(intel®)

Machine Learning-Based Advanced Analytics Using Intel® Technology

Learn about the machine-learning process, how it affects architecture choices, and how Intel® technology can help build a scalable, powerful, and reliable machine-learning solution

This reference architecture provides a fully functional technical picture for developing a cohesive business solution.

If you are responsible for...

- **Business strategy:**
  You will better understand how machine-learning technology can improve business capabilities and functions.

- **Technology decisions:**
  You will learn about the critical machine-learning architecture components and how they work together to create a cohesive business solution.

Executive Summary

Machine learning enables businesses and organizations to discover insights previously hidden within their data. Whether exploring oil reserves, improving the safety of automobiles, detecting fraud, or mapping genomes, machine learning is at the heart of innovation and business intelligence. Unleashing the power of machine learning, however, requires access to large amounts of diverse datasets, optimized data platforms, powerful data analysis and visualization tools, a highly scalable and flexible compute and storage infrastructure, and lastly, the right skill sets.

Intel's high-performance computing (HPC) reference architectures are optimized for machine learning. These solutions are built on a hardware foundation that includes compute, memory, storage, and network. They are designed to run optimized, scalable analytics and artificial intelligence (AI) workloads—including streaming, predictive analytics, machine learning, and deep learning.

Businesses can gain scalability, effectiveness, efficiency, and lower total cost of ownership (TCO) by using a machine learning-optimized solution built on Intel® architecture that is designed to handle analytics and AI workloads from the edge, in the cloud, or on-premises. Moreover, a machine learning-optimized architecture can reduce time to market (TTM) for intelligent solutions that can provide a competitive market edge.

Scalable Data and Analytics Platforms

Applications
Analytics-powered vertical and horizontal solutions

Data
Open-source Hadoop*-centric platform for distributed and scalable storage and processing

Machine Learning Optimized Intel® Architecture
Software-defined storage, virtualized compute, networking, and cloud

Machine-Learning Frameworks and Algorithms
Multilayered, fully optimized algorithms

Intel® Data Analytics Acceleration Library

Intel® Math Kernel Library

Performance and Security
Silicon and software enhancements to protect and accelerate data and analytics
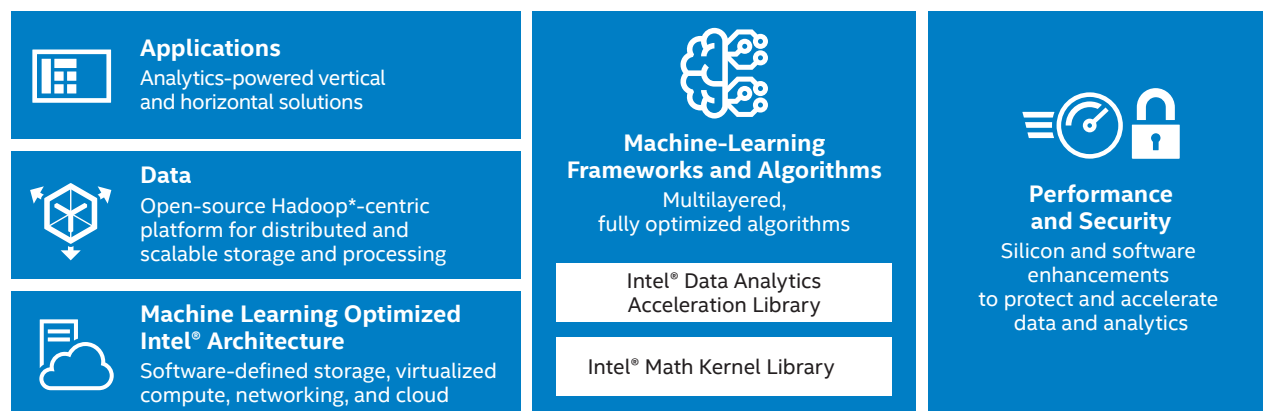
Figure 1. A fully optimized machine-learning solution is built on tightly integrated Intel® technologies for accelerated insight discovery at a lower cost of ownership.

Table of Contents

Figure 2. Machine learning is a branch of artificial intelligence (AI) that enables computers to self-learn and is an excellent advanced analytics tool.

In this guide, we explore the challenges associated with deploying machine learning, the business value machine learning can bring to the enterprise, and the machine-learning process. We then look deeper into which architectures are available and how Intel® technology can improve return on investment (ROI) at every layer of the solution architecture. The bottom line is that machine-learning solutions powered by Intel technology provide valuable business intelligence through accelerated model training, fast scoring, and highly scalable infrastructure.

Introduction

Building a machine-learning solution can be complex. There are many use cases, technologies, and tools available; sometimes it is hard to know where to start. Using this document, business owners, technical project managers, and solution architects can learn about the machine-learning process and the planning considerations necessary for choosing the right architecture. This document serves as the foundation for a set of best practice and design guides targeted for specific verticals and use cases.

How Machine Learning Intersects with Advanced Analytics and Artificial Intelligence

Machine learning is a branch of artificial intelligence (AI) drawing from the computer science field. Machine learning relies on statistical techniques that enable researchers, data scientists, engineers, and analysts to automate the creation of analytical models by constructing algorithms that can learn from and make predictions based on data—without being explicitly programmed. Machine learning also encompasses its subset, deep learning, which relies heavily on neural networks.

