

Pure Storage (PSTG)

Category: Storage

Est. Price Per Unit: ~\$150k+ per FlashBlade

What They Do

Product 101 and Where They Fit into the AI Stack



- Pure is often thought of as a commodity hardware vendor. The thesis is that it's actually a software company wrapping proprietary hardware to solve two constraints in the AI era: power and predictability.
- The key differentiator is **DirectFlash**. Competitors (NetApp, Dell) buy pre-packaged SSDs from Samsung/Hynix with their own internal controllers doing garbage collection and wear leveling independently. Pure buys raw NAND chips and manages them via software-defined controllers at the system level. This is somewhat analogous to Groq's method for inference chips with deterministic software defined networking.
- This architecture has three implications:
 - Power/density: removing individual controllers reduces power/TB by ~50%. They ship 75TB modules with a roadmap to 300TB.
 - QLC economics: Pure can use lower-cost, lower-endurance **QLC NAND** but achieve TLC-level reliability because their software manages writes globally, preventing uneven wear.
 - Deterministic latency: no random latency spikes from SSDs doing their own maintenance. Critical for keeping GPUs fed without tail latency stalls.

Alignment with Overall Thesis

- As LLM context windows grow (1M+ tokens), GPU HBM is too expensive to store the full KV cache. Pure allows offload to local flash with low enough latency to prevent stalls—effectively RAM expansion for inference.
- In training clusters, if one storage node hangs, the entire GPU cluster waits. Deterministic latency maximizes GPU utilization.
- For RAG architectures in finance/gov, indexing proprietary data on-prem avoids cloud egress fees and regulatory risk.

Business Model, Customers

- ~\$150k+ per blade. Bought when data buckets fill up.

- **Evergreen subscription** model: unlike competitors requiring forklift upgrades every 3-5 years, Pure upgrades controllers non-disruptively with replacements
- Pure isn't trying to beat AWS/Azure at general cloud storage. They're the premium tier for difficult workloads that customers don't want on the cloud for security, speed, or egress fee reasons. Pure's common customers are government departments, financial institutions (many trading firms with latency requirements), and customers with sovereign data requirements (increasingly common)
- Meta's AI Research SuperCluster uses Pure. Anecdotally, Meta may significantly expand contracts with Pure. Google Distributed Cloud (reportedly Google's fastest-growing division) partners with Pure and also confirms the tailwinds of hybrid and on-prem 'cloud' solutions (GDC is outpacing all other GCP divisions). GDC, for example, will partner with Pure as the foot in the door hybrid cloud vendor, and then GCP can offer the remaining on-prem solutions.
- RPO growth (24%) is outpacing revenue growth (17%)—signals a backlog of large commitments (even before the Meta upgrade)
- Acquired **Portworx**, the leading Kubernetes data layer. Gives Pure a software hook into applications, not just hardware.
- Competitive context: NetApp is software-on-commodity (bound by Samsung/Kioxia physics). Hyperscaler native storage is "good enough" but IOPS/\$ is poor. DIY NAND is theoretically possible but building an enterprise-grade flash translation layer took Pure 10+ years.

Comments on Team

- Founded in 2009.
- Founder is now Chief Visionary Officer and Director.
- Other co-founder is an advisor.

Early View of Moat Hypothesis

- 10+ years of flash translation layer IP. Building an FTL that handles QLC reliability at enterprise level is non-trivial.
 - Evergreen subscription creates switching costs.
-

Why They're Interesting, and Why Now

- The thesis shifts from "storage for databases" to "storage for AI context." We have to believe AI inference moves on-prem and hybrid—if that plays out, Pure becomes a hyperscaler partner rather than competitor.
- 1M+ token LLMs need KV cache offload. Pure's low-latency flash could serve as HBM expansion, though this use case is still emerging.
- Data center power is the new bottleneck. Pure's 50% power/TB advantage is tangible and measurable.

Enterprise AI Storage: Qualitative Signals

Enterprises are starting to think about storage more strategically for AI workloads. This is a shift from storage as commodity infrastructure to storage as a performance-critical component of the AI stack.

Moderator: Why are you Significantly Increasing (>16%) your Storage systems (NAS, SAN, SSD, Flash) spend over the next 12 months?

