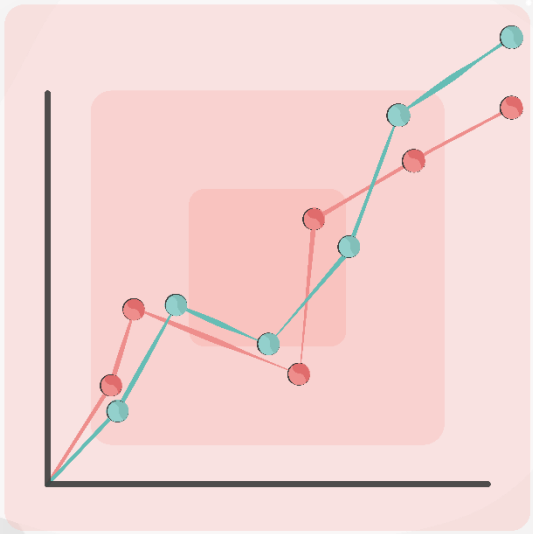
Data engineering

Module: Data pipelines

Topic: On-premise resource management



Welcome to today's webinar.

Ice breaker

Building Careers Through Education

Everybody Matters. Stronger Together. Student, Learner and Client Centric. Trust and Respect. Embrace Change.

Discussion…

- What's the most chaotic tech issue you've ever experienced at work or school?"
- True or False: On-premise systems are outdated and irrelevant in today's cloud-first world?
- What is one key skill or insight you hope to gain from today's session?

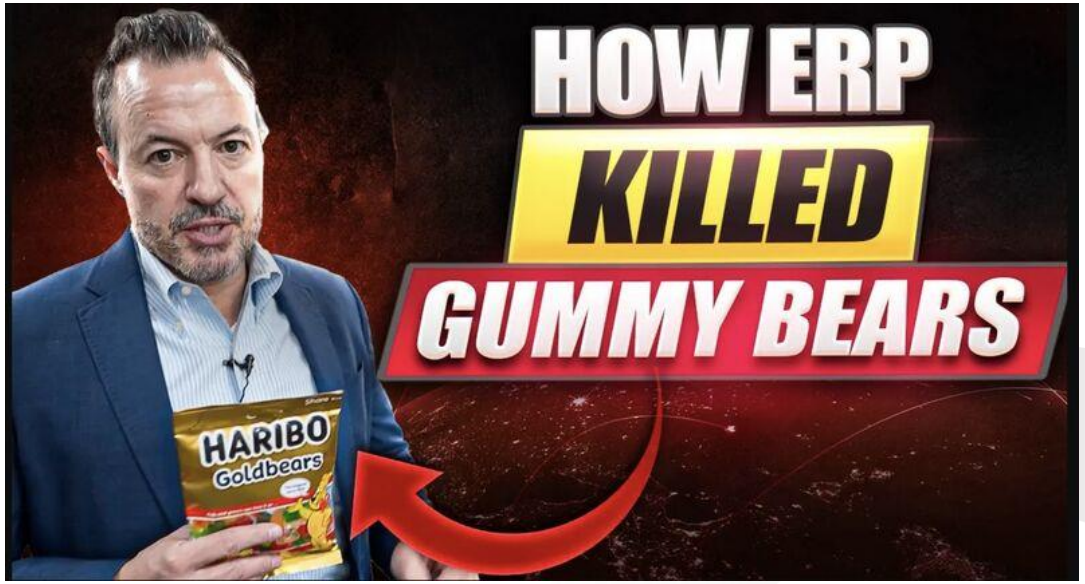
Submit your responses to the chat or turn on your microphone

BPP

Real-world case study

Haribo’s ERP Implementation Failure (2018)

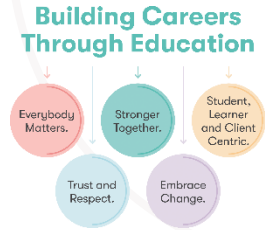
- Enterprise Resource Planning (ERP) launched across 16 factories in 10 countries
- Poor process mapping & no infrastructure stress testing
- Inventory tracking failed → empty shelves
- Sales dropped 25% after rollout
- **Root causes:** underplanning, no load balancing, no monitoring
- Highlights need for capacity planning & observability



