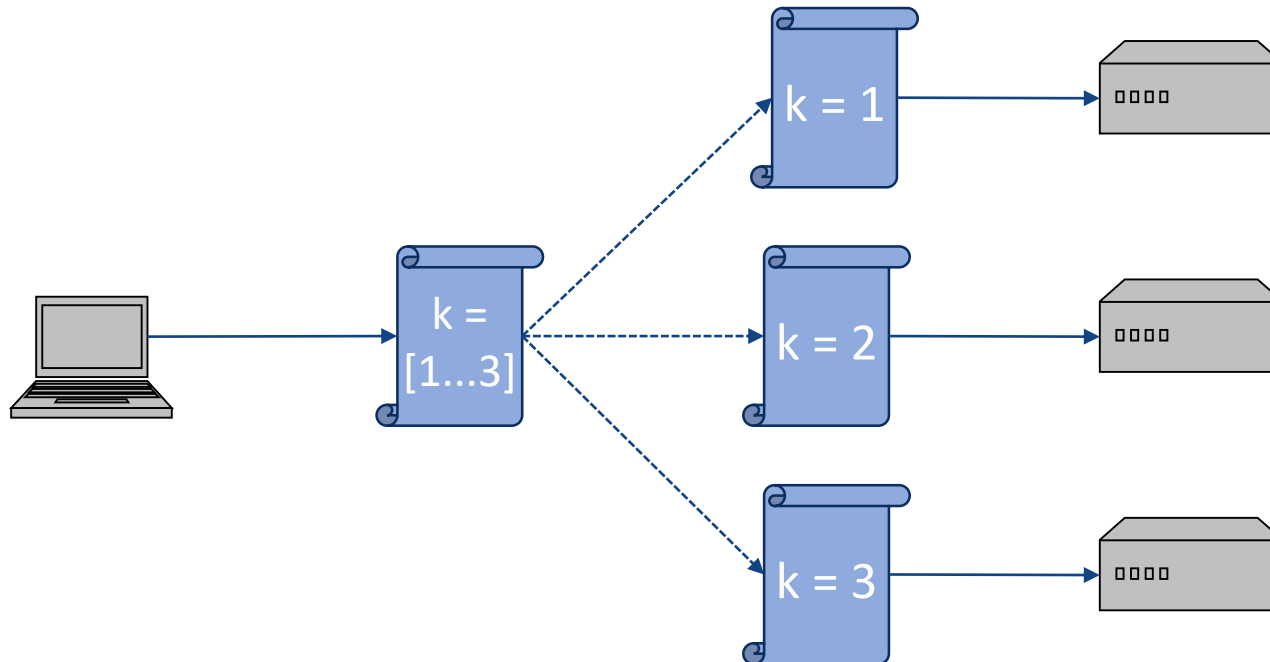
Parallelizing Experiments

- E.g. Send different experiments to separate machines



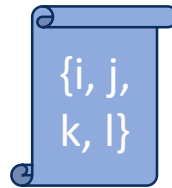
Parallelizing Experiments

- E.g. Send different experiments to separate machines



Parallelizing Experiments

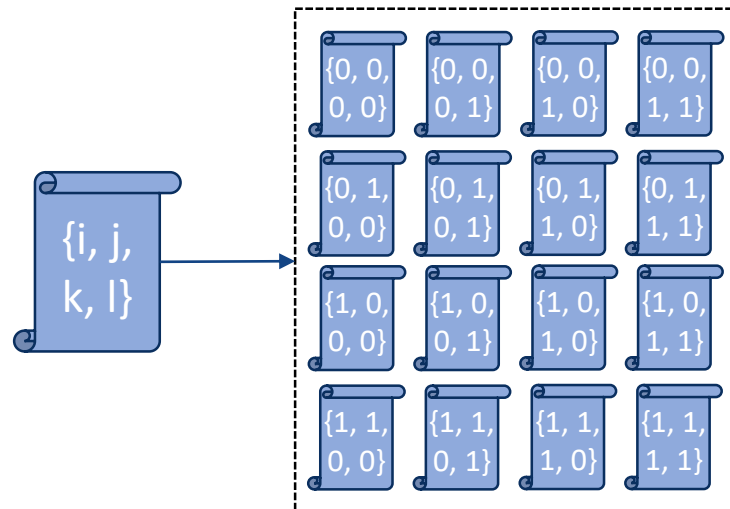
- E.g. Grid-search on execution parameters



“Search best parameters $\{i, j, k, l\}$ for Application A”

