



# **BUILD OUT YOUR MODERN DATA PLATFORM ONE USE CASE AT A TIME**

Generate real value from your information assets with  
Data Transformation Services

---



# CONTENTS

Executive summary.....	3
Factor data into your ongoing transformation.....	3
Start with a data strategy and discovery.....	3
Big-bang approach versus use-case approach.....	4
Data platform considerations.....	5
How to design and build out a data platform.....	5
Data platform architecture and functionality.....	5
What functionality does your use case and your business require?.....	5
As-a-service deployment in a hybrid cloud.....	7
Security and compliance.....	7
Reference projects.....	8
Service options: Speed up your design and deploy data and AI strategies with HPE Pointnext Services.....	9
Put the HPE end-to-end capability to work for you.....	10



## EXECUTIVE SUMMARY

As businesses continue to transform digitally, data must be factored into the overall strategy. Of course, data and AI technologies are changing rapidly, which adds complexity. Organizations, therefore, should be thoughtful in their approach, considering all factors from design and deployment to security and compliance. This white paper explores these and other considerations—and looks at the various components that go into a modern data platform from Cloud to the Edge. It also defines how HPE Pointnext Services can help strategize, design, deploy, manage, and evolve that platform over time.

## FACTOR DATA INTO YOUR ONGOING TRANSFORMATION

Any transformation should be intentional and fluid if there's any chance of success. Factoring in change is also a must. In the case of digital transformation, this process can certainly be challenging, especially with so many moving pieces, technology layers and dependencies, and evolving trends—including hybrid cloud, Internet of Things (IoT), edge, mobility, security, and application modernization.

Of course, data also belongs on that list. Any organization that can effectively harness their information to gain real-time insight into the business will have a distinct advantage. But the data landscape itself is changing dramatically. Technology that was cutting edge just a few years ago is now considered legacy by some experts. Meanwhile, data continues to grow exponentially. IDC's Global DataSphere measures the amount of data created and consumed in the world each year. The ratio of unique data (created and captured) to replicated data (copied and consumed) is roughly 1:9, but the trend is a slow migration toward less unique and more replicated data.<sup>1</sup>

For decades, data has also been confined to different repositories, which has made it incredibly difficult, time consuming, and costly to tap into its value. When organizations could derive intelligence or insight from their data, it was often backward focused. Data collected through batch processing and stored in a database has traditionally been useful in generating reports—but only about the past.

The emergence of AI and affordable access to extensive computing and storage resources, however, changes things. And adoption is growing. More than 75% of business analytics software will use AI capabilities by 2027.<sup>2</sup> Through continual advancements in analytics, and technology evolutions in general, organizations now have the means to address some of the challenges described above—and use their data to learn from and do so much more.

With a truly modern data platform, companies have the ability to address all data—from different sources, in a variety of formats, and in real time. You can remove the barriers that once confined data to siloed systems. And you now have the ability to look forward, as well as back, to predict what might happen in the future.

But this kind of progress requires more than a simple upgrade of your existing data infrastructure. Significant changes to the data landscape require a deep revision in the strategy and design of a modern, effective platform.

What does that look like exactly? At HPE, we believe that the modern data platform will look different for every business. There is no one-size-fits-all when it comes to addressing your unique needs and requirements. The process of moving forward in terms of your data transformation requires planning, a long-term data strategy, and a willingness to evolve incrementally as technology and opportunities shift and evolve.

## START WITH A DATA STRATEGY AND DISCOVERY

Strategy and discovery are crucial to the process as you journey toward a modern data platform. In many cases, this involves re-envisioning your strategy and making sure it is aligned with your organization's broader transformational objectives.

All organizations have massive amounts of data, yet few truly enter into the process of analyzing that information. Data scientists, for example, are still spending about 45% of their time on data preparation tasks, including loading and cleaning data.<sup>3</sup>

The goal of a modern platform is to improve these percentages by targeting the value you want to achieve from your data and focusing on manageable wins.