Advanced analytics and machine learning are intertwined in a number of ways (Figure 2). Both use computer programming along with various statistical and data-mining techniques including time series analysis, text analysis, random forest, decision trees, pattern matching, forecasting, visualization, semantic and sentiment analysis, network and cluster analysis, to name a few.

In summary, both machine learning and advanced analytics aim at analyzing data to discover patterns and draw insights.

Business Challenge: Using Machine Learning Effectively

Businesses in every industry can gain a competitive advantage and generate new revenue by delivering products and services that are more personalized, efficient, and adaptive. But doing so means enterprises must put the vast amount of data that is available to effective use—something that many struggle with. Experts predict that world data will grow 10X in 10 years,[1] yet less than 1 percent of all data is ever analyzed and used.[2]

The growing importance of big data makes machine learning a differentiating factor. However, the powerful potential of machine learning seems out of reach for many organizations.
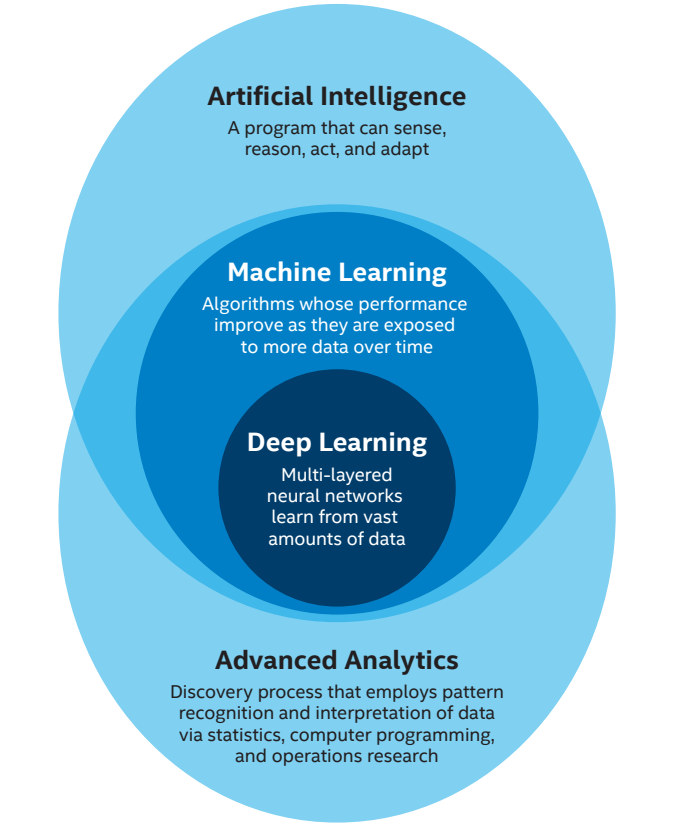
To be successful, the following foundational elements are necessary:

- Access to large amounts of diverse data in order to build robust and accurate inference models

- Optimized data and analytics pipelines running on high-performance compute platforms that are designed to manage and process massive volumes of data and that can execute machine-learning workloads at high speed in any environment from the edge to the cloud, both on- and off-premises

- A highly scalable, flexible infrastructure (compute, memory, storage, and network) on which to develop, train, and deploy machine-learning models

- A pool of appropriately skilled talent, such as data scientists, statisticians, data engineers, and solution architects with expertise in using open-source frameworks such as Apache Hadoop* and Apache Spark*

Machine Learning Provides Business Value to Multiple Industries

Traditional data warehousing and data marts can be inflexible. A key challenge with legacy data warehousing technologies is the cost of maintaining and accessing data efficiently. Moreover, these systems were not designed to process unstructured data or support blending of data from multiple structured and unstructured sources while adhering to data governance and data lineage policies.

Machine learning surmounts these issues and can drive business growth by producing fresh insights previously buried in mountains of big data. Through model training, the value of a machine-learning solution can continue to grow as it consumes new data and further improves accuracy, reaching levels that far surpass traditional systems—making machine learning a good investment for virtually any enterprise.

For example, in Europe, more than a dozen banks have replaced older statistical-modeling approaches with machine-learning techniques and, in some cases, experienced 10-percent increases in sales of new products, 20-percent savings in capital expenditures, 20-percent increases in cash collections, and 20-percent declines in churn.[3]

According to IDC, "cognitive computing, AI, and machine learning will become the fastest growing segments of software development by the end of 2018; by 2021, 90 percent of organizations will be incorporating cognitive/AI and machine learning into new enterprise apps."[4]

INCREASED ENGAGEMENT 3X

Intel IT developed a machine-learning system that doubled potential sales and increased engagement with Intel's resellers by 3X in a number of industries.[8]

The following examples prove that virtually every industry has become data-driven and can therefore benefit from machine learning:

- **Fraud detection.** Banking and insurance companies use machine learning to reduce fraud. Recently, the Intel® Saffron™ platform helped an auto insurance company examine 113,000 claims from one year in one state in less than a month. It revealed three potential fraud rings involving a radiology clinic and other medical providers that could be discovered only by highlighting hidden connections among the clinics.[5] As another example, one financial institution used machine learning to identify USD 1.5 million per month in fraud that their traditional rules-based system had not previously detected.[6]

- **Healthcare and life sciences.** Machine learning has significantly sped up genomic analysis. It can also help identify cancer and other diseases, put more data at caregivers' fingertips to improve quality of care, and improve diagnosis. For example, one cardiologist used machine learning to more accurately diagnose heart conditions with similar symptoms.[7]

- **Customer relationship management (CRM) and consumer behavior.** Enterprises can better predict customer purchases and reduce customer churn by delivering targeted and contextual offers. For instance, Intel IT developed a machine-learning system that doubled potential sales and increased engagement with Intel's resellers by 3X in a number of industries.[8]

- **Failure prediction for preventive maintenance.** Manufacturing, utilities, oil and gas, construction, transportation, and many more industries can use machine learning-based failure prediction to reduce the frequency and cost of maintenance.

