CHAPTER 21 / 4 MINUTE BRIEF

Multiple Markets

Read the full chapter → / Read this brief online →

The main idea

Once an organization serves several markets, some of its placement rules become conditions that no commercial agreement can change, and two locations with the same machines stop being interchangeable. The earlier examples kept those privacy, sector, and infrastructure rules in the background. Scale makes them harder to coordinate, but a small business using an overseas provider meets them too. Before moving work between locations, the capacity team needs to know which ones its specialists have cleared, and on what conditions.

The example organization from this book has run a European operation for years, under its own budget and its own agreement with Provider B, which is why Volume I, Chapter 1's ledger leaves it out. That bill put the web workers on Provider A, the serving layers behind the data platforms and AI assistant on Provider B, and the customer and order records in the organization's own data center. European customers are served from Provider B's European regions, chosen partly because its data-residence terms satisfied legal review, while Provider A keeps serving the Americas. The same React application, the same tier structure, the same databases behind it, and a topology that looks arbitrary until somebody asks why the European workers cannot simply fail over to Virginia. The duplicated stack shows up in the bill long before anyone mentions the law.

Explore this section in the full chapter →

Location is a capacity property

Once placement is constrained, geography becomes part of what "capacity" means. The tested worker capacity is unchanged, but the locations are no longer interchangeable: a region that could absorb the web tier's failover traffic may be unable to hold the data the tier serves. Chapter 18 asked what a reserve is available for, and each candidate location an-

swers differently. Capacity counts toward a workload only where it is permitted, available, and tested.

Then price the permitted arrangements, including quoted rates, chargeable data transfers, and the standby or duplicated services each topology requires. Volume I, Chapter 1's "where it runs" view was drawn with two providers and a data center; I would add permitted locations and their conditions to it before comparing costs, so that the cheapest row is also one we could operate.

Explore this section in the full chapter →

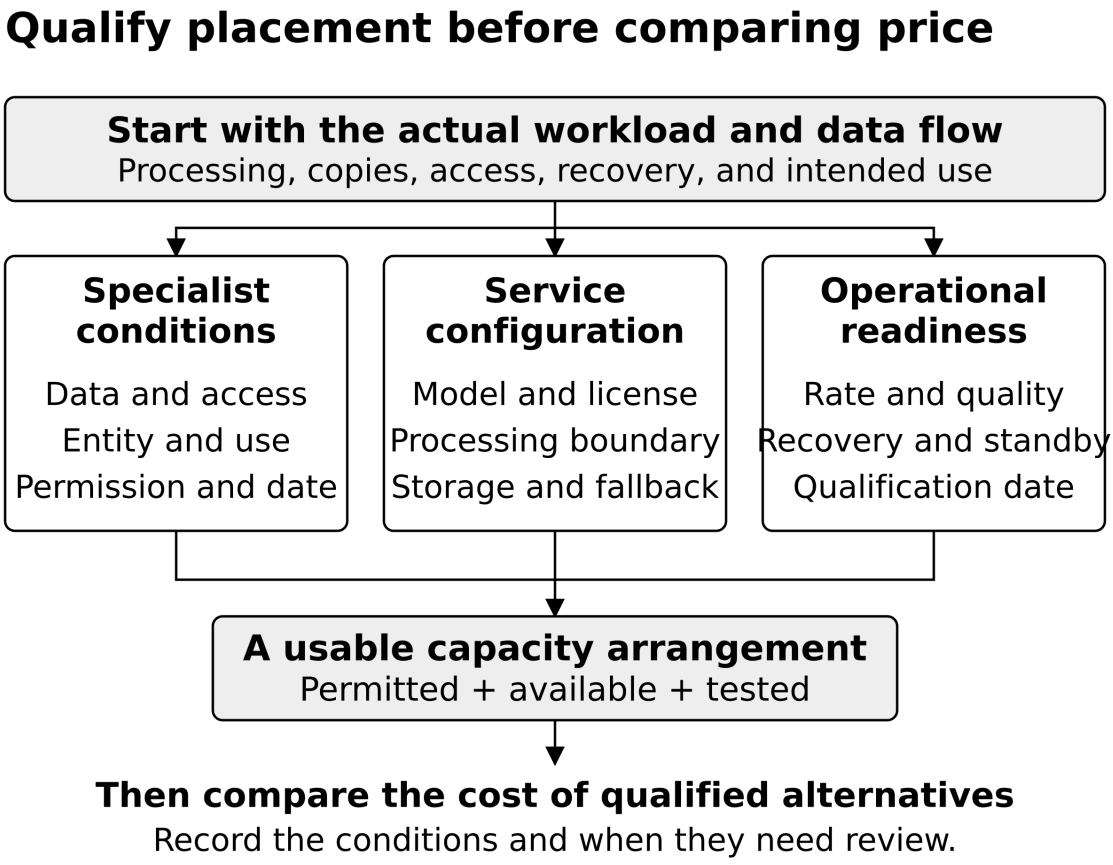


Diagram 21.1. Placement becomes usable capacity only when specialist conditions, service configuration, and operational readiness agree. Compare prices among those qualified arrangements, and retain each condition's review date.

