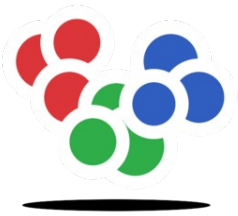


ccPlanning

ccplanning.net

WORKFORCE PLANNING MASTERCLASS

Capacity Planning, End to End

Turning a volume forecast into a monthly FTE plan and a recruitment schedule that respects ramp, attrition and budget

John Casey

Workforce Planning Expert · ccPlanning

A masterclass in the ccPlanning white-paper series · read the rest at ccplanning.net

Contents

1. What capacity planning is

Forecasting predicts the demand. Capacity planning is the discipline that answers it with people — the bridge between a volume forecast and a headcount budget, and the place where the biggest, earliest and most expensive decisions in the operation get made.

Scheduling and real-time management live in days and weeks. Capacity planning looks months and quarters ahead and asks one hard question: do we have — or can we recruit and train in time — enough productive people to meet the demand we are forecasting? It is the most strategic seat in the planning function, because the answer drives hiring, budget and, ultimately, whether the operation hits service at a cost the business can bear.

Planning happens on three horizons. The **strategic** horizon, one to three years out, sets budgets, sites and the operating model. The **tactical** horizon, three to twelve months, turns that into a hiring plan. The **operational** horizon, measured in weeks, is schedules and real-time. Capacity planning owns the top two, and each horizon hands constraints down to the next: get the strategic numbers wrong and no amount of clever scheduling at the bottom can rescue them.

Why is it a distinct job, rather than a corner of forecasting? Lead time. You can re-time a break in seconds, but a new agent takes weeks to recruit and months to reach full productivity — so the decision to staff for next quarter must be made this quarter. This paper walks the whole chain: from a forecast, to a required-FTE profile, to an honest plan that survives attrition, ramp and a finite recruitment pipeline, to a business case a finance director will actually approve. It is written for planners moving from operational into strategic work, and for the managers who depend on the number they produce.

> *You can re-time a break in seconds, but you cannot summon a trained agent on the day. That single fact is why capacity planning exists.*

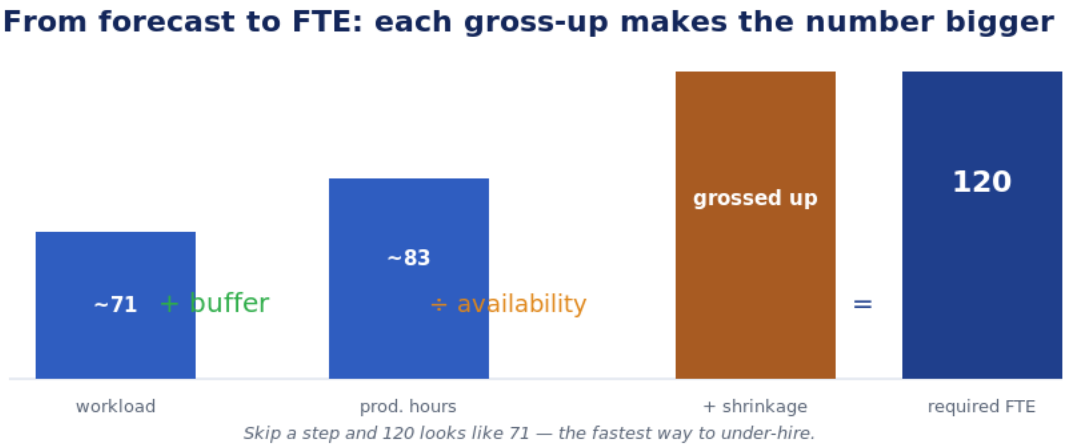
2. From forecast to FTE

You cannot run interval-level Erlang across every fifteen minutes of next year — you have no intraday profile that far out, and you do not need that precision. Capacity planning works in aggregate: it totals the workload over a month and converts it into the number of full-time-equivalent staff needed to carry it. That conversion is a chain of gross-ups, and the discipline is to apply every link, because each one makes the number bigger, and skipping any of them quietly under-staffs the plan.

The conversion chain

Start with **workload** — forecast contacts multiplied by forecast AHT — which gives the raw handling hours of work to be done. Add a **service buffer**, because a live channel cannot run at 100% efficiency; you need slack between contacts to hit service level, so you gross the workload up to the productive agent-hours actually required. (Large operations need a smaller

buffer than small ones, because big queues pool randomness more efficiently.) Then **convert hours to bodies** by dividing by the productive hours one agent really delivers — which brings in the largest gross-up of all: **shrinkage**. If 30% of paid time is lost to leave, training, sickness and breaks, each agent gives only 70% of their contracted hours to contacts, so you need more contracted FTE than the productive-hours figure alone suggests.



Each gross-up enlarges the number. A workload that looks like 71 heads can require 120 once the buffer and shrinkage are applied.