How Haribo’s ERP failure cost them millions,
Image source: *LinkedIn*

This case shows that even a beloved global brand can suffer major losses when on-premise systems are poorly managed.

e-learning recap

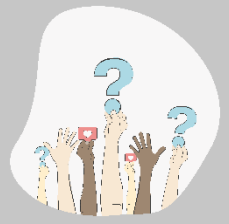


Reflecting on your learning…

The e-learning for this topic, covered the following areas:

1. Understanding On-Premise Infrastructure
2. Setting Up and Maintaining Infrastructure
3. Managing Computational Resources
4. Optimising Pipeline Efficiency
5. Load Balancing and Scheduling
6. Using Docker and Kubernetes On-Prem

Q&A discussion



- Do you have any questions about any of these areas?
- Did everything in the e-learning make sense?

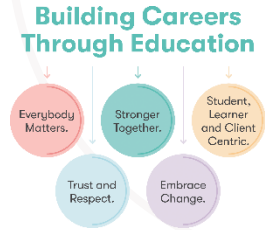
BPP

Webinar Agenda

Today, we will cover the following:

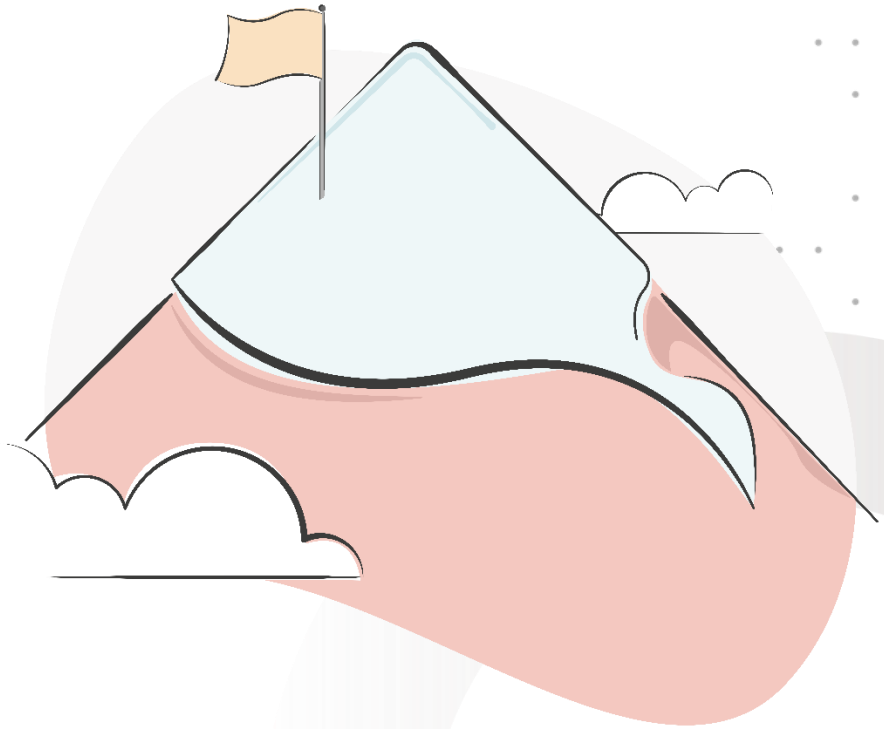
1. Core concept recap
2. Practical lab
3. Summary
4. Q&A

Session aim and objectives



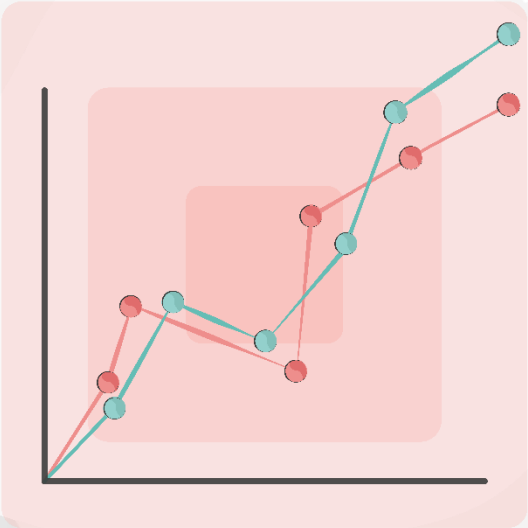
By the end of this session, you should be able to:

