

# Technical White Paper

## DATA CENTERS IN SPACE

### It's Happening Now



### As of Today

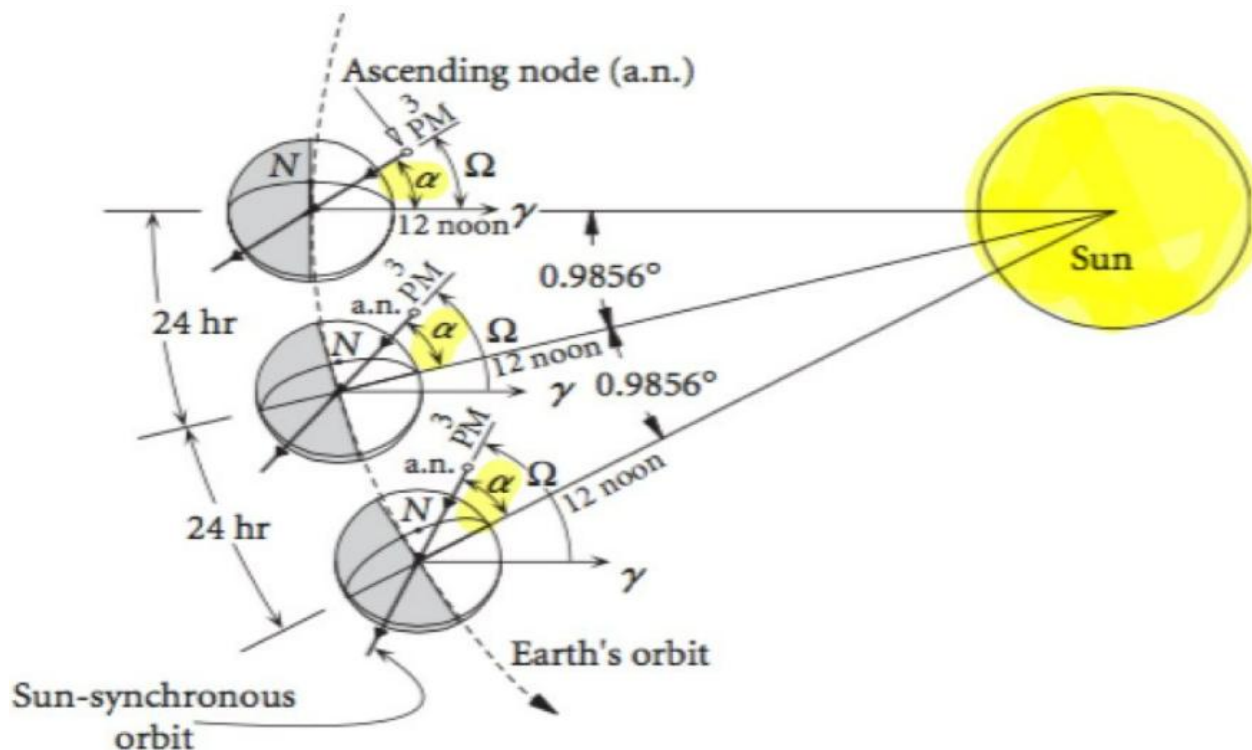
Small-scale prototypes of Space-Based (orbital) Data Centers are **already deployed** and operating in low Earth orbit (LEO) as of 2026, but large-scale, commercially competitive versions comparable to terrestrial Data Centers are not feasible now and are realistically projected for the 2030s at the earliest.

The concept involves placing computing hardware (servers, GPUs/TPUs for AI workloads) in satellites or constellations, primarily in Sun-Synchronous LEO (Sun-synchronous LEO (often abbreviated SSO) is a special type of Low Earth Orbit (LEO) designed so a satellite passes over any spot on Earth at roughly the same local solar time every day. This keeps lighting conditions (sun angle, shadows) extremely consistent, which is perfect for Earth-observation satellites, and increasingly useful for power-hungry applications like orbital Data Centers

Key advantages include near-constant solar power (no night cycle, ~36% higher irradiance than on Earth), radiative cooling in vacuum (no need for water, direct-to-chip, dielectric or massive air systems), reduced strain on terrestrial grids/land, and potential for only solar operation at scale.

