

DIGITAL TRANSFORMATION IS DATA-CENTRIC

Why and how to put data at the center of your vision.

You might have heard that “data is the new currency,” referring to its dominant role in today’s economy. What is it that gives data such importance? Like currency, the value of data comes not from its existence or accumulation, but from the opportunity it represents.

Data-centric organizations have the power to disrupt traditional ways of working and create new economies. The proof, and threat, surrounds us every day. The travel and hospitality industry was disrupted nearly overnight (pun intended) by peer-to-peer accommodations renting. Walmart is the reigning king of global retail, with 4x the revenue of Amazon, yet Amazon has twice the valuation owing to opportunity it can release from its high-quality customer data.¹ The most valuable companies are no longer giants of infrastructure and physical resources, but instead Apple and Alphabet, the biggest purveyors of data.² Lessons can be learned from industries such as automotive, in which, rather than digging in to do more of the same and preserve traditional car-sales businesses, automotive companies are reinventing themselves with revenue streams around mobility and data — which positions them for future revenue, rather than watching it float by from the comfort of a nice, disconnected car.³

Connectivity has given power to companies born in the modern era as data emerges as one of the most important assets. This connectedness and the avalanche of data will distinguish winners from losers, and the ability to make sense of information will be the deciding factor.² However, data alone is inherently powerless. The value of data is what can be released from it: intelligence generated through aggregation, correlation, analytics, and real-time combination of different datasets. That is, data becomes more valuable the more it is processed and used. It appreciates the more it’s shared, and it returns multiples of value when shared the right way, in the right place. Like currency, the potential value of data remains untapped by organizations that don’t take action to invest it.

The main challenges in tapping that potential value lie in siloed data, clouds, and organizations:

- **Siloed data.** Data is often spread across the enterprise in discrete silos to meet the needs of specific applications. However, this inhibits modern analytics-driven workflows that demand data be available and delivered quickly as a consistent whole, not fractured in data silos. Additionally, cold data, collected as the economics of storing it improved, had its potential frozen in time when it was captured and forgotten, unintegrated with the objectives of your business.
- **Siloed clouds.** Cloud has earned its place in enterprise IT, but that place has so far been separate and distinct. Separate isn't shared, isn't invested, and isn't valuable. Those data silos withhold data's value from your business, constraining innovation and degrading your users' experience.
- **Siloed organizations.** At the same time, organizational silos, inherited from traditional separation of IT builders and technology operators, prevent the agility of DevOps where the builders *are* the operators and the service is their objective.

Overcoming these silos and putting data to work in your business is key to success in the twenty-first century, and it is the next step in the evolution of your infrastructure. Indeed, you may already be on this evolutionary track by taking advantage of technologies such as hybrid cloud and containers. This progress toward digital transformation has organically come to encompass machine learning (ML) and artificial intelligence (AI) as methods by which massive, dispersed data is processed into intelligence that can be transferred to applications, users, and machines wherever they live.

This developmental trajectory leads toward the purposeful design of infrastructure that is built to take advantage of data and exploit the value that is inherent within it. This purposeful design is called a data-centric architecture. Such an architecture acknowledges and empowers the centrality of data in value creation, and it can help organizations ensure that data gets to the right place when it is needed for maximum impact. This paper provides a vision for evolving your infrastructure into a data-centric architecture.

It's a Data-Centric World

We live in a world in which intelligence, algorithms, and user behaviors have put data at the center of value creation for businesses, research, communities, and individuals. We experience the data-centric world as a simpler and faster way to do things, driven by large, fast-moving datasets that have fundamentally changed how we live our lives and conduct business. For example, the last time you used your smartphone for navigation, you accessed a large pool of data that had already been processed into relevant map information, to which your device added your location. Similarly, data enables always-on, affordable electricity, self-driving vehicles, better healthcare, the convenience of public transportation, and the manufacture and distribution of low-cost goods. Each of these examples is possible because data-centric organizations harness large datasets that are part of a widely available worldwide data pool. The imperative of data-centric organizations is to get data to the right place at the right time to create value from it. They regularly make choices informed by data, and they make investments to maximize their access to and use of data to improve results. If that sounds like you, then yours is a data-centric business.

