PROWESS

ACCELERATING YOUR DATA ANALYTICS JOURNEY

Dell EMC and Intel can help optimize your business to compete on data, analytics, and AI across the whole data life cycle.

Your organization is awash in data, arriving from different sources, in different formats, and destined for different uses. And most data is never analyzed or used. The challenges you face include the growing volume of data, its dispersion across different systems and, ultimately, the ability of people to find the data they need and use it to their competitive advantage. You aspire to take full advantage of all that data with sophisticated analytics and artificial intelligence (AI) techniques to extract the most insight from the data.

> "By 2020, enterprise execs need their analytics to be 75% faster and twice as sophisticated."
>
> — Gartner[1]

But how do you get there from here? This paper offers insights and suggestions to help you on your journey toward optimizing data for analytics and AI across the whole data life cycle.

Take a Holistic View of Your Data Life Cycle

The first step of your journey might be to take a step back and look at your whole data life cycle, from creation to archiving. Your data continuum might look something like Figure 1.

Analytics and AI acceleration are not just about optimizing data at one stage of the data life cycle, but are rather about looking for opportunities to optimize across the continuum.

Don't just try to perform analytics on the data as you find it. Take control of how the data is managed and optimized every step of the way. That optimization begins out at the edge, where data is created and ingested into your system.

