DELL Technologies

intel

Dell Precision Data Science Workstation Intel User and Installation Guide

intel 1 oneAPI AI ANALYTICS TOOLKIT

Contents

Introduction

Pages 4-5

The way forward
Choosing the right Dell Data Science Workstation

Pages 10-11

Powering up your performance
Setup and optimizations

Pages 12-13

Appendix
Ubuntu Installation Guide

Pages 14-17

Appendix
Intel AI Analytics Kit installation guide

Pages 18-23

References and resources

Pages 24-25

Introduction

Capturing opportunities when they present themselves and driving technological innovations is central to business success, particularly in today's unpredictable world. It's been said, "the measure of intelligence is the ability to change" then the smartest organizations prioritize their ability to adapt.

Artificial Intelligence (AI) based on machine learning (ML) and deep learning (DL) allows hardware to learn from previous computations to produce reliable, repeatable decisions and results. It gives organizations that all important adaptability, and can create new products and business initiatives, enhance existing markets as well as creating entirely new ones. But if AI relies on human guidance and intervention to achieve this, how useful can its impact really be?

By 2024, International Data Corporation (IDC) industry analysts believe the global machine learning market will grow to a total of $30.6 billion, attaining a compound annual growth rate (CAGR) of 43 percent.[1] In that same year, analysts also predict that 143 zettabytes of data will be created, captured, copied, and consumed throughout the world.[1]

For AI, machine learning, and deep learning developers, it's an intriguing business opportunity that challenges them to build high-performance applications with flexible deployment options. One hardware architecture can no longer do it all; we need to take Einstein's idea on board and change how we approach the technology and our mindset towards it.

Dell Technologies has developed the Data Science Workstations (DSWs), a tailored set of Precision workstations pre-bundled with hardware and software, designed for machine learning and deep learning developers. They're built with the latest Intel® Xeon® CPUs, ECC memory, professional graphics and advanced storage solutions.

We have created this step-by-step guide to provide users with a template to quickly set up the OS and Intel® oneAPI Data Science Software on their Dell DSWs. There are numerous DSW-certified and tested configurations, designed to meet the different development cycles required by our customers. This rigorous testing ensures that the configuration our customers purchase will meet, or exceed, their performance requirements right out of the box.

The Intel® oneAPI CPU-accelerated software and the Dell Data Science Workstations are designed to allow the developers to start working on large datasets with the optimized libraries on the same day they receive their DSW.

> Based on customer feedback, we learned that many data scientists are interested in how they can boost performance in the pre-processing and ETL phase of data science. With data analytics software tools from Intel, data scientists have seen 10x performance gains.

We've learned from our customers that many data scientists spend 70% to 80% of their time cleaning, organizing, and visualizing their data before going to production. Interestingly, most solutions in the market today are focused on the production phase where deep learning is prevalent rather than the data science phase in AI.

We've taken this feedback and developed a solution that focuses on machine learning and the pre-processing phase in AI. Using a combination of optimized Python tools and libraries from Intel, hardware instruction sets like AVX512, and increased memory capacity, you can significantly reduce the time it takes to develop the models and algorithms necessary to achieve the insights you need.

Intel® oneAPI optimized software reduces the overall cost of your entry into AI development, as it gives you the performance needed to utilize today's AI analytics software tools without a GPU.

The Dell data science workstation platforms are purpose-built for data science.

The data science workstation includes Intel AVX instruction sets and the ultimate memory capacity needed for data scientists to iterate on various datasets, visualize, and analyze complex workloads at scale. Data Science tasks need the highest memory

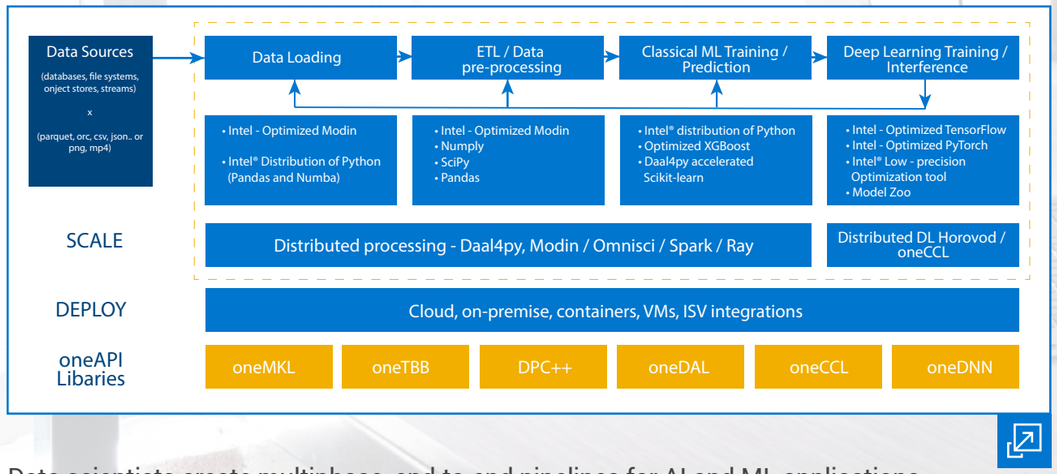
Dell Data Science Workstation platforms are purpose-built for data science. The Data Science Workstation includes Intel AVX instruction sets and the ultimate memory capacity needed for data scientists to iterate on various datasets, visualize, and analyze complex workloads at scale.