Expert: Our approach to storage has completely flipped over the last 18 months. We used to just buy more space, but now it's all about speed and intelligence. Our strategy has moved towards decentralization and feeding our new AI ML workloads in real time, because we are running hybrid clouds and supporting a remote workspace. **Fast resilient storage isn't a luxury, it's an engine for everything.** The real goal here isn't just about capacity, it's performance. We need to pull data faster, back it up quicker, and cut down on lag, especially in sensitive areas like banking and healthcare, where every millisecond counts. We're dealing with huge data spikes from things like video KYC tricks, AI model training, and massive telemetry logs. Our old gear just couldn't keep up without slowing down our new innovations. The push is being led jointly by our data engineering, AI ops, and compliance teams. So who are we buying from and why? **We're putting more money into Pure Storage, Dell EMC, and NetApp.** Pure storage is our top choice for anything that needs lightning speed, like our AI projects. Dell EMC PowerStore is great for handling unpredictable bursts of activity, like when we are processing a flood of insurance line. NetApp is essentially our hybrid setup. Its integration with Azure and AWS lets us easily move our burst data where it makes the most sense, either for cost or compliance. We chose them based on raw speed and reliability, but the increased spend now reflects better deduplication, better disaster recovery for both RPO and RTO, and easier management across all of our sites. We're also kicking the tires on vast data, and because they look very promising and managing our unstructured data, like for AI training. The numbers driving the increase, the growth is fueled by real activity. Let's talk about capacity. **We are adding 40% to 50% more storage year over year, just to keep up with logs, recordings, and saved AI models.** Talking about speed, our analytics pipelines need more than double the throughput because we've shifted from slow daily batch updates to real-time data. Let's talk about people. We're adding quite a few seats just in Q1, forcing us to plan for much bigger capacity overall. The workloads causing the biggest bumps are AI model training, running our DR as a service, and high-frequency transaction processing. Who are we kicking to the curb? **We've already retired our HPE and IBM storage here, the reason being performance and their interfaces were a lot clunkier, and they didn't play nicely with our cloud strategy.** We've also scaled back on Western Digital arrays because their data duplication wasn't as good as they advertised. The vendors we're keeping are the ones that offer modern features right out of the box.

Moderator: Why are you Significantly Increasing (>16%) your Storage systems (NAS, SAN, SSD, Flash) spend over the next 12 months?

Expert: Our strategy with SAN systems, well, storage systems actually, is that we are facing unprecedented data growth with connection with all the AI projects our customers and users have. The vendors that have also a proper strategy for supporting those AI projects are Pure Storage, HPE, and Tera. Functional requirements are basically allowing the data engineers to concentrate diverse data sets and prepare them to be used by the AI model either for creating the RAG database or by fine-tuning the LLM that will use the data, and the increases in spending with these vendors is basically driven by the growth of data, nothing else. I've already commented on the specific product features, and there might be vendors that are behind in AI, in my opinion, it's definitely Dell and Hitachi Vantara.

Moderator: You mentioned Pure Storage, HPE, and Tera as vendors supporting your AI projects. Could you elaborate on what specific features of Pure Storage led you to choose them over Dell and Hitachi Vantara, considering you find Dell behind in AI?

Expert: The three vendors I've mentioned were Pure Storage, HPE, and NetApp. They support our AI projects, and the specific features is there is of use that basically enables users to concentrate and refine the datasets they will be using, and also their low operating cost, because all those systems are designed to basically use less operator power than the others.

Key Risks

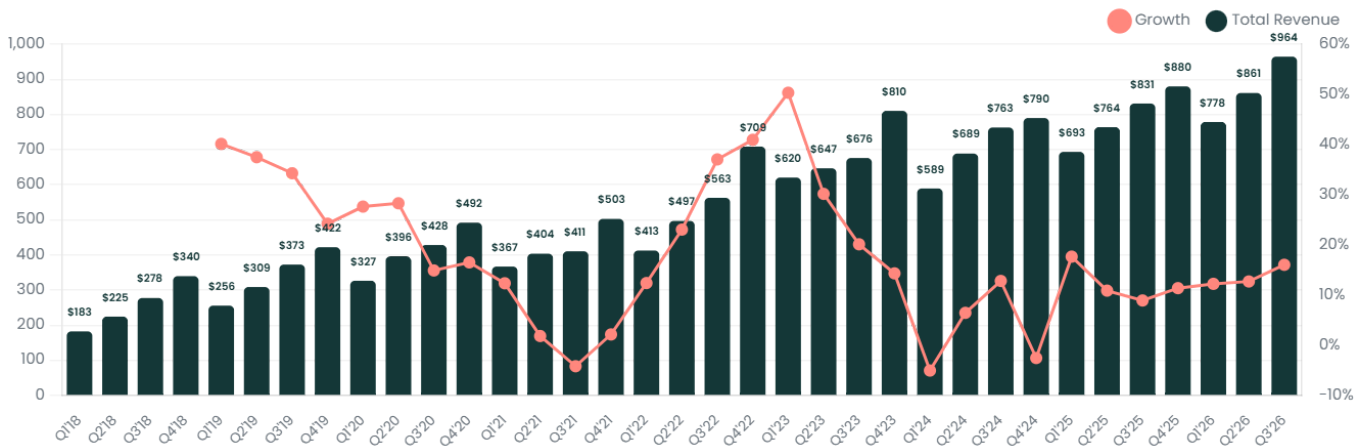
- If standard SSD controllers improve enough to handle QLC reliably, Pure's margin advantage erodes.
- If AI inference moves entirely to memory-heavy chips (Groq, Cerebras) that don't need storage offload, the AI tailwind weakens.
- Customer concentration—heavy reliance on Meta and potentially 1-2 others creates binary risk.