In addition to aligning your data strategy with your overall transformation strategy, it's important to identify all facets of your data landscape and evaluate their level of maturity in comparison to your target goals for AI adoption. Factors to consider in your planning include:

- Data transportation and ingestion at whatever speed (from batch to real time) from heterogeneous and dispersed data sources
- Data unification and virtualization

<sup>1</sup> [IDC's Global DataSphere Forecast Shows Continued Steady Growth in the Creation and Consumption of Data](#)

<sup>2</sup> ["Artificial General Intelligence Market: General Purpose Artificial Intelligence, AI Agent Platforms, and Software 2020–2025,"](#) Research and Markets, June 2020.

<sup>3</sup> ["The State of Data Science 2020. Moving from Hype Toward Maturity,"](#) Anaconda, 2020



- Data management and governance
- Data infrastructure management and governance
- Data and data infrastructure security
- Data transformation into information by a variety of analytical engines and data science frameworks that allow you to accommodate any format of structured and unstructured data
- Data consumption in the context of multiple purposes and through a variety of processes, channels, and instruments

Discovery, in turn, is the process of mining your data to determine what you’re dealing with, especially if you have data coming in from multiple sources and in multiple formats. You also want to define how well the data accessed and used matches your analytics strategy.

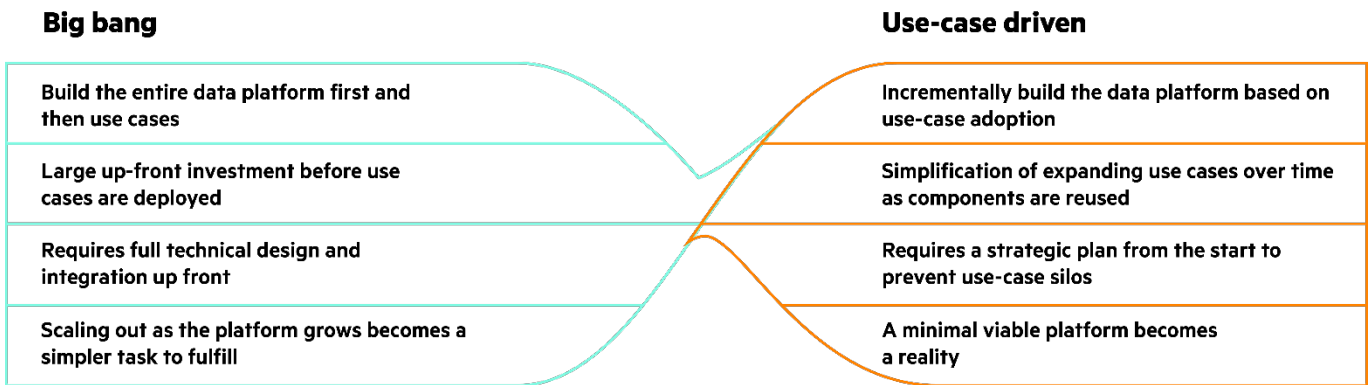
At this time, you might also evaluate the potential value that open data—information made available through public or government agencies—can bring to your business, and, if open data can be useful, determine how to integrate that data into your platform.

**Big-bang approach versus use-case approach**

During the strategy and discovery phase, consider the approach that is most feasible for your organization. In our experience at HPE, we initially saw many companies following a big-bang approach: first building the platforms and loading the data and only then developing the use cases. With a solid strategy in place, this approach can work, but it also requires a large up-front investment and can be risky if mistakes are made.

The use-case approach is an alternative method that is accessible to more businesses because it doesn’t require the extensive up-front investment and it best matches modern, agile approaches to solution implementation. This approach also requires that you have a long-term strategy in place guiding, prioritizing, and aligning your use cases. You want to avoid ending up with siloed data platform use cases that do not have the benefit of building upon or enriching each other.

Ideally, you want to start with one use case that forms the foundation of your evolving data platform. The first use case might start with the bare minimum of required data platform components. The second use case should then integrate with the first, adding more components to the platform. Over time, as you continue to add use cases, you enrich your data platform in terms of its overall functionality and the data that you can pool from.



**FIGURE 1.** Big bang versus use-case driven

It’s important to note that you have options in your approach. There is no one right answer, but HPE can certainly help you understand which approach serves your organization, given your priorities, funding, and capabilities to run such projects.



## DATA PLATFORM CONSIDERATIONS

### How to design and build out a data platform

While your long-term data strategy should be quite stable and aligned with your business strategy, the design and implementation of your modern data platform will be iterative in nature. Both should be approached incrementally so that you can see the value your investment brings to the business as quickly as possible. At the same time, it is important to remain open to change, recognizing that revision and iteration will likely be needed as you proceed, given the emergence of new technologies and products over time.