Data Science tasks need the highest memory span to fit the dataset in-memory for complex analysis before AI/ML. With Intel Optane Persistent Memory (PMem), these workstations can fit full datasets and perform at a fraction of the cost of a GPU-based, AI-ready workstation.

Intel® AVX instruction sets provide huge performance boosts to many linear algebra-based operations used in data science. With the Intel® AI Analytics software stack, data scientists can perform AI experiments on small machines like laptops and then scale out seamlessly to production servers, IoT and edge devices such as workstations, as well as the cloud.

The newest 3rd Generation Intel® Xeon Scalable processors enhance AI, cloud computing, security, and many other areas. Intel has optimized an array of software tools, libraries, and frameworks so that applications can easily take advantage of the latest hardware advances. The results are impressive.

Figure 1.



Data scientists create multiphase, end-to-end pipelines for AI and ML applications. The phases include data ingestion, data cleaning, data exploration, and feature engineering followed by prototyping, model building, and finally, model deployment.

Figure 2.

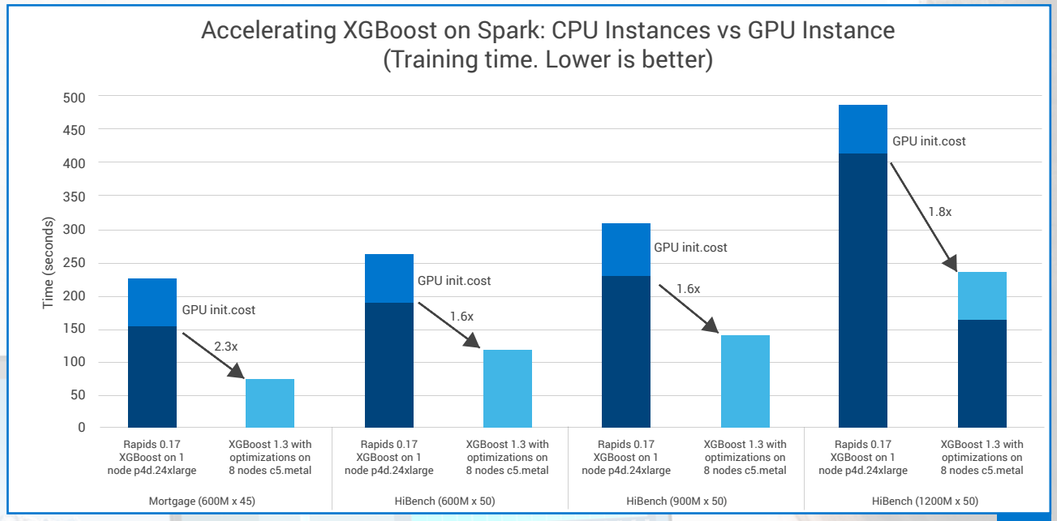
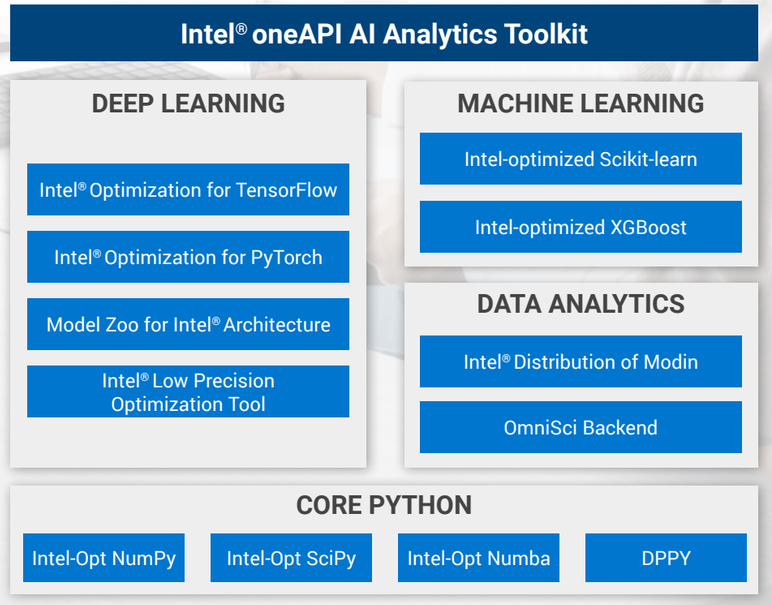


