OA

WHAT I READ THIS WEEK

When Does Space Become the Cheapest Place to Compute?

Starship can reduce the cost of reaching orbit. It cannot repeal thermodynamics.

OTTO ANALYTICS · THE WEEKLY READ · NO. 07 · JULY 31, 2026

> *Orbital AI is no longer a question of whether useful models can run in space. They already can. The economic question is when cheaper launch and abundant solar power outweigh radiative cooling, radiation, communications, servicing and finite hardware life. The winning metric is not cost per kilogram. It is cost per useful compute-year.*

The week Earth proved it still needs more compute

TERRESTRIAL CAPACITY IS STILL BEING MONETIZED AS QUICKLY AS IT ARRIVES

This was an unusual week to start thinking about data centers in orbit.

Amazon reported that AWS revenue grew 37% year over year to $42.2 billion, its fastest growth in 18 quarters. AWS operating income reached $16.6 billion. Microsoft reported 43% Azure growth, said annual Azure revenue had surpassed $100 billion and again stated that customer demand exceeds available capacity. Microsoft added another gigawatt of capacity during the quarter and said the additional supply was quickly monetized.

Those are not signs that terrestrial AI infrastructure has run out of economic value. They are signs that the market can still absorb new capacity as fast as large operators can energize it.

Then Elon Musk offered the longer view: more than 90% of AI compute would remain server-side for the next few years, but almost all compute would eventually move to space.

The first statement fits the evidence. The second is not a forecast we can validate today. It is a prompt to ask what must become true.

The $210,000 gallon of milk

LAUNCH ECONOMICS EXPLAIN WHY THIS QUESTION IS BEING ASKED NOW

A NASA paper compared the Space Shuttle at roughly $54,500 per kilogram to low Earth orbit with Falcon 9's advertised benchmark of roughly $2,720 per kilogram. A U.S. gallon of milk weighs about 3.9 kilograms. On payload mass alone, that is approximately $210,000 on the Shuttle versus $10,600 at the Falcon 9 benchmark.

Now add an illustrative Starship target of $200 per kilogram. The same mass falls below $800.

Payload-mass analogy only. The Starship case is an illustrative target, not an achieved commercial price. It excludes spacecraft construction, integration, insurance, orbital operations, communications, servicing and replacement.