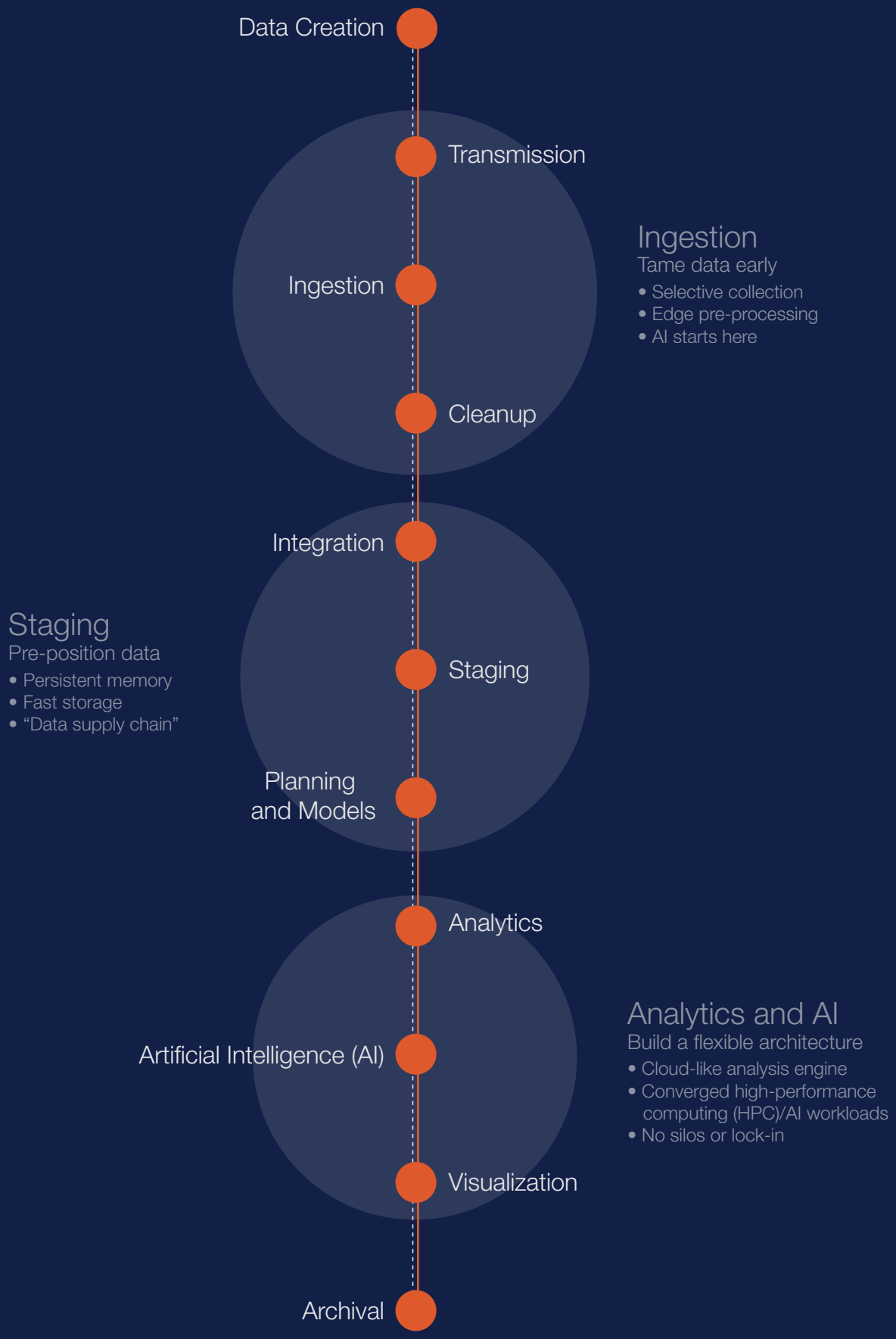

Figure 1. Optimizing data across the typical data life cycle

Focus on Optimizing Data Early in Its Life Cycle

Your infrastructure struggles to keep pace with volume and complexity of data. That's why it's important to tame the data starting at its source. You want to take control of data, compress it, shape it, and process it whenever possible, before ingestion into your system. Let's look at a few examples to see what this can mean.

One good example is when you are collecting video data from remote cameras. Decisions early in the process about frame rate and resolution at the front end can have huge implications for the amount of data to be transmitted, stored, protected, analyzed, and archived during its life cycle. Moreover, any pre-processing that can be done with edge computing devices—classification, compression, and selective saving, for example—can dramatically reduce the volume and increase the value of the streaming data that's ingested into the data center.

Another example is Internet of Things (IoT) data, which is one of the major new sources of data creation that many data centers need to deal with. Streams from hundreds or thousands of sensors and devices can be preprocessed at the intelligent edge, with only the important information transmitted for analysis. For instance, a data stream of temperature readings in a freezer collected every 60 seconds does not need to be transmitted and stored in full if what's important is only a record of *changes* in temperature.

Finally, consider the example of the financial-services industry, where streaming data arrives from multiple sources in multiple formats and is prone to be "bursty," rather than steady over time. Here, pre-processing work that classifies data by type can make bursts of data easier to ingest into the data center. This kind of pre-processing can even use AI techniques to learn pattern matching.

Dell EMC and Intel offer products and technologies that can help tame incoming data at the source:

- Dell EMC™ IoT Connected bundles to reduce the risks associated with IoT
- Dell ruggedized equipment including servers for pre-processing data in harsh field conditions
- Intel® QuickAssist Technology to add significant performance increases to compression and encryption of data
- Intel Atom® processors and Intel® Xeon® D processors for lighter intelligent-edge workloads
- Intel® Arria® 10 FPGA to tame bursty data streams with pattern-matching intelligence
- Intel® Movidius™ Vision Processing Units (VPUs) for the demanding workloads of modern computer-vision and AI applications at ultra-low power
- Intel® Ethernet 700 Series network interfaces to move data faster

Rethink Staging

Your current methods of staging data might be fairly traditional. Perhaps your storage media is divided into two tiers, with more important "warm" data stored on faster solid state (SSD) drives and more voluminous "cold" data relegated to slower physical media, such as spinning disks or tape. But wherever your data is stored, it must be loaded into memory for processing. Storage and memory are worlds apart in terms of cost and performance. Memory is fast, expensive, and volatile—gone if power is lost or the system restarts.

It's time to start thinking differently about memory and storage, because Intel® Optane™ technology bridges the gap between the two. On the storage side, Intel Optane DC SSDs raise the speed limit on fast read and write operations. Meanwhile on the memory side, Intel Optane DC persistent memory changes the rules of the game by putting large amounts of non-volatile data right on the memory bus by the processor. It's no longer a binary world; now storage need not be slow, and memory need not be volatile.