- Design and Maintain On-Premise Data Infrastructure
- Optimise Resource Usage for Data Pipelines
- Implement Scalable and Reliable On-Prem Workflows Using Modern Tooling



BPP

Recap of core concepts

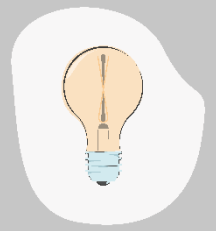


Defining on-premise infrastructure

What Is On - Premise Infrastructure?

- Physical servers, storage, and networking
- Self - managed environments
- No auto - scaling or managed services

Figure: *On-premise architecture, Source: Datapac*



Unlike the cloud, where services are abstracted and scalable, on - prem systems require manual setup, configuration, and maintenance.

BPP

Setup essentials

Building your infrastructure

- Plan for compute, storage, and redundancy
- Design secure network topology
- Harden OS and runtime environments

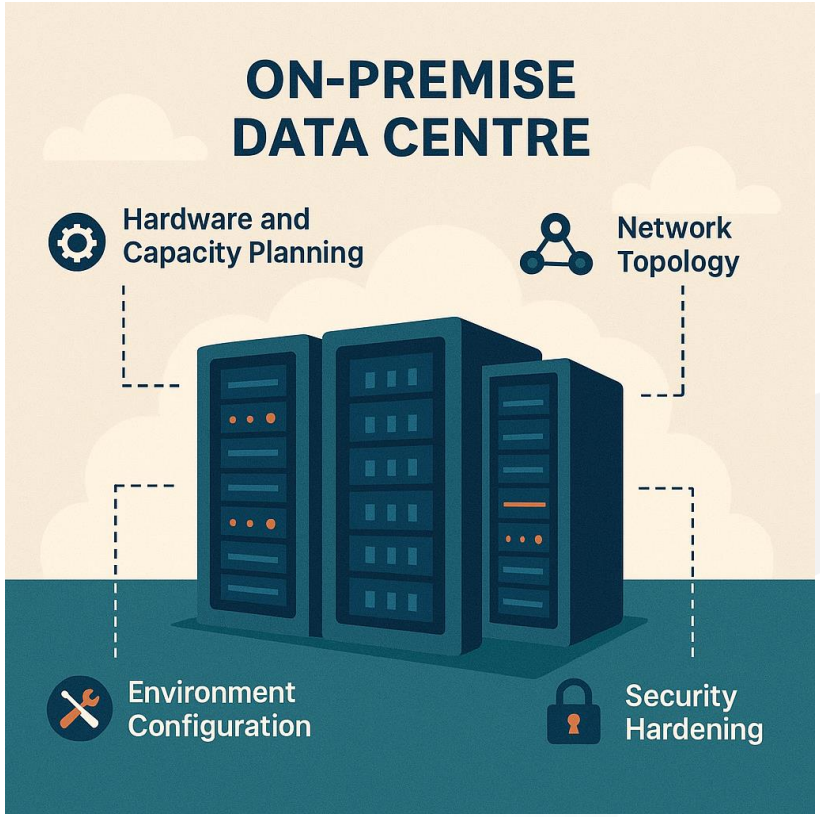


Figure: *Key Considerations for Setting Up an On-Premise Data Centre*

Maintenance matters

Keeping systems healthy

- Patch regularly and automate updates
- Monitor CPU, memory, and disk
- Test backups and document everything

