

Alphabet's next compute allocation

Internal ID: S&F-STR-042

Topic: Operations Strategy, Growth Strategy, Customer Strategy

Industry: Interactive Media & Services

Source: Strategy & Finance

Publication date: 5 October 2026

License: Not specified

Decision date: October 3, 2026

Information cutoff: October 3, 2026, 2:41 p.m. ET

Audience: MBA students and working professionals

The planning question

Alphabet's next infrastructure plan must support two demanding uses of AI compute. Google's consumer products—including Search, Gemini, and YouTube—need capacity for models, inference, and product improvement. Google Cloud needs capacity to deliver customer workloads, meet commitments, and compete for new enterprise demand. Both uses can strengthen the other: internal models and custom chips may improve Cloud's offer, while Cloud deployments may improve utilization and bring customer learning. The company does not publicly disclose the marginal economics needed to rank one workload against another.

At an October planning review, you are advising a cross-functional capacity committee. It must recommend a priority rule and a commitment pace for the next planning cycle. When capacity is scarce, whose workload goes first? How much should Alphabet commit before utilization, demand conversion, and unit economics become clearer?

The assignment is to recommend a decision rule, not to calculate a company-specific net present value from unavailable inputs. State what evidence would cause you to change the rule.

What the public results show

Company-reported facts. In Q2 2026, Alphabet reported \$119.796 billion in consolidated revenue. Search & other revenue was \$63.271 billion, compared with \$54.190 billion a year earlier; the company reported rounded growth of 17%. Google Cloud revenue was \$24.768 billion, compared with \$13.624 billion; the company reported rounded growth of 82%. Alphabet's Q2 results release also attributed Search query growth to popular AI features and reported 22 billion Gemini API tokens processed per minute and 950 million Gemini App monthly active users. These are management-reported usage indicators, not workload profitability measures. [P2, printed pp. 13-14 and 47-50; P3, pp. 1-2]

Search & other is an advertising revenue line. Google Cloud combines infrastructure and platform services and, beginning in Q2 2026, includes TPU system sales. Neither revenue line tells the committee the compute cost, marginal contribution, or return for an additional unit of capacity. The measures answer different business questions. [P2, printed pp. 13, 48; P3, p. 2]

Company-reported facts. Q2 segment operating income was \$39.544 billion for Google Services and \$8.814 billion for Google Cloud. Cloud's reported segment operating margin was 35.6% in Q2 2026, derived as \$8.814 billion divided by \$24.768 billion. In Q2 2025, the corresponding reported segment figures imply 20.7%. Shared AI research and development is

primarily reported in Alphabet-level activities rather than allocated to these segments. The segment figures therefore are not workload-level or fully allocated returns, and Cloud's reported margin is not directly comparable to an incremental return on new compute. [P2, printed pp. 48-50; P3, p. 2; calculation from reported figures]

Company-reported facts. Alphabet reported \$519.5 billion in remaining performance obligations, including \$513.9 billion for Cloud, and expected just over half of total backlog to be recognized over the next 24 months. The company says conversion timing depends on contract duration, delivery, and customer usage. The measure includes deferred revenue and future billings and excludes cancellable contracts. Alphabet expanded its backlog reporting in Q1 2026 to include contracts with original expected terms of one year or less. It is a demand indicator, not cash already collected or guaranteed profitable delivery. [P2, Note 2, printed pp. 13-14]

Capacity and capital at stake

Company-reported facts. Capital expenditures were \$80.598 billion in the first half of 2026, versus \$39.6 billion in the first half of 2025. First-half operating cash flow was \$84.859 billion. Management expected 2026 technical-infrastructure investment to increase significantly versus 2025, but the reviewed filing gave no numeric range. Data-center projects generally take months to years before assets enter service. [P2, printed pp. 51-52]

Derived values. First-half operating cash flow exceeded first-half capex by \$4.261 billion (\$84.859 billion – \$80.598 billion). In Q2 alone, \$44.9 billion of capex exceeded \$39.1 billion of operating cash flow by \$5.8 billion. The quarterly values are rounded. Neither comparison is a forecast of free cash flow or a funding shortfall. [P2, cash-flow statement printed pp. 8-9; MD&A pp. 46, 51-52]

Company-reported facts. Uncommenced leases, primarily for data centers, represented \$85.2 billion of future lease payments, with noncancelable terms of 1-26 years. A separate \$811.0 billion total for material purchase commitments and other contractual obligations included technical-infrastructure and inventory supply agreements, content licenses, and energy take-or-pay contracts. The two disclosures may overlap in economic purpose and should not be added as if they were disjoint compute obligations. Alphabet warns that long-duration agreements and third-party data-center and power backstops can leave it with liabilities or excess capacity that is difficult to redeploy. [P2, Note 4, printed pp. 22-23; MD&A pp. 53-54, 57; Note 10, pp. 29-30]

A competitive clock

Microsoft's FY2026 Q4 call reported that it added 31 data centers and another gigawatt of capacity in the quarter, with a plan to roughly double capacity in two years. Its \$678 billion commercial RPO included OpenAI; Microsoft said RPO excluding OpenAI grew 25%, and all sequential RPO growth came from customers outside frontier-model companies. Amazon reported Q2 2026 AWS revenue of \$42.2 billion, up 37%, and operating income of \$16.6 billion. These company-reported figures make delay a strategic question, but differences in fiscal periods, RPO definitions, segment mix, and accounting prevent direct return comparisons. [P4, full official transcript; P5, official Q2 release]

Decision alternatives

1. **Internal-capacity priority.** Protect Search, Gemini, YouTube, and shared model capacity first. This may preserve user experience, product differentiation, and advertising economics across a large installed base. The risk is that Cloud cannot reliably convert backlog and customer interest into capacity, revenue, and long-term relationships.
2. **Contracted and high-value Cloud priority.** Reserve capacity for workloads tied to firm customer commitments and attractive expected contribution. This may capture enterprise demand and avoid losing workloads to competitors. The risks are that backlog conversion and workload economics are not fully visible, and long-duration commitments may outlast customer demand or the technology.
3. **Staged portfolio gates.** Protect a minimum internal service and research envelope, meet delivery obligations, and release discretionary capacity in stages as utilization, customer usage, unit cost, and commissioning milestones are observed. This preserves learning and flexibility, but gates may slow scale, raise unit costs, or leave capacity unavailable when demand arrives.

Each alternative requires a clear rule for conflicts, such as a shortfall in commissioned capacity or a change in expected customer usage. The available public data cannot establish the correct answer.

Your assignment

Prepare a two-page recommendation for the committee. Specify:

- a priority rule for internal and Cloud workloads;
- the pace and reversibility of new commitments;
- the measures and thresholds that would trigger another investment tranche or a pause;
- how you would account for shared AI research, customer commitments, asset utilization, and redeployment risk;
- the strongest argument against your chosen approach and the evidence that would change your recommendation.

Separate reported facts from your own calculations and clearly label any teaching assumptions. Do not treat revenue growth, backlog, reported segment margin, or competitor capex as a direct measure of incremental workload return.

Source notes

Primary sources and exact locators are cataloged in the accepted source register; extracted facts and calculation conventions are in the evidence ledger. Relevant source IDs are P2-P5. No information after the stated cutoff is included.