What does it mean to be a data-centric business? A data-centric business creates value from data in its digital-transformation journey to achieve sustainable competitive advantage. This means that you create value from data, and then share that value with partners, customers, applications, and users by making information available to all. Here are some defining traits of data-centric businesses:

- They track metrics that give insight into performance, efficiency, and growing trends.
- They have core data stores, such as enterprise databases from SAP, Oracle, or Microsoft, and possibly also newer web-scale applications.
- They might also use data held in distributed places or gathered from third-party services.
- Increasingly, their business data consists of different types, including structured and unstructured: relational databases, data lakes, and imagery.

Digital Transformation Is All About Data

Your organization has likely felt the growing pains of being a data-centric business. For example, production IT is stretching in new directions. Historically, IT could accommodate business growth by adding more servers, more virtual machines (VMs), and more infrastructure to serve workloads in an on-premises data center. But now IT is moving toward encompassing external resources to acquire, process, and share data wherever those resources are. While IT organizations have embraced the public cloud, they do not move workloads wholesale. Rather, IT staff work within a hybrid environment, placing and accessing workloads and data wherever it makes the most sense in terms of technical and business needs. When organizations take stock of their data centers that have grown naturally around their needs, they often find that what they think of as their data center might even not be theirs at all — the applications, platforms, and infrastructure one traditionally might own and operate (on premises) can now be selectively subscribed to “as a service.” From infrastructure as a service (IaaS) to platform as a service (PaaS) to software as a service (SaaS), an organization can determine at what level it wishes to consume IT value as a service, layering its own resources on top as needed.

This extension of on-premises live environments to include a public-cloud footprint isn’t lift-and-shift, and it’s not experimentation. It’s a logical extension of the prevalence and centrality of data. After all, there’s no reason to have extra compute resources without data to feed the compute capacity; no reason to stand up new workflows, whether cloud-based or on-premises, without data as input and output. And it’s no longer enough to connect compute and storage in neighboring racks to a Fibre Channel or 10-gigabit network and expect that to make data available when it is needed. So, the task at hand is to unify the data platform across footprints to encompass both on-premises and cloud datasets.

This unification overcomes the silos described above and results in an organically grown unified cloud: a data-centric architecture that enables next-generation hybrid applications to run seamlessly across clouds. It also enables application mobility by making it easy to move enterprise apps to the public cloud and web-scale apps on premises, and to develop applications consistently across both.

Evolved Analytics

Another area in which data has caused businesses to stretch is that of analytics. Analytics used to be a siloed project: a distinct software package run as a batched application with its own compute and storage stack. It's now a ubiquitous, always-on activity. The advent of data lakes added the ability, at massive scale, to capture much more data, including unstructured and unexpected data, and it allowed organizations to glean insights based on questions they never thought to ask before the ubiquity of data-transformed analytics.

The traditional approach was to store business data and periodically mine it in batch processes. This approach was adequate for market intelligence or macro-trend identification. However, velocity and intelligence are required to feed constant decision-making processes, ML, and AI, which are the new normal and future of pre-processing for downstream automation.

This evolved mix of data, applications, and objectives for the analytics stack has come to be organized in a more efficient way, growing organically into a data hub. A data hub is a data-centric architecture designed not only to store, but to consolidate, share, and deliver any data — powering data protection, data warehouse, modern analytics, and AI — to any user or application when and where needed, in addition to eliminating silos.

The stretching of IT into a unified cloud and evolution of modern real-time analytics with a data hub are examples where data-centric design decisions are already being made in pockets; so why not apply the same principles to the whole business? These business realities, and the need to take full advantage of data, call for a new vision for your architecture, one that acknowledges that data is at the center of business value. What your data-centric business needs to thrive in this data-centric world is a data-centric architecture.

What Is a Data-Centric Architecture?