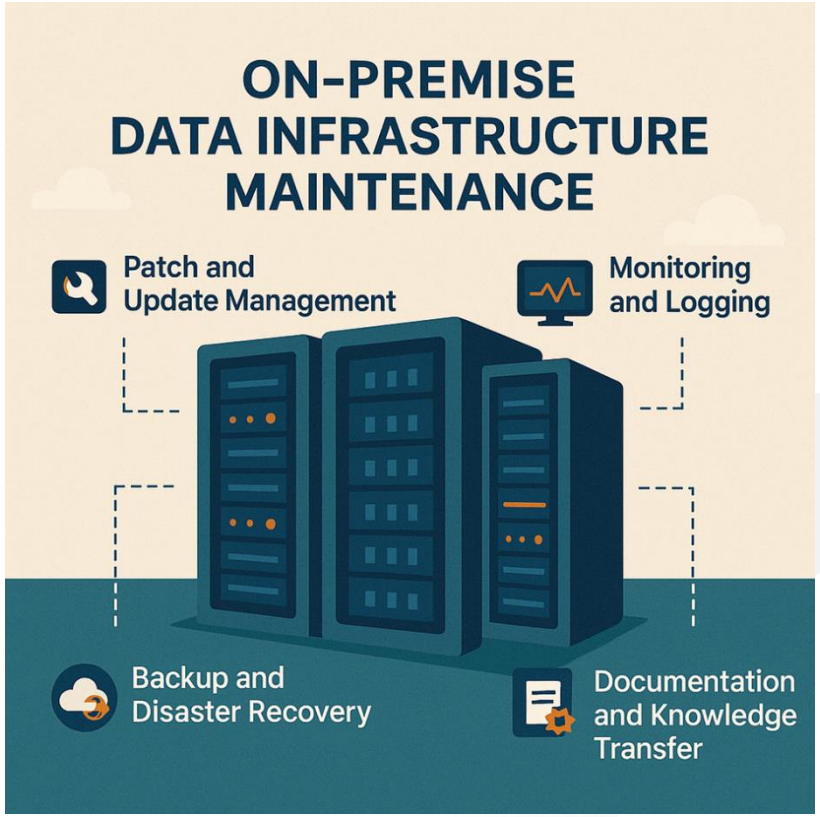


Figure: *Ensuring System Health through Regular Maintenance and Monitoring*

Resource awareness

Understanding system limits

- CPU, RAM, I/O, and network are finite
- No “scale up” button
- Poor planning = system crashes

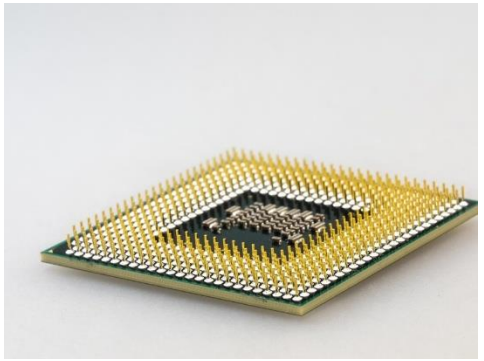


Figure: a Central Processing Unit (CPU)

Figure: a RAM stick, or memory module.

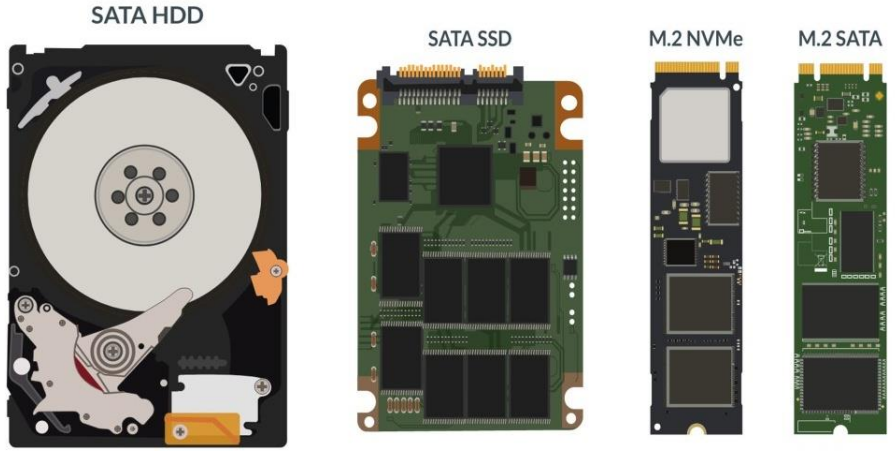
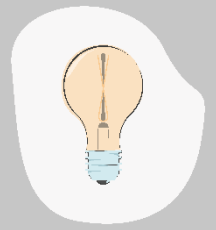


Figure: Figure: different types of hard disk storage.