You'll also need to make decisions about deployment. Will you deploy on-premises, in a private or public cloud, or in a hybrid environment? Will you move forward with open-source technology or commercial products or a mix of both? The beauty of the incremental approach combined with a hybrid cloud is that both enable you to dilute your investment over time and mitigate risk by remedying mistakes as you go and before they become problematic.

### Data platform architecture and functionality

Your data platform is essentially the underlying architecture that will determine how you prepare your data, manage and protect it, analyze and monetize it, comply with regulations, and use it to inform decision-making.

A data platform architecture has different layers, each comprising multiple components, functionalities, and features. The platform is designed to be modular, letting you customize and tailor it according to your requirements. At the beginning of your implementation, you may only need a small subset of the entire functionality to get started on your initial use case. Over time, you can grow the platform to deliver additional use cases and functionality. Figure 2 shows a high-level view of the different components within a modern data platform.

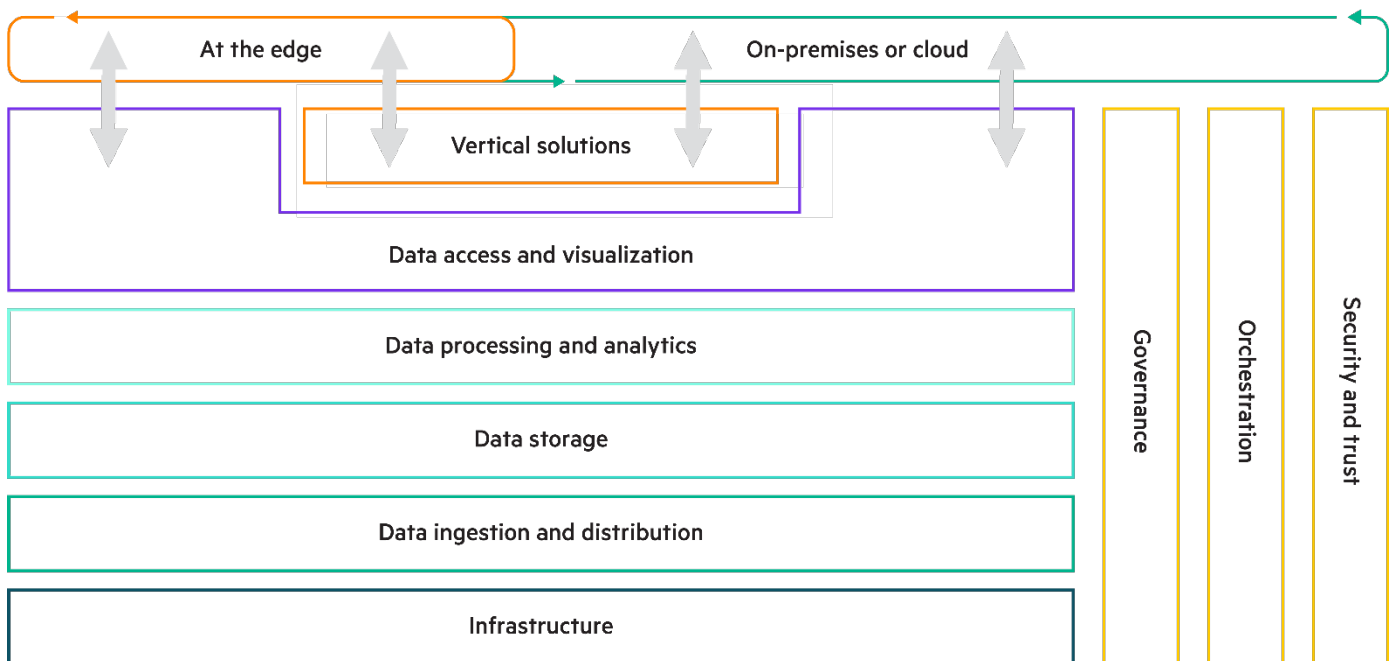


FIGURE 2. Data platform at a high level

### What functionality does your use case and your business require?

Desired functionality is what allows you to build out architectural detail. Once you've decided upon your use case and the outcomes you want to achieve, it's time to determine what functions and performance you need from your platform.