- **Smart machines.** Autonomous cars, image recognition for security systems, smart cities, and wearable devices all illustrate how machine learning can dramatically transform how machines behave and how they deliver and act on data.

- **Workforce management.** Industries such as travel, transportation, and utilities can use machine learning to optimize the management and productivity of their employees by being able to predict where workers are most likely to be needed.

- **IT operations.** Machine learning can improve activities such as network threat monitoring, data management, and IT support capabilities.

In all these cases, machine learning offers four important benefits that can lead to competitive advantage: velocity of insight, volume of processed data, operational efficiency, and the intelligence to learn autonomously.

Preparing to Deploy a Machine-Learning Solution

While the potential business value of machine learning is clear, organizations face several challenges when beginning their machine-learning journey. Overcoming these challenges is the key to digital transformation and taking full advantage of data. Some of the most pressing challenges associated with deploying machine learning include the following:

- **Data often exists in silos.** Raw data is the fuel for machine learning. Large enterprises often have a complex data landscape with limited interoperability between data repositories. A data integration strategy is necessary to handle the variety and volume of data required by machine learning.

- **Data is constantly changing.** The relevance of data may decrease during a machine-learning analysis, or combining one dataset with another may introduce data inaccuracies and other contextual issues.

- **Privacy and ethical issues must be addressed.** Personally identifiable information must be anonymized, and care must be taken not to build cultural bias into machine-learning models.

- **Finding the right talent can be difficult.** Experienced data scientists and data engineers with relevant business acumen are in short supply in all industries. Yet it is these experts who can extract a wide range of knowledge from data, understand the end-to-end process, and solve data science problems.[9] Machine learning typically requires strong mathematical and statistical skills if a solution is going to go beyond an off-the-shelf algorithm or a packaged set of APIs.

- **Desire for self-service machine learning is growing.** Data democratization, or the process whereby non-data scientists want to analyze data, is becoming more popular. It can be difficult to design a machine-learning solution that is powerful enough to generate the desired insights yet is also easy to use by those less experienced in data science.

If your team doesn't have extensive data science and data engineering expertise, there are many resources available that can help. Online classes, self-study of simple machine-learning algorithms, and a few straightforward cloud-based machine-learning projects can go a long way to helping demystify machine learning. Intel is a trusted advisor to a vast ecosystem of technology providers that are enabling AI workloads. Free training and downloadable open source-optimized libraries for building machine- and deep-learning solutions are available at Intel® AI Academy.

The Machine-Learning Process

As shown in Figure 3, machine learning is essentially straightforward: At a high level, it requires the following steps:

1. Ingesting and aggregating a large volume and wide variety of data

2. Preparing that data through cleansing, transformation, and normalization

3. Building an appropriate machine-learning model and learning from it

4. Delivering results such as forecasts and recommended actions

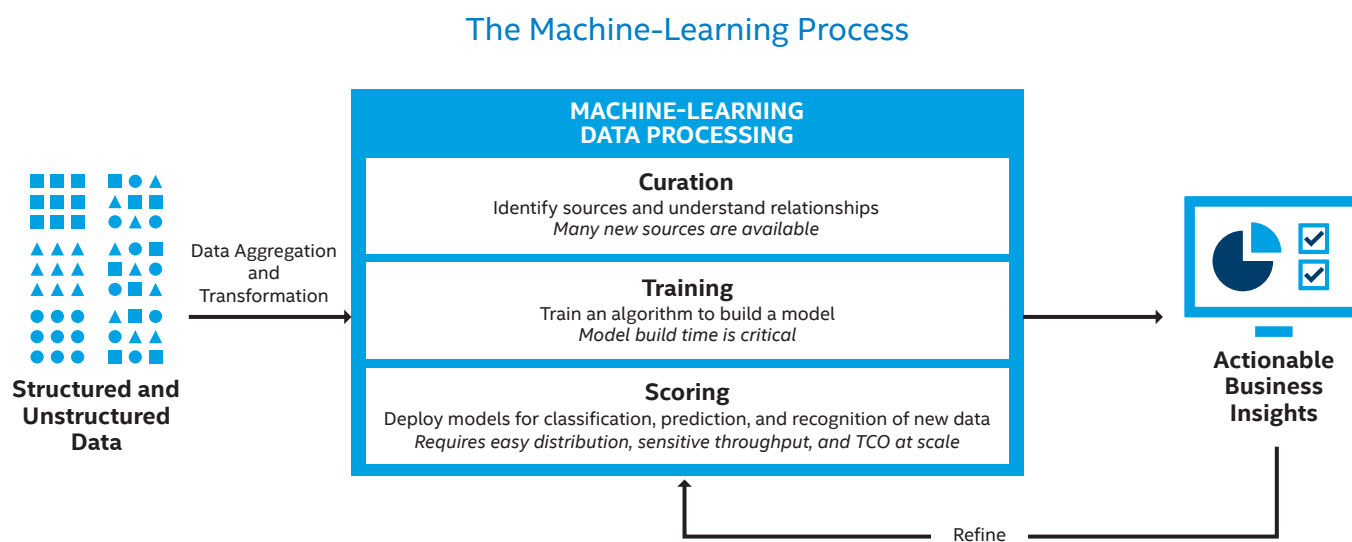5. Refining the learning model based on the accuracy of the results

The Machine-Learning Process

Figure 3. The basics of machine learning include data ingestion, data analysis and model training, and results generation. Self-learning capabilities continuously improve the model's accuracy.