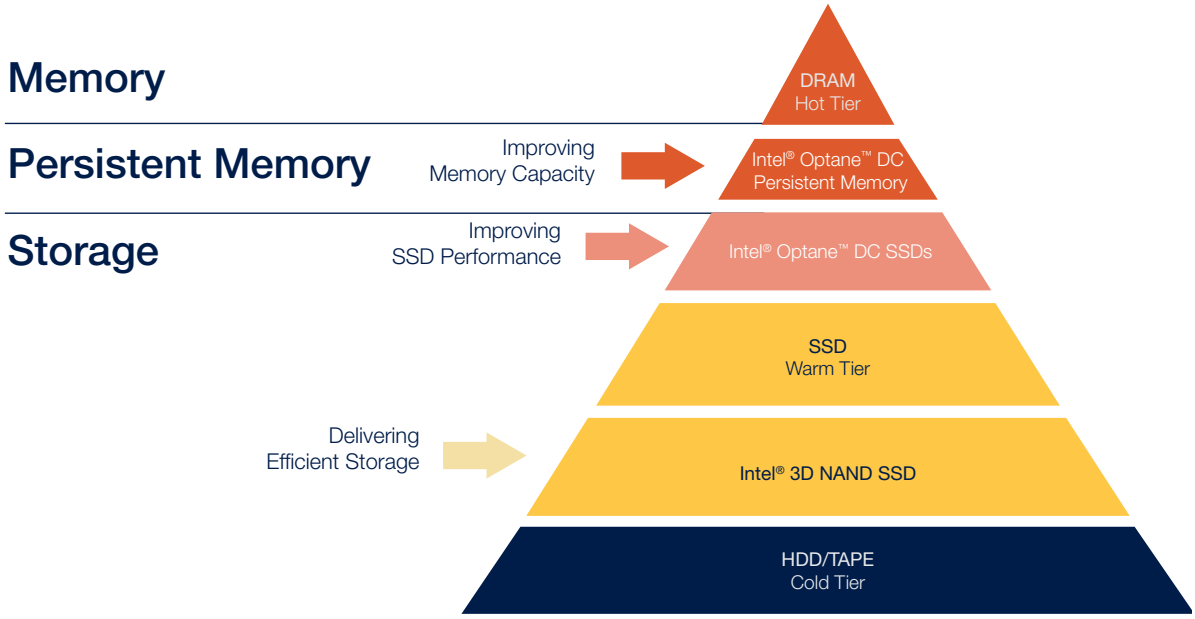

Figure 2. Bridging the gap between memory and storage

This new architecture calls for new ways of thinking about staging data. Rather than storing, think about *pre-positioning* data. It's like a supply chain for retail goods: everything is not kept in one giant warehouse until needed; rather, goods are distributed to multiple locations in anticipation of being needed in the vicinity. That's the way to think about data staging now. Think beyond hot and cold data tiers. Think about positioning data in the right place, in the right format, on the right media, for the right person or process to use it efficiently. The primary storage location for your data might now be right in the persistent memory of the server that will be using it.

Create a Scalable, Flexible Infrastructure for Analytics and AI

One challenge in planning for advanced analytics and AI is the diversity and rapid evolution of tools and frameworks. Different projects end up using different tools, and frameworks are often not compatible. Data tends to be optimized for a particular task, fragmented by department, and handled by specialized systems. As you build your platform for the future, it's important not to be wedded to a particular framework or algorithm, or to get stuck with hardware that can only do one thing.

Just as data silos need to be broken down and consolidated, so too with infrastructure and analytics silos. Your goal should be to build cloud-like, multipurpose, flexible analytics engines that can handle most any kind of workload.

One common form of fragmentation is between data analytics done on high-performance computing (HPC) infrastructures based on CPUs, and AI and machine learning (ML) projects carried out on special-purpose GPU-based nodes. Fortunately, HPC capabilities in the area of AI have been improving rapidly, allowing companies to move toward running AI workloads alongside their other workloads in their HPC data centers. This convergence of analytics and AI workloads is not only less complex than maintaining separate dedicated systems, but it also allows for better resource utilization by combining workloads in the same HPC infrastructure.

Dell EMC offers a variety of scalable Ready Solutions to help you build out your HPC infrastructure for analytics and AI:

- **Ready Solutions for AI, Machine Learning with Hadoop**. If you're already using Apache™ Hadoop® for big data storage and processing, you can achieve fast implementation of ML with a hardware and software stack optimized for performance.[2]
- **Ready Solutions for AI, Deep Learning with Intel**. If you want to get up and running quickly with deep learning (DL) in a CPU-based infrastructure, this recently released solution includes all the hardware, software, and services you need.[3]
- **Ready Solutions for HPC**. Dell EMC experts can help you design an optimized solution of high-performance technologies to help you succeed: workstations, servers, networking, storage, software, solutions, and services.[4]

Prepare for AI and ML Everywhere.

AI and ML are often seen as an advanced evolution of analytics beyond descriptive to predictive and prescriptive. This is true, but can lead to the false impression that AI exists only at the tail end of the data life cycle. In fact, AI emerges all across the data continuum and can be a valuable tool for optimizing data in your business.