## A Quick Breakdown of SSO

- ✓ LEO part: Altitude is low—typically 600–800 km (370–500 miles) above Earth’s surface. Orbital period is about 96–100 minutes, so the satellite circles Earth ~14–15 times per day.
- ✓ Sun-Synchronous part: The orbit is nearly polar (inclination around 97–99°, slightly retrograde). Because Earth isn’t a perfect sphere (it’s oblate), the orbital plane slowly precesses (rotates) eastward by about 0.986° per day. This exactly matches Earth’s yearly orbit around the Sun, so the satellite’s path stays “locked” relative to the Sun.



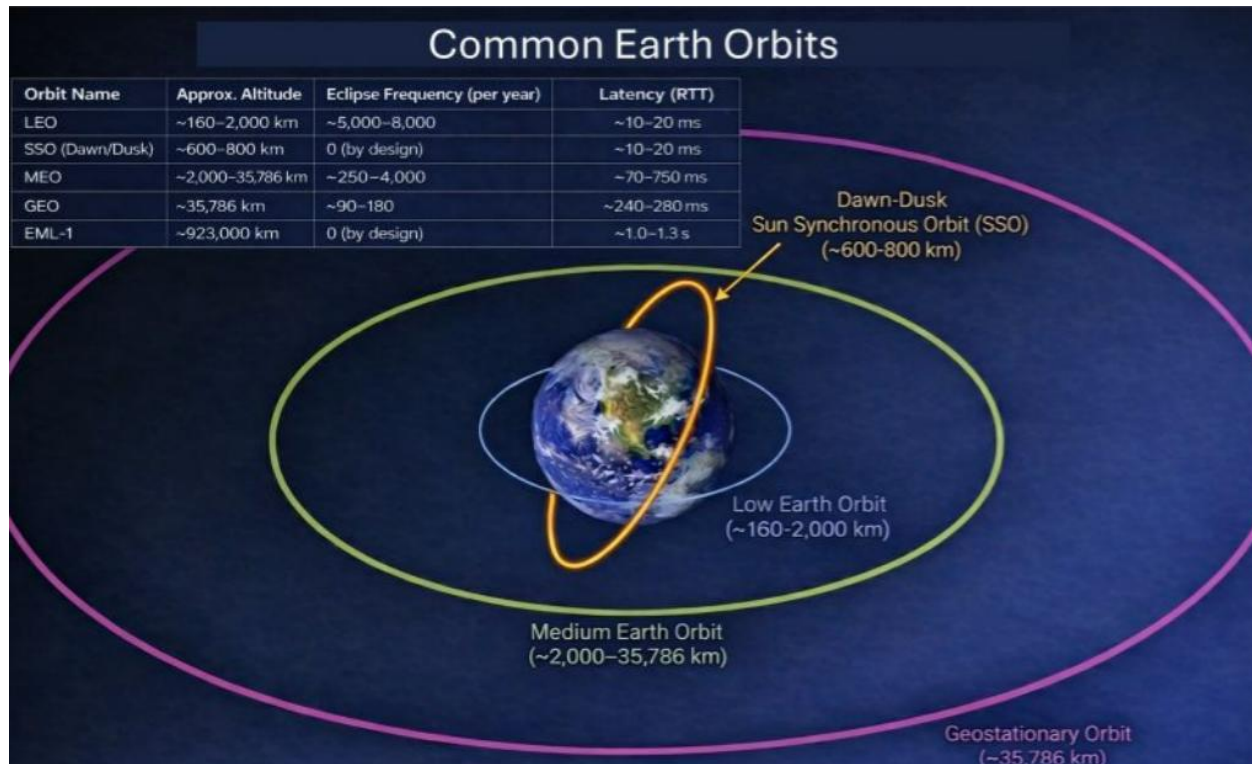
Classic diagram showing how the orbital plane precesses (precession) to keep the same local solar time.

## Why “Dawn-Dusk” SSO Matters (Especially for Orbital Data Centers)

Most SSOs are tuned for a specific local time (e.g., 10:30 a.m./p.m. for Landsat satellites). Most modern power-focused missions prefer the dawn-dusk variant:

- The satellite rides the terminator (the line between day and night).
- Result: almost continuous sunlight with zero or very few eclipses per year.

