Accelerate Your AI Journey with IBM Spectrum Storage







Contents

- 01 Executive Summary
- 02 Storage Key to Drive Value as HPC, Analytics and AI Converge
- 03 A Common Infrastructure Enables HPC, Analytics and AI Convergence
- **04** High Value Game-Changing AI/ML/DL Use Cases 08 Emerging
- 05 ... But a High Performance Information Architecture (IA) is Needed
- 06 AI Data Challenges, Storage Requirements and Solutions
- 07 Storage is Critical to Accelerate Complex Iterative AI Workflows
- 08 IBM Spectrum Storage Accelerates AI
- 09 Description of Key IBM Spectrum Storage Solutions for HPC and AI
- **10** Client Examples
- 11 IBM Spectrum Storage the Bedrock for HPC and AI
- 12 Learn more
- **13** IBM Legal Disclosures



Click to navigate



01	Executive Summary	03
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04
03	A Common Infrastructure Enables HPC, Analytics and AI Convergence	06
04	High Value Game-Changing AI/ML/DL Use Cases Emerging	08
05	But a High Performance Information Architecture (IA) is Needed	09
06	AI Data Challenges, Storage Requirements and Solutions	10
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13
80	IBM Spectrum Storage Accelerates AI	18
09	Description of Key IBM Spectrum Storage Solutions for HPC and AI	19
10	Client Examples	23
11	IBM Spectrum Storage – the Bedrock for HPC and AI	26
12	Learn more	27
13	IBM Legal Disclosures	27

TEV.

PAGE 3

IBM Spectrum Storage

01 Executive Summary

A profound digital transformation is underway as High Performance Computing (HPC) and Analytics converge to Artificial Intelligence/ Machine Learning/Deep Learning (AI/ML/DL). Across every industry, this is accelerating innovation and improving a company's competitive position, and the quality and effectiveness of its products/services, operations and customer engagement. AI/ML/DL initiatives are front and center in the C-suite at many enterprises.

While the promise and economic value of AI are immense, so are the challenges. As the volume, velocity and variety of data continue to explode, deploying and scaling AI across the enterprise is not easy. It requires holistic solutions across business, process, applications, data and infrastructure dimensions.

What's also needed is a well-designed, agile, scalable, high-performance Information Architecture (IA). "There is no AI without IA". At the foundation of an IA are high-performance storage systems and software optimized for every stage of the AI workflow.

You can collect, organize and analyze data and infuse AI everywhere with trust and transparency with IBM Spectrum Storage. Already proven in HPC, this Software-defined Storage (SDS) portfolio also accelerates your AI journey from ingest to insights.



01	Executive Summary	03
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04
03	A Common Infrastructure Enables HPC, Analytics and AI Convergence	06
04	High Value Game-Changing AI/ML/DL Use Cases Emerging	08
05	But a High Performance Information Architecture (IA) is Needed	09
06	AI Data Challenges, Storage Requirements and Solutions	10
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13
80	IBM Spectrum Storage Accelerates AI	18
09	Description of Key IBM Spectrum Storage Solutions for HPC and AI	19
10	Client Examples	23
11	IBM Spectrum Storage – the Bedrock for HPC and AI	26
12	Learn more	27
13	IBM Legal Disclosures	27

IBM Spectrum Storage

02 Storage Key to Drive Value as HPC, Analytics and AI Converge

Deep Learning (DL) and Machine Learning (ML), major large-scale, coupled, multidisciplinary, iterative modeling of components of Artificial Intelligence (AI), are rapidly growing complex systems. This from the value-accretive convergence of High Performance Computing (HPC) with Analytics.¹ Over the next five years, AI is expected to improve economic growth (90%), productivity HPC **Analytics** (86%), innovation (84%) and job creation $(69\%)^2$. This is profoundly enhancing how businesses and governments engage with their stakeholders.

AI adoption alone has grown an astounding 270% in the last four years and 40% of organizations expect it to be a game changer.³

Consequently, with 2018 revenues of \$28.1 billion, the relatively new AI market is rapidly growing at 35.6% annually.⁴ This is also spurring a 7.2% annual growth in the decades old HPC market (\$27.7B in 2018). In fact, HPCbased AI revenues are expected to grow 29.5% annually.⁵

These healthy projections could further increase as traditional enterprise (non-HPC) clients deploy HPC-like architectures to scale their AI/ML/DL workloads. This is Machine Learning already happening in several non-traditional HPC industries like Retail, Telecommunications, Cybersecurity and more. HPC and Analytics.

Data gathering techniques are improving with more sophisticated analytics from descriptive to predictive to AI/ML/DL (Figure 1). Increasingly, this requires ingestion, generation and retention of larger datasets at each stage of the analytic workflow. Likewise, HPC has evolved from storage systems and software. simulating simple physics at a component/part level to



01	Executive Summary	03
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04
03	A Common Infrastructure Enables HPC, Analytics and AI Convergence	06
04	High Value Game-Changing AI/ML/DL Use Cases Emerging	08
05	But a High Performance Information Architecture (IA) is Needed	09
06	AI Data Challenges, Storage Requirements and Solutions	10
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13
08	IBM Spectrum Storage Accelerates AI	18
09	Description of Key IBM Spectrum Storage Solutions for HPC and AI	19
10	Client Examples	23
11	IBM Spectrum Storage – the Bedrock for HPC and AI	26
12	Learn more	27
13	IBM Legal Disclosures	27



IBM Spectrum Storage

02 Storage Key to Drive Value as HPC, Analytics and AI Converge

Also, the volume, velocity and variety of data continue to already almost \$5 billion a year and expected to grow.⁷ explode. By 2025, worldwide data will grow 61% annually A common infrastructure for HPC, Analytics and AI enables to 175 zettabytes⁶ (about 175 trillion gigabytes). This is organizations to overcome these data management and compounding data management and storage challenges storage challenges. This results in accelerated innovation for HPC and AI-further driving up the demand for very highand improved quality and effectiveness of a company's performance storage systems and software. In fact, spending products/services, operations and customer engagement. on storage systems and software just for AI initiatives is

03 A Common Infrastructure Enables HPC, Analytics and AI Convergence

Organizations continue to gather, process and store larger datasets from all sources, such as sensors, instruments, log files and so on. Consequently, their HPC and Analytics workloads are becoming more compute and data intensive and increasingly look alike. This is placing similar management, performance and scale demands on IT infrastructures to the point where the same infrastructure can be used for both workloads.