Smart usage

Optimising resource consumption

- Chunk large datasets
- Use vectorised operations
- Avoid unnecessary I/O



These techniques reduce load on CPU, memory, and disk - keeping your system responsive and stable.

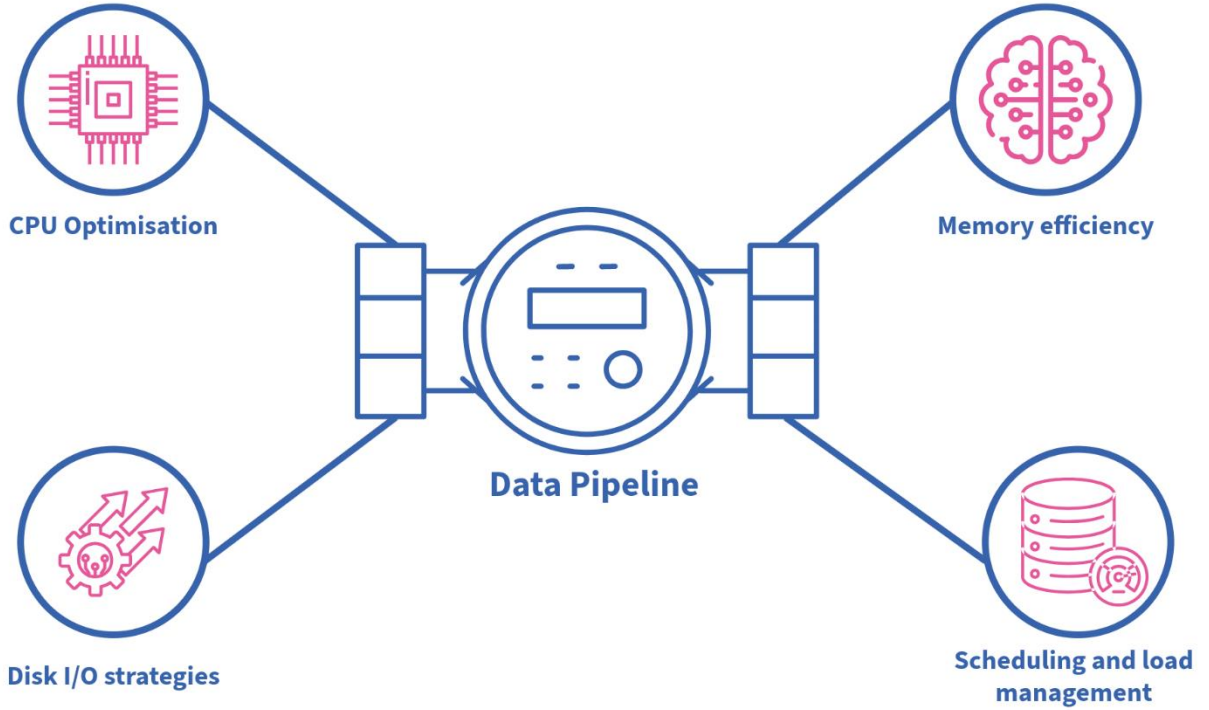
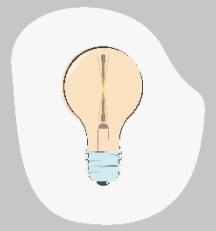


Figure: *Ensuring System Health through Regular Maintenance and Monitoring*

Scheduling strategy

Avoiding the midnight crash

- Stagger job start times
- Prioritise critical workloads
- Schedule off - peak execution



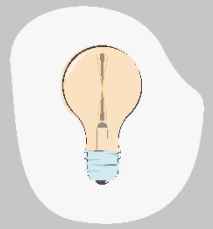
Tip: Don’t schedule all jobs at once—stagger them, prioritise critical tasks, and use off-peak hours to keep systems running smoothly.