- Solar panels get near-24/7 power; no massive batteries needed; less thermal stress on hardware.



## Current Status (Prototypes and Early Tests)

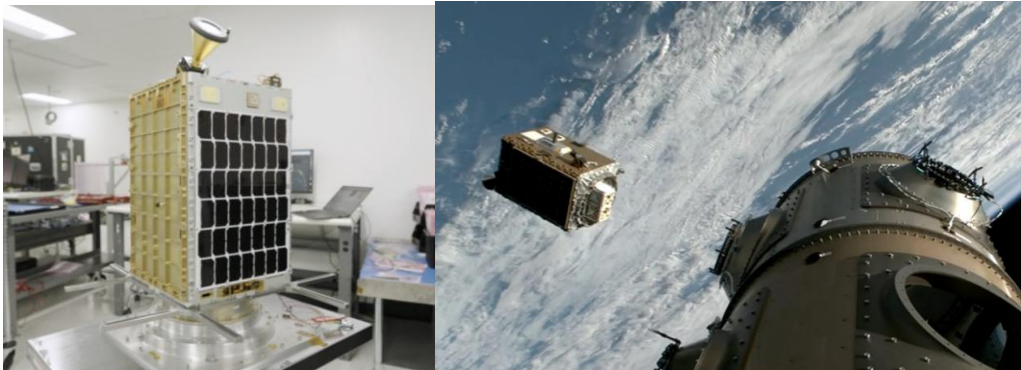
Deployed now: Starcloud launched Starcloud-1 in November 2025 a small satellite with an NVIDIA H100-class GPU. It has successfully trained a Large Language Model (LLM) and run versions of Google’s Gemini AI in space. Additional tests (e.g. Starcloud-2) are underway or planned for higher power and applications like **Bitcoin mining**.

### STARCLOUD Overview

Abilities / Capabilities (with NVIDIA H100 GPUs)

- Starcloud-1 (launched November 2025 on SpaceX): Proof-of-Concept satellite (~60 kg / 130 lbs, fridge-sized, based on Astro Digital’s Corvus-Micro bus). It carries the first terrestrial Data-Center-class NVIDIA H100 GPU **ever flown in space** (80 GB HBM3 memory). This delivers roughly 100x more GPU compute power than any prior space-based system.
- Key achievements (December 2025 onward):
  - First-ever LLM trained entirely in space (Andrej Karpathy’s nanoGPT on Shakespeare dataset).

- Ran inference on Google's Gemma model (and a version of Gemini).
  - Demonstrated real-time on-orbit AI processing for satellite data (e.g., Earth observation).
- Power & cooling: Solar-powered (continuous in dawn-dusk sun-synchronous orbit) + radiative heat dissipation into space. No water or grid needed.
- Future roadmap:
- Starcloud-2 (target ~2026–2027, ~450 kg, 8 kW): GPU cluster with multiple H100s + Blackwell GPUs, persistent storage, commercial services (e.g., partnering with Crusoe for cloud workloads).
  - Long-term: Gigawatt-scale constellations (e.g., 5 GW with ~4 km solar/radiator arrays) for massive AI training clusters. Claims ~10x lower energy costs and 10x CO<sub>2</sub> savings vs. Earth Data Centers.



### Height (Orbital Altitude)

- Starcloud-1: Operates in Low Earth Orbit (LEO) at 325 km altitude. Expected lifetime ~11 months before controlled deorbit.
- Future satellites/constellations: Planned for dawn-dusk sun-synchronous orbits (SSO) at 600–850 km (for near-continuous sunlight) or very low Earth orbit variants.

Lower LEO altitudes enable lower communication latency and easier deorbit but require more frequent station-keeping due to atmospheric drag.

### Latency

- ✓ Ground-to-satellite communication latency: Very low for LEO (~20–50 ms round-trip time / RTT in practice, similar to Starlink). They use high-bandwidth laser terminals for direct links to satellite constellations (hundreds of Gbps, as low as ~50 ms claimed). This is far better than geostationary orbits.