In addition, companies can integrate simulation operations and services end-to-end (Figure 2) across the stack on the same infrastructure from the edge to core to the cloud. Key layers include:

Functions, processes, and applications: Organizations use many applications to implement key business processes in and improves customer or patient outcomes. It also provides deeper insights into the consumer's use patterns or product each function. Increasingly, in many industries, data across effectiveness, addresses regulatory obligations, mitigates functions are being integrated and analyzed. This drives greater risks, and more. innovation, enhances efficiency of product/services operations



Figure 2: A Common Infrastructure enables HPC, Analytics and AI Convergence across many Industries

HPC and Analytics are converging to AI/ML/DL. Many HPC









01	Executive Summary	03
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04
03	A Common Infrastructure Enables HPC, Analytics and AI Convergence	06
04	High Value Game-Changing AI/ML/DL Use Cases Emerging	08
05	But a High Performance Information Architecture (IA) is Needed	09
06	AI Data Challenges, Storage Requirements and Solutions	10
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13
08	IBM Spectrum Storage Accelerates AI	18
09	Description of Key IBM Spectrum Storage Solutions for HPC and AI	19
10	Client Examples	23
11	IBM Spectrum Storage – the Bedrock for HPC and AI	26
12	Learn more	27
13	IBM Legal Disclosures	27

IBM Spectrum Storage

03 A Common Infra Convergence

applications have more data-driven computing capabilities and several data-driven Analytics applications are leveraging HPC technologies.

Data types: Organizations are managing an ever-growing volume and variety of data in both science-based simulation and data-driven computing. The data sets associated with traditional HPC (computer-aided engineering, oil and gas exploration, drug discovery, risk analytics, and more) are growing because of larger multidisciplinary analyses. At the same time, a company's operations and service functions are dealing with even larger volumes of unstructured data in the form of images, video, voice, text, sensor data, and others.

AI/ML/DL techniques are increasingly being used to analyze and interpret this huge volume and variety of data and provide dashboards/reports for actionable insights.

Infrastructure: To support very demanding compute and data-intensive workloads and achieve higher quality results faster, organizations need a faster and more scalable and capable infrastructure. Adding more hardware is not always possible because of costs, complexities, and the risk of cluster and data sprawl. A single common infrastructure for both types of workloads is desirable, feasible and facilitates data sharing and integration.

Key industry use cases that illustrate the value of integrating simulation, operations and services:

Automotive: Engineers use Computer-Aided Engineering (CAE) software to simulate vehicle collisions. They then conduct actual crash tests and collect data from tens of thousands of sensors for further analysis. They also analyze field defect data, service and warranty data and real-time telemetry from vehicles-in-use. This gives them deeper insights to **build better vehicles** that perform better on the road.

03 A Common Infrastructure Enables HPC, Analytics and AI



01	Executive Summary	03
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04
03	A Common Infrastructure Enables HPC, Analytics and AI Convergence	06
04	High Value Game-Changing AI/ML/DL Use Cases Emerging	08
05	But a High Performance Information Architecture (IA) is Needed	09
06	AI Data Challenges, Storage Requirements and Solutions	10
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13
80	IBM Spectrum Storage Accelerates AI	18
09	Description of Key IBM Spectrum Storage Solutions for HPC and AI	19
10	Client Examples	23
11	IBM Spectrum Storage – the Bedrock for HPC and AI	26
12	Learn more	27
13	IBM Legal Disclosures	27

IBN .

PAGE 7

IBM Spectrum Storage

03 A Common Infrastructure Enables HPC, Analytics and AI Convergence

Healthcare and Life Sciences: Genomic medicine pipelines are large and sophisticated workflows with dozens of compute and data-intensive tasks that span Next Generation Sequencing (NGS), Translational Medicine and Personalized Healthcare. To **develop new treatments**, institutions rely on HPC and increasingly analytics frameworks such as Apache Spark to execute hundreds of thousands of jobs to analyze petabytes of data including text and images that are often distributed across tens of thousands of files.

Financial Services: Firms can maximize the value from their existing businesses while creating new revenue streams. They analyze both structured and unstructured data including email and PDF files to **improve profits and investment outcomes**, as well as to find patterns and trends in their clients' and/or employees' activities that may signal investment opportunities or fraud.

The convergence of HPC and Analytics to AI/ML/DL is game-changing, producing many more high value use cases across every sector of the economy and society.



01	Executive Summary	03
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04
03	A Common Infrastructure Enables HPC, Analytics and AI Convergence	06
04	High Value Game-Changing AI/ML/DL Use Cases Emerging	08
05	But a High Performance Information Architecture (IA) is Needed	09
06	AI Data Challenges, Storage Requirements and Solutions	10
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13
08	IBM Spectrum Storage Accelerates AI	18
09	Description of Key IBM Spectrum Storage Solutions for HPC and AI	19
10	Client Examples	23
11	IBM Spectrum Storage – the Bedrock for HPC and AI	26
12	Learn more	27
13	IBM Legal Disclosures	27



IBM Spectrum Storage

04 High Value Game-Changing AI/ML/DL Use Cases Emerging

Across many industries, high value transformative use cases in personalized medicine, predictive maintenance, fraud detection, cybersecurity, and more are rapidly emerging (Figure 3). The economic impact of AI/ML/DL is immense.

AI is rapidly shaping the future of work by primarily driving greater value with higher revenues (72%) and then secondarily with cost savings (28%). ⁸ With AI, organizations can predict and shape future outcomes, empower people to do higher value work, automate decisions, process and experiences, and reimagine new business models.

However, for Analytics and AI/ML/DL to become a crucial integral part of an organization, numerous challenges must be overcome. In a recent survey ⁹, 77% of respondents indicate that "business adoption" of big data and AI initiatives continues to be a challenge. Moreover, only 31% have a data-driven organization and fewer (28%) have a data culture where all stakeholders and decision-makers focus on the information conveyed by the existing data and make decisions and changes accordingly.

To operationalize and amplify the value of AI, it is imperative that clients consider best practices and solutions that address these challenges holistically across several dimensions: Business, Process, Applications and Technology.^{10,11,12} Doing so will enable clients to maximize value.



Figure 3: High Value AI Use Cases across many Industries







01	Executive Summary	03
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04
03	A Common Infrastructure Enables HPC, Analytics and AI Convergence	06
04	High Value Game-Changing AI/ML/DL Use Cases Emerging	08
05	But a High Performance Information Architecture (IA) is Needed	09
06	AI Data Challenges, Storage Requirements and Solutions	10
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13
08	IBM Spectrum Storage Accelerates AI	18
09	Description of Key IBM Spectrum Storage Solutions for HPC and AI	19
10	Client Examples	23
11	IBM Spectrum Storage – the Bedrock for HPC and AI	26
12	Learn more	27
13	IBM Legal Disclosures	27

05 ...But a High Performance Information Architecture (IA) is Needed

To maximize value from AI, customers must collaborate to *collect* data and make it simple and accessible and then *organize* the data to create a business-ready analytics base. Then, they must *analyze* this data to build Analytics and AI models with trust and transparency, and finally operationalize and *infuse* these AI models everywhere across the business.

This intricate, multistep, ladder-like journey and workflow is hard to implement. The biggest AI deployment challenges are with Data and Infrastructure: data volumes (50%), data quality and management (47%) and skills (44%).¹³

A high-performance Information Architecture (IA) is required to address these challenges. **"There is no AI without IA".** High-performance storage systems and software solutions with choice, scalability and flexibility are critical for a well-designed IA.