Beyond these basics, you can expect to spend time on several steps for any machine-learning project:

- **Define the business problem and objectives.** What are you trying to solve? What business goals are you trying to meet? All too often, IT departments start building machine-learning infrastructure with no clear goal of what it will be used for—leading to overprovisioning and failed projects. According to a Gartner survey, only 15 percent of enterprises have matured their big data projects from proof of concept (PoC) to production.[10] A sound machine-learning solution starts with an understanding of the desired business outcome and how that will relate to both the data and the known capability of the proposed architecture. Choose a relatively small and simple business problem to solve initially, to gain experience, prove return on investment (ROI), and then gradually expand.

- **Explore your data.** Also referred to as curation or data inventory, this step involves determining what data is necessary for the machine-learning project and where it will come from. Options include structured data such as that stored in enterprise resource planning (ERP) and CRM databases; semi-structured data such as JSON* and XML* documents; and unstructured data such as emails and social media posts. Some of the necessary data may already exist at the enterprise, while other sources may need to be acquired. If Internet of Things (IoT) data is part of the project, you may need to investigate whether the existing network can handle the data volume.

- **Get the data ready.** Not all data is created equal, and integrating a variety of data sources is a significant effort. In fact, three quarters of the effort in building out a machine-learning project involves acquiring and preparing the data to do the training.[11] Deduplication, data cleansing, normalization, and transformation all take time. However, output generated by machine learning is only as good as the data input, so getting this step right is important. You should plan for an almost continuous process where training data arrives in batches and the machine-learning algorithm runs on a shifting window of recent batches. In this way, a new batch of data triggers the generation of a new, more accurate model.

- **Selecting the data.** This step is critical. Once the data is ready, you must also decide which data will be used as training datasets for the machine-learning model. Without a well-defined data strategy coupled with stringent data preparation, governance, and lineage policies (see

"Governance and Security" later), the risk of failed projects increases significantly. Models and subsequent insights are intrinsically tied to ingested data, so we consider this a fundamental step to enabling successful machine-learning workloads at scale.

- **Choose the right algorithm.** There are many machine-learning algorithms to choose from (see Table 1), and usually it is a good idea to try a suite of standard algorithms on your problem and discover which algorithm performs best. Some use supervised learning, where training datasets are used to train the algorithm. Others use unsupervised learning, where no training datasets are provided. Still others use a combination (semi-supervised learning), such as analyzing a photo archive where only some of the images are labeled and the majority are unlabeled.

- **Test and validate.** This step includes defining and building the architecture that will support the analysis, and validating the chosen model and algorithm. You may have to re-run tests with varying algorithm parameter values to achieve the desired accuracy. For supervised learning, testing and validation include training and scoring the results. For unsupervised learning, validation is not performed on labeled features in the dataset; however, you can use a set of baseline golden results that have been manually labeled and human-validated for correctness. In addition, you can cross-validate the model using another model created for the problem in supervised learning mode. For clustering problems, variations within similar groups can be used to validate if the groups are accurate. Similar to the getting the data ready, this step may take some time but is essential for success.

- **Deploy.** Once you are confident the model is generating accurate results, it's time to put it into production and generate business value. Exactly what form the business value takes depends on your use case. Supply-and-demand analysis may enable better purchasing decisions. Predictive failure analysis may prevent costly downtime in the factory. Stored information, such as image classification, may be used by other applications. This step also entails deciding on how to make the results available for optimal consumption. A final recommendation is to develop a process for transitioning machine-learning projects into production, so that as the use of machine learning matures in your enterprise, you can easily scale.

- **Refine.** Machine learning is inherently iterative. The algorithms can learn from their mistakes, ingest new data,

Table 1. A Sampling of Machine-Learning Algorithms and Usages

Learning Type	Problem Categories	Typical Use Cases	Algorithm Examples
Supervised	• Regression analysis • Classification analysis	• Recommendation engines • Time series prediction	• Random forest • Linear regression • Support vector machines
Unsupervised	• Association • Clustering	• Grouping customers by purchasing behavior • Customers who buy X also tend to buy Y	• k-means • Apriori
Semi-supervised	• All of the above	• Image classification	• All of the above

and continually improve. You may have to take a similar approach to the architecture—as new technology or new data sources become available, refining and scaling the solution can generate even higher business value.

Mapping the Architecture to the Machine-Learning Process

When designing your machine-learning architecture, you must consider each step of the machine-learning process. The resulting architecture must be flexible enough to adapt to new sources of data, elastic enough to handle varying workloads, and powerful enough to crunch through terabytes of data. You also need to consider overarching architecture components that provide security and governance, as well as create a solution that supports Agile and DevOps methodologies.

