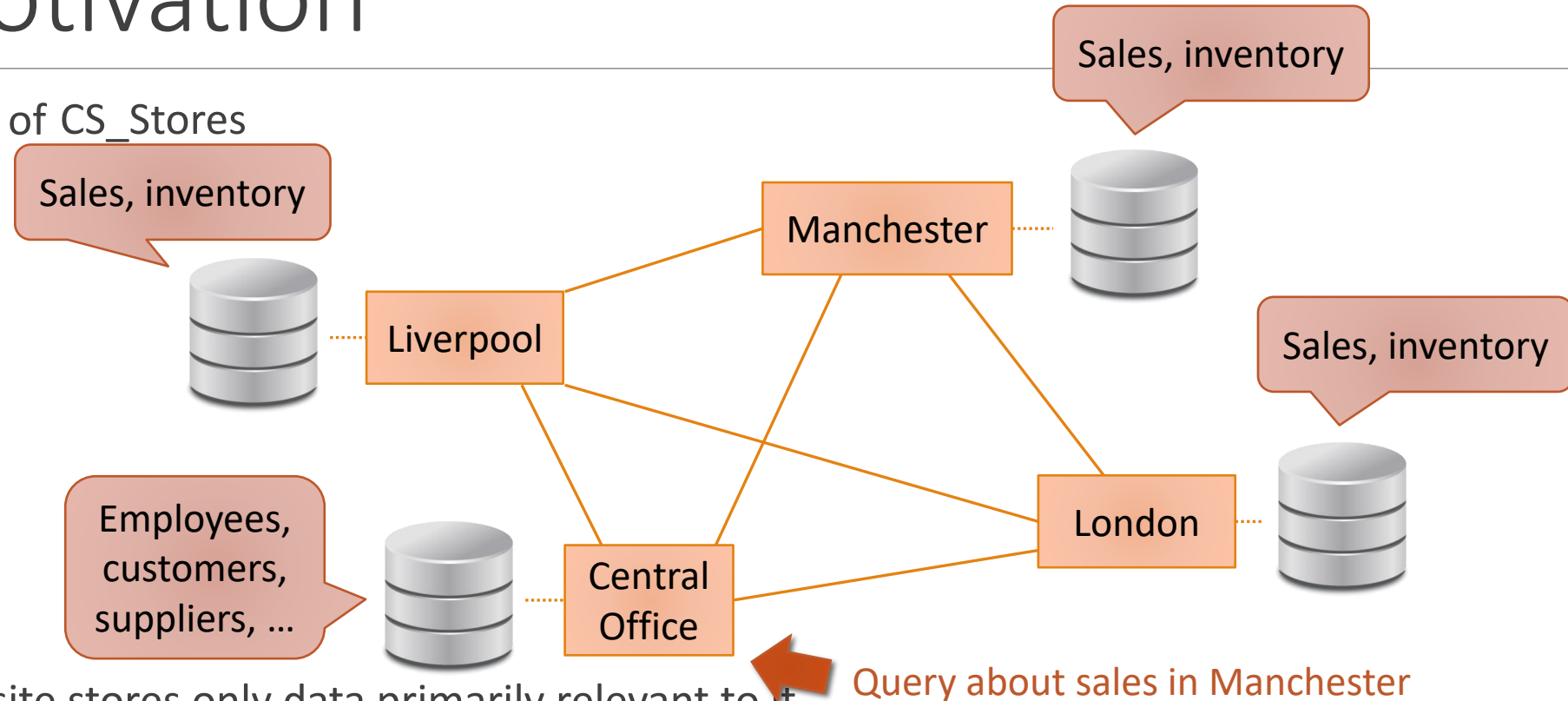# Introduction to distributed databases

Overview over this video

In this video, we will see what a distributed database is and why they are important
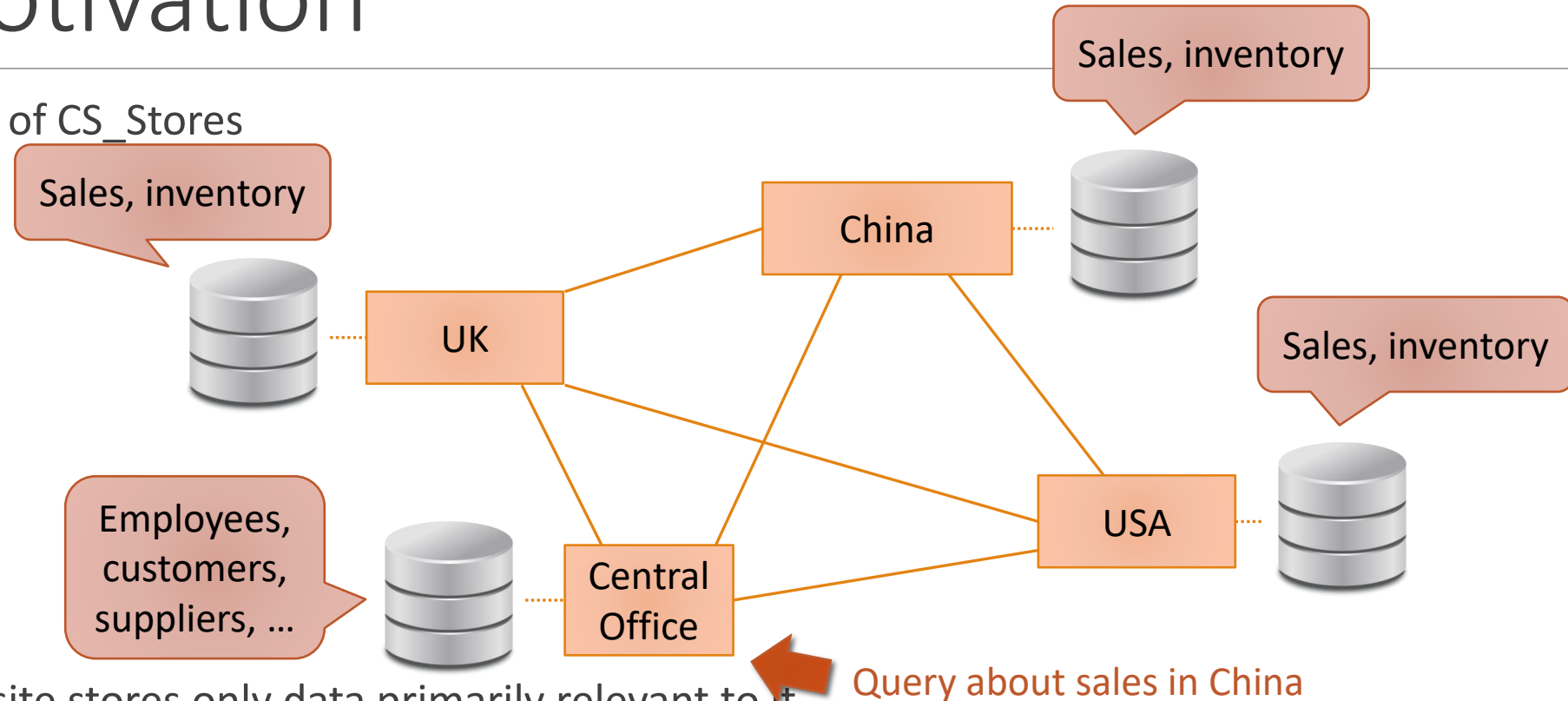
Motivation

