Accelerate Data Science Operations with Workstations Powered by Intel

intel.

What core capabilities do workstations need to perform data science tasks?

The types of operations and tasks performed daily, weekly, and monthly by data scientists necessitate system architectures geared to the distinct challenges of their work.

Traditional compute approaches, such as cloud computing and GPU architectures, don't adequately handle core tasks.

Handling massive datasets while minimizing latency is another advantage of a workstation enabled for data locality. Expansive memory capacities are needed to handle data often in the 3 TB to 6 TB range.

Workstations tailored to task central to data scientists—data exploration; extract, transform, and load (ETL) operations; model evaluation and visualization tasks—can accelerate operations and streamline workflow.

Gain the advantages of a workstation purpose-built for data science

Data scientists have to perform a wide array of core tasks, sometimes relying on a generic laptop for basic tasks and depending on cloud instances for more demanding operations. There is a better approach.
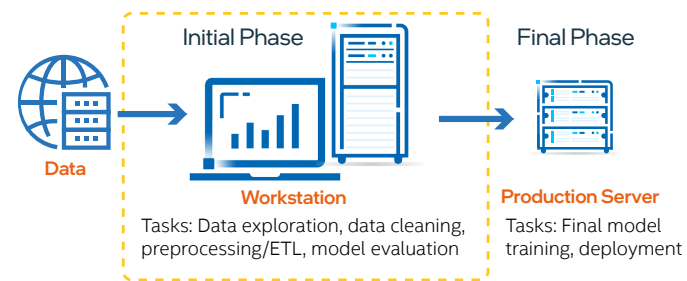

Figure 1. Typical AI/ML workflow.

Use highly interactive systems

Supporting the workflow of data scientists requires systems that are highly interactive and can handle massive volumes of data, using tools designed for single-node processing. GPUs have no useful role in this workflow. Hardware, software, libraries, and toolkits from the Intel portfolio offer the versatility and the right mix of capabilities to build a workstation optimized for data science tasks.

Tap into hardware capabilities that maximize data science operations

Workstations offered by Intel include a slate of processors infused with AI capabilities, persistent memory options, and fast SSDs. Compared to typical high-performance computing cluster nodes or premium laptops, workstation memory and drives can be more effectively configured to accommodate larger volumes of data.

Capitalize on open standards and a cross-architecture development platform

To support the end-to-end processes of AI development teams, data scientists can benefit from open standards toolkits and the Intel® oneAPI unified programming model. Optimized libraries and frameworks deliver powerful, cross-platform interoperability.

Discover workstation solutions that improve workflow, accelerate data science tasks, and help build AI solutions faster.

Guidelines for choosing the right workstation for the right tasks appear on page two.

Successful AI solutions are grounded in hardware. Intel® Xeon® processors and Intel Core™ processors power workstations with deep capabilities for diverse uses. No GPU required.

Data scientists can choose from a diverse selection of capable workstations from Intel to meet the requirements of simple to demanding data science tasks.
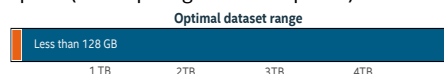
Mobile workstations
Configuration for data science enthusiasts.
Single-socket Intel Core i9 -10900k processor, 3.7 GHz, 10 cores/20 threads
- Cache sizes: 20 MB, 2.5 MB, 320 KB
- 128 GB memory (4 x 32 GB)
- 2TB SSD

Entry level

Primary use cases
Best for basic data science projects, meeting baseline memory and storage requirements for budget-sensitive uses. Intel Core processors excel where workloads scale with raw clock speed (not requiring AVX-512 speeds).

Optimal dataset range

Less than 128 GB			
1 TB	2TB	3TB	4TB

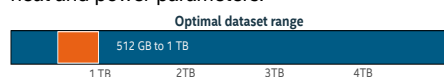**Approximate price as configured**: USD 6,000[1]

Mid-tier workstations
Balanced core count and frequency.
Single-socket Intel Xeon W 2295 processor, 3.0 GHz, 18 cores/36 threads
- Cache size: 24.75 MB, 18 MB, 1 MB
- 4 memory channels
- 512 GB memory (8 x 64 GB)
- 2 TB SSD
- Intel Optane™ SSD 905P Series (960 GB) (AIC PCIe x 4 3D XPoint)

Mid-tier

Primary use cases
Achieves an equitable balance between core count and processor frequency, providing cost-effective performance within moderate heat and power parameters.

Optimal dataset range

	512 GB to 1 TB		
1 TB	2TB	3TB	4TB

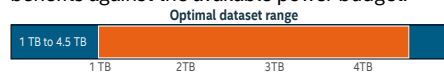**Approximate price as configured**: USD 16,000[1]

Top-tier workstations
Demanding applications with broad memory span.
Dual-socket Intel Xeon Gold 6258R processor, 2.7 GHz, 28 cores/56 threads
- Cache size: 38.5 MB, 28 MB, 1.75 MB (Intel Xeon processor L2/L3 cache hierarchy)
- 1024 GB (1 TB) memory (16 x 64 GB DDR4 ECC RDIMM)
- 2 TB SSD

Top-tier

Primary use cases
Suits applications in which requiring the memory span exceeds 3 TB, ranging as high as 6 TB. Also favors applications in which server-based optimization and core scaling are essential. This configuration requires balancing benefits against the available power budget.

Optimal dataset range

1 TB to 4.5 TB			
1 TB	2TB	3TB	4TB

Approximate price as configured: USD 30,000[1]

The workstation configurations and use cases shown are a representative sampling.

Explore other choices at Workstations Powered by Intel.

Improve data science workflow with preloaded, optimized libraries and frameworks from Intel.

Intel oneAPI Toolkits
Unify coding efforts with a single programming model that delivers exceptional cross-architecture performance

Built on open standards, **Intel oneAPI** simplifies development and deployment of data-centric workloads. The product family includes compilers, performance libraries, analyzers, debuggers, and more. Unifies coding for CPUs, GPUs, FPGAs, and other accelerators. Some of the components that are particularly valuable to data scientists are described in more detail in the following subsections.
Check out the **Installation Guide for Intel oneAPI Toolkits** for Linux, Windows, and macOS.

Intel Distribution for Python
Accelerate compute-intensive applications—from data analytics to machine learning—that use NumPy, SciPy, scikit-learn, and more

An essential tool for data scientists, this distribution delivers faster application performance on Intel platforms. Implement and scale production-ready algorithms for scientific computing and machine-learning workloads, and accelerate applications with native Intel Performance Libraries. **Download the free package** and unleash a faster Python on your data.

Intel AI Analytics Toolkit
Boost performance across data science and AI pipelines with this toolkit, built using oneAPI libraries

Enable significant acceleration for data analytics and machine-learning workflows, as well as direct access to Intel analytics and AI optimizations across the entire AI pipeline, from preprocessing through machine learning. **Download the Intel oneAPI AI analytics Toolkit**.

Intel oneAPI HPC Toolkit
Enhance high-performance computing applications with data-centric libraries, a powerful compiler, and advanced analysis tools

This add-on to the Intel oneAPI Base Toolkit provides valuable tools for data scientists building, analyzing, optimizing, and scaling HPC applications. The toolkit incorporates the latest techniques in vectorization, multithreading, multi-node parallelization, and memory optimization. **Download the Intel oneAPI HPC Toolkit**.

Find out more about Intel's commitment to enhancing the AI journey for data scientists at software.intel.com/ai.

intel