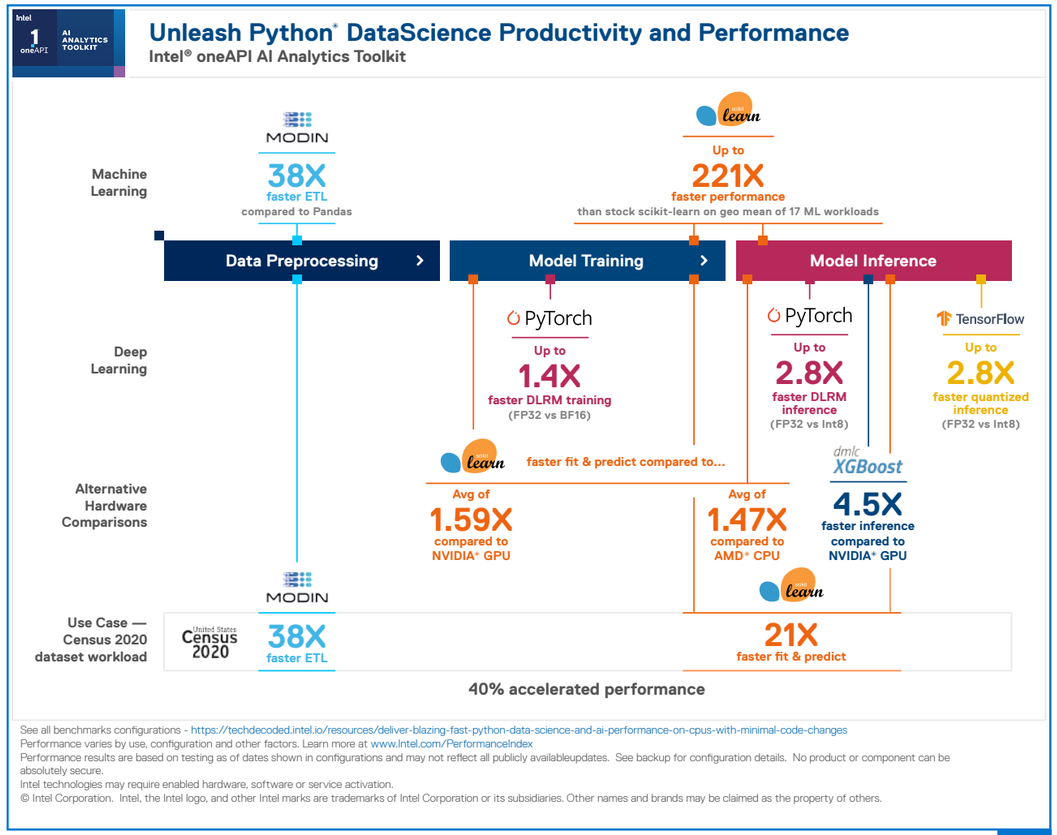
Figure 2 shows a scenario where the optimized XGBoost running on a Spark cluster of eight c5.metal instances (Intel® Xeon Platinum 8275CL processors) outperforms an equivalently priced p4d.24large instance (NVIDIA A100 processors).

Figure 3.



The Intel® oneAPI AI Analytics Toolkit (AI Kit) provides high-performance APIs and Python packages to accelerate the phases of these pipelines.

Figure 4



See all benchmarks configurations - https://techdecoded.intel.io/resources/deliver-blazing-fast-python-data-science-and-ai-performance-on-cpus-with-minimal-code-changes
Performance varies by use, configuration and other factors. Learn more at www.Intel.com/PerformanceIndex
Performance results are based on testing as of dates shown in configurations and may not reflect all publicly availableupdates. See backup for configuration details. No product or component can be absolutely secure.
Intel technologies may require enabled hardware, software or service activation.

Intel® oneAPI Python Data Science Productivity and Performance benchmarks.

Maximise performance with the Intel® oneAPI AI Analytics Toolkit

Achieving peak performance with Python across a variety of architectures is critical to the success of modern data science. Python is a powerful, scalable, and easy-to-use language, but it's not built from the ground up for blazing-fast performance.