Chain of CS_Stores

Each site stores only data primarily relevant to it

Distributed DBMS provide access to data at all sites

Motivation

Chain of CS_Stores

Each site stores only data primarily relevant to it

Distributed DBMS provide access to data at all sites

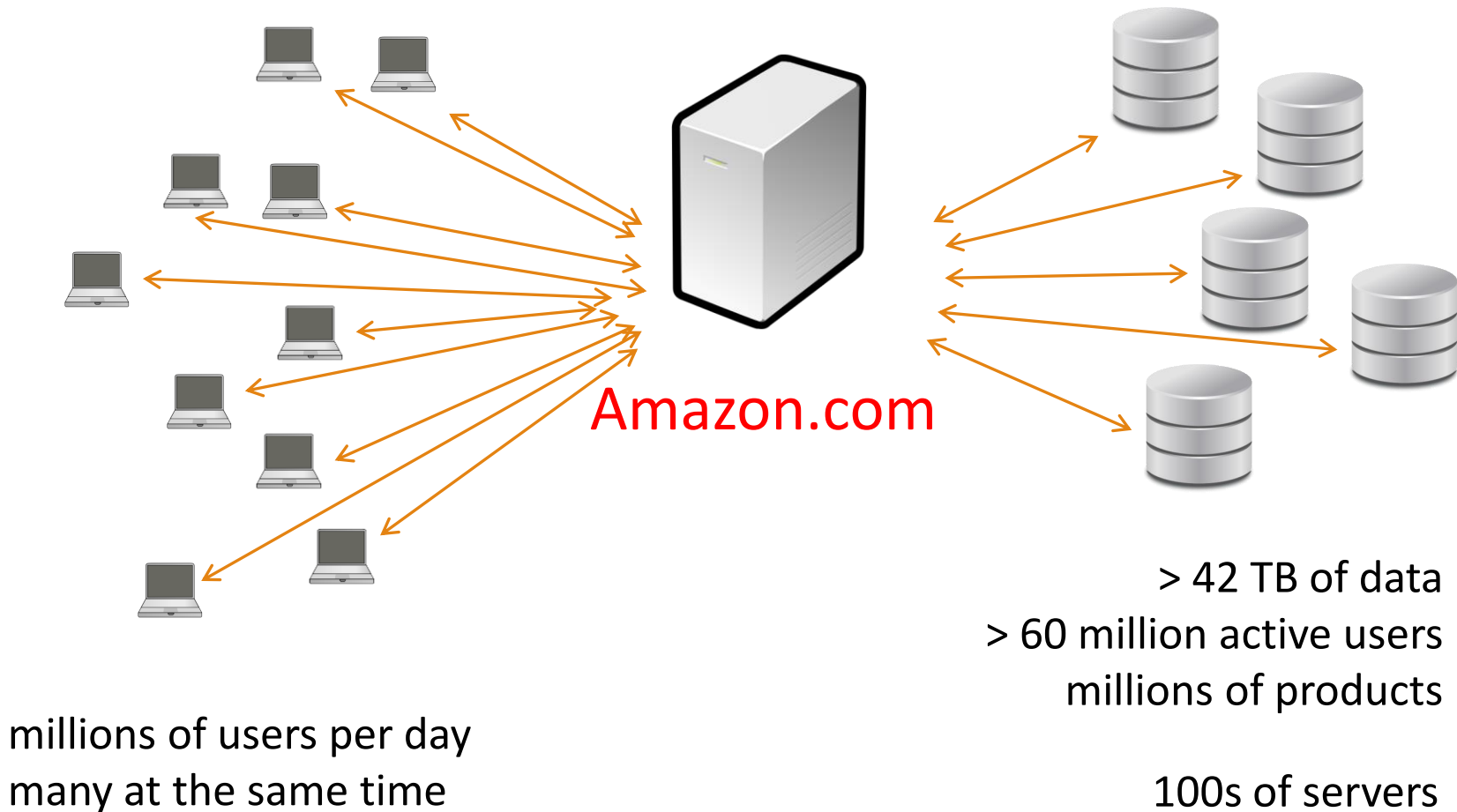# Other Applications

More general: large organisations/companies
- …with different branches or offices
- …with different sub-companies
- …or simply so large that one computer can't handle all the request

Providing access to large datasets to many users (e.g., for an online store)
- Distribute data over several computers – not necessarily identical (in software or hardware)
- Computers could be at geographically separate locations (but also possible that they are the same place)
- Possible advantages:
  - Balance workload & network traffic
  - Easier to extend capacity or scale to higher number of users

From "introduction to databases" video

Amazon.com

> 42 TB of data
> 60 million active users
