



Accelerate Insights with a Holistic Analytics Infrastructure

Build a future-ready data ecosystem that empowers fast, complex analytics across the business with Intel

Executive Summary

Businesses today need to make increasingly nuanced and time-sensitive decisions, so it is vital that their IT teams support them to do more with their data, faster. To enable this, it's important for IT teams to think holistically. This means taking into account the entire hardware, software and solutions framework to ensure capabilities are consistent and integrated across the organization. At the hardware and infrastructure level, this means carefully considering the choice of processors, memory capacity, storage media, network technologies and cluster architecture that will optimize analytics speed. Meanwhile, it's also important to consider how much improvement can be gained by adding the right library to your software stack and optimizing industry frameworks and operating systems to boost speed and efficiency. Using a fully tuned solution stack from a blueprint, like Intel® Select Solutions, can help accelerate analytics.

Business Challenge: Deeper Insight, More Decisions, Faster

The average enterprise executive today demand more of their data. They want to be able to draw deeper insights, faster and with more accuracy, and their expectations are rising all the time. Advanced analytics is a crucial tool in the IT leader's mission to fulfill these expectations, but just having the capabilities to meet today's needs isn't enough. Research by PwC¹ revealed that by 2020, these executives need their analytics to be 75 percent faster and twice as sophisticated (which PwC defines² as being twice as broad and twice as accurate).

This indicates that the business is fully aware of the potential benefits that analytics can bring, especially when using more advanced techniques like machine learning, natural language processing (NLP) and intelligent agents. These capabilities are key to enabling some of today's more complex business use cases. For example:



Making faster decisions: Scenarios where processing and understanding real-time data is critical are common across multiple industries – from navigational systems for automated driving, to real-time patient monitoring in hospitals.



Solving more complex challenges: It's often necessary to consider multiple layers and sources of information when making a single decision – from large-scale strategic considerations like analyzing market conditions for a new product launch, down to identifying the best solution to offer a customer in your online store.



Doing both at once: Often it's not a case of choosing one or the other, and decisions must be both fast and complex, as is the case in fraud detection and security, for example.

While business leaders are clear on the need to embrace these capabilities within their own organizations, they often struggle to know how to get to this advanced state of analytics competency from where they are today. Shrinking time to insight is a challenge, and they often turn to their peers in the IT team for help. This in turn presents a great opportunity for IT to play a strategic role in the evolution of the business and its future success.

Holistic Analytics Acceleration

Responding to these demands and making the most of the opportunity to positively impact business demands requires a holistic approach to IT, data and analytics strategy.

First, engage with your business users, data scientists and machine learning experts to identify the use cases and applications they need you to support. You can gain more insight on how to approach this stage of the process in the [Intel white paper](#) "Five Steps to Delivering the Data-Driven Business".

With the business need(s) identified, the next priority is to analyze and prepare your data ecosystem and IT solution stack. This should cover four key areas:



Hardware: High-performance compute, storage and network technologies



Accelerators: Acceleration libraries for math, compression, storage and parallelization



Software: Optimized versions of industry frameworks, open source and applications



Templates: Solutions blueprints for rapid deployment



Let's explore each of these areas in more detail.



Hardware: High-performance compute, storage and network technologies

A comprehensive analytics strategy must address the distributed nature of data from the edge to the cloud, and examine the key infrastructure capabilities that are needed to support a broad range of workloads – from traditional analytics and business intelligence (BI) through to real-time streaming analytics and artificial intelligence (AI) techniques like machine learning. It must deliver all this with increasingly stringent levels of data protection, in line with standards like the General Data Protection Regulation (GDPR) in the European Union and other recent and emerging requirements.

Harnessing intelligence quickly from a distributed data model that covers the edge, the cloud and the data center, requires a modern infrastructure that is performance optimized for a broad spectrum of analytics workloads. Intel has developed foundational technologies that are critical to efficiently manage, secure and rapidly harness data for competitive advantage. Three key areas to consider when designing your hardware infrastructure are:

i. Identifying high-performance workload-optimized, world-class silicon for your range of analytics

The hardware you choose for your infrastructure will be a crucial part of your future success. As analytics capabilities and requirements continue to evolve, you'll need to ensure your underpinning technologies have the flexibility and scalability to evolve with them, adding new workloads or use cases to your existing infrastructure as they emerge. It will also need to deliver the performance needed to cope with increasingly large data sets, types and sources, in order to deliver the real-time results that the business and its customers demand.