A data-centric architecture is an approach to designing an end-to-end environment across compute, network, storage, and cloud, optimized for ubiquitous and fast consumption of data to create value. A data-centric architecture is characterized by five key pillars:

- Fast, shared data
- On-demand and automated
- Globally reliable and secure
- Hybrid cloud by design
- Constantly on and improving

Your data-centric architecture is built around your organization, by you. As you evolve toward a data-centric architecture, plan to develop the enabling pillars shown in Figure 1 and discussed below as fundamental requirements of your architecture.

Data-Centric Architecture

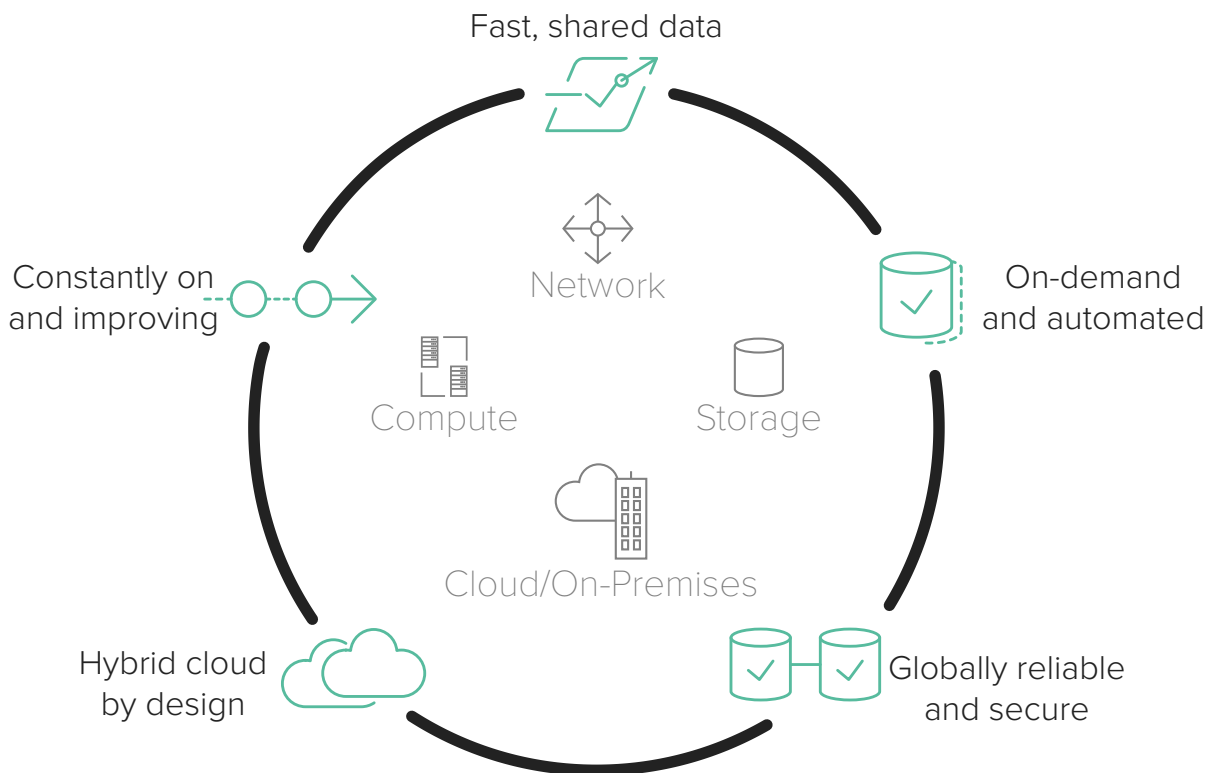


Figure 1. The five defining pillars of a data-centric architecture.

1. Fast, Shared Data

A data-centric architecture enables high-velocity processing of high-volume data, making it accessible to applications, users, and machines wherever they live, and resulting in more efficient and agile processes. You don't always need "fast," but if your environment isn't capable of fast, you won't be able to take advantage of expanding technologies like ML and AI to address the real-time advantages of converting your world of data into actionable information, and you will be left making "keeping up" investments later.

The sharing of data is fundamental to realizing its current and future value. Sharing occurs between people, applications, as pre-processed information passed to downstream processes, or as ML algorithms feeding AI. This shared information combines to enable new insights.