- ✓ Advantages for workloads:
  - In-orbit processing shines here: Raw satellite data (e.g., 10 GB/s from SAR imaging) is analyzed on-board → low-latency insights delivered to Earth (minutes vs. hours for full downlink). Great for Earth observation, real-time AI (crop monitoring, wildfires, etc.).
  - AI training: Highly latency-tolerant (batch/offline jobs). Ideal for space-based compute.
  - Inference/serving: Suitable for many cloud workloads; not ultra-low-latency edge (e.g., <10 ms gaming), but competitive for most AI use cases via Starlink-like networking.
- ✓ Intra-cluster: Very low latency within a satellite or constellation (direct laser/mesh links).

Overall, Starcloud-1 has already proven the core concept works. Larger systems are in development for commercial orbital AI cloud services

### **Axiom Overview**

Deployed Now: Axiom Space's Orbital Data Center (ODC) is a pioneering initiative to build the world's first commercial, scalable, cloud-enabled data processing and storage infrastructure directly in space.

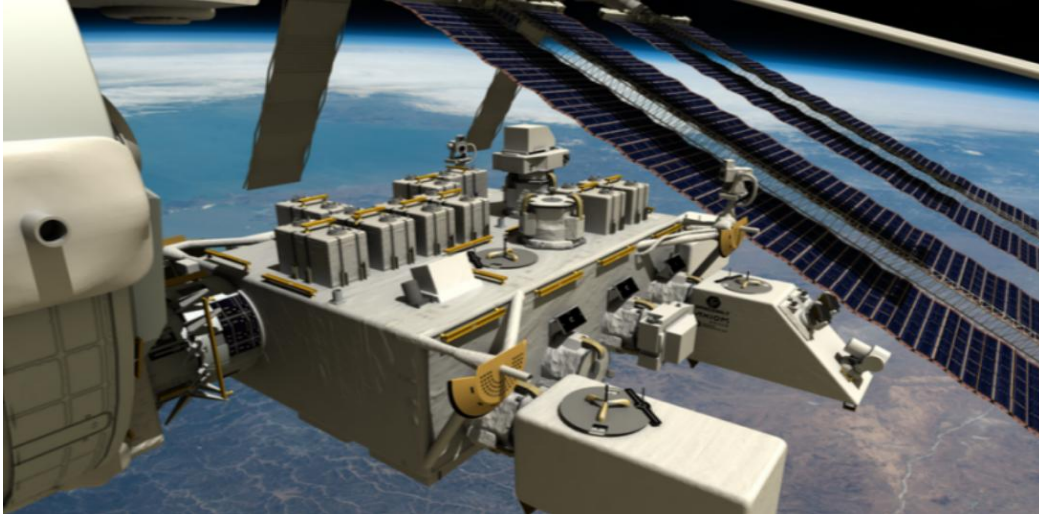
As of May 2026:

- Prototype: AxDCU-1 (Axiom Data Center Unit-1), Launched August 2025 aboard SpaceX CRS-33 and deployed on the International Space Station (ISS) in fall 2025. This shoebox-sized unit, powered by Red Hat Device Edge, tests terrestrial-grade edge computing, cloud workloads, AI/ML, and space cybersecurity in orbit.



- First dedicated free-flyer nodes: ODC Nodes 1 and 2 — Successfully launched January 11, 2026, aboard SpaceX's Twilight mission. These are integrated into

the first tranche of Kepler Communications' optical relay satellite constellation and are now operational in LEO, marking the shift from prototypes to initial production capability for in-space cloud computing.



Conceptual view of an expanded orbital data center module attached to the ISS (Axiom Space and partners are planning larger nodes like this for 2027+)

These nodes enable real-time processing of satellite data without constant downlinking to Earth

### Power and Cooling

- **Power:** Relies on abundant, near-continuous solar energy in orbit (no terrestrial grid limitations). Current nodes operate at kilowatt-scale, with modular solar arrays providing scalable power. This is a core advantage over Earth data centers, which face energy constraints for AI workloads.
- **Cooling:** Uses passive radiative heat rejection into the vacuum of space (no atmosphere means no convection, but large surface-area radiators efficiently dissipate heat). This provides a natural, low-energy cooling solution compared to power-hungry Earth-based systems (e.g., no water or massive HVAC required). Early prototypes and nodes are designed around thermal management suited to the orbital environment; future scaling will leverage even greater external real estate and heat rejection capacity on platforms like Axiom Station.