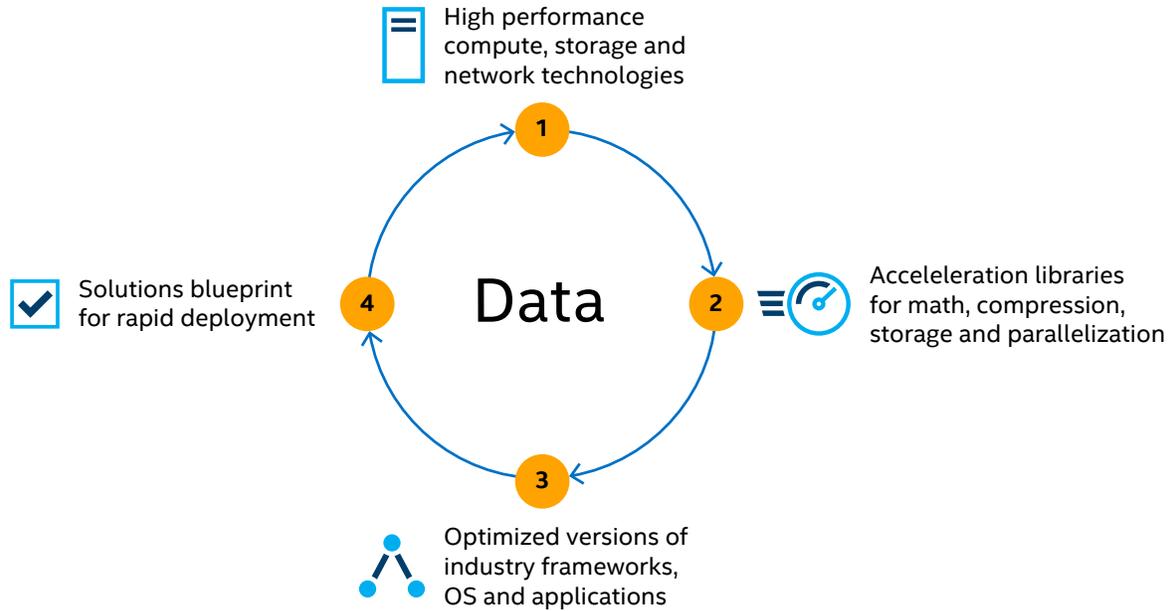


Figure 1: Holistic analytics acceleration requires optimization throughout the solution stack

Intel's technology suite includes the powerful new [Intel® Xeon® Scalable processor](#) for a broad range of analytics workloads. Additional infrastructure technologies include:

- [Intel® Field Programmable Gate Arrays](#) (Intel® FPGAs) offer further acceleration, flexibility and integration as well as helping lower the total cost of ownership (TCO) of your infrastructure. By introducing the ability to re-program hardware to support emerging workloads, these devices help future-proof your advanced analytics investments.
- Technologies optimized for artificial intelligence (AI) workloads, such as [Intel® Nervana™ Neural Network Processor \(NNP\)](#), a new class of hardware that is a purpose-built neural processing unit (NPU) for the most intensive deep-learning training and inference.
- [Intel® Optane™ DC Persistent Memory](#) keeps more data closer to the CPU and retained in-memory during power-down cycles, eliminating the latencies caused by I/O bottlenecks, fetching data from slower SSDs and speeding up restarts.
- [Intel® Ethernet connectivity](#) delivers rapid networking for low-latency performance.
- [Intel® Omni-Path Fabric \(Intel® OP Fabric\)](#) delivers high-performance interconnect for HPC workloads.

ii. Choosing complementary technologies for latency-sensitive, memory-intensive data applications

The core infrastructure can't operate independently. It must integrate smoothly with other technologies that support mission-critical applications, both now and in the future.

Intel offers one of the industry's broadest performance-optimized data center technology portfolio to support an analytics-enabled infrastructure for a range of analytics workloads – from real-time, in-memory SAP HANA* or Oracle Exadata* databases, to streaming analytics programs like Storm* and Flink*, big data Hadoop* deployments, and machine learning applications.

iii. Deciding on a suite of technologies for platform and data protection, encryption and reliability

Keeping data safe as it resides in and moves around your network is critical. This is especially true when it comes to advanced analytics, which uses large volumes of complex data that may often include sensitive information about the company, its employees or its customers. Making sure data is secured and encrypted reliably and in real-time, must be a high priority for every infrastructure manager.

When assessing your organization's security credentials, it pays to consider three key areas:

1. **Securing the platform:** The first step in any security effort must be ensuring the underlying hardware can provide a solid and trusted foundation. You can do this by taking advantages of technologies like Intel® Trusted Execution Technology (Intel® TXT).
2. **Protecting the data:** Encryption is essential to preventing attackers from accessing data while it is in-use, in-flight, or at-rest. This can be achieved with technologies such as Intel® Key Protection Technology (Intel® KPT), a new platform security feature with the Intel Xeon Scalable processor family, which helps secure keys in hardware
3. **Maintaining performance:** While robust security is imperative, delivering it must not be allowed to hinder or slow down the delivery of the services and applications that the business demands. Using technologies like Intel® Advanced Vector Extensions 512 (Intel® AVX-512), a set of instructions that can accelerate performance for compute-intensive workloads, can help avoid the need to compromise.



Accelerators: Acceleration libraries for math, compression, storage and parallelization

Even with the greatest hardware resources in the world, advanced analytics applications can be further enhanced with acceleration technologies and libraries to help deliver results faster. For example, installing the Intel® Math Kernel Library (Intel® MKL) in an Apache Spark* solution can accelerate the Alternating Least Squares (ALS) machine learning algorithm by four times³ without any other hardware or software changes. Alternatively, organizations looking to develop computer vision solutions can use Intel's Open Visual Inference and Neural Network Optimization (OpenVINO™) toolkit to emulate human vision.