HPE can work with you to dig deeper into the design detail until we arrive at the right mix of products and functionality to integrate into your overall data platform build. This is precisely why every data platform is unique, tailored to your use cases and objectives. Figure 3 shows a more thorough view of the platform, including specific functionality.



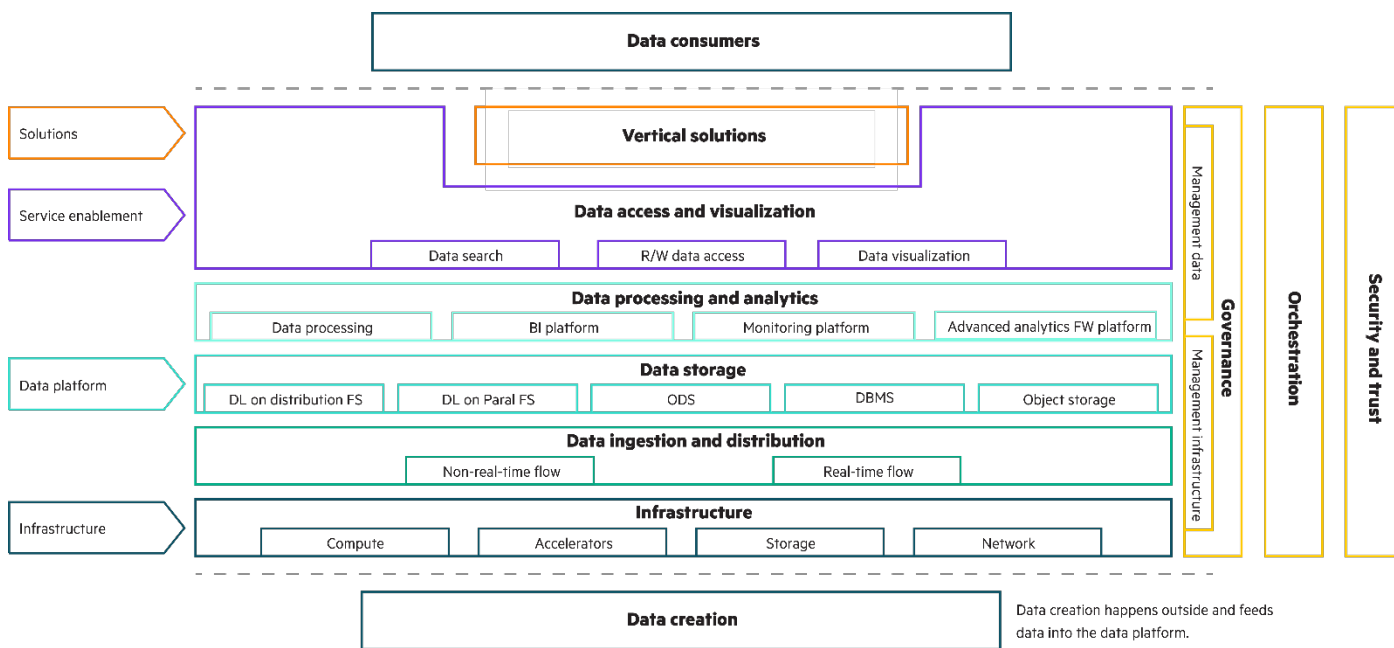


FIGURE 3. Data platform with more functional detail

### Industry-specific data platform scenarios

While your use case will help define what your data platform looks like, so will your industry. Different verticals have unique use cases. While all data platforms have a bulk set of components and specifications, some components will vary pending the use case. The vertical element is therefore a crucial part of the discussion and planning. Sample verticals and use cases include:

- Highly autonomous driving:** This scenario supports several different use cases, one of which is development. The information that auto manufacturers collect comes from hundreds of sensors in and on the cars they're testing. These sensors generate many terabytes of data with every test drive. Big Data batches are typically generated to help learn and refine development. Therefore, manufacturers need a data platform that can support high ingest rates of large files and can also prioritize ingest of the most interesting data first.
- Healthcare and life sciences:** Many use cases in healthcare and life sciences assume data protection and lineage. Data platforms must address specific privacy laws and compliance regulations. At the same time, many in healthcare and life sciences are also looking for a semantic database that will allow specific search functionality to support decision-making around issues such as procedures, infections, and research.
- Smart cities:** This scenario typically requires large numbers of parallel ingest feeds from connected devices all around the city. The data collected varies, ranging from tiny readings such as current temperature to large video feeds. Smart cities rely heavily on control loops to affect change around the city in real time. And data management controls are frequently required.
- Factory of the future:** Many manufacturers are highly distributed, with multiple factories spread out geographically. They may have different data platforms running in each factory. Much of the processing is often done locally at the edge (i.e., video analytics for quality control), while subsets of that data may be sent back to headquarters for centralized processing.