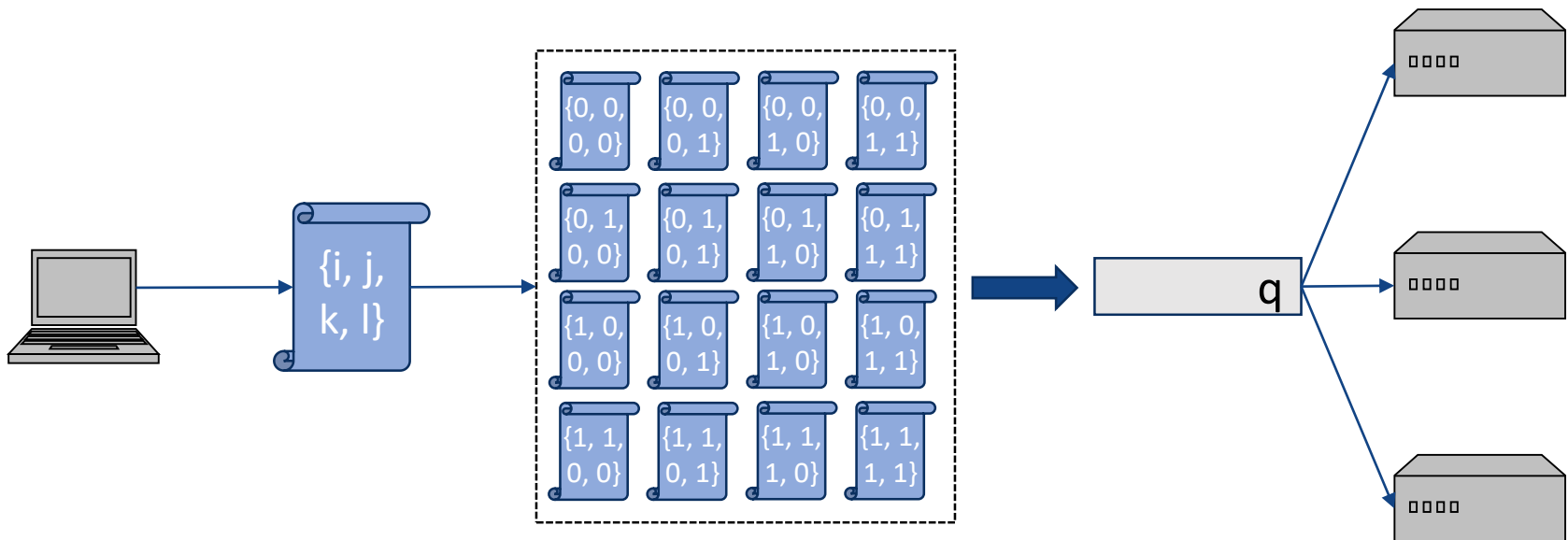
Parallelizing Experiments

- E.g. Grid-search on execution parameters



Parallelizing Experiments

- E.g. Grid-search on execution parameters
 - “Search best parameters $\{i, j, k, l\}$ for Application A”