Figure 4 shows a comprehensive machine-learning architecture. If you are just getting started with machine learning, it's likely that you won't build the whole thing at once. As mentioned earlier, it is often best to start small and scale up as the number of machine-learning use cases multiply.

- **Data acquisition.** Machine learning is built on a foundation of data, so choose data ingestion tools that support a wide variety of data sources. The architecture should include both real-time streaming and batch processing capabilities. A well-designed data pipeline is fast, reliable, and elastic.

- **Data processing.** Plan for data transformation, normalization, cleansing, and encoding. In the case of supervised learning, the architecture should also enable selection of training datasets. The choice of architecture must reflect the needs of the machine-learning use case. Considerations include whether data is coming in discrete chunks or continuously; how much throughput is required (if high throughput is necessary, consider a Lambda architecture and/or an in-memory database); and how data integration will be done.

- **Feature engineering and data modeling.** Sometimes called feature analysis, here is where you turn your inputs into something the algorithms can understand. This might involve simplifying the data, filtering it, or creating new features. Feature engineering can be done manually or using automated feature-extraction tools.

- **Model fitting.** A machine-learning model is a combination of the algorithm and the training data. It is a mathematical representation of the data. Examples of algorithms include random forest, least squares, and logistic regression. As mentioned earlier, the choice of algorithm depends on your machine-learning use case. Choosing the appropriate algorithm will require some trial and error and is highly dependent on the previous step (feature engineering and data modeling), because an algorithm's performance (accuracy) is affected by the data it runs on. While it is possible to write your own machine-learning algorithms, it is usually beneficial to invest in one or more libraries of algorithms that can be adapted to suit your needs. Besides machine-learning toolkits, another useful architecture component is a key-value store for machine-learning model metadata.

- **Model training.** In this step, you use a training dataset to "educate" the model by treating the training dataset as new data and withholding the output values from the algorithm. Predictions from the trained model on the inputs from the training dataset are compared to the actual output values of the training dataset. The model can learn from its mistakes, and many training runs may be necessary to increase the model's accuracy—meaning that the architecture should be capable of supporting the compute and storage needs of the training process.

- **Model validation.** This is the process of using a testing dataset to evaluate a trained model. The testing dataset is a separate portion of the same dataset from which the training dataset is derived. Four primary validation techniques are
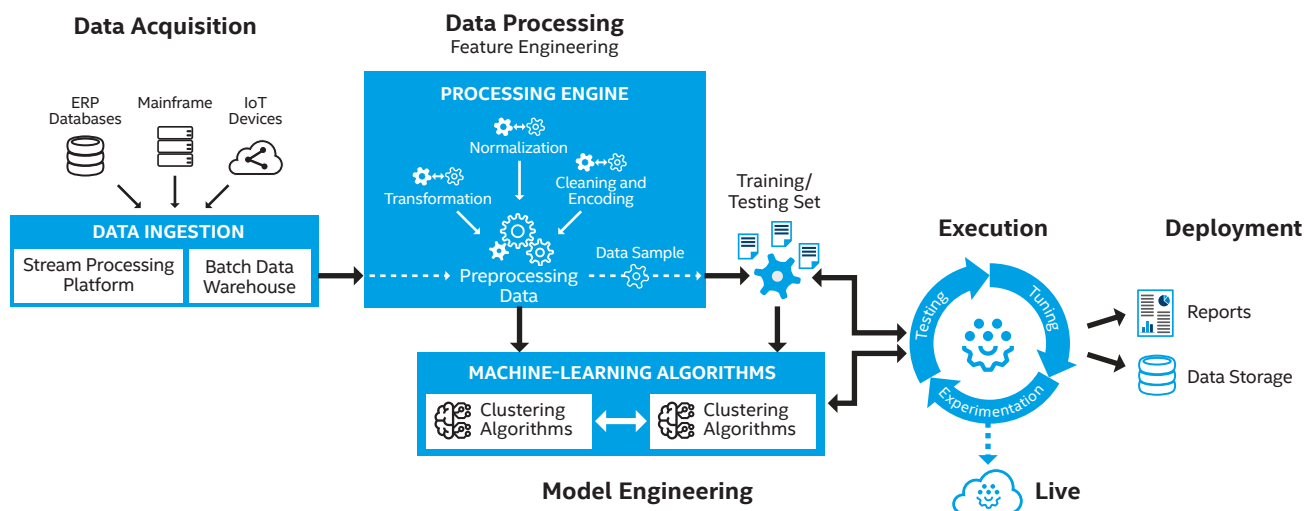
Comprehensive Machine-Learning Architecture

Figure 4. A comprehensive machine-learning architecture supports the entire machine-learning process.

available: predictive modeling, training error, test error, and cross-validation. As you build your architecture, look for tools that support comprehensive model validation.

- **Deployment.** The execution portion of the architecture must be powerful enough to support repeated cycles of experimentation, testing, and tuning. Investing in high-performance and easily scalable compute and storage enables the architecture to grow with the needs of your enterprise. Also consider that given different data, the same model may behave quite differently. Therefore, the architecture should be able to scale automatically to avoid increases in latency.

- **Monitoring.** Choosing tools that include monitoring capabilities can help with model-optimization efforts.

Operational Efficiency

Machine learning is a relatively new field, and innovation occurs rapidly. Therefore, machine-learning projects require Agile, DevOps-style workflows. Teams need to be able to swiftly re-evaluate models and use continuous integration/continuous delivery (CI/CD) to put the latest information or technology to work and prevent model decay.