It's clear that different industries will have very different data requirements. HPE has experience working with all verticals; and we have extensive expertise working through a variety of utilization scenarios—like control loops that define and manage desired processes—across industries. An example of a data platform delivered for some of our customers is illustrated in the following figure.



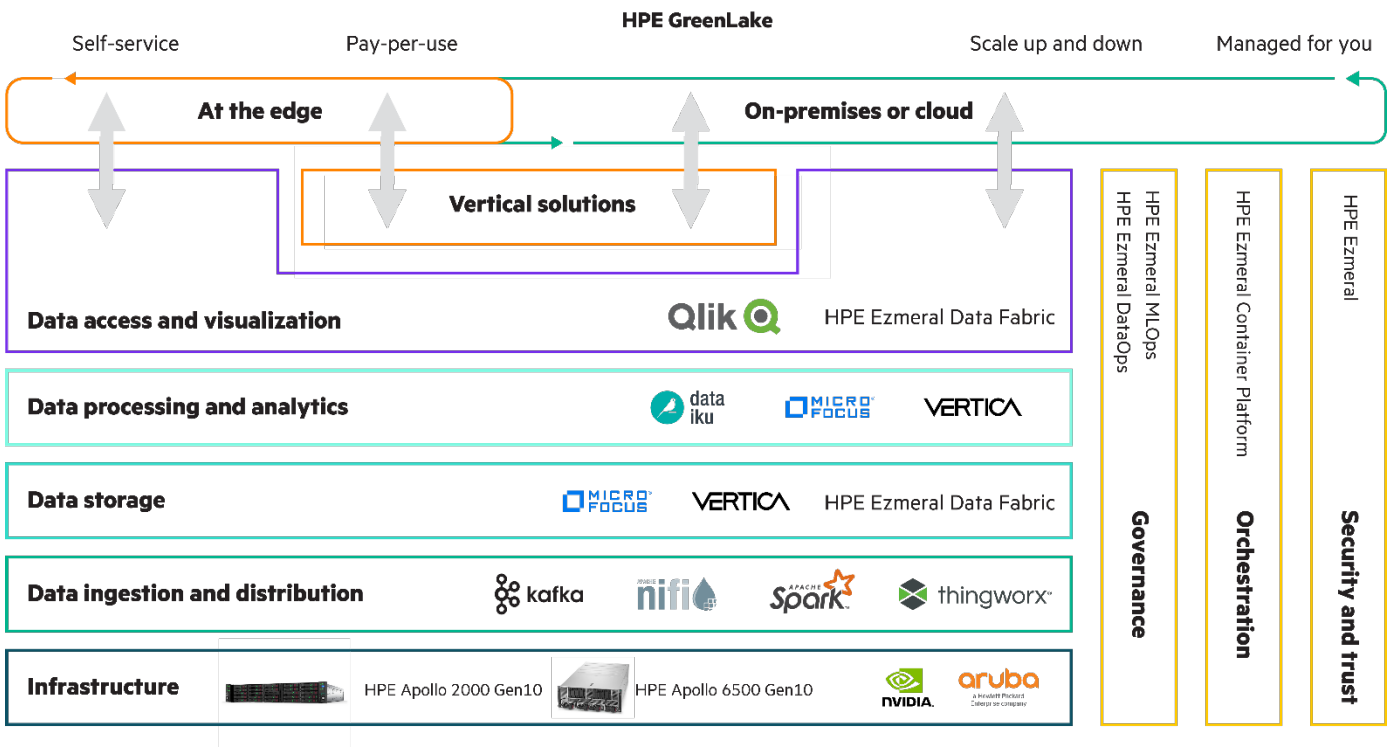


FIGURE 4. Example of on-premises data platform also consumable via HPE GreenLake

**As-a-service deployment in a hybrid cloud**

HPE can deploy your modern data platform as a standalone project or through an as-a-service model.