06 AI Data Challenges, Storage Requirements and Solutions

By using proven HPC storage software and systems for AI deployments, organizations can reducedatacosts, consolidate compute and storage siloes, simplify system administration, improve efficiency, and more. They must also ensure security, affordability, performance, scalability, compliance, flexibility to manage service level agreements (SLAs), and support different hardware and file systems on-premises and in multi-cloud environments, and more. Figure 4 depicts these **Data Challenges**, *Storage Requirements* and **Solutions** that are now detailed:

Data Volume: AI workflows iteratively process vast amounts of unstructured data in a short timeframe – an essential attribute since large and often rapidly changing data sets are required to deliver accurate algorithms. This data volume, velocity and variety drive significant storage requirements in *scalability, performance and affordability.*

PAGE 9



01	Executive Summary	03
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04
03	A Common Infrastructure Enables HPC, Analytics and AI Convergence	06
04	High Value Game-Changing AI/ML/DL Use Cases Emerging	08
05	But a High Performance Information Architecture (IA) is Needed	09
06	AI Data Challenges, Storage Requirements and Solutions	10
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13
80	IBM Spectrum Storage Accelerates AI	18
09	Description of Key IBM Spectrum Storage Solutions for HPC and AI	19
10	Client Examples	23
11	IBM Spectrum Storage – the Bedrock for HPC and AI	26
12	Learn more	27
13	IBM Legal Disclosures	27



IBM Spectrum Storage

06 AI Data Challenges, Storage Requirements and Solutions



To do this effectively, they not only need *performance*, scalability and affordability of the infrastructure but also enterprise-grade features that ensure security, Figure 4: Data Challenges, Storage Requirements and Solutions compliance and data durability. Backing up a multifor HPC, Analytics and AI petabyte training data set is often cost- and timeprohibitive. So, the storage system needs to be selfprotecting. This requires *intelligent* storage software solutions which can also help with the cumbersome curatorial, data cleansing tasks, and help run and monitor compute and data-intensive workloads efficiently and reliably from the edge to the core to multiple clouds.

An effective storage system must be both scalable and affordable, two attributes that don't always co-exist in enterprise storage. Historically, highly scalable systems have been more expensive on a cost/capacity basis. However, newer architectures allow computing and storage to be integrated more pervasively and costeffectively throughout the AI workflow. This enables clients to scale and accelerate AI, even as the volume, velocity and variety of data continue to grow.

Data Quality and Management : As data is quickly acquired repeatedly and managed throughout the AI workflow, it is hard to maintain good data quality and

security. Currently, tedious semi-manual, error-prone processes are used to cleanse and

prepare data, remove bias, prioritize it for analysis, and provide clear traceability. In fact, Data Scientists typically spend about 79% of their time with cumbersome data preparation and cleansing tasks.¹⁴ They also often spend Days/weeks in an iterative/cyclic process of building, tuning and selecting the correct features for the training model. This is very expensive.

Skills: This is in severe short supply and is throttling successful AI deployments. While the entire AI ecosystem (universities, solution providers, clients, etc.) is investing in education, training and certification, an *intelligent* infrastructure can streamline data flow from ingest to



01	Executive Summary	03
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04
03	A Common Infrastructure Enables HPC, Analytics and AI Convergence	06
04	High Value Game-Changing AI/ML/DL Use Cases Emerging	08
05	But a High Performance Information Architecture (IA) is Needed	09
06	AI Data Challenges, Storage Requirements and Solutions	10
0-		
0.7	Storage is Critical to Accelerate Complex Iterative AI Workflows	13
07	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI	13 18
07 08 09	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI	13 18 19
07 08 09 10	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples	13 18 19 23
07 08 09 10 11	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples IBM Spectrum Storage – the Bedrock for HPC and AI	13 18 19 23 26
07 08 09 10 11 12	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples IBM Spectrum Storage – the Bedrock for HPC and AI Learn more	13 18 19 23 26 27
07 08 09 10 11 12 13	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples IBM Spectrum Storage – the Bedrock for HPC and AI Learn more IBM Legal Disclosures	13 18 19 23 26 27 27 27

06 AI Data Challenges, Storage Requirements and Solutions

insights and ensure that *scalability*, *performance*, *affordability*, *security* and *compliance* requirements are met. This greatly improves the productivity of highly skilled Data Scientists, Data Engineers, Data Architects, Data Stewards and others throughout the AI workflow.

Data Fragmentation: AI algorithms become more accurate and efficient the more they get trained on large volumes of data from many sources, including valuable enterprise data. However, about 80% of data is not searchable¹⁵ because it is often fragmented or locked in siloes and not easily accessible. Also, since a lot of data is being generated at the edge or in the cloud, there needs to be streamlined flow of data to/from the cloud as information is generated and analyzed.

So, the infrastructure must have the *flexibility* to simplify and optimize these complex data flows for performance. Intelligent metadata management can deal with data stored in multiple geographic locations and environments. Wherever possible, moving the algorithms to where the data resides (data gravity) can accelerate the AI workflow and eliminate expensive data movement costs especially when reusing the same data iteratively.

Storage **Solutions** to address the preceding Data Challenges and Requirements must be:

Parallel: As clients add more storage capacity (including Network Attached Storage -NAS), they are realizing that the operating costs (including down time and productivity loss) of integrating, managing, securing and analyzing exploding data volumes are escalating. To reduce these costs, many clients are using high performance scalable storage with parallel file systems which can store data across multiple networked servers. These systems facilitate high-performance access through concurrent, coordinated input/ output operations between clients and storage nodes across multiple sites/clouds.





01	Executive Summary	03
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04
03	A Common Infrastructure Enables HPC, Analytics and AI Convergence	06
04	High Value Game-Changing AI/ML/DL Use Cases Emerging	08
05	But a High Performance Information Architecture (IA) is Needed	09
06	AI Data Challenges, Storage Requirements and Solutions	10
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13
07 08	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI	13 18
07 08 09	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI	13 18 19
07 08 09 10	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples	13 18 19 23
07 08 09 10 11	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples IBM Spectrum Storage – the Bedrock for HPC and AI	13 18 19 23 26
07 08 09 10 11 12	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples IBM Spectrum Storage – the Bedrock for HPC and AI	13 18 19 23 26 27

06 AI Data Challenges, Storage Requirements and Solutions

Hybrid: Different data types and stages in an AI workflow have varying performance requirements. The right mix of storage systems and software is needed to meet the simultaneous needs for scalability, performance and affordability, on premises and on the cloud. A hybrid storage architecture combines file and object storage to achieve an optimal balance between performance, archiving, and data governance and protection requirements throughout the workflow.