2. On-Demand and Automated

A data-centric architecture provides scalable, elastic services to applications and developers on-demand, with policy-based automation that efficiently stands up services while minimizing and even eliminating the need for repetitive intervention. This benefits a business from two directions:

- First, it enables an IT team to create and iterate services to be made available to its user community (DevOps), whether internal employees or external customers. The same is true of a SaaS company building and rolling out services.
- Second, it allows user self-service for employees or customers (whether business-to-business [B2B] or consumer) to point-and-click their way to the services they need. These can be routine tasks to request provisioning of resources, revenue-generating purchases of new services, or features that a business makes available.

3. Globally Reliable and Secure

A data-centric architecture delivers self-healing and protection capabilities to provide business continuity while helping to secure data. This encompasses a spectrum of availability:

- First, data must always be available to the business — no exceptions — and a data-centric architecture will deliver data to the business for everything up to and including mission-critical workloads.
- The midpoint on the spectrum delivers reliability of the business itself with backup and recovery (B/R), emphasizing the speed of recovery, disaster recovery (DR), and business continuity (BC). BC can have a range of service-level agreements (SLAs) from simple recovery or failover to a live environment in hours, to "I never knew it was down," immediate hot-site readiness.
- That BC plan overlaps with security. In addition to securing workflows by having a failover plan, a data-centric architecture will secure the data itself with encryption. There are technologies to consider and operational efficiencies to manage.

As a result, for reasons ranging from recovery after a malicious attack (or a benign mistake) to ongoing regulations and compliance, a data-centric architecture provides assurance that your data is a trusted single source of truth.

4. Hybrid Cloud by Design

A data-centric architecture enables the co-existence and interworking of on-premises, hosted, as-a-service, and public cloud data resources, allowing applications to take advantage of the most appropriate mix and mobility across resources to meet specific business needs.

The future is multi-cloud — an approach that isn't lift-and-shift or just colocation; it is designing for the transparent inclusion of, and mobility of, data in different places and from various sources. This approach allows data to both serve the BC plan and also be processed together for the creation of new intelligence, enabling everything from analytics insights to ML algorithms that feed downstream machine-speed processes.

5. Constantly on and Improving

A data-centric architecture improves and upgrades while continuously operating at the speed demanded by applications and users, minimizing technology obsolescence concerns. Accept no downtime, whether planned or unplanned. Machine processes and systems now rely on your IT environment; the days of off-hours windows when employees are not accessing data are over. Machines don't sleep; they inhale massive streams of data at high velocity, and if you skip a beat, they know. A data-centric architecture will grow and improve at a top-speed sprint. Have a plan for a constantly running environment and for non-disruptive upgrades of your data-centric architecture.

These attributes of a data-centric architecture are delivered by your entire environment: compute, network, and storage. It's an end-to-end vision. The people and machines accessing data won't be conveniently centralized adjacent to the server and storage in your data center. They'll be distributed, accessing different information with various tools at a range of velocities and connecting to that data over network hops that might not all be under your control. To make processed data available at velocity, at the point where it is part of your value creation, takes more than just adding more compute or new networking to go faster.

How to Get to a Data-Centric Architecture

Today's organizations and their data centers are characterized by a constant need for velocity and the need to move increasing amounts of data over the network. This is true because growing numbers of interconnected apps and smart devices make use of output from increasingly high-powered processing capabilities in the data center. That's more streams of dense data making a round trip from the edge to the core. However, there will be network latency, processing limitations, and other issues that affect your ability to make data available where and when it is needed, much of which is beyond your control. It's not enough to be hyper connected; you already are. What's next, then? The answer is to optimize what you can, and data ingress, processing, and egress from storage are three critical actions at the pivot point of data's round trip from edge to core and back. Use storage as the fulcrum to "add fast" to your data-centric architecture.