Governance and Security

Data is an important asset, and with the advent of the General Data Protection Regulation (GDPR), protecting data becomes even more important. As you build your machine-learning architecture, integrate security measures and governance processes into each layer. This can help mitigate the risk of breaches in data, learning and classification modules, and output.

- **Governance.** In addition to the best practices mentioned earlier for getting the data ready and testing/validation, you must establish processes that track data lineage—where it comes from, where it is stored, who has access to it, and where it goes next. These processes must be inherently auditable. Map the existing data flow and develop standard data taxonomies across your organization. Plan for metadata collection, integration, usage, and repository maintenance. Essentially, think of your data governance as an x-ray for your data, revealing source to destination along with the various processes and rules involved and how the data is used.

- **Security.** Along with data lineage, data security is paramount to a mature machine-learning architecture. Security includes authentication (such as with Kerberos*), authorization, and encryption. You must apply authentication and access controls across the entire framework, from ingestion to report delivery. Like data governance, your data security measures must be auditable. You should plan for encryption for both data at rest and data in transit. Depending on your industry and your data, you may also need to anonymize personally identifiable information. Cloudera has published an excellent paper covering these topics.

Trusted infrastructure relies on hardware-based security built directly into Intel® Xeon® Scalable processors. Trusted infrastructure comprises a suite of solutions that includes Intel® Cloud Integrity Technology (Intel® CIT), Intel® Trusted Execution Technology (Intel® TXT) with One-Touch Activation, and Intel® Platform Trust Technology (Intel® PTT). For storage, Intel technologies such as Intel® Advanced Encryption Standard New Instructions (Intel® AES-NI) help protect data at rest and in motion.

Intel® Technology Building Blocks for Your Machine-Learning Solution

An effective machine-learning solution requires scalability and elasticity; significant compute power; adequate, low-latency storage; and high-bandwidth Ethernet fabric. Without this high-performance combination, most machine-learning projects of any significant size will not generate optimal business value. Adjacent technologies and software libraries can benefit varying use cases. Projects can also benefit from Intel's collaboration with the ecosystem.

The exact composition of a machine-learning solution will vary, depending on the machine-learning algorithms, how much automation needs to be built into the system, and which frameworks and tools are being used. However, whatever the workload, using a machine-learning solution based on Intel® architecture (Figure 1 and Figure 4) enables businesses to gain scalability, effectiveness, efficiency, and lower TCO while reducing time to market (TTM) for intelligent solutions—all of which contributes to an edge over their competitors.

Intel® Xeon® Processors: The Core of an Effective Machine-Learning Solution

Servers equipped with Intel® Xeon® processors help keep costs affordable while delivering exceptional performance, agility, reliability, and security. Intel Xeon Scalable processors offer highly parallel performance and efficient use of local memory bandwidth through hardware-based technologies such as Intel® QuickAssist Technology (Intel® QAT), Intel® Hyper-Threading Technology (Intel® HTT), Intel® Turbo Boost Technology, Intel Advanced Encryption Standard New Instructions (Intel AES-NI), Intel® Advanced Vector Extensions 512 (Intel® AVX-512), and Intel® Speed Shift Technology. Machine-learning workloads that are optimized for the latest generation of Intel Xeon processors can execute much faster than non-optimized code,[12] and can significantly speed metrics such as training throughput and time to train.[13]

Built-in Intel® field-programmable gate array (Intel® FPGA) modules for acceleration can be reprogrammed in a fraction of a second with a datapath that matches your workload's key algorithms. The Acceleration Stack for Intel® Xeon® CPU with FPGAs is a new collection of software, firmware, and tools that allows software developers to leverage the power of Intel FPGAs much more easily than before.

The Acceleration Stack for Intel Xeon CPU with FPGAs offers many benefits:

- **Workload optimization.** Ensure Intel Xeon CPU cores serve the highest value processing

- **Efficient performance.** Improve performance/watt

- **Real-time performance.** High-bandwidth connectivity and low-latency parallel processing

- **Developer advantage.** Code re-use across Intel FPGA data center products

Microsoft's Bing* Intelligent Search is powered by Intel FPGAs, which enable Bing to use machine learning and reading comprehension to rapidly provide intelligent answers that help users find what they're looking for faster, instead of a list of links for the users to manually check.[14]

Additional High-Performance Intel Technologies

It's no secret that machine-learning workloads can test the limits of hardware and software. Intel provides technologies that can stand the test:

- **Memory and storage.** As model sizes increase, it is important to keep data close to memory to reduce latency while processing large datasets. Beneficial Intel technologies include Intel® 3D XPoint™ technology, Intel® Optane™ technology, and rugged, high-performance PCIe*- and NVMe*-based Intel® Solid State Drives (Intel® SSDs).

- **Network.** Effective machine-learning solutions require a high-performance, low-latency fabric like Intel® Omni-Path Architecture (Intel® OPA) to maximize memory capacity and floating-point performance and accelerate results. Intel® Ethernet Server Adapters (10 GbE, 25 GbE, and 40 GbE) speed data transmission.