Software-defined: It is hard to support and unify many siloed storage architectures and optimize data placement to ensure the AI workflow runs smoothly with the best performance from ingest to insights. With no dependencies in the underlying hardware, Software-defined Storage (SDS) provides a single administrative interface and a policy-based approach to aggregate and manage storage resources with data protection and scale out the system across servers. It also provides data-aware intelligence to dynamically adapt to real-time needs and orchestrate IT resources to meet critical service level agreements (SLAs) in parallel, virtual and hybrid multicloud environments. SDS is typically platform agnostic and supports the widest range of hardware, AI frameworks, and APIs.

Integrated: A lot of AI innovation is occurring in the cloud. So, regardless of where the data resides, on-premises storage systems with cloud integration will provide the greatest flexibility to leverage cloud-native tools. Since over 80% of clients are expected to use at least two or more public clouds¹⁶, there will be a need for smooth and integrated data flow to/from multisite/multicloud environments. This requires more intelligent storage software for metadata management and integrating physically distributed, globally addressable storage systems.



01	Executive Summary	03
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04
03	A Common Infrastructure Enables HPC, Analytics and AI Convergence	06
04	High Value Game-Changing AI/ML/DL Use Cases Emerging	08
05	But a High Performance Information Architecture (IA) is Needed	09
06	AI Data Challenges, Storage Requirements and Solutions	10
		4.0
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13
07	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI	13 18
07 08 09	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI	13 18 19
07 08 09 10	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples	13 18 19 23
07 08 09 10 11	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples IBM Spectrum Storage – the Bedrock for HPC and AI	13 18 19 23 26
07 08 09 10 11 12	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples IBM Spectrum Storage – the Bedrock for HPC and AI Learn more	13 18 19 23 26 27
07 08 09 10 11 12 12 13	Storage IS Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples IBM Spectrum Storage – the Bedrock for HPC and AI Learn more IBM Legal Disclosures	13 18 19 23 26 27 27 27



IBM Spectrum Storage

07 Storage is Critical to Accelerate Complex Iterative AI Workflows

Compared to the much-hyped focus on compute-intensive AI training and inference tasks, there is little appreciation of the complexities and importance of data/storage management. As organizations move from experimentation and prototyping to deploying AI in production, they must embed AI into current HPC/Analytics workflows and leverage new/existing data repositories.

The AI/Analytics workflow (Figure 5) has more green data/storage management blocks and fewer blue reporting or compute-intensive applications boxes. Managing all this data at every phase of this complex, iterative workflow is one of the most vexing challenges in deploying AI.

As data flows through the workflow, storage requirements at each stage varies. Each phase has unique demands for scalability, performance, affordability, location and more. Different stages of the workflow exhibit unique I/O characteristics and benefit from complementary storage infrastructure. The infrastructure must be carefully optimized at every stage of the workflow from ingest to insights :

Transient storage: Software-defined, throughput-oriented and generally stored in Solid State Devices (SSD). This type of storage is optimized to process data in real time. Real-time is useful when customers are processing data from a streaming source, such as the data from financial markets or telemetry from connected devices. As the name suggests, the storage is temporary.

Global ingest: Capacity tier, throughput-oriented and globally accessible. These storage systems and tools are optimized to work with cloud-based data, such as data from cloud buckets. These tools are hosted in the cloud, allowing customers to save money on

1. Collect/Ingest of mixed data, primarily unstructured, with raw throughput is the first step of the / AI workflow. This stage benefits from the flexibility of SDS at the edge and demands high throughput. Three major storage types used for data acquisition include:



01	Executive Summary	03
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04
03	A Common Infrastructure Enables HPC, Analytics and AI Convergence	06
04	High Value Game-Changing AI/ML/DL Use Cases Emerging	08
05	But a High Performance Information Architecture (IA) is Needed	09
06	AI Data Challenges, Storage Requirements and Solutions	10
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13
07 08	Storage is Critical to Accelerate Complex Iterative AI WorkflowsIBM Spectrum Storage Accelerates AI	13 18
07 08 09	Storage is Critical to Accelerate Complex Iterative AI WorkflowsIBM Spectrum Storage Accelerates AIDescription of Key IBM Spectrum Storage Solutions for HPC and AI	13 18 19
07 08 09 10	Storage is Critical to Accelerate Complex Iterative AI WorkflowsIBM Spectrum Storage Accelerates AIDescription of Key IBM Spectrum Storage Solutions for HPC and AIClient Examples	13 18 19 23
07 08 09 10 11	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples IBM Spectrum Storage – the Bedrock for HPC and AI	13 18 19 23 26
07 08 09 10 11 12	Storage is Critical to Accelerate Complex Iterative AI WorkflowsIBM Spectrum Storage Accelerates AIDescription of Key IBM Spectrum Storage Solutions for HPC and AIClient ExamplesIBM Spectrum Storage – the Bedrock for HPC and AILearn more	13 18 19 23 26 27



IBM Spectrum Storage

07 Storage is Critical to Accelerate Complex Iterative AI Workflows



Figure 5: A Typical Complex Iterative AI Workflow Dominated by Data/Storage Management Tasks

infrastructure and expert resources because they can rely on the infrastructure and expertise of the provider hosting the workflow. Cloud object storage provides geographically dispersed object repositories that support global ingest, transient storage and cloud archive of object data.

Fast ingest/real time analytics: High throughput performance tier generally stored in SSD devices. This type of storage must be able to handle a massive number of digital "events" – everything from changes in a car's engine to millions of retail

01	Executive Summary	03
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04
03	A Common Infrastructure Enables HPC, Analytics and AI Convergence	06
04	High Value Game-Changing AI/ML/DL Use Cases Emerging	08
05	But a High Performance Information Architecture (IA) is Needed	09
06	AI Data Challenges, Storage Requirements and Solutions	10
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13
07	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI	13 18
07 08 09	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI	13 18 19
07 08 09 10	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples	13 18 19 23
07 08 09 10 11	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples IBM Spectrum Storage – the Bedrock for HPC and AI	13 18 19 23 26
07 08 09 10 11 12	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples IBM Spectrum Storage – the Bedrock for HPC and AI Learn more	13 18 19 23 26 27
07 08 09 10 11 12 13	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples IBM Spectrum Storage – the Bedrock for HPC and AI Learn more IBM Legal Disclosures	13 18 19 23 26 27 27

IBM Spectrum Storage

07 Storage is Critical to Accelerate Complex Iterative AI Workflows

transactions – in real time or near-real time to help support a wide and growing range of applications and customized services. In addition to fast ingest, these systems must be able to handle in-memory indexing for fast and efficient lookups, near real-time analytics on all ingested data with online analytical processing (OLAP), integrated machine learning capabilities to "learn" from previous events, and high availability and replication to provide continuous value to the business.

Clean, Format, Label: With data streaming in from hundreds of sensors, a single source (vehicle, plant equipment, building, gene sequencing machine, etc.) can produce terabytes of data each day. However, Data Scientists typically do not look at just one source. They may have to look at numerous sources and as time goes on, might have multiple Analytics/AI models, with multiple versions and hundreds of different data subsets.