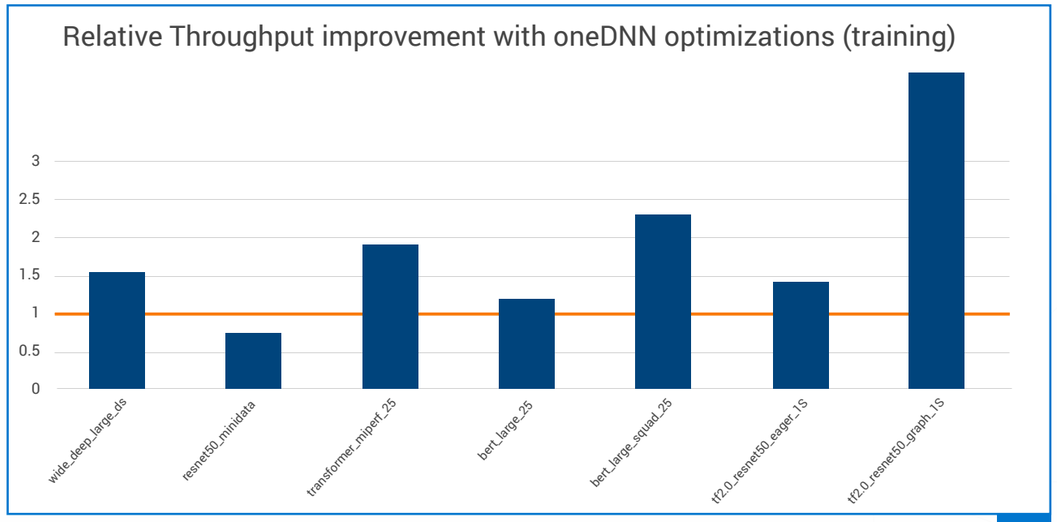
Many developers rely on vendor-specific optimizations for frameworks and libraries such as TensorFlow, PyTorch, or scikit-learn to achieve the speeds they need on their hardware of choice.

However, when new hardware emerges or the market evolves to require deployment on additional hardware types, they have to find new optimizations to get similar results.

That's where the Intel® oneAPI AI Analytics Toolkit (AI Kit) comes in.

The toolkit maximizes performance from preprocessing to ML and DL training and inference, on Dell Technologies and Intel hardware architectures commonly used today, as well as those set to become industry standards in the future.

Figure 5



For training, up to 2.4x performance gains were observed with previous TensorFlow 1.x graph models and the newer TensorFlow 2.x eager execution-based models.

The Intel® oneAPI AI Analytics Toolkit provides optimized Python libraries, deep learning frameworks, and a lightweight parallel dataframe—all built using oneAPI libraries—to maximize performance for end-to-end data science and AI workflows on Dell Data Science Workstations.

The toolkit allows data scientists and AI developers to get the latest Intel deep learning and machine learning optimizations from a single resource with seamless interoperability and out-of-the-box performance.

1. The way forward – Choosing the right Dell Data Science Workstation

Our Dell Precision Data Science workstations come in a variety of form factors to work seamlessly in your environment. It all depends on what your needs are.

Find more configuration information by visiting the Dell Precision Workstation Solutions webpage.

Precision for your Data Science requirements

Precision 7560:
Dell's most powerful 15" mobile workstation is AI ready, with Intel® Xeon® CPUs.

Precision 7760:
The most powerful Precision mobile workstation is AI ready, with Intel® Xeon® CPUs.

Precision 5820 Tower:
Ideal for cognitive solution development and inference applications, with Intel® Xeon® CPUs.

Precision 7920 Tower and Rack:
Handles learning model training and larger solution frameworks with ease, with Intel® Xeon® CPUs and Intel® Optane™ technology.

2. Powering up your performance – Setup and Optimizations (optional)

The following optimizations are recommended to increase the performance and productivity for machine learning and deep learning workloads.

Step 1. Disable idle user power management configuration

To complete this action, we recommend disabling power management settings for Idle User case, since in most cases during installation and development, the machine will download the software/dataset over the internet or run ML/DL workflow unattended for an extended period of time.

Disable power management settings with the command below.

```
/usr/bin/gsettings set org.gnome.settings-daemon.plugins.power sleep-inactive-ac-timeout 0
```

Step 2. Disable Hibernate settings

There is no advantage in enabling “hibernate” on Dell Data Science workstations, quite simply because it is unnecessary and consumes valuable storage space. Disable it as per below.

```
sudo systemctl mask hibernate.target
```

Step 3. Remove swap file

In addition, we recommend removing the “swap file”. Swap file sizes are typically 1.5x the size of installed RAM. On large RAM systems this can be a prohibitive amount of SSD storage.