Like a shared kitchen, your infrastructure needs coordination - stagger jobs, prioritise tasks, and schedule smartly to avoid a system traffic jam.

Tooling up

Docker & Kubernetes On-Prem

- Manual provisioning and networking
- Local storage and registries
- Monitor everything yourself



Docker and Kubernetes are powerful - but on - prem, you're responsible for the plumbing.

Logos for two tools common to a on-premise data infrastructure: Docker and Kubernetes

Real - World Lessons

Building Careers Through Education

Everybody Matters. Stronger Together. Student, Learner and Client Centric. Trust and Respect. Embrace Change.

Case Studies in Action…

Manufacturing: Staggered jobs → no more failures,

Image source: itn.co.uk

Hospital: Rehearsed updates → enabled rolling upgrades.

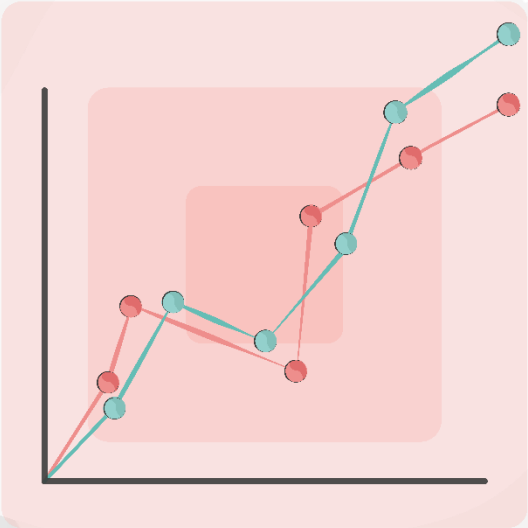
Image source: Tagmedstaffing.com

University: Added quotas → fairer GPU usage.

Image source: Universitiesac.co.uk

BPP

Practical lab



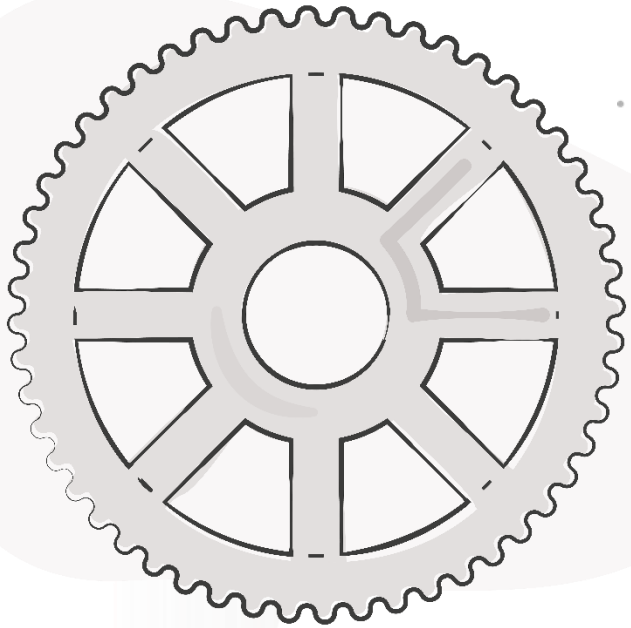
Exercise part 3

Building Careers Through Education

Everybody Matters. Stronger Together. Student, Learner and Client Centric. Trust and Respect. Embrace Change.

Next phase of design based on US data

- **Update Database Schema:** Modify the schema to accommodate data from the American version of the software application.
- **Re-implement Data Pipeline:** Develop a pipeline to clean, validate, and load UK, French, and American data into the revised schema.
- **Use Test Data:** Utilise provided test data for 10 American users and their login timestamps.
- **Future Phases:** Prepare to integrate data from additional countries w varying formats.
- **Documentation:** Establish comprehensive documentation for the schema and pipeline for future implementation by other teams.



Practical challenge

Files Provided in the Hub:

- US User Data.csv: Contains 10 sample records.
- US-User-LoginTS.csv: Contains login timestamps for January 2025..