While the public cloud experience is valuable, the reality is that 70% of apps and data still remain outside the cloud<sup>4</sup> for reasons of data gravity, latency, IP protection, performance, and app entanglement. Many organizations want to bring that same cloud speed, agility, and as-a-service model to the masses of apps and data that aren't in the cloud.

Fortunately, HPE GreenLake brings this experience directly to your apps and data wherever they reside—the edge, colocations, or data centers. With a pay-per-use, scalable, point-and-click, self-service experience that is managed, HPE GreenLake delivers services for high-performance compute, Big Data, machine learning operations (MLOps), and foundational IT infrastructure workloads. In fact, MLOps developers and data scientists can rapidly build, train, and deploy ML models—from pilot to production, at any scale.

HPE GreenLake provides a viable as-a-service option to organizations to manage Big Data and data platforms when necessary—all while controlling costs, data governance, and security.

**Security and compliance**

While enterprises have embraced open data and opensource analytics tools in a wide variety of functions, their security practices don't always keep pace. In a recent survey, 30% of respondents claimed their organizations do not have mechanisms in place to secure open source tools used for data science and ML.<sup>5</sup>

Security in the data landscape should be a consideration from day one. It runs throughout your data platform regardless of whether you're working with just a few components or a full-featured system.

This is especially important as you move away from siloed data systems to a consolidated data platform. Because your data is valuable, consolidating it within a single platform, while beneficial for many reasons, also creates some risk. That's why built-in security, at every layer of the platform, helps safeguard your information assets and address issues along the way.

<sup>4</sup> IDC Cloud Pulse Q1 2020

<sup>5</sup> "The State of Data Science 2020. Moving from Hype Toward Maturity," Anaconda, 2020



Compliance is also a complex issue. Some industries and markets have very strict data governance requirements, regulated by national legislation and international standards. The process can be difficult. In healthcare, for example, how do you prevent someone from accessing patient information? And how do you put mechanisms in place that enable sharing of that data, but with added controls so only what's required is actually shared?

At the same time, laws are not consistent. Regulations are heavily influenced by location—they vary by region, country, state, and city. That complexity is further compounded for multinational organizations that commonly share information across borders.

In short, these are difficult, but incredibly important, issues that all organizations must consider and plan for. It's likely that you'll need guidance and expertise when it comes to integrating both security and compliance into your comprehensive data landscape and data platform build.

## REFERENCE PROJECTS

HPE has vast experience helping organizations across industries and across the globe with their data transformation initiatives. The data platform is the engine on which AI and data-driven use cases can be built. The following examples demonstrate the flexibility and breadth of opportunity that a well-architected data platform can provide.

**Financial services—U.S.:** In the process of assisting this financial services company with its digital transformation, the bank also wanted a solution that would help them better address compliance and regulatory requirements. They were looking to create a unified data repository, and they also wanted to avoid high costs associated with the public cloud. They ultimately landed on the HPE Ezmeral Container Platform. On top of that, they're using HPE GreenLake, and containers as a service, to run the company's Splunk software.

This example demonstrates how modern technologies such as containerized applications and container platforms can be used to provide a scalable architecture for use cases, regardless of whether you're dealing with one large use case or multiple smaller use cases. With HPE GreenLake, this company gained the advantages of a cloud-like as-a-service experience on-premises.

**Public sector—Europe:** A European customer wanted to accelerate AI/MLOps use case implementation and deployment. The client needed a proof of concept (POC) and technical assistance automating HPE Ezmeral software tenants and application deployment with their Dev/Ops tools, integrating with both a secure Hadoop data lake through Spark and an existing Kubernetes platform.

The department adopted HPE Ezmeral software and other third-party software, using HPE Pointnext Services to build the overall solution. The solution includes integrated KubeDirector applications for data scientists and a scalable Spark cluster with connection to the underlying data lake. Now they can more efficiently deliver new use cases on a scalable platform and access everything as a code.

**Healthcare—U.S.:** A U.S.-based customer needed a single solution to support its data scientists with faster analytics and its developers with an efficient AI/ML platform that could be provided as a self-service offering. Using the HPE Ezmeral Container Platform and HPE Pointnext Services, the customer was able to build out the AI/ML solution.