- **Scalable data and analytics platforms.** These can be layered on the core solution, which can then efficiently run individual analytics applications. Apache Hadoop-based data lakes, such as Cloudera Enterprise*, support distributed and scalable storage and processing. New Intel® Scalable System Framework (Intel® SSF) hardware and software in combination with code modernization delivered an observed 50x machine-learning performance improvement in a Colfax Research case study.[15]

Ecosystem Collaboration Results in Benefits for All

Intel collaborates with many fellow travelers in the machine-learning ecosystem to drive innovation and business value. Collaboration teams often include academic researchers, OEMs, communications service providers (CSPs), ISVs, and system integrators. Here are a few examples:

- BigDL*. Intel has made significant technical contributions to the Spark community over the years, including leading an open-source initiative called BigDL, a distributed library for deep-learning. Unlike a number of other libraries for building deep-learning solutions, BigDL is native to Spark. With BigDL, you can create deep-learning solutions as standard Spark programs that run on existing Spark or Hadoop clusters. By using infrastructure already in place instead of deploying a new cluster with an unfamiliar architecture, BigDL accelerates time to value, reduces TCO, and improves ease of use. BigDL has strong support in the industry, including Microsoft Azure*, Cloudera, Amazon Web Services* (AWS*), JD.com, Databricks, Cray, and GigaSpaces, among others.

- Intel® Math Kernel Library **(Intel® MKL).** This library optimizes code with minimal effort for future generations of Intel® processors. The routines are optimized specifically for Intel processors, and are compatible with a wide choice of compilers, languages, operating systems, and linking and threading models. It is available in a number of distributions, such as Python*, YUM, APT-GET, and conda*. Intel MKL features highly optimized, threaded, and vectorized math functions that maximize performance on each processor family. It uses industry-standard C and Fortran APIs for compatibility with popular BLAS, LAPACK, and FFTW functions—no code changes are required. Recent improvements include small matrix multiplication performance in GEMM and LAPACK, and enhanced ScaLAPACK performance for distributed computation. An Intel MKL license includes Priority Support that connects you directly to Intel engineers for confidential answers to technical questions.

- Intel® Distribution for Python*. Powered by Anaconda*, this distribution accelerates computational packages such as NumPy, SciPy, and scikit-learn*. Data scientists can use it to easily implement and scale performance-packed, production-ready algorithms for data analysis, while domain experts who are not programmers can simply download, install, and obtain immediate optimized performance.

- Intel® Data Analytics Acceleration Library (Intel® DAAL). This library helps speed big data analytics by providing highly optimized algorithmic building blocks for all data analysis stages (preprocessing, transformation, analysis, modeling, validation, and decision making) for offline, streaming, and distributed analytics usages. It is designed for use with popular data platforms including Hadoop, Spark, R*, and MATLAB*. Intel DAAL helps improve prediction quality and speed, can increase the size of datasets without increasing compute resources, and optimizes data ingestion and algorithmic compute together for high performance. As with Intel MKL, you can connect privately with Intel engineers for technical questions about Intel DAAL.

- Intel® MPI Library. This library focuses on enabling MPI applications to perform better for clusters based on Intel architecture. Designed and developed for high scalability, the library is optimized for Intel OPA. The Intel MPI Library is available as part of Intel® Parallel Studio XE Cluster Edition and as a free stand-alone version. A license purchase includes Priority Support.

- **Intel® nGraph™ library.** This is an open-source library for developing frameworks that can efficiently run deep-learning computations on a variety of compute platforms. Frameworks currently directly supported include TensorFlow*, MXNet*, and neon™ framework; CNTK*, PyTorch*, and Caffe2* are supported indirectly through ONNX*. Intel nGraph library lets you explore and experiment without resorting to hand-tuning compilers to achieve high performance. You can perform kernel fusion, efficient memory buffer allocation, and improved data layouts, as well as intelligently enable distributed training across a variety of CPU, neural network processor (NNP), and graphics processing unit (GPU) hardware without writing new libraries from scratch.

- A Kaggle* Competition, sponsored by Intel and MobileODT*, challenged research teams to use AI to improve the precision and accuracy of cervical cancer screening.

- TensorFlow is a leading machine-learning and deep-learning framework. Optimizing TensorFlow for Intel architecture has resulted in an up to 85x speedup on common neural network models.[16]

- Intel is collaborating with industry experts such as Dell EMC, Cloudera, and DataRobot to streamline machine-learning deployments, improve interoperability, and help customers bridge the skills gaps in infrastructure, software, models, algorithms, and data science.[17]

A Sample of Machine-Learning Software Frameworks and Tools

There are many libraries, frameworks, and tools available—both open-source and proprietary—that can be integrated into a machine-learning solution. When choosing tools, one important consideration is interoperability. The more frameworks with which a particular tool is compatible, the better the investment.

Forums are available, and system integrators can help you determine which resources might be best suited to your workload and how to deploy them. Here are a few examples:[18]

- **Data science tools.** Many tools offer easy-to-use graphical user interfaces (GUIs) that enable even those with minimal knowledge of algorithms to build machine-learning models. Here are few to explore: R, Python*, KNIME, Gawk*, Weka*, Scala, SQL, RapidMiner*, scikit-learn*, and the Apache* ecosystem.

- **Machine-learning frameworks.** Machine-learning frameworks can scale across distributed systems. Through code modernization, these frameworks are optimized to take advantage of parallelization at the thread, data, and vector levels, as well as innovations in memory and storage. Some examples include Caffe*, Torch*, Theano*, Apache Singa*, Apache Mahout*, Shogun*, Apache Spark MLlib*, TensorFlow*, Oryx 2*, and Accord.NET*.