Dealing with raw data is also not very convenient since it was generated and formatted without considering analysis requirements. Raw data often contains semantic errors, missing entries, or inconsistent formatting, so it needs to be "cleaned and formatted" prior to analysis. It is also important to accurately track data provenance, i.e., where each piece of data comes from and whether it is still up to date, since data often needs to be re-acquired in the future to run updated experiments.

Classification and Metadata Tagging: Standard metadata provides information that is adequate for a cursory understanding of data. Custom tags enable deeper analysis from

2. Organize to clean, format, label data; classify and tag metadata; extract, transform, load (ETL) with random I/O; and govern. These stages require intelligent storage software/ systems to scan vast data, discover key patterns or anomalies (including security breaches, lack of compliance, etc.), report them and then aggregate, normalize, classify data, and enrich it with useful metadata with extremely high throughput for both small and large I/O operations.



01	Executive Summary	03
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04
03	A Common Infrastructure Enables HPC, Analytics and AI Convergence	06
04	High Value Game-Changing AI/ML/DL Use Cases Emerging	08
05	But a High Performance Information Architecture (IA) is Needed	09
06	AI Data Challenges, Storage Requirements and Solutions	10
		4.0
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13
07	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI	13 18
07 08 09	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI	13 18 19
07 08 09 10	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples	13 18 19 23
07 08 09 10 11	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples IBM Spectrum Storage – the Bedrock for HPC and AI	13 18 19 23 26
07 08 09 10 11 12	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples IBM Spectrum Storage – the Bedrock for HPC and AI Learn more	13 18 19 23 26 27
07 08 09 10 11 12 13	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples IBM Spectrum Storage – the Bedrock for HPC and AI Learn more IBM Legal Disclosures	13 18 19 23 26 27 27

IBM Spectrum Storage

07 Storage is Critical to Accelerate Complex Iterative AI Workflows

high-volume, heterogeneous data by resolving data management problems and allowing users to easily search, discover, and understand data relationships to get the most value.

ETL/Data Processing: Extract, transform, load (ETL) is the process to copy data from one or more sources to a target system which holds the data in another way or context from the source(s). ETL makes it possible to extract data from multiple siloed source systems, enforce data quality and consistency standards, integrate and deliver data so that developers can build applications and end users can analyze and get insights. SSDs and Hybrid systems are generally used for this operation.

Governance: As Analytics and AI are used more to answer important questions, the need to govern all IT assets (data, applications, processes, infrastructure, etc.) and ensure security, privacy and regulatory compliance increases. In fact, recent research indicates that by 2022, 65% of enterprises will task CIOs to transform and modernize governance policies to seize the opportunities and confront new risks posed by AI, Machine Learning, and data privacy and ethics.¹⁷ Every organization must constantly check for good data quality from source systems and prevent unethical uses, inaccurate outcomes and prejudices perpetuated by faulty data and biased algorithms.

Training requires a high-performance storage tier that can support these highly parallel

3. Analyze Data/Infuse AI by training and developing AI applications using an experimental and iterative approach. Data scientists conduct many experiments with several algorithms to arrive at a practical training model which expresses the relationships between various types of data. Developers use these models to build new applications or to feed them the appropriate data from the AI application. This training model is tested and updated regularly even when in production and especially when new data is gathered to ensure that business objectives and inference/prediction accuracy goals are being met.



01	Executive Summary	03
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04
03	A Common Infrastructure Enables HPC, Analytics and AI Convergence	06
04	High Value Game-Changing AI/ML/DL Use Cases Emerging	08
05	But a High Performance Information Architecture (IA) is Needed	09
06	AI Data Challenges, Storage Requirements and Solutions	10
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13
07 08	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI	13 18
07 08 09	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI	13 18 19
07 08 09 10	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples	13 18 19 23
07 08 09 10 11	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples IBM Spectrum Storage – the Bedrock for HPC and AI	13 18 19 23 26
07 08 09 10 11 12	Storage is Critical to Accelerate Complex Iterative AI Workflows IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples IBM Spectrum Storage – the Bedrock for HPC and AI Learn more	13 18 19 23 26 27

IBM Spectrum Storage

07 Storage is Critical to Accelerate Complex Iterative AI Workflows

processes with extremely high throughput and low latency. Different storage systems are required for Hadoop or Spark data lakes, ML/DL and archiving.

Hadoop/Spark data lakes: Throughput-oriented performance and capacity are the main requirements. Hadoop/Spark data lakes enable organizations to scale to massive volumes of data and provide additional fault tolerance. It supports a variety of different event schemas and customers can change the attributes of an event without impacting the analytics database. Another advantage is tools such as Spark SQL can be used to work with the data lake directly. This approach scales to a massive amount of data and provides a sound solution for long-running batch queries. Hard Disk Drives (HDD) are the preferred storage medium for this stage of operation.

ML/DL: Requires high-throughput low latency with small, random I/O. Retraining of models with inference may not require as much throughput, but still demands extremely low latency. SSD and Non-Volatile Memory Express (NVMe) technology can help lower storage latency.

Archiving: Requires a highly scalable capacity tier for active archive data that is throughput oriented, large I/O, streaming and sequential writes to HDD/Cloud. Migrating cold data to tape can further reduce storage costs.

IBM Spectrum Storage delivers the fastest path to AI from ingest to insights.

x517B62 x63B63B ×1004AF **xE1BAAC** x1A2F1E x867DE8 kD1297E x02E1F4 xC9F3AB k978FC7 x4384E4 986732 k1DCE50 x713AE2 x4ED8C8 x47EDDD k5F5EF0 k3ACBA2 x853C49 (39CA6A (E17C2B 301567 k9AA3D0 (997E5D k3870B2



DEFO

6CA

COF

ADA:

4ER

25C2

790

01	Executive Summary	03
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04
03	A Common Infrastructure Enables HPC, Analytics and AI Convergence	06
04	High Value Game-Changing AI/ML/DL Use Cases Emerging	08
05	But a High Performance Information Architecture (IA) is Needed	09
06	AI Data Challenges, Storage Requirements and Solutions	10
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13
08	IBM Spectrum Storage Accelerates AI	18
09	Description of Key IBM Spectrum Storage Solutions for HPC and AI	19
10	Client Examples	23
11	IBM Spectrum Storage – the Bedrock for HPC and AI	26
12	Learn more	27
13	IBM Legal Disclosures	27



IBM Spectrum Storage

08 IBM Spectrum Storage Accelerates AI

IBM Spectrum Storage is a comprehensive SDS portfolio that helps affordably manage and integrate all types of data in a hybrid on-premises and/or multicloud environment with **parallel** features that increase performance and business agility. Already proven in HPC, IBM Spectrum Storage comes with licensing options that provide unique differentiation and value at every stage of the AI workflow from ingest to insights (Figure 6).

