

Old Training Material Front Page

Welcome to the [C3.ai](#) Digital Transformation Institute!

You have been given a grant as part of the new [C3.ai](#) Digital Transformation Institute (DTI)! To make the start of your DTI experience as fast as possible, we have assembled a set of resources to:

1. Introduce researchers of all stripes to the C3 AI Suite.
2. Help researchers determine what level of training they will need to leverage [C3.ai](#)'s resources.
3. Point researchers directly to relevant documentation they will need.
4. Provide worked examples of different research workflows, and how they may be ported into the [C3.ai](#) environment or may use [C3.ai](#)'s resources.

If you have questions not covered by this guide, please contact the DTI team at the email help@c3dti.ai.

Introduction to the C3 AI Suite

The C3 AI Suite is a data analytics engine designed to make the ingestion and analysis of heterogeneous data sources as painless as possible. The C3 AI Suite joins data from multiple sources into a single unified federated data image. With the federated data image defined, C3 AI then provides an API to access that data and, in the case of time-series data, perform numerous transformations and computations, all producing normalized time-series data at regular intervals.

If you want more background on the C3 AI Suite, there is a one-hour [C3.ai](#) DTI webinar describing its capabilities [here](#).

C3 AI also supports R and [Python Jupyter notebook](#) analysis of the federated data image. These notebooks provide a great way for researchers to analyze data close to where the data is stored. While C3 AI supports many data science capabilities familiar to researchers, some expected functionality may be missing. For these cases, C3 AI supports implementing new data processing functions in Python and JavaScript.

Like any other API, porting your own workflows will take some care and time to learn properly. Please leverage this guide to make understanding the C3 AI Suite and porting your workflow as quick and easy as possible.

Services available from C3 AI

- [Covid-19 Datalake](#): This unified federated Datalake includes data from numerous sources.
- C3 AI computing platform.
- [C3 AI Integrated Development Studio](#).
- C3 AI Jupyter Notebooks.
- [C3 AI Marketplace](#).
- [C3 AI UI Framework](#): Useful for creating dashboards.

How does C3 AI differ from traditional HPC systems?

- Traditional HPC systems are similar to Hardware as a Service (HaaS), while C3 AI is more like a Platform as a Service (PaaS). Users are encouraged to work within the platform's API to achieve the best performance out of the C3 AI Suite.
- C3 AI offers a state-of-the-art data integration system as the basis for all Data Science operations. This is in contrast to HPC systems where all components of data management and the analysis pipeline must be installed and managed independently.

What types of software can be run on C3 AI?

- Nearly any Python module may be installed and used through pip or conda.
- Nearly any R package may be installed and used within the R Jupyter environment.

What types of software cannot be run on C3 AI?

- General binary executables are not supported by C3 AI out of the box.
- MPI-based Python software.
- Packages which must be built from scratch on the platform or require specific hardware drivers.
- Python modules which require special built binaries may not run as well.

What training resources are available?

See [this page](#) for a comprehensive list and categorization of the available training materials. This includes [C3 AI Documentation](#), DTI introductions, and DTI created examples and exercises.

How do I get started?

Use this guide to determine what training you need to utilize C3 AI resources effectively. We have identified four levels of usage of the C3 AI Suite. For each we include basic examples of workflows which might fall into that level, pros and cons of operating on that level, and a list of training resources we recommend researchers complete on the DTI training environment before starting their C3 AI allocations. This will ensure researchers will be able to use their allocation as efficiently as possible.

Examine the high level overviews of each level below then click the section titles to go to more in-depth discussions related to each level, like the recommended training.

Level 1: Use Public C3 AI COVID-19 Data Lake Only

For many researchers, accessing the public API for the COVID-19 Federated Data Image will be enough for their research goals. The public API provides fetch access to many datalake objects, metrics access to some time series data such as case data, and allows you to pull local copies of those objects and metrics results into your local compute environment.

C3 AI Suite access is not needed for Level 1!

Level 2: Full C3 AI COVID-19 Data Lake Access

Full access to the datalake offers access to all stored COVID-19 Datalake data while still allowing the researcher to use whatever analysis framework they choose with their own compute resources. This level offers the fastest startup time while still ensuring access to all data. Once you learn how to query data in C3, that data can be streamed to your compute resources where you can use your language and tools of choice.

Level 3: Define and Use C3 Types to Integrate Data

Some researchers will want to write their own C3 AI package and leverage more of the C3 AI Suite. C3 AI allows researchers to define their own types and methods to integrate their data into the C3 AI Suite – either independently or alongside the COVID-19 Datalake. This allows researchers to use C3 AI data analytics methods, such as timeseries metrics, just as they would on other Datalake data. Researchers will also have the ability to share their data with other researchers in the DTI by sharing their package. Adding another researcher's package as a dependency to your package will also bring another researcher's data into your package as well.

Level 4: Advanced C3 AI Platform Usage

Some researchers will want to bring state-of-the-art ML workflows to C3 AI. C3 AI can support such workflows but extra work may be needed.

Am I Ready to use the C3 AI Suite?

Follow our [DTI Readiness Checklist](#) to determine if you're ready to get started using the C3 AI Suite.

Covid-19 Datalake

As part of the initial [C3.ai](#) DTI, C3 AI is curating the Covid-19 Datalake. Follow the link above for more detailed information about this Datalake.

Accessing C3 AI

This section introduces the process to access C3 AI resources. Generally speaking, once you receive your grant, the DTI team will reach out and discuss with you what your needs are. The process will be:

1. Determine which researchers will require access to a C3 AI environment.
2. Each researcher will be given a [C3 AI Developer Portal](#) login.
3. Each research group will be given a tag on the [C3.ai](#) DTI training cluster.
4. Once training is complete, discuss with the DTI team what your needs for a C3 AI cluster will be.
5. The [C3.ai](#) DTI will work with C3 AI to stand up a new tag for your research.
6. Access to that tag will be granted to your researchers.
7. Research can then proceed until your allocation is exhausted!

DTI Office Hours

In addition to the materials in this guide, the DTI will also hold regular office hours to field any question from researchers, and perform tutorials and demonstrations. Please see the link above for the current schedule.

DTI Training Videos

We've recorded videos describing the C3 AI Suite concepts and usage. You can find these videos throughout our documentation. This page is a collation of all the videos in one place.

C3 AI Allocation Management (PLANNED)

This section introduces how researchers will be expected to manage their allocation while on the C3 AI Suite.

This section will be expanded once the DTI team understands how this procedure will look to the researcher.

Special Compute Resource Information

Here you can find information about the special compute resources available to [C3.ai](#) DTI researchers.

Help! This guide doesn't solve my problem!

No problem! Please send an email to help@c3dti.ai with a description of your issue and one of our team will work with you to resolve it.

Feedback

If you feel aspects of this guide are incomplete or inaccurate, please send an email to help@c3dti.ai with the issue or suggestion, and we will work to incorporate it to make the documentation better. We appreciate the new perspective more eyes can bring to a software project!

Your DTI Team

NCSA

Jay Roloff - Executive Director

Matthew Krafczyk - Data Analyst

Yifang Zhang - Data Analyst

Berkeley

Larry Rohrbach - Executive Director

Eric Fraser - Chief Technology Officer

Greg Merritt - DevOps Lead

Matt Podolsky - Managing Director of Research Technology