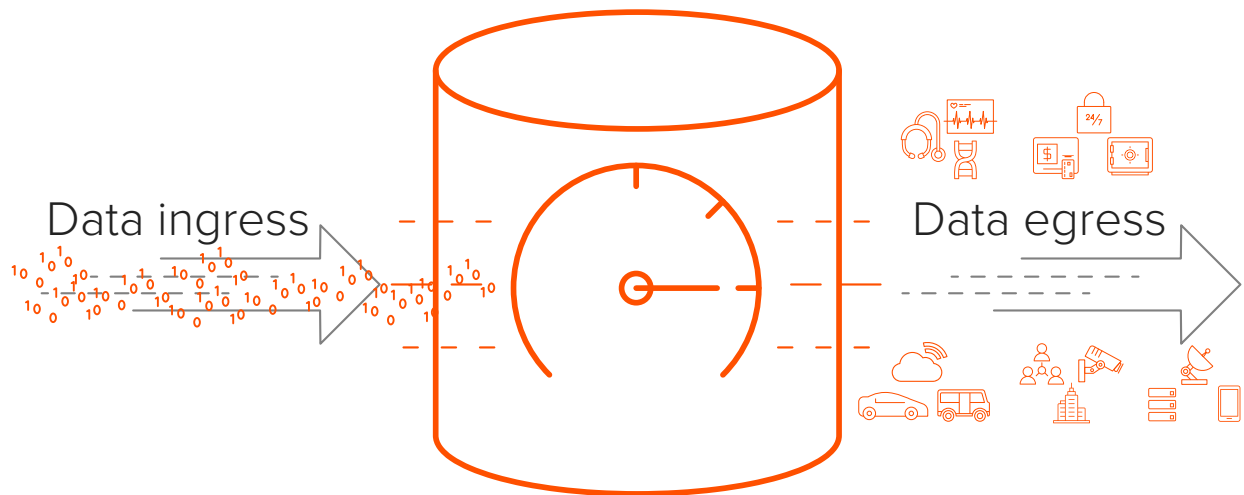


Figure 2. Use storage as the fulcrum to “add fast” to your data-centric architecture.

You have been adding more processing cores as you upgrade your compute base to modern chips; you have sought to optimize your network with 10- or 100-gigabit Ethernet (GbE) network interface controllers (NICs) and more-capable switches. In the evolutionary cycles of compute, network, and storage, it has become storage’s turn to help you lever up the capabilities of your data-centric architecture. Think of storage as the fulcrum that helps lift your infrastructure into the next stage of its evolution.

Pure Storage®, a leader in enterprise solutions, supports each of the data-centric architecture pillars:



- We instigated the all-flash array revolution, and **fast, shared data** is a fundamental principle at Pure. For the world of unstructured data, we deliver the multi-dimensional performance of FlashBlade™ for small and large files of varied types. We remove data-transfer obstacles and add performance at scale with DirectFlash™ software, flash modules, and fabric with NVM Express® (NVMe™) and NVMe over Fabrics in FlashArray™. A shared footprint is created between on-premises and the cloud with Cloud Block Store. Data-reduction efficiencies are built-in at each step for deep consolidation.



- At Pure, we feed an **on-demand and automated** data-centric architecture with building blocks of standardized file, block, object, VM, and container services, orchestrated with an API-first portfolio. This approach makes it possible for organizations to offer their users a catalog of repeatable offerings that are provisioned based on the data type, workload, and SLA that users need — all automated through software-defined orchestration. The result is a programmable data-centric architecture, from which DevOps offers services and users consume them. This is true in the enterprise and for SaaS providers standing up B2B services as their products.



- Pure addresses the need to be **globally reliable and secure** as an availability continuum because the measure of a data-centric architecture is the ability to make the right information available in the right form to the right place at the right time. With six-nines availability of FlashArray for mission-critical workloads, your data will always be accessible, so the next step is to plan for BC in the event of a catastrophic event.⁴

For that, Pure delivers instant recovery with FlashBlade RapidRestore, and recovery-time objective (RTO) and recovery-point objective (RPO) zero — across your data center, metro region, or globally across three data centers — with Purity ActiveCluster on FlashArray. Pure1® Cloud Mediator means you don't need a third site, and setup takes minutes. Finally, Purity software continuously protects your single-source-of-truth data with encryption that's built-in, always-on, and always in-line, and it costs your operations nothing: no impact on performance, no administrative overhead, and no key management.