These features enable "**infinite scalability**" with minimal environmental impact on Earth.

## Future Roadmap

Axiom Space aims to grow the ODC network into a full orbital cloud ecosystem:

- Short-term (by end of 2027): At least three operational nodes, including an optically interconnected AxODC Node on the ISS (in collaboration with Spacebit for higher-capacity, interconnected computing).
- Medium-term: Additional free-flyer nodes in LEO, maturing technologies for large-scale secure processing. Integration with Axiom Station (their commercial space station, which will host the first tranche of ODC capability — ODC T1).
- 2030 and beyond: Scale from current kilowatts to megawatts of processing power. Create a dense mesh network of ODC nodes connected via radio and next-gen optical relays, supporting space and terrestrial users with autonomous, resilient, Earth-independent cloud services. This includes dynamic software updates, AI model deployment, and sovereign data management.

The vision positions LEO as a global space marketplace, with ODCs reducing reliance on ground infrastructure while supporting defense, Earth observation, telecom, and AI applications.

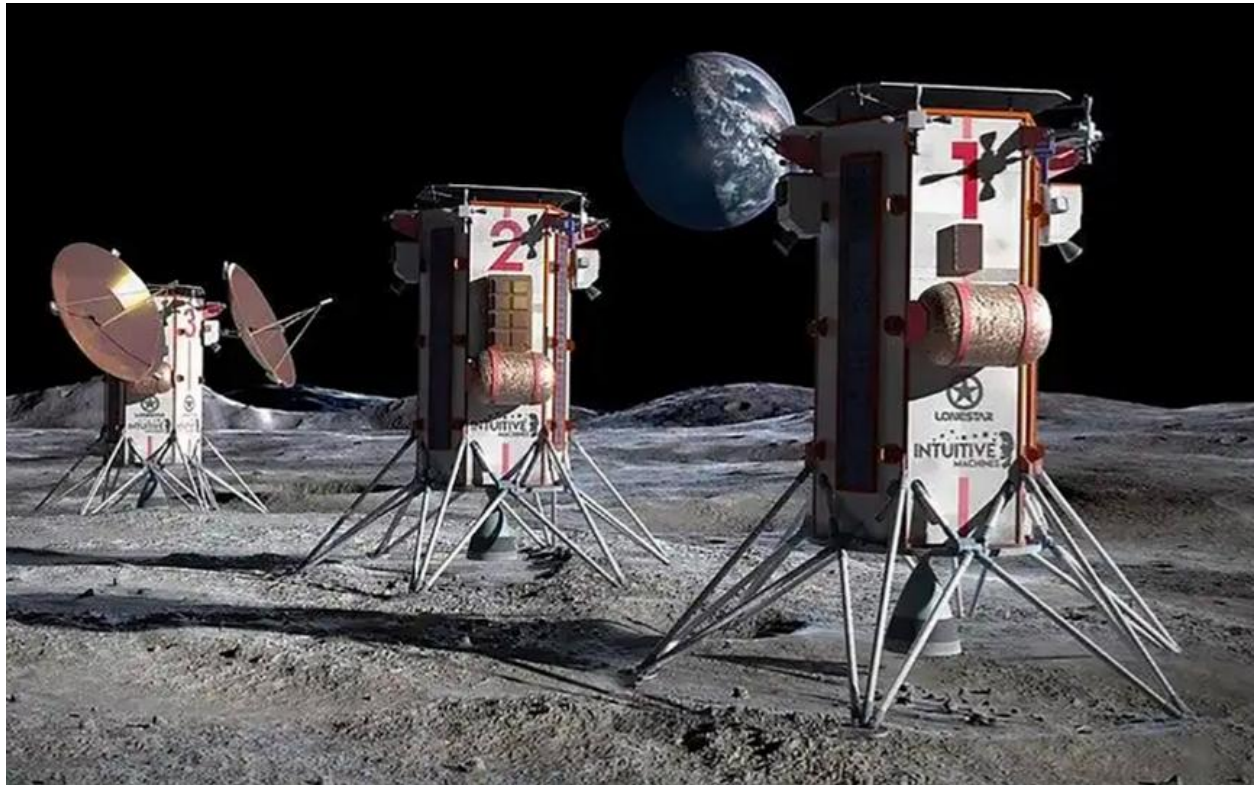
## YES, There's More

Other players like Lonestar Data Holdings have tested small data storage payloads on lunar missions (e.g., SSD-based archival on the Moon's surface), but these are more about secure/off-Earth backup than active high-power computing.