A worked month

Suppose a month carries 10,000 productive handling hours of work. Apply a service buffer at, say, 85% occupancy and you need roughly 11,800 productive agent-hours. Each agent offers about 140 contracted hours that month — but at 30% shrinkage only about 98 of those reach contacts. Dividing 11,800 by 98 gives roughly 120 required FTE — not the 71 the raw work alone implied. The gap between 71 and 120 is not padding; it is the buffer and the shrinkage the operation genuinely loses, and a plan that omits either is short before it starts.

Per month, and sense-checked

Volume, AHT and shrinkage all move through the year, so required FTE is a profile, not a single annual number — an average hides the months you will be desperately short. And whatever the bottom-up build says, cross-check it top-down against a simple contacts-per-FTE ratio from history, adjusted for known change. If the two are wildly apart, an assumption is wrong, and that reconciliation catches the spreadsheet error before it reaches the budget. The required-heads profile that falls out of this step is the spine of everything that follows.

3. The three drags: attrition, shrinkage, ramp

The required-FTE number tells you how many people you need in seats and productive. It never tells you how many to recruit, because you are not filling a static tank — you are holding a level of water in a leaky bucket. Three drags pull effective capacity below the headline number, and a plan that forgets any one of them comes up short.

- **Attrition — the leak.** Lose 30% of agents a year and you must hire that many just to stand still, before any growth. Attrition is not flat or random: new starters leave faster than tenured staff, some months are worse than others, and a pay review or a competitor opening nearby can spike it. Forecast it with the same care as volume.
- **Shrinkage — the permanent tax.** Each head delivers far less than their contract, so shrinkage is a standing multiplier on how many bodies you need; a few points of drift adds real FTE and real cost.
- **The ramp — the delay.** A recruit in week one is in training, then ramping for weeks or months. Counting a bum on a seat as a full, productive head overstates what you actually have.

So the real recruitment number is required productive heads, grossed up for shrinkage, **plus** enough hiring to cover both attrition and the ramp lag. It is always bigger than the headline requirement — and the danger is the capacity death-spiral: under-plan attrition and your headcount erodes month after month while the forecast climbs, until you are permanently chasing a gap you cannot recruit your way out of fast enough.

> *Required heads is the start of the plan, not the answer. Attrition leaks the bucket, shrinkage taxes every head, and the ramp delays new ones.*

4. The recruitment pipeline and lead times

The gap that matters is not the one you can see this month — it is the one a full lead time away. If it takes three months to advertise, hire, onboard and train someone to productivity, a shortfall you spot in June had to be actioned in March. Pipeline planning is as much about timing the hire as sizing it, and the whole discipline is built on planning backwards from the date you need people on the phones.

The deadline that matters is not when demand arrives — it is when recruitment must start to land trained people in time.

Yield and ramp inflate the number

Two things enlarge the figure you actually recruit. First, yield: not every applicant becomes a productive agent — some decline the offer, some no-show on day one, some fail training — so

if you need ten productive heads and your funnel yields 60%, you must start around seventeen people in the pipeline. Plan recruitment on the yield, never on the target headcount. Second, the ramp: a trained agent out of the classroom might be 50% effective, climbing to 100% over weeks, and during that time they consume capacity — a coach, a buddy — as well as adding it. Model effective FTE by applying a productivity curve to each starter cohort and summing the real contribution, rather than counting graduates as full heads.

Respect your own onboarding machine

Trainers, classrooms and floor-walking buddies are finite, so a plan that needs forty starters next month is worthless if you can only train fifteen at a time — sometimes training throughput, not demand, is the binding constraint. Put it together by back-planning from the month you need the heads: subtract the ramp, the training, the hiring lead time and the yield, and that tells you when the requisition opens and how big it must be. The classic failure is seasonal — realising in November that you are short for December, when the pipeline can no longer deliver trained people in time. For Christmas, you hire in autumn.

5. Building the plan: required vs available

For all its spreadsheet complexity, a capacity plan is at heart two rows across the months: required heads, from the forecast-to-FTE build, and available heads — what you will actually have after attrition and ramp. The gap between those two lines, month by month, is the entire job.

Where available dips below required, you have a hiring problem to solve in time.

Where the available line dips below the required line, you have a hiring problem to solve in time.

Build the available line honestly

The required line is the demanding part to calculate but the easy part to believe. The available line is where plans quietly lie. Start with today’s productive heads and roll forward: each month subtract forecast leavers, add the effective (ramped) contribution of cohorts already hired, and apply shrinkage. That gives a realistic line — not the fantasy of "everyone we have, fully productive, forever." Where available falls below required you have a service risk; where it rises

above, you have wasted cost. Both are worth seeing early, while there is still time to hire into a gap or slow recruitment ahead of a surplus.

Levers, granularity and keeping it live

Hiring is the main lever for a shortfall, back-planned for lead time and yield — and the hiring plan literally falls out of the gaps once they are visible, because the capacity plan and the hiring plan are two views of the same model. But hiring is not the only lever: you can also reduce attrition, cut shrinkage, add overtime, defer off-phone work, or flex with temps, and a good plan shows which lever and when, not just "hire more." Build it at the grain you actually hire and deploy at — skill and site — because a total that balances can still hide a language shortfall or one site over and another under. And re-run it every month: a capacity plan built once and filed is wrong by week two, because actuals always drift from plan.