> uctur d/Uns Structur Data

Figure 6: IBM Spectrum Storage Solutions Optimize the Workflow from Ingest to Insights Customers can build, optimize and deploy their workflows with IBM Spectru m Storage capabilities in several ways: Integrated into IBM storage systems: IBM FlashSystem, IBM Storwize family or IBM Elastic Storage Server (ESS) 1. 2. Implemented as software-only solutions running on servers that can be any server of the customer's choice and not limited

- to IBM
- are detailed next.



3. Deployed in the cloud for trouble-free management and maintenance. Key members of this portfolio to accelerate HPC/AI



01	Executive Summary	03
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04
03	A Common Infrastructure Enables HPC, Analytics and AI Convergence	06
04	High Value Game-Changing AI/ML/DL Use Cases Emerging	08
05	But a High Performance Information Architecture (IA) is Needed	09
06	AI Data Challenges, Storage Requirements and Solutions	10
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13
08	IBM Spectrum Storage Accelerates AI	18
09	Description of Key IBM Spectrum Storage Solutions for HPC and AI	19
10	Client Examples	23
11	IBM Spectrum Storage – the Bedrock for HPC and AI	26
12	Learn more	27
13	IBM Legal Disclosures	27
==		

09 Description of Key IBM Spectrum Storage Solutions for HPC and AI

Together, the IBM Spectrum Storage portfolio provides a The high-level functions, features and benefits of each of these high-performance infrastructure, advanced data/storage key offerings are: management, data protection and other services for HPC/AI:

- scalable common data lakes.
- powered by IBM Spectrum Scale.
- 3. vast amounts of data efficiently and securely.
- data to get better quality insights faster.

PAGE 19

1. IBM Spectrum Scale – a high-performance clustered file system starting from the edge for ultra-fast ingest and very fast access to extract and retrieve data as and when needed for processing, and for high performance file storage and

2. IBM Elastic Storage Server (ESS) – a variety of servers

IBM Cloud Object Storage at the edge for global ingest and geographically dispersed repositories, and for storing

4. IBM Spectrum Discover – a modern metadata management software that provides data insight at petabyte scale for unstructured data. IBM Spectrum Discover easily connects to IBM Cloud Object Storage and IBM Spectrum Scale to rapidly ingest, consolidate and index metadata for billions of files and objects. This helps winnow appropriate/relevant

5. IBM Spectrum Archive which enables direct file access to data stored on tape for very affordable archives.



Figure 7: High Level Architecture of IBM Spectrum Scale and IBM ESS Multisite Deployment

IBM Spectrum Scale is a high-performance clustered file system that supports many protocols and provides concurrent highspeed file access to applications executing on multiple nodes of clusters (Figure 7). Transparent cloud tiering enables nondisruptive, intelligent policy-based migration of data between flash, disk, tape, and cloud storage tiers, allowing users to manage and access both on-premises and multicloud data via a unified view. Key benefits include:

• Extreme scalability with a no-bottleneck architecture to scale performance for extreme throughput and lowlatency access typically not provided by network-attached storage









01	Executive Summary	03
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04
03	A Common Infrastructure Enables HPC, Analytics and AI Convergence	06
04	High Value Game-Changing AI/ML/DL Use Cases Emerging	08
05	But a High Performance Information Architecture (IA) is Needed	09
06	AI Data Challenges, Storage Requirements and Solutions	10
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13
80	IBM Spectrum Storage Accelerates AI	18
09	Description of Key IBM Spectrum Storage Solutions for HPC and AI	19
10	Client Examples	23
11	IBM Spectrum Storage – the Bedrock for HPC and AI	26
12	Learn more	27
13	IBM Legal Disclosures	27
T		

▙▟▟▋₹▙▖

PAGE 20

IBM Spectrum Storage

09 Description of Key IBM Spectrum Storage Solutions for HPC and AI

- business and regulatory requirements.

IBM Cloud Object Storage (COS) provides highly affordable onpremises and cloud-based object storage. It significantly reduces storage costs and provides continuous availability, reliability, protection and fast access to data assets through an innovative any-to-any-to-any parallel system architecture (Figure 8).

COS first encrypts all data and then slices them into segments **IBM Spectrum Discover** is modern metadata management for distributing to storage nodes – Slicestor[™]. There is no single software (Figure 9) that provides data insight for exabyte-scale point of failure and any application can access any Accesser[™] unstructured storage. It easily connects to multiple file and (through an IP address) and the access spreads the data to any object storage systems both on-premises and in the cloud to Slicestor. Most alternative object storage systems do not support rapidly ingest, consolidate and index metadata for billions of an access layer for throughput expansion. files and objects, providing a rich metadata layer on top of these storage sources. This metadata enables data scientists, storage Transmission and storage of data are inherently secure. No administrators and data stewards to efficiently manage, classify complete copy of the data resides in any single storage node, and gain insights from massive amounts of unstructured data and only a subset of nodes needs to be available in order and:

to fully retrieve the data. This method eliminates the high overhead associated with traditional RAID-based storage and

infrastructure while automatically moving file and object data to the optimal storage tier as quickly as possible

• Global collaboration with data-anywhere access that spans storage and locations to accelerate applications across the data center or around the world

• Data integrity and security through authentication, encryption, security and replication options to meet

• Data aware intelligence to grow and share storage the complexity of managing mirrors, replication and disaster recovery required in a traditional data center.



Figure 8: IBM Cloud Object Storage – Architected for High Throughput and Concurrency







01	Executive Summary	03
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04
03	A Common Infrastructure Enables HPC, Analytics and AI Convergence	06
04	High Value Game-Changing AI/ML/DL Use Cases Emerging	08
05	But a High Performance Information Architecture (IA) is Needed	09
06	AI Data Challenges, Storage Requirements and Solutions	10
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13
80	IBM Spectrum Storage Accelerates AI	18
09	Description of Key IBM Spectrum Storage Solutions for HPC and AI	19
10	Client Examples	23
11	IBM Spectrum Storage – the Bedrock for HPC and AI	26
12	Learn more	27
13	IBM Legal Disclosures	27

09 Description of Key IBM Spectrum Storage Solutions for HPC and AI

- curation.
- Increase productivity and enable comprehensive real-time insight by combining system metadata with custom tags to increase storage administration and data consumer productivity with dashboard and customizable reporting.



PAGE 21

• Save time and automate cataloging of unstructured data by capturing metadata as it is created. It supports multiple file and object storage systems from IBM and others. Fast searching enables rapid discovery of data assets and content-based tagging allows systematic metadata

• Integrate easily with security (automatically identify and classify sensitive data) and leverage extensibility using simple SQL query commands, the Action Agent API, custom tags and policy-based workflows to orchestrate deeper content inspection and activate data in HPC, Analytics and AI workflows.

