CHAPTER 20 / 3 MINUTE BRIEF

Commitments Between Organizations

Read the full chapter → / Read this brief online →

The main idea

"We expect to have it in six weeks" is useful information. It becomes a different kind of information when a project plan turns it into a date and a customer announcement turns that date into a promise. By the time the estimate returns to the capacity team, several people may be relying on a level of certainty that the supplier never offered.

Nobody needs to have intended that transformation. Each person used the best information available to move their own work forward while the original answer's conditions became harder to see. We need those conditions to survive the journey, whether the arrangement is between companies or between teams whose shared company logo conceals different customers, budgets, and obligations.

Explore this section in the full chapter →

A discount and a resource are different things

Azure and AWS arrangements, as of October 2026:

Arrangement	What we owe	What we have when supply is short
Spend commitment: Azure savings plan, AWS Savings Plans[1][2]	A fixed amount per hour for one or three years	No capacity; unused hours expire but are still paid for

[1]Microsoft Learn, "What are savings plans?" and "Automatically renew your savings plan", checked October 4, 2026. Hourly commitment for one or three years; unused hours expire; no cancellation or refund; current savings plan prices apply; renewal turned off by default, with pay-as-you-go rates after expiry. The $10/$7 illustration is hypothetical, not an Azure offer.

[2]Amazon Web Services, Savings Plans User Guide, "What are Savings Plans?", "Compute Savings Plans and Reserved Instances", "Understanding how Savings Plans apply to your usage", and "Returning a purchased Savings Plan", checked October 3, 2026 (return conditions checked October 4, 2026). Hourly commitment with prices fixed for the term; no capacity reservation; no cancellation; unused hours not carried over; the seven-day return conditions.

Arrangement	What we owe	What we have when supply is short
Consumption commitment, such as Microsoft's MACC[3]	A total over the term, perhaps with milestones; any shortfall is charged	No capacity; only eligible spending reduces the balance
Instance commitment: Azure Reserved VM Instance, AWS regional Reserved Instance[4][5]	The reservation price for one or three years, upfront or monthly	No capacity; the discount covers only what we can deploy
Capacity reservation: Azure on-demand capacity reservation, AWS On-Demand Capacity Reservation[6][7]	The on-demand rate for the reserved quantity, used or not; discounts can apply	The reserved size and quantity in one region or zone; Azure's SLA remedy is a service credit

[3]Microsoft Learn, "Track your Microsoft Azure Consumption Commitment (MACC)", checked October 4, 2026. Commitment, milestones, shortfall charged as Azure prepayment credit, and alerts 90, 60, and 30 days before expiry. The page does not address capacity availability; each agreement's terms govern.

[4]Microsoft Learn, "Prepay for Azure virtual machines to save money" and "Transition guide for retired Azure Reserved VM Instances", checked October 3, 2026 (transition guide checked October 4, 2026). Reserved VM Instances are a billing discount, not a capacity guarantee, and can be combined with capacity reservations; capacity priority requires single-subscription scope; purchases and renewals for several older series ended July 1, 2026, with existing reservations running to term. Vendor documentation; terms change.

[5]Amazon Web Services, Amazon EC2 User Guide, "Reserve compute capacity with EC2 On-Demand Capacity Reservations", "Concepts for Amazon EC2 Capacity Reservations", "View the state of a Capacity Reservation", "Capacity Reservation pricing and billing", and "Regional and zonal Reserved Instances (scope)", checked October 3, 2026 (future-dated terms and eligible families checked October 4, 2026). Billing and discounts for Capacity Reservations; zonal versus regional Reserved Instances; future-dated lead times, minimum commitment, eligible families, and states. Product limits change.

[6]Microsoft Learn, "On-demand capacity reservation in Azure" and "Create a capacity reservation", checked October 3, 2026 (overview checked October 4, 2026). Region or zone holds billed at the VM rate whether used or not, with discounts applicable; no SLA for capacity priority; the SLA's worked service-credit example and customer-calculated claims; ND-series and NCads H100 v5 unsupported. Microsoft's SLA text governs, and supported sizes change.

[7]Amazon Web Services, Amazon EC2 User Guide, "Reserve compute capacity with EC2 On-Demand Capacity Reservations", "Concepts for Amazon EC2 Capacity Reservations", "View the state of a Capacity Reservation", "Capacity Reservation pricing and billing", and "Regional and zonal Reserved Instances (scope)", checked October 3, 2026 (future-dated terms and eligible families checked October 4, 2026). Billing and discounts for Capacity Reservations; zonal versus regional Reserved Instances; future-dated lead times, minimum commitment, eligible families, and states. Product limits change.

[8]Amazon Web Services, Amazon EC2 User Guide, "Capacity Blocks for ML" and "Capacity Blocks pricing and billing", checked October 4, 2026. Up to eight weeks ahead and 64 instances per block; no cancellation; price set at purchase by supply and demand, charged upfront; no Savings Plans or Reserved Instance discounts. Supported types and Regions change.

[9]Amazon Web Services, Amazon EC2 User Guide, "Reserve compute capacity with EC2 On-Demand Capacity Reservations", "Concepts for Amazon EC2 Capacity Reservations", "View the state of a Capacity Reservation", "Capacity Reservation pricing and billing", and "Regional and zonal Reserved Instances (scope)", checked October 3, 2026 (future-dated terms and eligible families checked October 4, 2026). Billing and discounts for Capacity Reservations; zonal versus regional Reserved Instances; future-dated lead times, minimum commitment, eligible families, and states. Product limits change.

Dated capacity: AWS Capacity Blocks for ML, future-dated Capacity Reservations[89]	Blocks: the full price upfront, no cancellation. Future-dated: the on-demand rate for a minimum term, usually 14 days	The reserved instances for the booked period; a future-dated request can be delayed or declined before it starts

Explore this section in the full chapter →

The internal promise is still a promise

I would keep the agreement short enough that the people relying on it can read it, but specific enough to establish the resource or service, quantity and unit, placement, readiness date, duration, quality or failure condition, and payer. It also needs the authority to change the arrangement and the demand assumption behind it. If the plan works only for a staged rollout, that condition is doing more work than another list of administrative approvals.

Explore this section in the full chapter →

Evaluate the price of certainty

In a simplified pricing illustration, variable purchase costs $10 per unit-period. A take-or-pay agreement offers $7 per unit-period for 100 units over twelve periods, with no resale, cancellation, or other charges. Assume use never exceeds 100 units in any period, so there are no overage charges. We would commit to $100 \times 12 \times \$7 = \$8{,}400$, the same amount that 840 unit-periods would cost at the variable price.

There are 1,200 possible unit-periods, so break-even use is $840 \div 1{,}200 = 70$ percent, which is simply $\$7 \div \$10$. At 600 unit-periods, variable purchase would cost $6,000, so the agreement costs $2,400 more; at all 1,200, variable purchase costs $12,000 and the agreement costs $3,600 less. That $3,600 is the most it can gain, while demand that disappears entirely would lose the whole $8,400. The $7 rate has not changed between the cases; the amount of useful demand has changed its value to us.

Explore this section in the full chapter →

Carry the conditions into the review

Agreement gives us a set of conditions to watch. Product changes the rollout, a new request mix changes the service rate, a supplier revises delivery, or qualification reveals less

throughput than expected, while the borrowed workers approach their return date regardless. Any one of these changes can reopen the combined decision; Chapter 22 turns them into triggers agreed in advance.

Someone has to own those connections, not just the inputs. In the constructed portfolio, the capacity lead maintains the joined view and brings the choice back before the available options expire; the people with authority over the promises still decide which ones can change.

Explore this section in the full chapter →