Independence payload mounted on the Odysseus lander on the lunar surface during the successful 2024 IM-1 mission.

These are not full Data Centers, they are proof-of-concept satellites demonstrating that commodity or near-commodity hardware (with software mitigations for radiation) can survive launch, operate, and perform AI tasks in the space environment. Multiple companies (Starcloud, Axiom, Kepler Communications, etc.) have hardware in orbit running production-like workloads at small scale.



Concept of multiple expanded Lonestar Data Center modules on the lunar surface (future roadmap vision), showing a growing lunar data infrastructure network

Google (Project Suncatcher): Moonshot for TPU-equipped satellite clusters with optical inter-satellite links. Two prototype satellites planned for launch in early 2027 to test compute, comms, and radiation tolerance.

SpaceX/xAI: Filed FCC plans in early 2026 for up to 1 million satellites as an orbital data center constellation. Elon Musk has claimed the lowest-cost AI compute could shift to space within 2–3 years.

## Here is the Real Timeline

- Now–2027 (Demonstration Phase): Small prototypes and validation missions are here and expanding. Focus is on thermal management, radiation effects, inter-satellite optical links (tens of Tbps demonstrated in labs), and clustered formations. Early niche uses (e.g., in-orbit processing of Earth-observation data or edge AI) are possible.
- 2028–2030 (Early Commercialization): Micro Data Centers or small constellations for high-value, low-latency niche applications (e.g., secure computing, space-based AI training for specific workloads). Market reports project the in-orbit Data Center sector growing from hundreds of millions today toward billions.
- Mid-2030s+ (Scalable/Competitive): Gigawatt-scale AI infrastructure becomes economically viable if reusable launch costs drop to ~\$200/kg (enabled by Starship scaling to high flight rates). Google’s analysis and others project cost parity with terrestrial Data Centers’ energy costs around 2035. The EU has assessed them as technically and environmentally competitive in the long term, beyond 2035.

## Could Full Technologies Be Deployed Now?

Limited yes for small/experimental scale; no for practical, large-scale deployment.

- **What works now:** Radiation-tolerant (or software-resilient) computing hardware, solar arrays, radiative cooling via large panels/radiators, free-space optical links, and LEO satellite ops are mature enough for prototypes. Tests confirm feasibility at kilowatt levels.
- **What blocks large-scale now:**
  - ✓ Economics: Launch costs are still too high for massive hardware deployment (current ~\$2,500+/kg vs. needed <\$200–300/kg). A 1 GW system could require ~1 square mile of solar arrays.
  - ✓ Engineering: High-density heat dissipation relies on inefficient radiative cooling (requires enormous radiators). Radiation causes soft errors (mitigated but degrades chips over time). No easy maintenance/repairs. Orbital debris, solar storms, and constellation management add complexity.

- ✓ Other: Data latency/bandwidth to Earth users (though hybrid architectures with Starlink-like networks help at 20–40 ms), regulatory hurdles (FCC filings ongoing), and environmental concerns (launches, space junk).

## Conclusion

In short, the real timeline is incremental with demos happening today, early niche commercial use in the late 2018-29, and transformative scale in the 2030s if launch costs and supporting tech (e.g., in-space assembly, better thermal designs) continue progressing. It's driven by the AI energy crunch on Earth, but physics and economics mean it's not a near-term replacement for ground-based Data Centers. Hype exists, but credible studies and company roadmaps point to the mid-2030s for real impact.

This space is moving fast and with upcoming SpaceX IPO reportedly that could raise Two (2) Trillion. Who Knows What the Pace Will Be.