When you use field-programmable gate arrays (FPGAs) in edge computing to pre-process data before ingestion using pattern matching, that is a form of AI. When you use smart network interface controllers (NICs) in your network that learn better ways to allocate packets as the data moves, that is also a form of AI. When your network security learns to identify suspicious patterns that might be malware or an intrusion, that too is a form of AI. The point is that AI is not just a destination your data will reach at the end of its life cycle, it's a technique you can use to optimize your data every step of the way.

Dell EMC and Intel are your strategic partners for developing your data strategy. Turn to them for technologies, services, and a robust ecosystem of partners in each stage of the data life cycle—from creation and ingestion to analysis and insights.

Intel® Technologies for Analytics and AI

When you build your HPC infrastructure from the Dell EMC portfolio, you can also take advantage of these Intel technologies:

- **2nd Generation Intel® Xeon® Scalable processors with Intel® Deep Learning Boost (Intel® DL Boost)**. Intel Xeon Scalable processors are built specifically for the flexibility to run complex AI workloads on the same hardware as your existing workloads. Intel DL Boost is a group of acceleration features, introduced in 2nd Generation Intel Xeon Scalable processors, that provide significant performance increases to inference applications built using leading DL frameworks such as PyTorch, TensorFlow™, Apache MXNet™, PaddlePaddle®, and Caffe.[5]
- **Intel® Math Kernel Library (Intel® MKL)**. Easily accelerate math-intensive workloads, including ML.[6]
- **Intel® Machine Learning Scaling Library (Intel® MLSL)**. Efficient implementation of communication patterns used in DL, optimized to drive scalability.
- **Intel® Intelligent Storage Acceleration Library (Intel® ISA-L)**. Improve bandwidth and decrease disk usage in storage subsystems.
- **Nauta**. A distributed DL platform that makes use of Kubernetes® and Docker® technologies.[7]
- **Intel Optimized Frameworks**. BigDL, Caffe, and TensorFlow have been optimized by Intel for 2nd Generation Intel Xeon Scalable processors.[8] Intel also has an optimized distribution for Python®.[9]

Beyond providing hardware and software technologies, Intel's software team of more than 10,000 people and its ecosystem of partners can also provide you with expertise and deep engagement to fully optimize your system.

Learn More

Dell EMC Ready Solutions: www.dellemc.com/en-us/solutions/ready/

2nd Generation Intel Xeon Scalable processors: www.intel.com/content/www/us/en/products/processors/xeon/scalable.html

Intel Optane DC persistent memory: www.intel.com/content/www/us/en/architecture-and-technology/optane-dc-persistent-memory.html

For more info on Intel enterprise solutions visit: www.intel.com/yourdataonintel and www.intel.ai/

Dell EMC is a member of the **Intel® AI Builders Program**, an ecosystem of industry leading independent software vendors (ISVs), system integrators (SIs), original equipment manufacturers (OEMs), and enterprise end users, which have a shared mission to accelerate the adoption of artificial intelligence across Intel® platforms.

[1] PwC. "PwC's Global Data and Analytics Survey 2016." July 2016. www.pwc.com/us/en/services/consulting/analytics/big-decision-survey.html.

[2] Forrester. "The Total Economic Impact™ Of Dell EMC Ready Solutions For AI, Machine Learning With Hadoop." August 2018. www.dellemc.com/en-ca/collaterals/unauth/analyst-reports/products/ready-solutions/dell_emc_ready_solutions_for_ai_forrester_total_economic_impact_study.pdf.

[3] Enterprise Strategy Group. "Dell EMC Ready Solutions for AI: Deep Learning with Intel." April 2019. www.dellemc.com/en-us/collaterals/unauth/analyst-reports/products/ready-solutions/esg-technical-review-deep-learning-with-intel-april-2019.pdf.

[4] Dell EMC. "Solutions for High Performance Computing and AI." www.dellemc.com/en-us/solutions/high-performance-computing/index.htm#scroll=off.

[5] Intel. "Increasing AI Performance and Efficiency with Intel® DL Boost." www.intel.ai/increasing-ai-performance-intel-dlboost.

[6] Intel. "The Fastest and Most-Used Math Library for Intel®-Based Systems." https://software.intel.com/en-us/mkl.

[7] Intel. "Nauta." www.intel.ai/nauta/.

[8] Intel. "Framework Optimization Training." https://software.intel.com/en-us/ai/frameworks.

[9] Intel. "Accelerate Python* Performance." https://software.intel.com/en-us/distribution-for-python.