Step 4. Is swap active?

```
sudo swapon --show
```

If it is, it will show something similar to the below

```
NAME      TYPE   SIZE USED PRIO
/swapfile file 215.3G   0B   -2
```

Step 5. Deactivate the swap

```
sudo swapoff -v /swapfile
```

Step 6. Remove it from /etc/fstab

```
sudo sed -i '/^\/swapfile/d' /etc/fstab
```

Step 7. Delete the actual file

```
sudo rm /swapfile
```

Appendix: Ubuntu Installation Guide

In order to start a fresh install you will need to disable “Secure Boot” in the bios. See steps below”

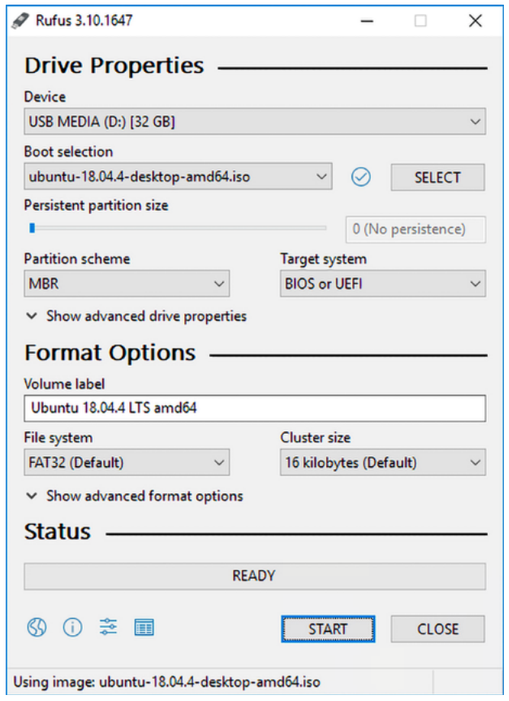
These steps require reformatting the storage drive and will wipe out any existing drive content, so be sure to save any important drive content before starting.

Unless instructed otherwise, do NOT use ‘sudo’, and use a standard user account in a Bash Linux shell (i.e., terminal window).

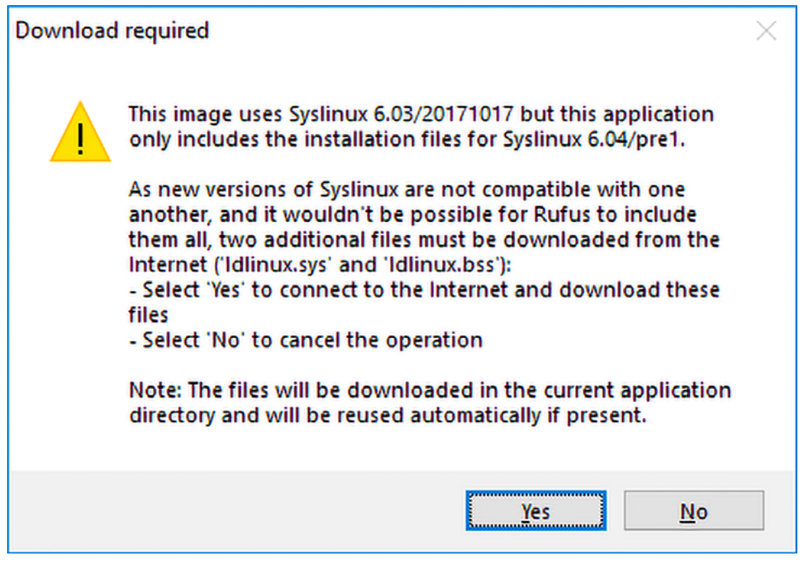
Step 1. Download Ubuntu 20.04 media here.

Step 2. Create installation media

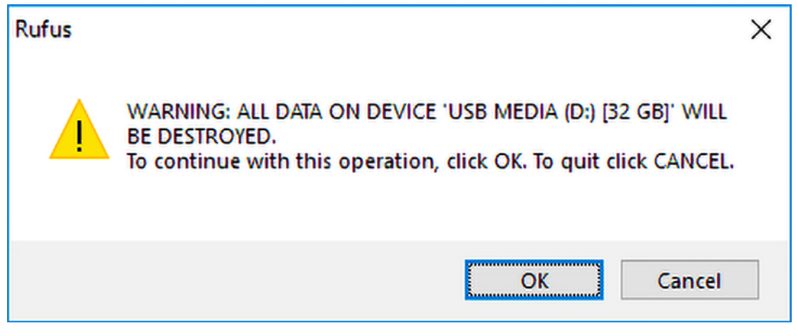
A. Download and run **Rufus**.
B. Insert your USB media device.
C. Configure as shown below and click **Start**.