- **Data management.** Data-management issues often arise when deploying machine learning in production. Invalid data can cause outages in production; therefore, data monitoring, validation, and correction are essential. Look for a data-management tool that supports lifecycles of complex collaborative data science workflows, including raw data and metadata. Examples include Talend* and Dremio*.

- **Data ingestion and integration.** Choose tools that cover all the basics, including data discovery, data profiling, data improvement, and data transformation. Apache NiFi*, StreamSets Data Collector*, Gobblin*, Sqoop*, Flume*, and Kafka* are all worth investigating. You may also want to explore Integration Platform as a Service (iPaaS), a suite of cloud services enabling development, execution, and governance of integration flows connecting any combination of on-premises and cloud-based processes, services, applications, and data within a single organization or across multiple organizations.[19]

- **Programming APIs.** As machine learning-based solutions evolve, there are numerous collections of APIs to explore. Which ones to choose depends on the capabilities you need, such as prediction, face recognition, image processing, or speech recognition. Consider feature lists, ease of use, language, and extensibility.

- **Compute engines.** These are managed services that enable you to easily build a broad range of machine-learning models. They are offered by several major cloud service providers.

- **Visualization.** Visualization applications are a specific type of data science tool that enable interactive visual exploration of data. Be sure to choose ones that can handle the huge and very fast-changing datasets associated with machine learning.

- **Dashboards.** There are several options available for utilities and monitors for machine-learning projects. Most dashboards include visualization tools to facilitate parameter tuning.

Summary

Machine learning is becoming a business necessity; it can help organizations quickly build models that enable development of intelligent solutions—which in turn create new revenue streams and differentiate those organizations from their competitors. But whether exploring underground oil reserves or improving the safety of automobiles, organizations need an architecture that supports business needs, can scale over time, and provides the necessary performance and security. Intel collaborates with ecosystem fellow travelers to provide a machine-learning architecture that can be tuned for predictive and prescriptive analytics as well as other machine-learning use cases.

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

Take the Next Step

Now that you have a good understanding of a generic machine-learning architecture, you are ready to build the specific solution that's right for your workload. You may find the following resources helpful on your machine-learning journey.

• **Intel® AI Academy**

1   Seagate, April 2017, "Data Age 2025." seagate.com/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf
2   Forbes, Sept. 2015, "Big Data: 20 Mind-Boggling Facts Everyone Must Read." forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#867048017b1e
3   McKinsey, 2015, "An executive's guide to machine learning," mckinsey.com/industries/high-tech/our-insights/an-executives-guide-to-machine-learning
4   IDC, November 2017, "IDC FutureScape: Worldwide Analytics and Information Management 2018 Predictions," idc.com/getdoc.jsp?containerId=US42619417
5   Intel, November 2017, "AI's Role in Fighting Fraud." itpeernetwork.intel.com/ais-role-fighting-fraud
6   SAS Institute, "Fraud detection and machine learning: What you need to know," sas.com/en_us/insights/articles/risk-fraud/fraud-detection-machine-learning
7   Intel, "Artificial Intelligence Lends a Hand to Cardiologists." intel.com/content/www/us/en/healthcare-it/solutions/ai-helps-cardiologists
8   Intel, October 2016, "Data Mining Using Machine Learning to Rediscover Intel's Customers," intel.com/content/www/us/en/it-management/intel-it-best-practices/data-mining-using-machine-learning-to-rediscover-customers-paper
9   BusinessTech, September 2017, "Using machine learning and AI to add value to business," businesstech.co.za/news/technology/199794/using-machine-learning-and-ai-to-add-value-to-business
10 Gartner, October 2016, "Survey Reveals Investment in Big Data Is Up but Fewer Organizations Plan to Invest," gartner.com/newsroom/id/3466117
11 TechTarget, January 2018, "Machine learning models require DevOps-style workflows," searchbusinessanalytics.techtarget.com/feature/Machine-learning-models-require-DevOps-style-workflows
12 Intel, "Inside Intel: The Race for Faster Machine Learning," intel.com/content/www/us/en/analytics/machine-learning/the-race-for-faster-machine-learning
13 Intel, "Artificial Intelligence with New Intel® Xeon® Scalable Processors: Most Agile AI Platform," intel.com/content/www/us/en/benchmarks/server/xeon-scalable/xeon-scalable-artificial-intelligence
14 Intel Newsroom, March 2018, "Intel FPGAs Accelerate Artificial Intelligence for Deep Learning in Microsoft's Bing Intelligent Search." newsroom.intel.com/editorials/intel-fpgas-accelerating-artificial-intelligence-deep-learning-bing-intelligent-search
15 The Next Platform, August 2016, "Intel SSF Optimizations Boost Machine Learning," nextplatform.com/2016/08/16/intel-ssf-optimizations-boost-machine-learning
16 Intel, 2017, "Optimize TensorFlow* for Intel® Xeon® and Intel® Xeon Phi™ Processors," software.intel.com/en-us/events/hpc-devcon/2017/ai?multiplayer=5646175620001
17 Intel IT Peer Network, November 2017, "Dell EMC & Intel: Collaborating to Help Customers Jumpstart Their Machine Learning Use Cases," itpeernetwork.intel.com/dell-emc-intel-jumpstart-machine-learning
18 Inclusion in this list does not indicate endorsement by Intel; there are many other options available.
19 Gartner, "IT Glossary," gartner.com/it-glossary/information-platform-as-a-service-ipaas