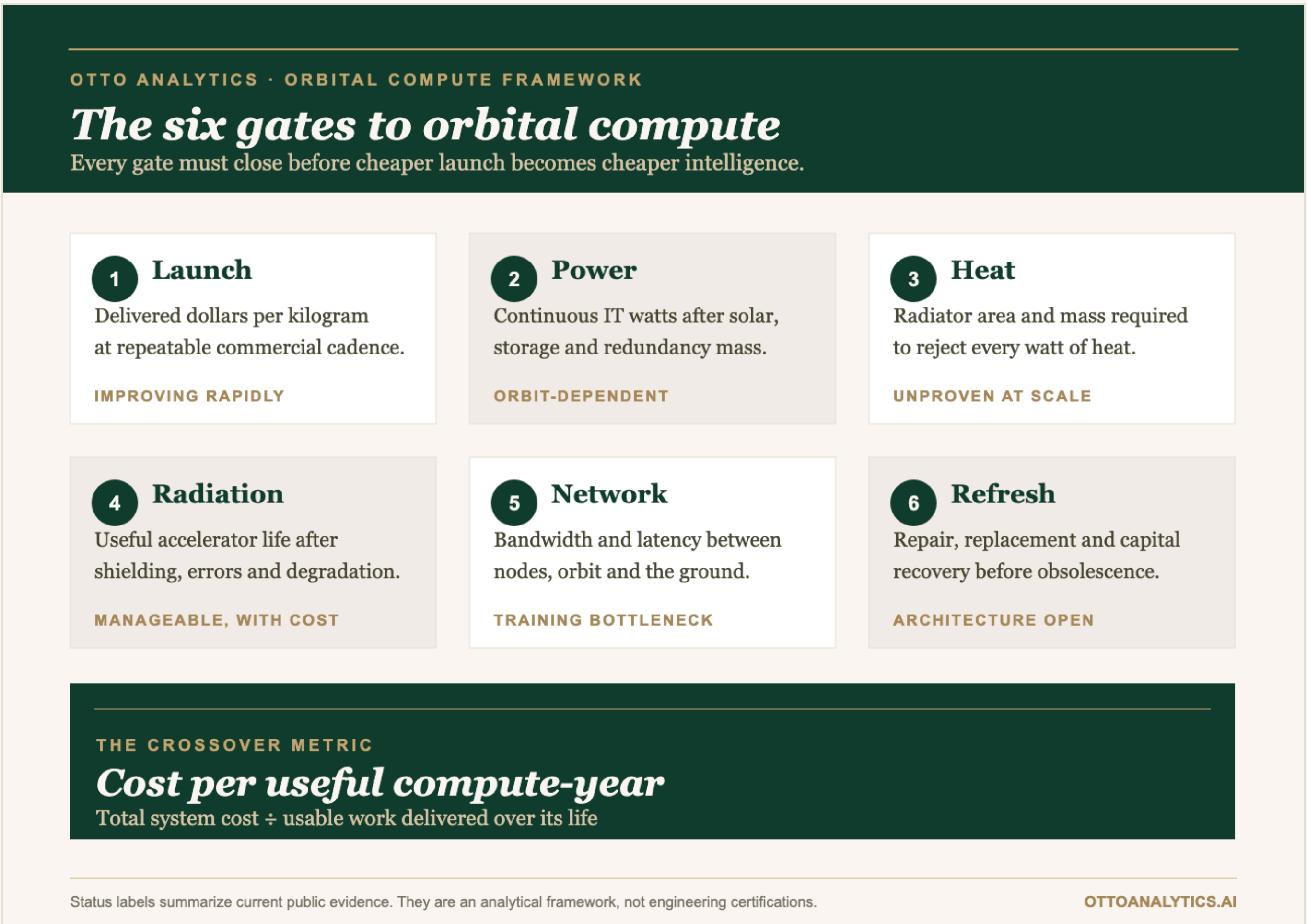
That final figure is an assumption, not an achieved commercial price. The analogy is useful because it shows the order-of-magnitude change in the cost of moving mass. It does not show the cost of operating a data center in orbit.

Cheap launch opens the door. It does not close

the business case.

The six gates

EVERY TEST HAS TO CLOSE AT THE SAME TIME



Status labels summarize current public evidence. They are an analytical framework, not engineering certifications.

GATE 1

Launch

> ### Can enough mass reach the right orbit cheaply and repeatedly?
> The relevant figure is sustained dollars per delivered kilogram at commercial cadence, including integration and the cost of replacing failed or obsolete equipment. Until cadence and price are demonstrated, Starship economics belong in the assumption column.

GATE 2

Power

Can each launched kilogram produce enough usable electricity?

Space offers high solar irradiance and, in selected orbits, near-continuous sunlight. Solar arrays, storage, power electronics, structural support and redundancy still add mass. The useful comparison is continuous IT power per launched kilogram after those systems are included.

GATE 3

Heat

Can the system reject waste heat without becoming mostly radiator?

Space is cold, but a vacuum does not carry heat away through air or water. Waste heat must be transported to large surfaces and radiated. The U.S. Government Accountability Office calls cooling at large data-center scale unproven.

GATE 4

Radiation

Can commercial accelerators survive long enough?

Radiation can corrupt data and degrade hardware. Shielding adds weight. Error correction and redundancy consume power and capacity. Useful hardware life therefore becomes part of the unit economics.

GATE 5

Network

Can the workload tolerate the communications topology?