Explore this section in the full chapter →

Consider segmentation by market

Segmentation is an option to evaluate, not a requirement to infer from a country boundary. I would identify which workloads need separate placement or operations, then look for what still connects them: identity, telemetry, support tooling, backups, and the deployment pipeline can each carry data or access across the line. Separation can limit the effect of a local rule change; it cannot guarantee independence. Where lawful sharing is possible, weigh its savings against the cost of separating later.

Explore this section in the full chapter →

Segmentation costs utilization

Separated systems will usually not reach the efficiency of one pooled system. That is a fact of life, and it does not contradict the earlier chapters: deliberately unused capacity can protect a service or satisfy a constraint.

The mechanism is the minimum configuration that availability and recovery requirements impose: replicas, standby capacity, test environments, and the people and systems to operate them. If two deployments must stay separate and each needs its own standby, idle capacity in one cannot meet the other's obligation. Show those duplicated minimums and why they exist. Judging each market against an unconstrained global pool would mistake a placement requirement for an efficiency failure.

Explore this section in the full chapter →

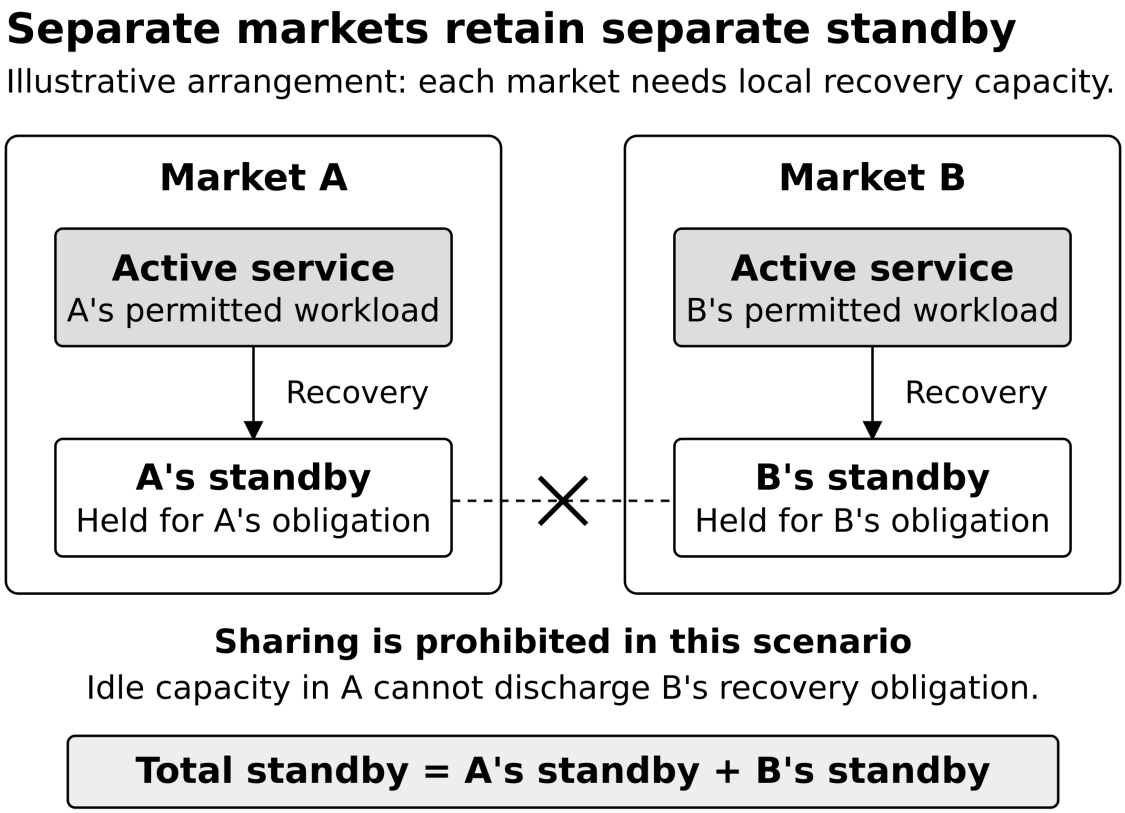


Diagram 21.2. In this illustrative arrangement, sharing is prohibited and each market needs its own standby. The unused capacity protects a local recovery obligation; it cannot be counted again as the other market's reserve.