Gaps in Understanding / Key Questions

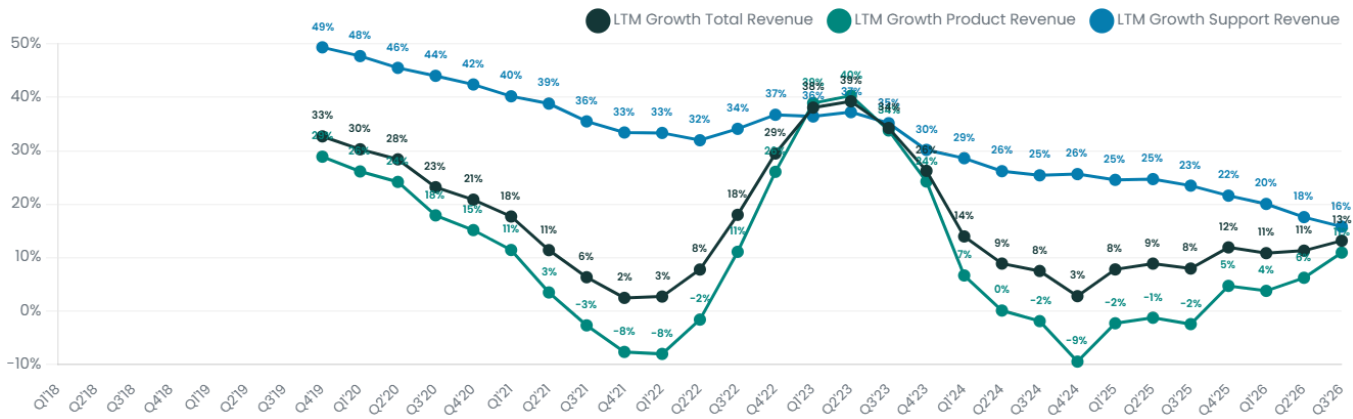
- How defensible is FTL IP if Samsung/Kioxia invest heavily in controller software?
- Depth of Meta relationship—is Pure truly tool of record or one of several vendors?
- Portworx adoption—is Kubernetes storage actually sticky or commoditizing?