Modern AI training clusters behave like one giant computer. Recreating that dense, high-bandwidth fabric across moving satellites is harder than connecting independent inference nodes. Inference and space-native preprocessing may become credible before frontier-scale training.

GATE 6

Refresh

Can the hardware be serviced before it becomes obsolete?

An orbital operator must either design a disposable system, build standardized replaceable modules or develop reliable in-space servicing. Each choice changes capital recovery, insurance, debris risk and residual value.

Cost per useful compute-year

THE BETTER DENOMINATOR

Cost per useful compute-year includes the capital required to build and launch the system, the percentage of time it delivers usable work, the performance lost to radiation protection and networking, and the years it operates before failure or obsolescence.

A 2026 orbital-data-center study modeled 34 to 59 kilograms of total deployed mass per kilowatt for a representative system. At roughly 40 kilograms per kilowatt, its terrestrial benchmark allowed only $250 to $1,000 per kilogram for launch and spacecraft construction combined, before communications, operations, utilization and lifetime penalties. That allowance remained below the public Falcon 9 launch benchmark before the spacecraft itself was built.

This is why a lower rocket price matters so much. It is also why rocket price alone is not enough.

Two moving cost curves

ORBIT IS COMPETING WITH THE TERRESTRIAL SYSTEM OF THE FUTURE

Orbital compute is not competing with the data center that exists today. It is competing with the terrestrial system that will exist when orbital infrastructure is ready. That distinction raises the hurdle.

Terrestrial operators are improving accelerator performance, liquid cooling, facility utilization, power procurement and the software that schedules workloads. They can replace a failed server, upgrade a network or redirect capacity without launching anything. Even when power is constrained, operators can build near new generation, add onsite supply or move workloads between regions.

The orbital curve is improving too. Lower launch cost expands the amount of solar generation, radiator area, shielding and redundancy that can be deployed economically. Better optical links can reduce the network penalty. Standardized spacecraft and servicing could lengthen useful life and lower replacement cost.

The crossover therefore cannot be underwritten from one rocket-price assumption. It depends on which curve improves faster for a particular workload.

This is also why mass per kilowatt may matter more than mass per server. A system that launches cheap accelerators but requires enormous power, thermal and shielding support can lose the advantage before performing useful work. Conversely, a workload that avoids a large terrestrial power bill or eliminates the need to transmit huge raw datasets to Earth can justify a higher launch cost.

Orbital economics will not arrive as one universal price point. They will arrive as a series of narrow crossovers.

The first workload is already telling us what comes next

PROCESS THE DATA WHERE IT IS CREATED

The earliest commercial use is unlikely to be moving a frontier training cluster off Earth. It is more likely to be processing data that is already in space.

In May, NASA reported that its open-source Prithvi geospatial foundation model had been demonstrated aboard the Kanyini satellite and a payload on the International Space Station. The model performed flood and cloud detection in orbit. Processing Earth-observation data before transmission can reduce the amount of raw information sent to the ground and shorten the time to an answer.

That is the natural entry point. The data starts in orbit, the output is smaller than the input and latency to terrestrial users is not the only objective.

A LIKELY WORKLOAD LADDER

What moves first?

1. Space-native preprocessing

Spacecraft autonomy and Earth-observation analysis where the raw data already originates in orbit.

2. Specialized inference

Communications-integrated, sovereign and loosely coupled workloads that can tolerate intermittent links.

3. General cloud workloads

Batch jobs with low communication intensity and a clear energy or location advantage.

4. Frontier training

Likely last, unless orbital networks approach the economics and density of terrestrial cluster fabrics.

The crossover will happen in slices

DATA GRAVITY DETERMINES WHERE THE FIRST MARKETS FORM

The first orbital compute market may look less like a cloud region and more like edge computing in space.

Data has gravity. If a satellite collects a high-resolution image, processing it in orbit can shrink the amount of information that must be sent to Earth. The customer pays to launch the compute, but may save communications capacity and receive a faster answer. If the raw data begins on Earth, the economics reverse: the operator must pay to send information up before the model can do anything useful.