Explore this section in the full chapter →

The monthly review across markets

Suppose the organization starts measuring the reuse of AI infrastructure across markets, and maximizing reuse becomes a success criterion. Capacity that looks idle on one market's chart may be unavailable to another, and a criterion written as if the deployment were unified will mark correctly segmented systems as failures. So the review keeps the constraint register beside the efficiency numbers, with specialists confirming the conditions behind any cross-market comparison. When a rule or interpretation changes, note its effective date, the options it removes or adds, and the time needed to respond, so nobody discovers an expired condition during a failed recovery or a purchase review.

Explore this section in the full chapter →

One practice or two

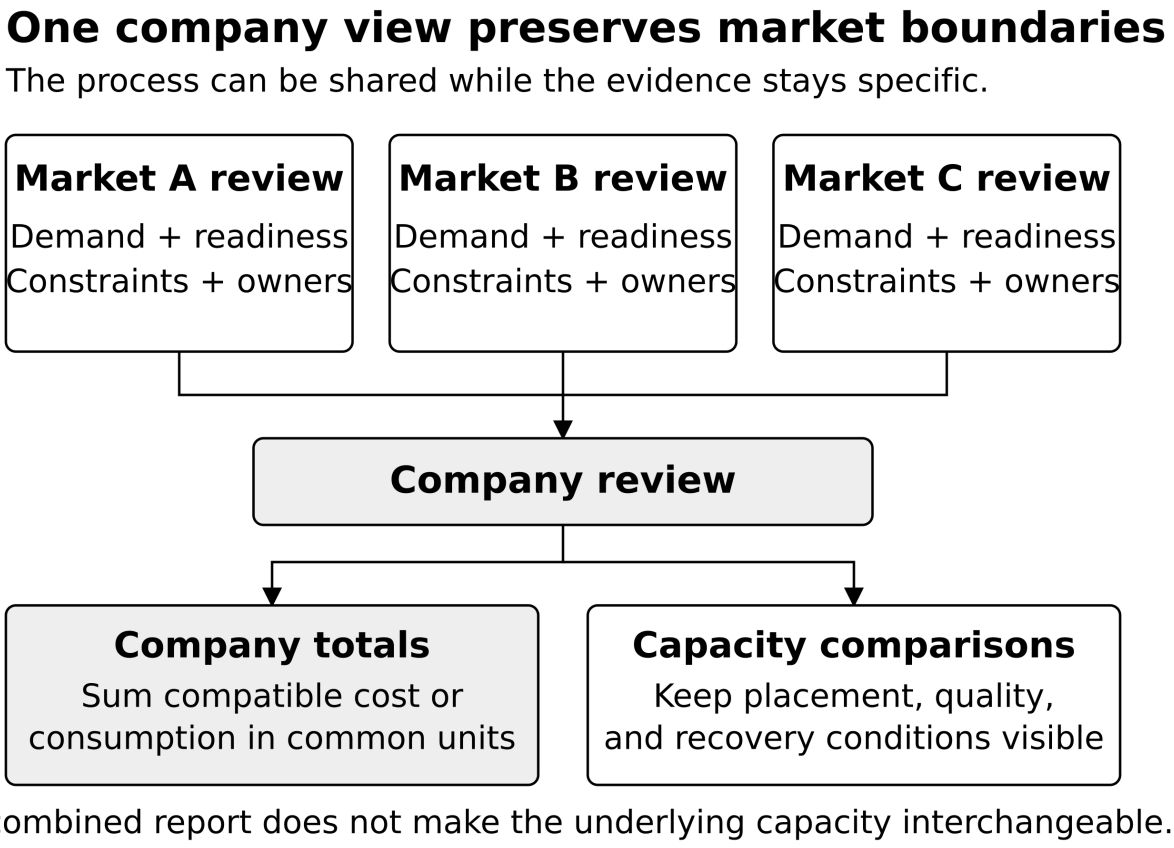


Diagram 21.3. Separate market reviews can feed one company view. Compatible cost or consumption units can be summed; capacity comparisons must preserve the placement, quality, and recovery conditions that make each market's numbers meaningful.

Explore this section in the full chapter →