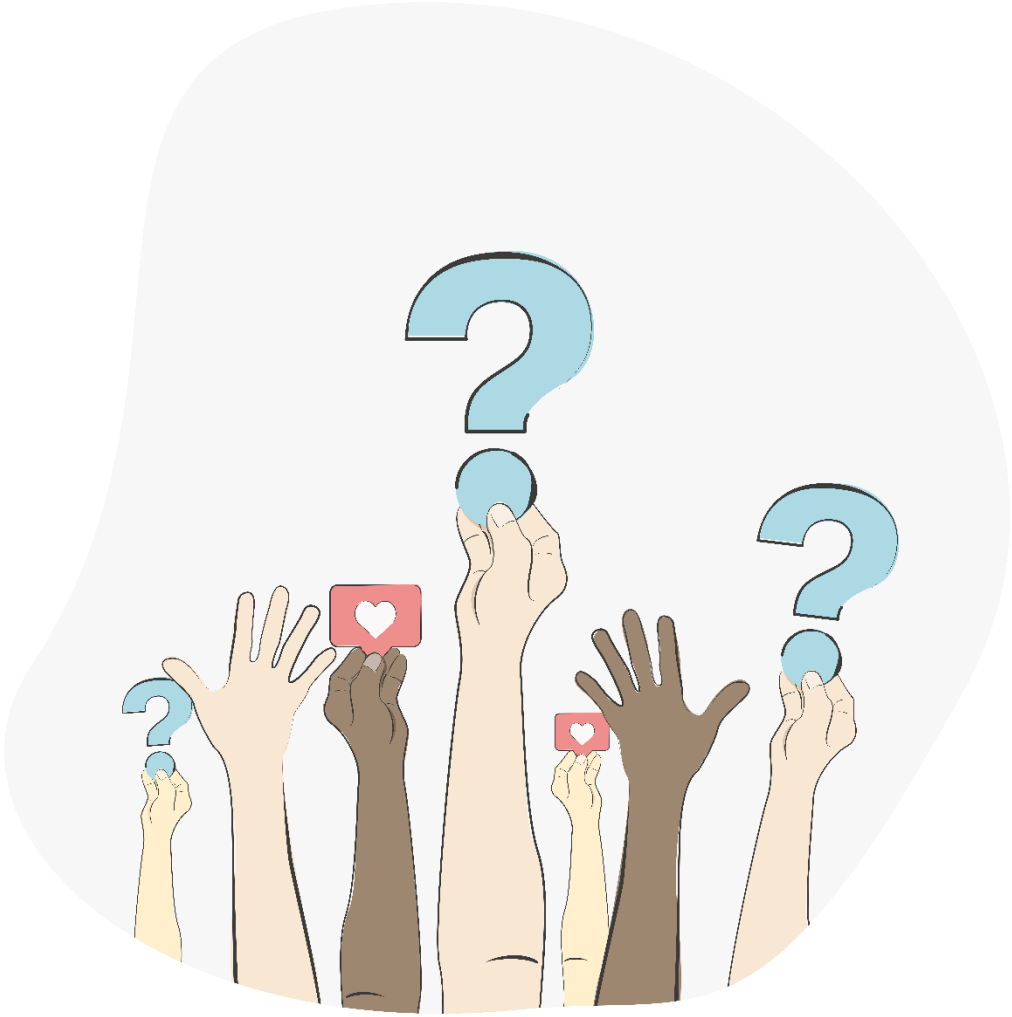
BPP

Key Learning Summary

- Running your own servers means owning everything from cabling to backups - and learning to think like a systems engineer.
- Know what happens under the hood when pipelines run - and why unmanaged resource use can bring everything else crashing down.
- Small adjustments like chunking data, limiting threads, or compressing files can make a big difference on tight on-prem systems.
- **Timing really matters -** when and how you run your jobs can prevent bottlenecks and keep shared systems running smoothly.
- Docker and Kubernetes still work on-prem, but you have to build the cloud-like environment around them yourself.
- Real-world success comes from visibility, good habits, and learning from the things that go wrong before they break everything.

BPP

Any questions or feedback?



Building Careers Through Education

Everybody Matters. Stronger Together. Student, Learner and Client Centric. Trust and Respect. Embrace Change.

BPP

Thank you