> *Required versus available, by month, kept live. The plan is a habit, not a document.*

6. Scenarios and sensitivity

The further out you plan, the wider the uncertainty — and a year-long plan rests on estimates of volume, AHT and attrition, each of which can be a little wrong, with the errors compounding over the months. Presenting one precise headcount number pretends to a certainty you do not have. Good capacity planning works in ranges and tested scenarios.

Run the plan three ways — a conservative low case, your best expected estimate, and a stretch high case — so you can say "180 heads if growth lands as expected, 165 if it is soft, 200 if it is strong," and let leadership decide how much risk to carry. Then test sensitivity: flex one input at a time and watch the required FTE. Usually one or two assumptions — most often attrition or AHT — swing the answer far more than the rest, and those are the numbers to forecast hardest, monitor most closely and caveat loudest. Model the big known unknowns as discrete what-if overlays — a product launch, a site closure, a major client win or loss — so leadership can see the capacity and cost implication of a decision **before** they commit to it.

You rarely staff to the worst case (too expensive) or the best (too risky). The scenarios let you choose consciously: plan to expected, hold a flex buffer of overtime or temps for the high case, and keep a slow-down lever for the low. That is a deliberate hedge rather than a single number you are quietly hoping is right. Resist the urge to model every permutation, though — a wall of twenty cases is as useless as one point estimate. Low, expected, high, plus a couple of named what-ifs, is enough for a real decision.

7. Presenting the plan

A perfect plan that does not get the budget approved is just a spreadsheet. The plan’s audience is finance and HR, and they do not think in FTE and shrinkage; they think in cost, risk and return. Getting the heads approved means translating the plan into their language and treating it as a business case, not a technical document.

Lead with the ask and the headline reason — "we need to hire 24 by September, here is why" — not the model; the spreadsheet is evidence held in reserve for the questions. Then make the strongest move in any capacity case: quantify the cost of inaction. The compelling version is not "hiring costs X," it is "**not** hiring costs more" — missed service, abandoned contacts, lost revenue, overtime burn, attrition from overload. Once the do-nothing path has a price tag, the hire often becomes the obviously cheaper option. Frame everything in their metrics: "12 FTE short" means little upstairs, but "service drops to 65%, around 4,000 abandoned calls a month, an estimated X in lost sales" lands.

Show the low/expected/high scenarios so leadership sees the risk, but give a clear recommendation rather than a menu with no guide. And because of lead times, make the decision deadline impossible to miss: "to have them productive for peak, recruitment must start in week 30, so a decision is needed by week 28." Concrete deadlines force decisions; vague need-by dates drift until it is too late to act.

> *The strongest capacity case is rarely "hiring costs X." It is "not hiring costs more" — with a number attached.*

8. The operating rhythm

Capacity planning is not an annual event; it is a rhythm. Re-baseline the plan every month against the latest actuals — attrition running hot, a forecast moving, a training class slipping — so the gaps you are acting on are real, not three months stale. Anchor that monthly cycle to the year: know when each peak's recruitment must start, when budget rounds fall, and when leave and training demand will spike, and work backwards from each.

A closing checklist for a plan you can defend:

- **Required FTE built per month** — workload, service buffer and shrinkage gross-up, sense-checked top-down.
- **Available line built honestly** — today's heads rolled forward with leavers, ramped cohorts and shrinkage.
- **Gaps read by month**, with the lever and timing named for each — not just "hire more."
- **Recruitment back-planned** from go-live for ramp, training, lead time and yield, within training throughput.
- **Built by skill and site**, not one pooled total that hides a local shortfall.
- **Scenarios and sensitivities** — low/expected/high, with a deliberate hedge and named what-ifs.
- **Costed and presented as a business case**, with the cost of inaction and a hard decision deadline.
- **Re-run every month** against actuals.

Do those eight things and you have moved from a back-office calculator to a partner the business plans its year around. That is the real prize of capacity planning done end to end.

Appendix. The FTE conversion, in formulas

The conversion from a month's workload to required contracted FTE, in one place:

- **Workload (productive hours) = forecast contacts × forecast AHT** (in hours).
- **Required productive agent-hours = workload ÷ target occupancy** (the service buffer; occupancy < 1).
- **Productive hours per agent = contracted hours × (1 − shrinkage).**
- **Required FTE = required productive agent-hours ÷ productive hours per agent.**
- **Gross hires to plan ≈ growth in required FTE + attrition over the period + the ramp lag**, divided by funnel yield.

Worked, for one month: 10,000 productive hours ÷ 0.85 ≈ 11,800 required agent-hours. 140 contracted hours × (1 − 0.30) ≈ 98 productive hours per agent. 11,800 ÷ 98 ≈ 120 required FTE.

About the author

John Casey has spent more than 30 years in contact-centre workforce planning, including roles as a workforce-planning manager and operations director, and now writes and teaches at ccplanning.net. The views in this paper are his own. It represents general professional experience and is not affiliated with, nor does it represent, any current or former employer.

ccplanning.net · Free guides, calculators and the rest of the white-paper series.