IBM Spectrum Archive enables users to automatically move infrequently accessed data from disk to tape (Figure 10) for lower costs while retaining ease of use and avoiding the need for proprietary tape applications. This functionality is provided by the IBM Linear Tape File System. IBM Spectrum Archive works in conjunction with Linear Tape-Open (LTO) technology to simplify data access from an IBM tape cartridge. It allows users to run any application designed for disk files against tape data without worrying that the data is physically stored on tape. Key benefits include :

- Easy to use and integrated with Elastic Storage Server/ IBM Spectrum Scale. Access and manage files on tape as easily as if they were on hard disk with a Graphical User Interface, drag and drop files to and from tape without requiring device-specific software and share file data across platforms.
- Figure 9: IBM Spectrum Discover Expedite Time to Locate Data and Gain Faster Insights
- Fast access to data and display directories and files without the need for proprietary backup software.

01	Executive Summary	03
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04
03	A Common Infrastructure Enables HPC, Analytics and AI Convergence	06
04	High Value Game-Changing AI/ML/DL Use Cases Emerging	08
05	But a High Performance Information Architecture (IA) is Needed	09
06	AI Data Challenges, Storage Requirements and Solutions	10
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13
08	IBM Spectrum Storage Accelerates AI	18
08	IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI	18 19
08 09 10	IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples	18 19 23
08 09 10 11	IBM Spectrum Storage Accelerates AIDescription of Key IBM Spectrum Storage Solutions for HPC and AIClient ExamplesIBM Spectrum Storage – the Bedrock for HPC and AI	18 19 23 26
08 09 10 11 12	IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples IBM Spectrum Storage – the Bedrock for HPC and AI Learn more	18 19 23 26 27
08 09 10 11 12 13	IBM Spectrum Storage Accelerates AI Description of Key IBM Spectrum Storage Solutions for HPC and AI Client Examples IBM Spectrum Storage – the Bedrock for HPC and AI Learn more IBM Legal Disclosures	18 19 23 26 27 27

09 Description of Key IBM Spectrum Storage Solutions for HPC and AI

- Simple file portability with flexible data sharing and easily exchange tapes between different operating systems and software.
- data availability.



• Optimize tape resources and reduce restore times by creating optional storage tiers and enable optimal utilization of tape resources with policy-based migration.

does not need the access performance of primary disk.

• Lower costs and reduce the storage expense for data that The IBM Spectrum Storage portfolio is the bedrock of the Information Architecture that helps clients worldwide unlock and accelerate business value from HPC/AI.

• Gain high availability and expand capacity without impacting

Figure 10: IBM Spectrum Archive – Direct, Intuitive and Graphical Access to Data Stored in IBM Tape





01	Executive Summary	03
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04
03	A Common Infrastructure Enables H Analytics and AI Convergence	PC, 06
04	High Value Game-Changing AI/ML/D Use Cases Emerging	DL 08
05	But a High Performance Informati Architecture (IA) is Needed	i on 09
06	AI Data Challenges, Storage Requirements and Solutions	10
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13
08	IBM Spectrum Storage Accelerates	AI 18
09	Description of Key IBM Spectrum Storage Solutions for HPC and AI	19
10	Client Examples	23
11	IBM Spectrum Storage – the Bedroc for HPC and AI	k 26
12	Learn more	27
13	IBM Legal Disclosures	27
		PAGE 23

10 Client Examples

These examples demonstrate how the IBM Spectrum Storage portfolio is enabling customers worldwide meet the challenges of managing data and deliver substantial business value across various industries/applications :

UNIVERSITY OF BIRMINGHAM

Driving innovative research forward by taking control of data

Challenges

- Maintain reputation as a premier research institution by ensuring that data is always available • to a growing number of users running increasingly complex simulations
- Address data storage challenges •
 - Meet rising growth in demand 0
 - Reduce risk of data loss 0
 - Compliance with data protection rules 0
 - Foster ground-breaking research 0

Solution/Results

- •
- the information they require.

Business Impact

•

IBM Spectrum Scale and IBM Spectrum Protect increase transparency around data location and who accesses it, and improves its mobility within a diverse IT environment Multiple layers of protection against data loss, ensuring that researchers can always access

Greater control over data enables maximizing value while complying with data protection regulations at low cost and without disruption



01	Executive Summary	03
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04
03	A Common Infrastructure Enables HPC, Analytics and AI Convergence	06
04	High Value Game-Changing AI/ML/DL Use Cases Emerging	08
05	But a High Performance Information Architecture (IA) is Needed	09
06	AI Data Challenges, Storage Requirements and Solutions	10
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13
08	IBM Spectrum Storage Accelerates AI	18
09	Description of Key IBM Spectrum Storage Solutions for HPC and AI	19
10	Client Examples	23
11	IBM Spectrum Storage – the Bedrock for HPC and AI	26
12	Learn more	27
13	IBM Legal Disclosures	27

10 Client Examples

UNIVERSITY OF PITTSBURGH MEDICAL CENTER (UPMC)

data.

Challenges

- To deliver life-changing care, clinical and research teams must have fast, reliable access to data. Need to shorten storage response times and increase availability
- Greater demand for storage as sophisticated clinical imaging systems generate increasingly large files and patient records get progressively more detailed

Solution/Results

- IBM Spectrum Storage Suite for unlimited access to a broad range of SDS solutions IBM Spectrum Scale supports synchronous replication in a cluster across two data centers, and provides a global namespace that enables multiple users to access data simultaneously IBM FlashSystem for deduplication and compression features
- •
- The first FlashSystem A9000R currently deployed is 90 percent utilized and still provides a data reduction ratio of 5:1. The next model is provisioned to over 2 PB and can provide over 6:1 data reduction

Business Impact

PAGE 24

Enabling quality patient care and improving user experience by delivering rapid access to vital

2400% growth in storage under management, supported with zero increase in IT headcount 50% TCO saving for primary storage and 5:1 data reduction ratio maximize efficiency 50% shorter patient record seek times and 99.9% decrease in response times enhance care



01	Executive Summary	03
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04
03	A Common Infrastructure Enables H Analytics and AI Convergence	PC, 06
04	High Value Game-Changing AI/ML/D Use Cases Emerging	L 08
05	But a High Performance Informati Architecture (IA) is Needed	on 09
06	AI Data Challenges, Storage Requirements and Solutions	10
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13
08	IBM Spectrum Storage Accelerates A	AI 18
09	Description of Key IBM Spectrum Storage Solutions for HPC and AI	19
10	Client Examples	23
11	IBM Spectrum Storage – the Bedroc for HPC and AI	k 26
12	Learn more	27
13	IBM Legal Disclosures	27
		PAGE 25

10 Client Examples

L7 INFORMATICS

Building a high-performance Genomic Cloud to support ground-breaking research

Challenges

- struggle to process vast volumes of data
- •
- •

Solution/Results

- analysis and long-term storage
- flexibility, simplicity, and scalability.