The provider deploys applications in minutes versus weeks, and the platform leverages the existing data sources securely, without the need for migrating petabytes of data to a new data platform. The solution's multi-tenancy capabilities for various teams allows isolation of data and workloads, while the data connections to existing HDFS and NFS data sources mean that data does not need to be replicated in order to use it. The HPE Ezmeral Data Fabric is used for storing data locally for workloads requiring higher IO throughput.





## SERVICE OPTIONS: SPEED UP YOUR DESIGN AND DEPLOY DATA AND AI STRATEGIES WITH HPE POINTNEXT SERVICES

HPE Pointnext Services provides a complete set of methodologies and services to help IT teams lead their data transformation journey to plan for, build, and consume a modern data platform. They are designed and organized via a three-phase approach to solutioning in both the data and AI domains: explore, experiment, and evolve.

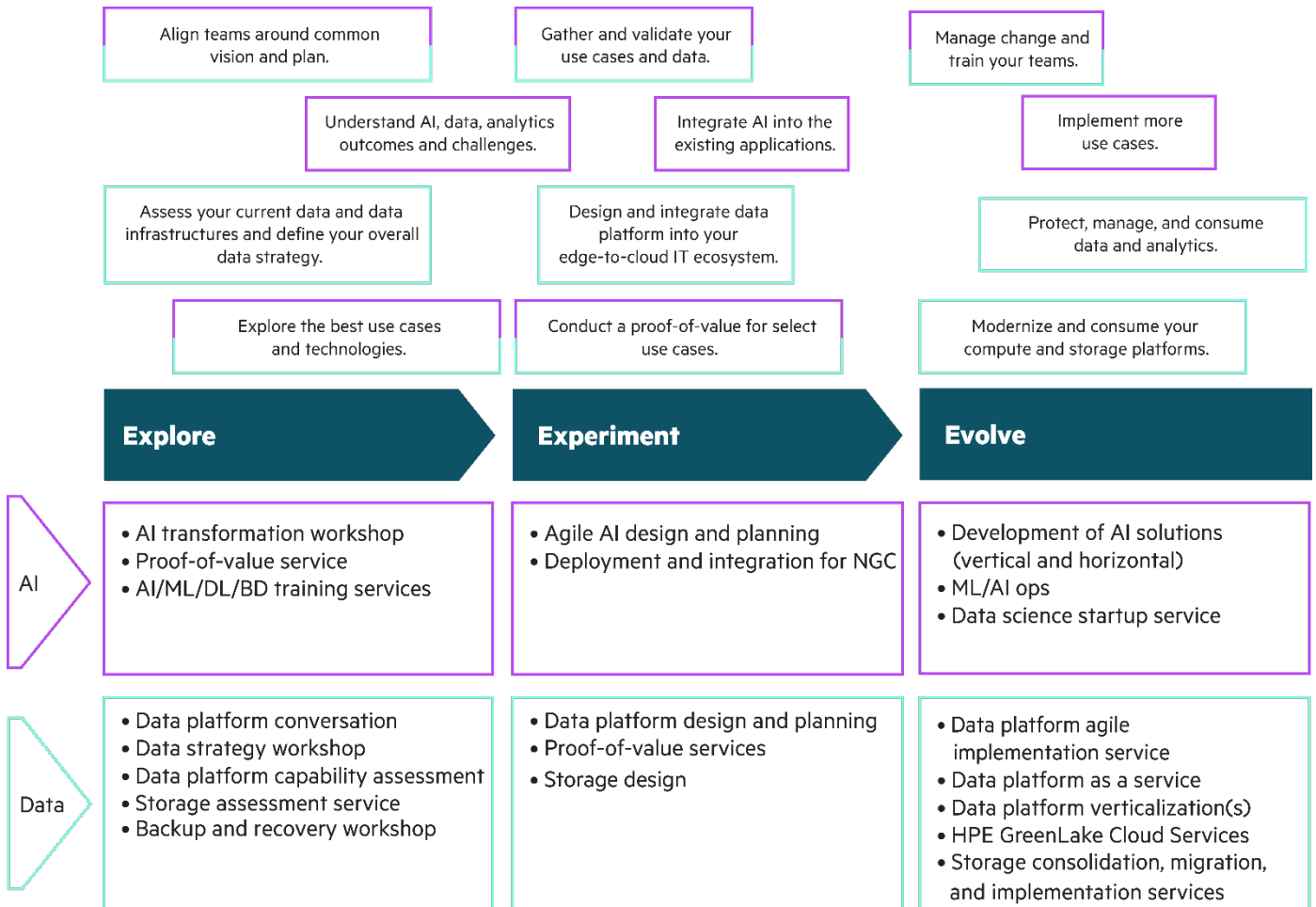


FIGURE 5. 3-step approach with data transformation services from HPE

- Explore:** This phase is for customers just starting to consider and understand the kind of value that data and AI can bring to their organization. It explores questions like: Where do we start? What are some quick-win projects? How do we build internal consensus and awareness for data and AI? How can we gain buy-in from decision-makers and build a business case?
- Experiment:** Once a customer knows where they want to go, the experiment phase is the natural next step. We help customers translate their plans into executable solutions/applications, experiment with these solutions, and experience the real value that data and AI can generate for their business.
- Evolve:** Becoming a true data-driven organization means extensively adopting the use of data and AI strategically throughout your organization. Our services can help you develop additional use cases, deploy into production, and integrate them into the entire IT environment, operationalize data and AI solutions and systems, and support the continuous evolution in line with your technology transformation, changing demands, and development of new business models.



The uniqueness of HPE's proposition is based on our ability to unite technology innovation with business innovation through collaboration with our valued customers and to offer different consumption models, such as on-premises and as-a-service models like HPE GreenLake. We also rely on our vast partner network to help ensure a comprehensive offering and end-to-end services and solutions. Here are some of the HPE Pointnext Services to help you get started:

- **Data strategy workshop** is a one-day workshop designed to help you discover AI and data analytics needs and opportunities, select priority use cases, and create a high-level plan.