D. Click Yes.



E. Click OK



Continued →

Appendix: Ubuntu Installation Guide

In order to start a fresh install you will need to disable “Secure Boot” in the bios. See steps below”

Step 3. Configure system BIOS.

A. Power up or reboot system. When the Dell logo appears, immediately tap F2 repeatedly to load the one-time boot menu.

B. Disable Secure Boot
   (1) Settings - Secure Boot - Select Disabled

C. Change SATA setting to AHCI
   (1) Settings - System Configuration - SATA Operation - Select AHCI

D. If yours is a Precision Mobile Workstation, skip this step and go to step 3.0 (e).
   (1) Settings - Advanced Boot Options - Check Enable Legacy Option ROMs
   (2) Settings - Boot Sequence - Select Legacy
   (3) **Note:** This will cause systems equipped with NVMe hard disk carriages to produce a flashing yellow disk health light. This is normal and is the result of using the Legacy boot option.
   (4) **Note:** If this is a Precision Mobile Workstation please keep UEFI selected.

E. Click **Apply, OK, Exit.**

Step 4. Install Ubuntu 20.04.

A. Insert USB installation media device created in Appendix Step 2.
B. Power up or reboot system. When the Dell logo appears, immediately tap **F12** repeatedly to load the one-time boot menu.
C. Under UEFI Boot, select your USB installation media device and press <Enter>.
D. After the installer loads, select preferred language and click **Install Ubuntu**.
E. If you boot into Ubuntu, you will need to double click on the Install Ubuntu 20.04.4 icon.
F. Select keyboard layout and click **Continue**.
G. If asked about WIFI, select **I don’t want to connect to a WIFI network right now**.
H. If asked about updates and other software, select **Minimal Installation**.
I. Uncheck **“Download updates while installing Ubuntu.”**
J. Uncheck **“Install third-party software for graphics and Wi-Fi hardware and additional media formats.”**
K. Click **Continue**.
L. Select **“Erase disk and install Ubuntu.”**
M. Click **Install Now**.
N. Confirm writing changes to disk by clicking **Continue**.
O. Select **time zone** and click **Continue**.
P. Enter name, computer name, username, password and click **Continue**.
Q. Ubuntu will install for the next several minutes. When it finishes, click **Restart Now**.
R. Remove USB installation media device.
S. Log into the system with username and password entered during installation.

Step 5. Proceed with instructions in Dell Data Science Workstation Installation Guide, Section 3 to install the Intel oneAPI Data Science Software stack.

Appendix: Intel Installation Guide

Intel AI Analytics Kit 2021.4 Update 2 for OEMs installation guide.

Intel AI Kit is an accelerated integrated environment of popular Machine Learning Python libraries that is optimized and curated by Intel. It provides Intel optimized versions of packages for the best possible performance on Intel Architecture.

Components of This Toolkit

- PyTorch*: The Intel® Math Kernel Library for Deep Neural Networks (Intel® MKL-DNN) is included in PyTorch as the default math kernel library for deep learning. See this article on the Intel® Developer Zone for more details.
- Intel® Optimization for TensorFlow*: This version integrates primitives from the Intel® Math Kernel Library for Deep Neural Networks (Intel® MKL-DNN) into the TensorFlow runtime for accelerated performance.
- Intel® Distribution for Python*: Get faster Python application performance right out of the box, with minimal or no changes to your code. This distribution is integrated with Intel® Performance Libraries such as the Intel® oneAPI Math Kernel Library and the Intel® oneAPI Data Analytics Library. The distribution also includes daal4py, a Python module integrated with the Intel® oneAPI Data Analytics Library as well as the Python Data Parallel Processing Library (PyDPPL), a light weight Python wrapper for Data Parallel C++ and SYCL that provides a data parallel interface and abstractions to efficiently tap into device management features of CPUs and GPUs running on Intel® Architecture.
- Intel® Distribution of Modin*, which enables you to seamlessly scale preprocessing across multi nodes using this intelligent, distributed dataframe library with an identical API to pandas.
- Low Precision Optimization Tool: Provide a unified, low-precision inference interface across multiple deep learning frameworks optimized by Intel with this open-source Python library.
- Intel® Extension for Scikit-learn*: a seamless way to speed up your Scikit-learn application using the Intel® oneAPI Data Analytics Library (oneDAL). Patching scikit-learn accelerates stock scikit by single line change.

Standard Python installations are fully compatible with the AI Kit, but the Intel® Distribution for Python* is preferred.

Installation
This instance was built using Ubuntu 20.04 LTS. If you are on Windows 11, jump to addendum to setup the environment and come back here to Step 1.

Step 1: Update your system

Launch a terminal (CTRL+ALT+T) and enter:

```
sudo apt update && sudo apt upgrade -y
```

Step 2: After the update a reboot will be required, enter:

```
sudo reboot
```

Note: if you are in WSL2 exit the session and relaunch.