Business Impact

- •
- •
- •

PAGE 25

To advance the understanding of the human genome, scientists at research institutions

A single instrument run can produce over 100 GB of data

Transferring this data from lab equipment to short and long-term storage is hard

IBM Spectrum Scale provides high-performance data storage that can be scaled quickly and easily. Built-in tiering capabilities allow a lot of flexibility in how the data can be moved around, enabling users to seamlessly migrate data from lab instruments up to the cloud for

Developed a genomics-specific data management solution that offers cost-effectiveness,

96% reduction in the runtime of a standard genome analysis pipeline 1/3 the price of using commodity solutions to perform the same work at scale 2 weeks from conceptual design to a fully functional environment in the cloud



01	Executive Summary	03	
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04	
03	A Common Infrastructure Enables HPC, Analytics and AI Convergence	06	
04	High Value Game-Changing AI/ML/DL Use Cases Emerging	08	
05	But a High Performance Information Architecture (IA) is Needed	09	
06	AI Data Challenges, Storage Requirements and Solutions	10	
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13	
08	IBM Spectrum Storage Accelerates AI	18	
09	Description of Key IBM Spectrum Storage Solutions for HPC and AI	19	
10	Client Examples	23	
11	IBM Spectrum Storage – the Bedrock for HPC and AI	26	
12	Learn more	27	
13	IBM Legal Disclosures	27	
TEL			

IBM Spectrum Storage

11 IBM Spectrum Storage – the Bedrock for HPC and AI

As HPC and Analytics grow and converge, many clients are implementing high-value AI/ML/DL use cases in several industries. For this they need a reliable partner with deep HPC, Analytics and AI/ML/DL expertise to overcome the many challenges with deploying and scaling AI across their enterprise.

IBM is this reliable partner and provides comprehensive solutions to create a robust and scalable Information Architecture (IA). This IA is needed to operationalize AI and maximize its value, especially as AI workflows become more data and compute intensive.

IBM Spectrum Storage is the bedrock of this IA. With this industry-leading Software-defined Storage (SDS) portfolio (on-premises or multicloud), clients can collect/ingest, organize and analyze data and infuse AI throughout their organization with trust and transparency.

Already a proven leader in HPC, IBM Spectrum Storage comes with flexible licensing options that provide unique differentiation, value, optimization and investment protection at every stage of the AI workflow.

You can depend on IBM Spectrum Storage to accelerate HPC/AI deployments from ingest to insights to maximize business value.



01	Executive Summary	03		
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04		
03	A Common Infrastructure Enables HPC, Analytics and AI Convergence	06		
04	High Value Game-Changing AI/ML/DL Use Cases Emerging	08		
05	But a High Performance Information Architecture (IA) is Needed	09		
06	AI Data Challenges, Storage Requirements and Solutions	10		
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13		
08	IBM Spectrum Storage Accelerates AI	18		
09	Description of Key IBM Spectrum Storage Solutions for HPC and AI	19		
10	Client Examples	23		
11	IBM Spectrum Storage – the Bedrock for HPC and AI	26		
12	Learn more	27		
13	IBM Legal Disclosures	27		
TRM				

IBM Spectrum Storage

12 Learn More

To learn more about IBM S your IBM representative.

13 IBM Legal Disclosures

© IBM Corporation 2020

IBM, the IBM logo, ibm.com and Watson are trademarks or registered trademarks of International Business Machines Corp., other countries, or both. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml

Other company, product and service names may be trademarks or service marks of others. This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided. It is the user's responsibility to evaluate and verify the operation of any other products or programs with IBM products and programs. The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation. The performance data discussed herein is presented as derived under specific operating conditions. Actual results may vary.

To learn more about IBM Spectrum Computing, click ibm.com/spectrum-computing or contact



01	Executive Summary	03
02	Storage Key to Drive Value as HPC, Analytics and AI Converge	04
03	A Common Infrastructure Enables HPC, Analytics and AI Convergence	06
04	High Value Game-Changing AI/ML/DL Use Cases Emerging	08
05	But a High Performance Information Architecture (IA) is Needed	09
06	AI Data Challenges, Storage Requirements and Solutions	10
07	Storage is Critical to Accelerate Complex Iterative AI Workflows	13
80	IBM Spectrum Storage Accelerates AI	18
09	Description of Key IBM Spectrum Storage Solutions for HPC and AI	19
10	Client Examples	23
11	IBM Spectrum Storage – the Bedrock for HPC and AI	26
12	Learn more	27
13	IBM Legal Disclosures	27



IBM Spectrum Storage

13 IBM Legal Disclosures

Sources

- Tony Hey, "Convergence of HPC, Big Data and Machine Learning: A Science and Engineering Perspective", https://www.nitrd.gov/ 1. nitrdgroups/images/c/cd/HEC-BD-Tony-Hey-10302018.pdf
- https://eiuperspectives.economist.com/sites/default/files/EIU_Microsoft%20-%20Intelligent%20Economies_AI%27s%20 2. transformation%20of%20industries%20and%20society.pdf
- https://futureiot.tech/gartner-ai-adoption-growing-despite-skills-shortage/ 3.
- https://www.idc.com/getdoc.jsp?containerId=US45334719 4.
- https://www.enterpriseai.news/2019/11/21/hyperion-ai-driven-hpc-industry-continues-to-push-growth-projections/ 5.
- https://www.networkworld.com/article/3325397/idc-expect-175-zettabytes-of-data-worldwide-by-2025.html 6.
- IDC: Worldwide Storage for Cognitive/AI Workloads Forecast, 2018–2022, April 2018 https://www.idc.com/getdoc. 7. jsp?containerId=US43707918
- 8. Management Review, 2017.
- NewVantage Partners, "Big Data and AI Executive Survey 2019 Executive Summary of Findings", 2019. 9.
- 10. https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Notes%20from%20the%20 frontier%20Modeling%20the%20impact%20of%20AI%20on%20the%20world%20economy/MGI-Notes-from-the-AI-frontier-Modeling-the-impact-of-AI-on-the-world-economy-September-2018.ashx
- 11. https://blogs.oracle.com/bigdata/7-machine-learning-best-practices
- https://www.newgenapps.com/blog/best-practices-machine-learning-models-applications 12.
- CrowdFlower, "2016 Data Science Report" 14.
- 13. Ritu Jyoti, "Accelerate and Operationalize AI Deployments Using AI Optimized Infrastructure", IDC Technology Spotlight, June 2018.
- 15. https://www.fudzilla.com/news/45867-ibm-ceo-80-percent-of-data-is-not-searchable
- https://www.gartner.com/smarterwithgartner/why-organizations-choose-a-multicloud-strategy/ 16.
- https://www.idc.com/getdoc.jsp?containerId=prUS44420918 17.

Sam Ransbotham, David Kiron, Philipp Gerbert, and Martin Reeves, "Reshaping Business with Artificial Intelligence", MIT Sloan