- **Data platform capability assessment** evaluates the attributes of your data platform to determine the readiness of data and analytics services and technologies to satisfy your organization's needs and help you achieve your business objectives.
- **Data platform design and planning** is based on agile and enterprise architecture methods in design and planning. We combine innovation with a structured approach to technical and economical evaluation. We focus on continuous improvement through incremental planning and development and help reduce risk through complexity management.
- **Agile data platform implementation** leverages an agile (incremental) method to implement and deploy your data platform solution on the basis of a road map and implementation plan outlined up front through our advisory services (above) or your internal analysis. We build your data platform by splitting the job into multiple chunks. Incremental implementation of your data platform enhances your ability to control the quality of the solution, ensure alignment with your objectives, and get an immediate measure of the respective ROI and benefits.

## PUT THE HPE END-TO-END CAPABILITY TO WORK FOR YOU

Your data landscape is complex. But don't let that get in the way of taking full advantage of your information assets. Integrate data into your larger transformation strategy and begin to take control of the insight and value that exists within.

Build out a modern data platform at your pace and on your terms by targeting one use case or a series of use cases that can deliver a fast and solid return on your investment. Build on your strategy, moving incrementally, so you can adjust and evolve with change along the way.

And remember, you're never alone. HPE is uniquely positioned to help you strategize, design, build, and ultimately benefit from a modern data platform. We bring an end-to-end solution portfolio and professional skills to see you through—not just from beginning to end, but across every layer of the platform. From infrastructure to software, the HPE Ezmeral Data Fabric, and a broad range of services—including HPE Pointnext Services and Financial Services—we can help you achieve the outcomes you're seeking while building up an agile data platform as well as a very competitive advantage.

## LEARN MORE AT

[hpe.com/services/ai-data](https://hpe.com/services/ai-data)

Make the right purchase decision.  
Contact our presales specialists.



Chat



Email



Call



HPE support



Get updates

© Copyright 2021 Hewlett Packard Enterprise Development LP. The information contained herein is subject to change without notice. The only warranties for Hewlett Packard Enterprise products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Hewlett Packard Enterprise shall not be liable for technical or editorial errors or omissions contained herein.

The NVIDIA logo is a trademark and/or registered trademark of NVIDIA Corporation in the U.S. and other countries. All third-party marks are property of their respective owners.

a50003944ENW, April 2021