Step 3: Launch a terminal (CTRL+ALT+T) and enter:

```
wget https://registrationcenter-download.intel.com/akdlm/irc_nas/18273/Intel-AIkit-2021.4.1-Linux-x86_64.sh
```

Note: the download is approximately 1.8GB, a stable network connection is desired.

Step 4: Change the permissions of the resulting download so that it is an executable.

```
chmod 755 Intel-AIkit-2021.4.1-Linux-x86_64.sh
```

Should the version change the product page is here.

Step 5: Install the Intel AIkit:

```
./Intel-AIkit-2021.4.1-Linux-x86_64.sh
```

Follow the steps and make sure to choose yes to the final question “Do you wish the installer to initialize the Intel-AIkit”

Step 6: You can close the terminal and launch a new one or enter:

```
source .bashrc
```

This will result in the base conda environment being activated.

Enter:

```
conda env list
```

*Base * /home/user/intel-aikit* should be the result

Enter:

```
conda list
```

You should see a list of packages with many coming from the Intel channel. These are the optimized packages.

Note: This collection of packages has been validated against each other for functionality. At the time of publishing, this is a stable environment and should be treated as the base package environment. Any modifications to this collection should be done in a cloned environment so that changes can be isolated and you can easily get back to a known good.

Appendix: Intel Installation Guide

Intel AI Analytics Kit 2021.4 Update 2 for OEMs installation guide.

Intel AI Kit is an accelerated integrated environment of popular Machine Learning Python libraries that is optimized and curated by Intel. It provides Intel optimized versions of packages for the best possible performance on Intel Architecture.

Getting Started
In this section we will:

- Create a new environment and add some additional packages.
- Clone the Intel samples repository.
- Run a sample project using a Jupyter Notebook.

Create New Environment

Step 1: Launch a terminal (CTRL+ALT+T) and enter:

```
conda create –-clone base –-name
```

for example:

```
conda create -–clone base –-name jupyter
```

Step 2:

```
Conda env list
```

You should see two environments now. The * denotes the active environment.

Activate the new environment:

```
Conda activate jupyter
```

If you enter 'conda env list' the asterisks should be next to the new environment name and the prompt should indicate the environment name as well.

Step 3: Add missing packages.
We need several packages for the getting started samples, let's add those knowing that if we run into issues we can always delete this environment and get back to our clean base environment.
Enter:

```
conda install -c conda-forge jupyterlab
```

Note: there will be some warnings but they are safe to ignore.
Enter:

```
Conda install -c intel neural-compressor
```


Note: this takes a while as conda will check for package conflicts.
Step 4: Clone the Intel oneAPI Samples Repository, Git will likely not be installed so to install it enter:

```
sudo apt install git
git clone https://github.com/oneapi-src/oneAPI-samples.git
current repo size is ~410M
```

Step 5:
Enter:

```
jupyter lab
```

The browser should auto launch. Then on the panel on the left navigate to:

```
~/oneAPI-samples/AI-and-Analystics/Getting-Started-Samples/INC-Sample-for-Tensorflow/
```

Double click:

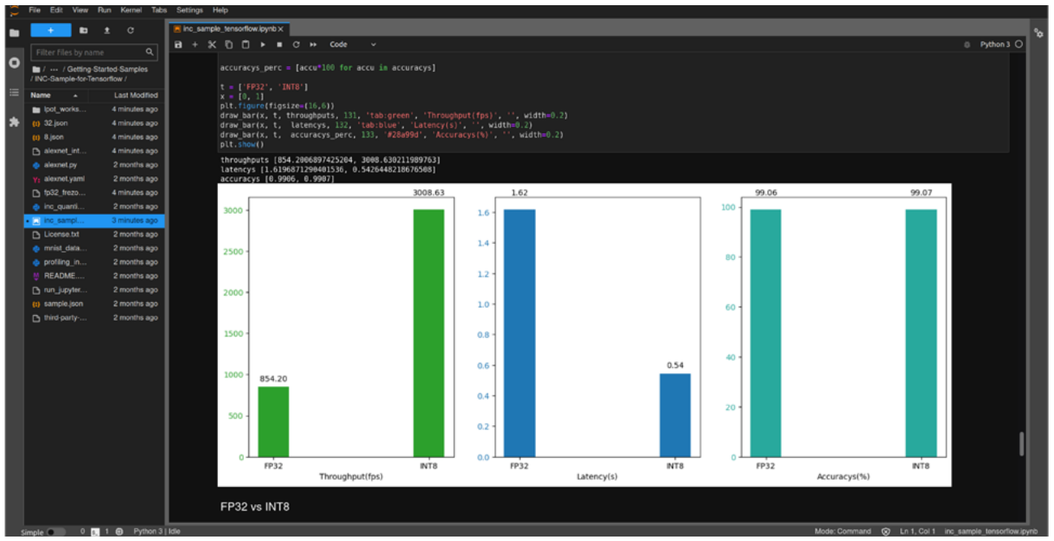
```
Inc_sample_tensorflow.ipynb
```

Appendix: Intel Installation Guide

Intel AI Analytics Kit 2021.4 Update 2 for OEMs installation guide.