Select Financial Graphs

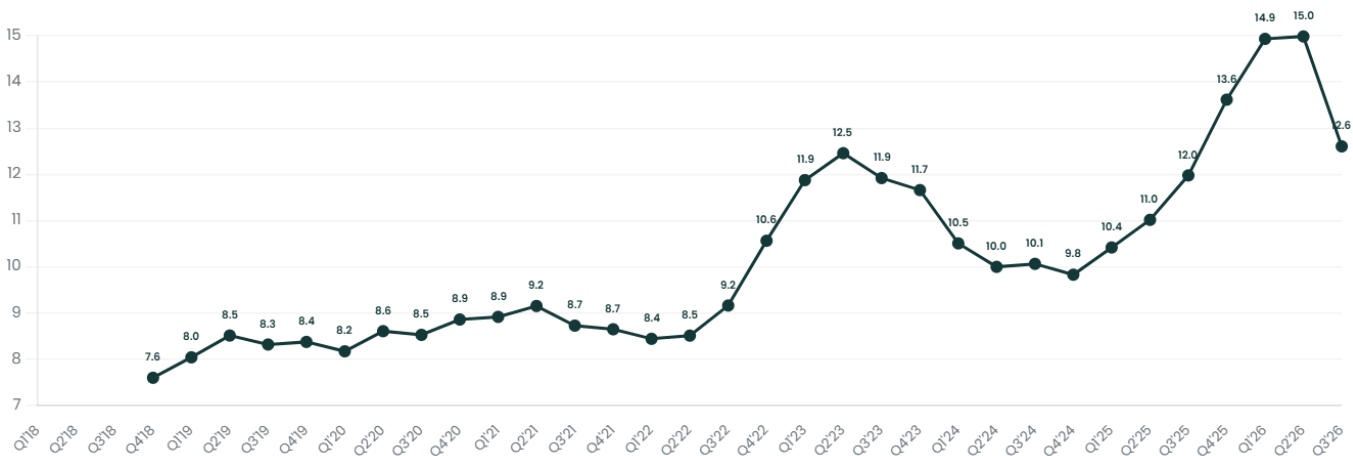
R1: Total Revenue & YoY Growth



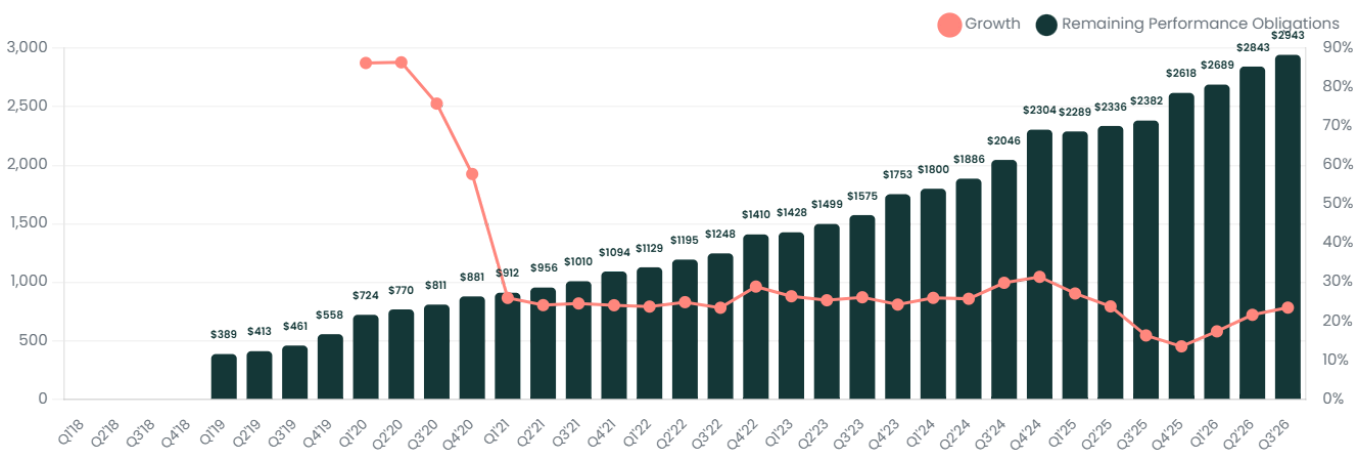
R3: LTM Revenue Growth (Total, Product, Support)



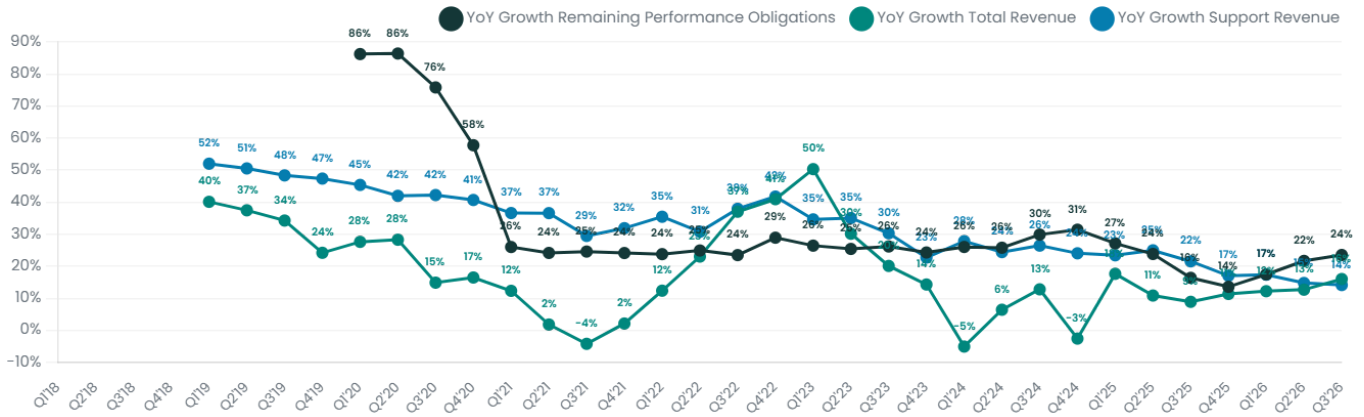
I1: Inventory Turns



SUB2: Total RPO & YoY Growth



SUB3: RPO vs Revenue vs Support Growth



V1: EV/Sales NTM



Patent Analysis

Storage Software Innovation Leader

Pure Storage shows exceptional patent citation rates—peaking at 178 citations/patent in 2020 (extremely high). This suggests their flash translation layer and storage software IP is highly influential and frequently referenced by others in the industry.

“Patent analytics/30_pure_storage.png” could not be found.

Interesting Topics to Read

- DirectFlash vs. commodity SSD architecture
- QLC vs. TLC NAND economics and reliability
- KV Cache offload for large context LLMs
- Pure Evergreen subscription model
- Portworx and Kubernetes storage
- NetApp competitive positioning