Intel's developer-friendly software initiatives are also optimizing the performance of popular machine learning and deep learning developer frameworks on Intel® technology-powered hardware platforms. This enables you to support the requirements of your enterprise developers for advanced analytics projects such as machine learning. Meanwhile, we are continuing to optimize tools, libraries and frameworks such as BigDL*, Spark MLib*, Caffe*, Theano*, Torch*, CNTK*, and Tensor Flow*. Advanced workloads across popular machine learning developer frameworks can also benefit from Intel® software libraries, such as Intel® Data Analytics Acceleration Library (Intel® DAAL) and the Intel® Distribution for Python*.



Software: Optimize versions of industry frameworks, open source and applications

Having access to - and easy compatibility with - industry-standard tools, frameworks and applications can help analytics software developers work smarter and faster, focusing their time on development rather than resolving compatibility issues or reinventing the wheel. Analytics software platforms from industry-leading independent solution vendors (ISVs) and open source technologies are optimized for Intel® architecture to deliver excellent performance for analytics workloads, offering greater choice in the platform to suit unique analytics needs.

Optimizations for Intel architecture across industry-leading analytics software platforms create freedom of choice in terms of solution stack, covering SAP HANA, SAS, Oracle Exadata, Microsoft, and Cloudera, depending on the type of analytics workload in use (for example, traditional, big data, or machine learning). Meanwhile, Intel contributes to leading Open Source analytics platforms and frameworks such as Apache Hadoop* and emerging platforms Apache Spark*, Storm*, Flink*, Shark*, and NoSQL*. This helps ensure that Intel innovations integrate smoothly across the open source solution ecosystem to help minimize software platform lock-in.

Intel's developer-friendly software investments optimize the performance of popular developer frameworks on Intel architecture-based platforms, further supporting enterprise developer requirements for advanced analytics projects like machine learning and deep learning.



Templates: Solutions blueprint for rapid deployment

With all the components in place, the next challenge is determining the precise configuration and combination to achieve optimal performance from your advanced analytics workloads. Intel has worked with leading ecosystem players to develop workload-optimized solutions and reference architectures that meet the holistic infrastructure demands of today's data-dependent organizations. Combining hardware, software and other infrastructure components in proven configurations, Intel® Select Solutions are tested to deliver optimum performance, improved security, reliability, and agility.

[Intel Select Solutions for Microsoft SQL Server*](#) can help minimize the time, cost, and complexity required for an enterprise to evaluate hardware and software integrations for advanced analytics use cases. It is not only optimized to meet current performance needs but also includes technologies to help reduce the complexity of meeting evolving data center needs over time.

Intel has also worked closely with SAP to develop a combined solution to optimize data management and advanced analytics for in-memory databases. By using Intel® Optane™ DC Persistent Memory, the solution enables more data to be kept closer to the CPU, even after a power-down. This helps reduce latency and so enhance real-time use cases like streaming analytics.

Conclusion

With lines of business pushing to adopt or enhance specific analytics use cases, it can be tempting to prioritize these top-of-mind requirements. However, taking a moment to step back and consider your analytics capabilities and future potential holistically may help drive much more business value over time. From enabling analytics at the edge that helps enhance patient care in hospitals, to helping data scientists push the boundaries of their profession by cutting their time cost to develop new algorithms, a wide-ranging analytics strategy is key.

Approaching this challenge can become more manageable by focusing on the four key areas of hardware, software, accelerators and templates. Working with Intel and our broad network of ecosystem players can help ensure you build the infrastructure that will empower your business to achieve faster, more sophisticated analytics.

Learn More

- **Solution Brief:**
[Create Real-time Business Value with Advanced Analytics](#)
- **Solution Brief:**
[Intel® Select Solution for Microsoft SQL Server](#)
- **Solution Brief:**
[Streaming Analytics in the Real-Time Organization](#)
- **Video:**
[Transforming the Enterprise with a Future Proof Infrastructure](#)

Find the solution that is right for your organization. Contact your Intel representative or visit www.intel.com/analytics

Solution Provided By:



¹ PwC's Global Data and Analytics Survey, July 2016: Big Decisions TM. Global Base: 2,106 senior executives. <https://www.pwc.com/us/en/services/consulting/analytics/big-decision-survey.html>

² <https://www.pwc.com/us/en/services/consulting/library/analytics-speed-sophistication.html>

³ Performance benchmark of Intel® MKL against F2JBLAS and OpenBLAS on run on system powered by Intel® Xeon® E5-2697A. Full details: <https://blog.cloudera.com/blog/2017/02/accelerating-apache-spark-mllib-with-intel-math-kernel-library-intel-mkl/>

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at intel.com

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit www.intel.com/benchmarks

Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.

All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest Intel product specifications and roadmaps.

Intel, the Intel logo, Xeon, Optane, Nervana, Movidius, and OPENvino are trademarks of Intel Corporation or its subsidiaries in the U.S. and/or other countries.

*Other names and brands may be claimed as the property of others.