Intel AI Kit is an accelerated integrated environment of popular Machine Learning Python libraries that is optimized and curated by Intel. It provides Intel optimized versions of packages for the best possible performance on Intel Architecture.

As you execute the notebook you will see that your Data Science Workstation supports the latest set of instructions and has the ability to take advantage of INT8 as opposed to FP32 increasing throughput and reducing latency while maintaining identical accuracy.



Please explore other Getting Started installation guides with Intel ›

Addendum: WSL2 on Windows 11:

Step 1: Tap the Windows Key and enter update, windows update will be the first match. Go through that process and reboot.

Step 2: Launch windows terminal with administrative privileges. Tap the Windows key and enter terminal. The down carrot on the right will expose the admin privilege option.

Step 3: At the prompt enter: wsl - install (two dashes)
This will automatically install Ubuntu 20.04

Step 4: Restart -- The install actually takes place here and after the reboot, this takes a couple of minutes.

Step 5: Ubuntu will launch automatically, follow the steps to create a new user. It's a good idea to have the identical username for Windows as you do Ubuntu.

Step 6: Jump back up to step 1 and follow along, even the update steps.

Note: a reboot in WSL2 is exiting the WSL2 session. To relaunch WSL2 tap the Windows key, enter terminal and then click on the down carrot and choose Ubuntu. This is a great way to have multiple terminal sessions in a tabbed environment.

References & Resources

References

1. Online installation step by step - https://youtu.be/SXTS_hURiAQ -0
2. Online installation step by step - https://youtu.be/VC_ucczGR_Y
3. (on Intel site) https://software.intel.com/content/www/us/en/develop/tools/oneapi/ai-analytics-toolkit/download.html?operatingsystem=linux
4. Installation guide for Package Managers (YUM, APT, NuGet, etc): https://software.intel.com/content/www/us/en/develop/documentation/installation-guide-for-intel-oneapi-toolkits-linux/top/installation/install-using-package-managers/yum-dnf-zypper.html
5. Installation guide for installing with Conda Package Manager: https://software.intel.com/content/www/us/en/develop/articles/installing-ai-kit-with-conda.html#gs.0lmif8
6. Installation via Docker: https://hub.docker.com/r/intel/oneapi-aikit

Resources

1. Artificial Intelligence (AI) technologies powered by Dell Precision workstations https://www.delltechnologies.com/en-us/ai-technologies/index.htm
2. Here's a link to educational videos for Intel oneAPI AI Analytics Toolkit:
   a. Intel Distribution for Python: https://www.intel.com/content/www/us/en/developer/videos/fast-scalable-analytics-ml-distribution-for-python.html#gs.e9q58g [intel.com]
   b. Intel Distribution for Modin: https://techdecoded.intel.io/essentials/achieve-high-performance-scaling-for-e2e-machine-learning-and-data-analytics-workflows/?elq_cid=3630800_ts1634864550409&erpm_id=5537033_ts1634864550409#gs.e9q4km [techdecoded.intel.io]
   c. Intel Optimization for PyTorch: https://techdecoded.intel.io/essentials/optimize-the-latest-deep-learning-workloads-using-intel-optimized-pytorch/?elq_cid=3630800_ts1634865655884&erpm_id=5537033_ts1634865655884#gs.e9qlev[techdecoded.intel.io]
   d. Intel Optimization for TensorFlow: https://techdecoded.intel.io/essentials/intel-optimization-for-tensorflow-tips-tricks-for-aihpc-convergence/?elq_cid=3630800_ts1634865961834&erpm_id=5537033_ts1634865961834#gs.e9qpir [techdecoded.intel.io]
   e. Intel Optimization for XGBoost: https://techdecoded.intel.io/essentials/maximize-your-cpu-resources-for-xgboost-training-and-inference/?elq_cid=3630800_ts1634866041986&erpm_id=5537033_ts1634866041986#gs.e9qqo8 [techdecoded.intel.io]

DELL Technologies

intel

Learn more about Dell Precision workstation solutions

Contact a Dell Expert

intel 1 oneAPI AI ANALYTICS TOOLKIT