- The reason for being **hybrid cloud by design** is to position and access data seamlessly, whether on-premises or in the cloud, with the automated ease that's the hallmark of cloud services. Pure delivers deep enterprise software-defined automation and orchestration on-premises with VMware vSphere® Virtual Volumes™ (VVs), coupled with the Evergreen™ Storage Service (ES2). Pure also unifies on-premises and public-cloud environments with Cloud Block Store and Purity CloudSnap™, extending your data-centric architecture to the cloud, with the ability to back up to the cloud with StorReduce technology.



- Our Evergreen Storage architecture and ownership model has been engineered into our products from day one because we saw the need for customers to be **constantly on and improving** before anyone called that a data-centric architecture. We deliver non-disruptive upgrades with data in-place across our portfolio. This means an in-place upgrade of Pure Storage controllers and flash, all while data is running at 100-percent performance — eliminating the need for planned downtime or legacy migration inefficiencies. As a result, we remove both the product and operational overhead that prevented customers from keeping their data-centric architectures fresh before Pure Storage.

Everything we do at Pure is meant to make the power of data available to you. Data center *and* distributed, cloud *and* on-premises, velocity *and* scale. “*And*” is hard, but these are all part of your data-centric architecture, so Pure gives your data-centric architecture “and.”

We deliver reliable, fast scale to handle all types of data, from engineered solutions for SAP, Oracle, and Microsoft, to new web-scale applications, and a data hub for the consolidation of different types of workloads and data types. We provide your single source of truth, API-driven for automation, and extending your operational control from on-premises to the cloud. We make your data available when, where, and how it creates value for you.

The difference with Pure doesn't end with implementation. Our support philosophy is that we work until you're back to 100 percent. How many times have you reached out to your storage technology provider and suffered through their efforts to qualify you out of getting help as they work to prove it's not their problem? That's not the Pure way. Our customers are family, and we stay with family until they're whole.

In addition, our products are architected from the ground up to be Evergreen. Controllers or shelves can be removed and replaced with newer, faster, higher-capacity generations on the fly and without downtime. Similarly, the Purity operating system was built from the start to deliver better storage management and true non-disruptive upgrades.

Your Data-Centric Architecture Is Within Reach

Your business, and the world it inhabits, are data-centric, which means new complexity, but also new opportunity. You seize that opportunity when you make daily decisions to create value for your business by taking advantage of data. Incremental change has brought us to this point, and the path continues forward to a data-centric architecture that enables data to become information and intelligence. That data-centric architecture has defining characteristics that you can work toward as you modernize and evolve your infrastructure; storage plays a key role in that evolution.

Pure Storage was founded on the first principles of making data effortless and efficient. We didn't set out with a vision of a product — our vision was of you, and how we could reduce the complexity, exertion, and inefficiency that plagued your legacy storage. Our vision was of the data platform just working, and working fast. We envisioned you getting your time back and building value with your data.

FOR MORE INFORMATION

Contact a Pure Storage representative today to discuss the evolution of your infrastructure toward a data-centric architecture.

© 2019 Pure Storage, Inc. All rights reserved. Pure Storage, Pure1, the P Logo, DirectFlash, Evergreen, FlashArray, FlashBlade, and Purity CloudSnap are trademarks or registered trademarks of Pure Storage, Inc. in the U.S. and other countries. All other trademarks are registered marks of their respective owners.

- ¹ VentureBeat. "What Amazon taught us this week: Data-centric companies will devour competitors." June 2017.
<https://venturebeat.com/2017/06/24/what-amazon-taught-us-this-week-data-centric-companies-will-devour-competitors/>.
- ² Forbes. "Why Data Is The Most Important Currency Used in Commerce Today." March 2018.
www.forbes.com/sites/michelleevans1/2018/03/12/why-data-is-the-most-important-currency-used-in-commerce-today/#5aada46654eb.
- ³ KPMG. "Data is the new currency." November 2018.
<https://home.kpmg/im/en/home/insights/2018/11/data-is-the-new-currency.html>.
- ⁴ Pure Storage. Entry-Level Storage web page.
www.purestorage.com/products/entry-level-storage.html.