millions of products

100s of servers

millions of users per day
many at the same time

Distributed Databases

Distributed Database:
- ◦ Collection of multiple **logically interrelated** databases
- ◦ **Distributed** over a computer network

B

Node/site = computer in the network, where a database is stored

A

Network link

C
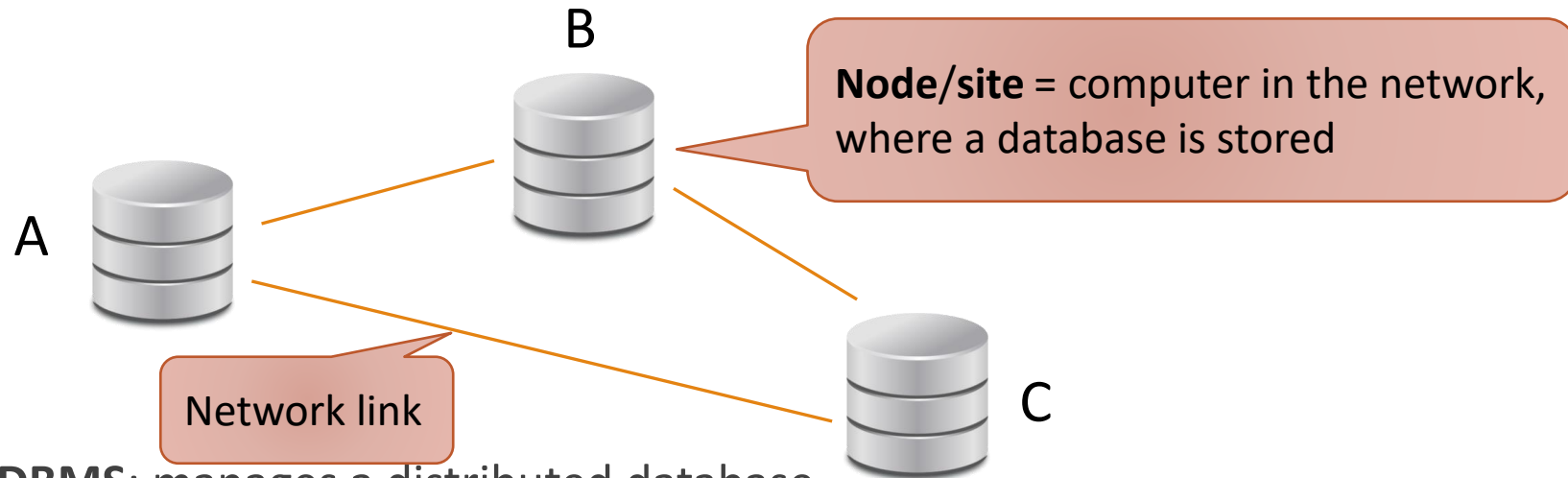
Distributed DBMS: manages a distributed database

Advantages

Performance improvements
- Answer queries faster by distributing tasks over the nodes
- Reduces CPU time, disk accesses, communication cost, … at individual nodes

Scalability
- Easier extension of the system: capacity, performance, …
- Essentially just add a new node

Resilience
- Data can be replicated at geographically separate sites
- Catastrophic failures don't affect the entire system

Summary

A distributed database is a database made up of multiple logically interrelated databases that is based on multiple different computers, distributed over a network