That simple difference separates plausible early workloads from more speculative ones.

The same pattern should apply within AI. Independent inference jobs can be distributed across many nodes. Frontier training requires thousands of accelerators to communicate continuously as one system. A satellite constellation may have abundant aggregate compute while still lacking the tightly coupled network needed for training.

A successful niche can still become a platform. Space-native inference creates demand for standardized compute payloads, optical links, autonomous operations and replacement services. That volume can improve manufacturing cadence and produce real data on radiation, utilization and hardware life. Each solved support layer reduces the cost of serving the next workload.

The long-term opportunity is therefore not based on moving every terrestrial server rack into orbit at once. It is based on orbital edge workloads funding the infrastructure that gradually expands the addressable market.

The investment map

A SECOND INFRASTRUCTURE LAYER, NOT AN IMMEDIATE SUBSTITUTE

Near term, the hyperscaler results point back to Earth. New terrestrial capacity is being monetized, customer demand remains above supply and the cloud companies still have reasons to build.

Longer term, orbital compute could create a second infrastructure layer rather than a substitute. Launch providers would be only one part of it. The stack would also require high-efficiency power electronics, deployable solar arrays, thermal materials, radiation-tolerant compute, optical links, autonomous operations, modular servicing and manufacturing at satellite cadence.

Vertical integration matters. A company that combines low-cost launch, mass-produced spacecraft, optical communications and compute payloads can remove margins and coordination risk between layers. That is the strategic optionality in SpaceX's position. It is not proof that the economics have closed.

What we would watch

EVIDENCE BEFORE EXTRAPOLATION

- Sustained launch price and cadence, not a one-time payload record.
- A high-density AI payload operating continuously for at least one full year.
- Published power and radiator mass per delivered kilowatt.
- Demonstrated radiation resilience on commercially relevant accelerators.
- High-utilization optical networking under real orbital conditions.
- A credible refresh or servicing architecture.
- A paying customer for a workload that did not originate in space.

The thesis weakens if mass per kilowatt remains too high, useful hardware life is too short, inter-satellite bandwidth cannot support valuable workloads, or terrestrial power and cooling improve faster than orbital systems mature.

Directionally right is not economically complete

THE FIRST PROFITABLE WORKLOAD IS THE MILESTONE

Musk may be directionally right that an increasing share of compute eventually moves off Earth. The mistake would be treating cheaper launch as the entire proof.

This week's cloud results show why the near-term buildout remains terrestrial: useful capacity can be plugged in, sold quickly and serviced continuously. Space must offer enough energy and deployment advantage to compensate for everything it makes harder.

The crossover arrives when the full orbital system delivers a lower cost per useful compute-year for a specific workload. That will probably happen first where the data already lives above us.

The last data center on Earth is not the relevant milestone. The first workload that is economically better in orbit is.

THAT IS WHAT WE READ THIS WEEK.

Otto Analytics

DOWNLOAD THIS ISSUE (PDF)

THE WEEKLY READ RETURNS NEXT WEEK.

SOURCES AND ASSUMPTIONS

Amazon Q2 2026 results, July 30, 2026.

Microsoft FY2026 Q4 earnings call, July 2026.

NASA, The Recent Large Reduction in Space Launch Cost.

U.S. GAO, Data Centers in Space, April 28, 2026.

NASA, Prithvi Becomes First AI Geospatial Foundation Model In Orbit, May 7, 2026.

Turyshev, Orbital Data Centers: Spacecraft Constraints and Economic Viability, April 29, 2026.

Van Berkel, The Cost and Network Limits of Space-Based AI Compute, July 15, 2026.

Visual assumption: one U.S. gallon of milk is treated as 3.9 kilograms. The Starship case uses $200 per kilogram as an illustrative future target cited in current technical literature. It is not an achieved SpaceX commercial price. All three launch figures describe payload-mass economics, not delivered orbital-data-center cost.