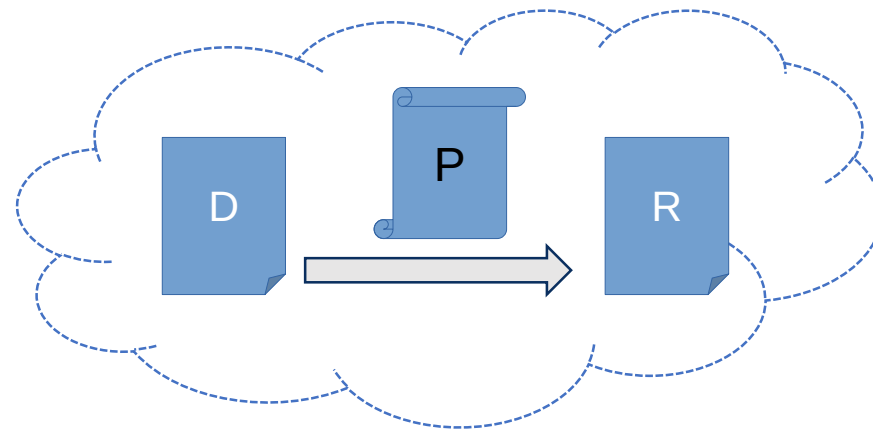


Parallelizing Experiments

Even data sets can be split

Parallelism

- Processing Data-sets

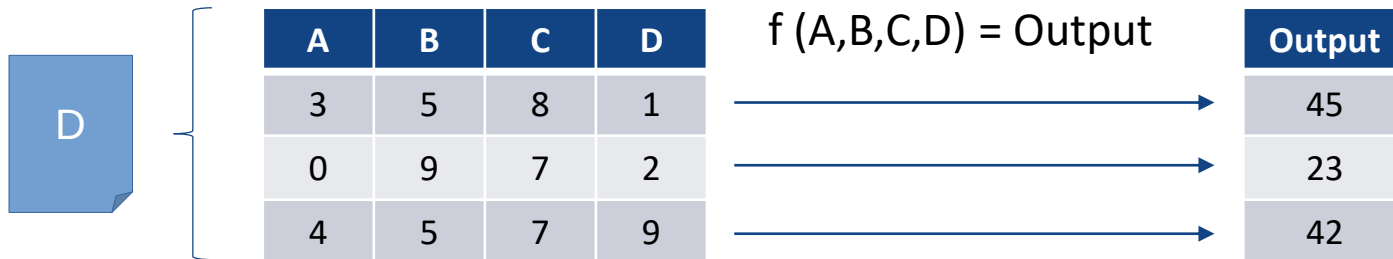


Split the data-set into batches

E.g. “Apply process P to rows in data-set D”

Parallelism

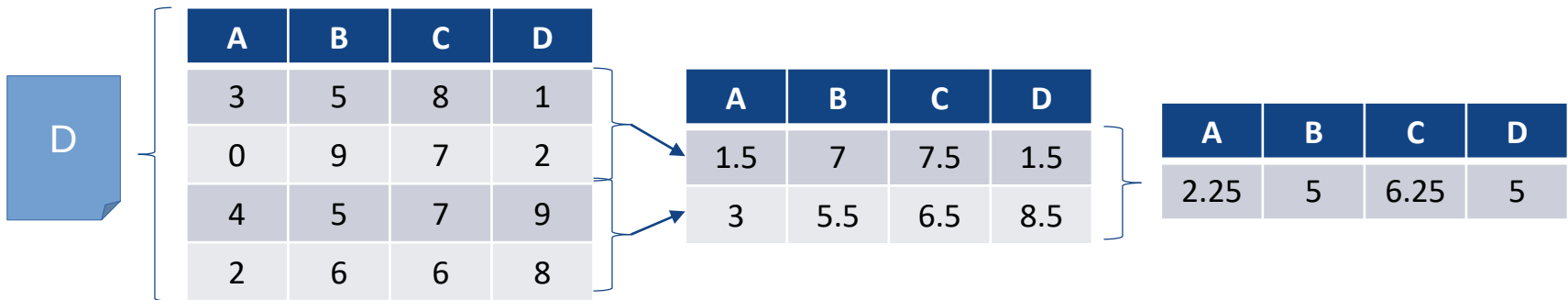
- Processing Data-sets
 - Independence of data



Parallelism

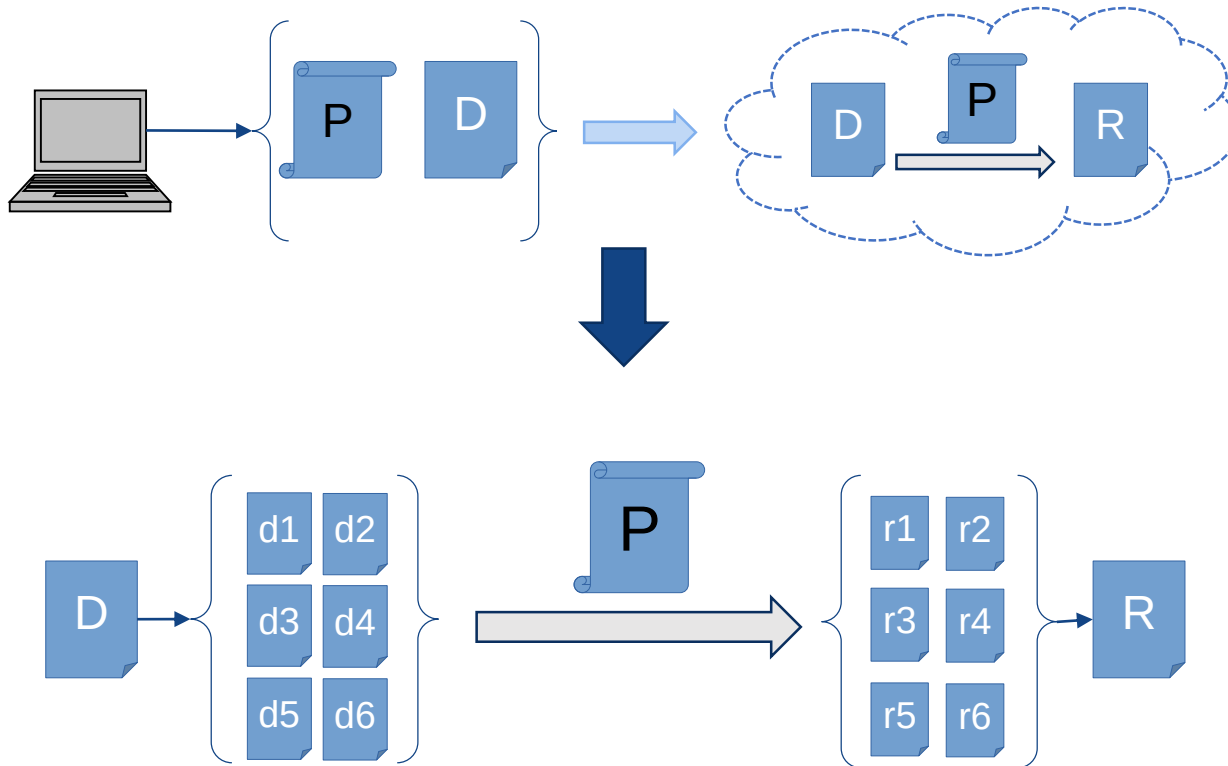
- Processing Data-sets
 - Distributing process in subsets of data, then apply again

$$f(A,B,C,D) = \{\text{avg}(A), \text{avg}(B), \text{avg}(C), \text{avg}(D)\}$$



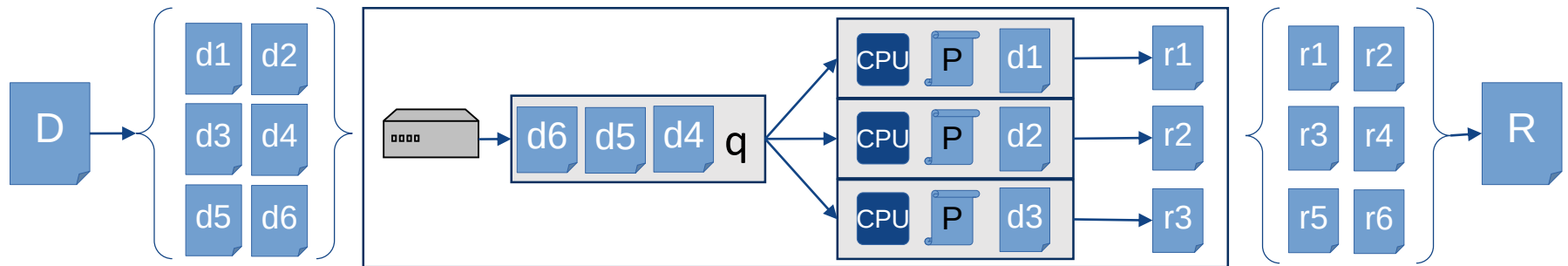
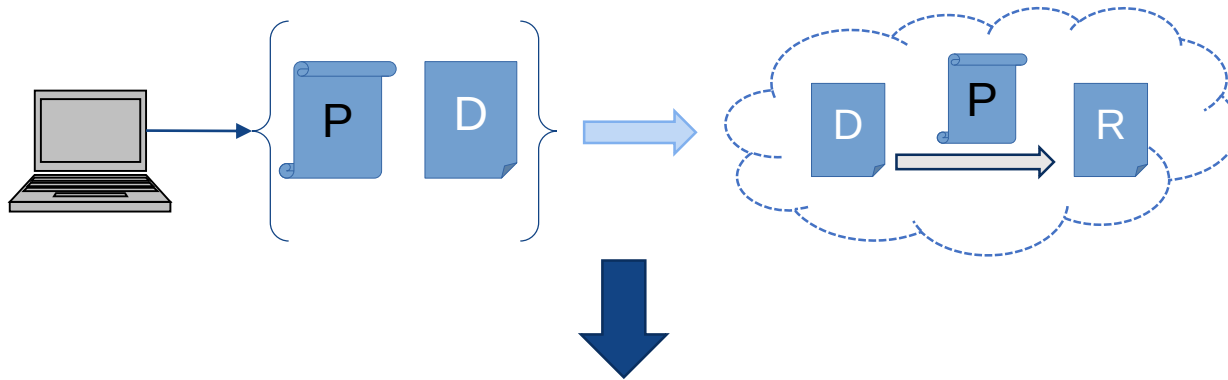
Parallelism

- Processing Data-sets



Parallelism

- Processing Independent Data-sets
 - Split the data-set into batches
 - E.g. “Apply process P to rows in data-set D”



Hands-On

- Examples of parallelism
 - Exercises Sequential vs. Parallel, compare TIME and check on HTOP

Summary

- Basics on running Parallel experiment executions
- We can distribute/split/break experiments
 - Run independent parts in parallel
 - Leverage the amount of available resources
- We must carefully analyze our experiments and system to optimize the execution