

Virtual Dataset - VDS

The purpose of this page is to briefly describe the new HDF5 Virtual Dataset (VDS) feature and provide a gateway to available documentation. The page includes the following sections:

- [Virtual Dataset Overview](#)
- [Virtual Dataset User and Resource Documents](#)
- [HDF5 Library APIs](#)
- [Expected Updates and Additional Documentation](#)
- [Tools](#)
- [Virtual Dataset Design](#)

Virtual Dataset Overview

With a growing amount of data in HDF5, the need has emerged to access data stored across HDF5 files using standard HDF5 objects, such as groups and datasets, without rewriting or rearranging the data.

While the ability to build hierarchical structures across existing HDF5 files has been available in HDF5 for quite some time through the mounting and external link features, the ability has not existed to present data stored in several HDF5 datasets and files as a single HDF5 dataset and to access the data via HDF5 APIs without rewriting and rearranging the data.

To address this, The HDF Group has implemented a new feature called the HDF5 Virtual Dataset (VDS).

The feature is a logical next step in the development of HDF5 that enables HDF5 users to access and work with data stored in a collection of HDF5 files using well-known tools and existing HDF5 applications and higher-level libraries such as h5py, MATLAB, and IDL without changing the way the data is collected and stored.

The following examples illustrate situations that will benefit from use of virtual datasets:

- Synchrotron centers such as DLS and DESY will be generating and storing terabytes of experimental data per day in HDF5 files. Because of the nature of the experiments and hardware constraints, the data representing, for example, an X-ray image will be stored across different HDF5 datasets in multiple HDF5 files. With VDS, the whole image may be accessed by an application without any specific knowledge of where data for each part of the image is stored.
- Climatologists who study and analyze climate variations (temporal changes at a given location) will be able to use the VDS feature to describe and access “data rods” – data of interest stored in a series of HDF5 files organized by time stamps – without rewriting the data into new HDF5 file. The “data rods” will be accessible as a regular HDF5 dataset via their applications without any special knowledge “coded” into the applications.

Virtual Dataset User and Resource Documents

HDF5 VDS User's Guide (*This document is not yet available.*)

Until the HDF5 VDS User's Guide becomes available, users may find the following resources helpful:

[RFC: HDF5 Virtual Dataset](#) (PDF) Includes several sections illustrating the use of virtual datasets (VDS) and discussing the VDS programming model, some feature constraints, and several use cases.

Note: The current version of this document reflects the design, strategies, and general approach employed in the VDS feature, but the API implementation had to be modified from the specification. An expected update will correct this divergence.

HDF5 Library APIs

New VDS Functions

H5P_SET_VIRTUAL	Sets the mapping between virtual and source datasets
H5P_GET_VIRTUAL_COUNT	Retrieves the number of mappings for the virtual dataset
H5P_GET_VIRTUAL_VSPACE	Retrieves a dataspace identifier for the selection within the virtual dataset used in the mapping
H5P_GET_VIRTUAL_SRCSPACE	Retrieves a dataspace identifier for the selection within the source dataset used in the mapping

H5P_GET_VIRTUAL_DSETNAME	Retrieves the name of a source dataset used in the mapping
H5P_GET_VIRTUAL_FILENAME	Retrieves the filename of a source dataset used in the mapping
H5P_SET_VIRTUAL_PRINTF_GAP	Sets maximum number of missing source files and/or datasets with printf-style names when getting the extent of an unlimited virtual dataset
H5P_GET_VIRTUAL_PRINTF_GAP	Returns maximum number of missing source files and/or datasets with printf-style names when getting the extent for an unlimited virtual dataset
H5P_SET_VIRTUAL_VIEW	Sets the view of the virtual dataset to include or exclude missing mapped elements
H5P_GET_VIRTUAL_VIEW	Retrieves the view of a virtual dataset

Supporting Functions

H5S_IS_REGULAR_HYPERSLAB	Determines whether a hyperslab selection is regular
H5S_GET_REGULAR_HYPERSLAB	Retrieves a regular hyperslab selection

Modified Functions

H5P_SET_LAYOUT	Specifies the layout to be used for a dataset Virtual dataset, H5D_VIRTUAL, has been added to the list of layouts available through this function.
H5P_GET_LAYOUT	Retrieves the layout in use for a dataset Virtual dataset, H5D_VIRTUAL, has been added to the list of layouts.

Expected Updates and Additional Documentation

The following additional documentation will be posted as it becomes available:

- Update: "RFC: HDF5 Virtual Dataset" (see below).
The current document reflects the design, strategies, and general approach employed in the VDS feature, but the API implementation had to be modified from the specification. The update will correct this divergence.
- HDF5 VDS User's Guide material
- Presentation materials describing the VDS feature

Tools

No new tools are necessary to examine or manipulate virtual datasets. Where necessary, existing HDF5 tools have been updated to be aware of the new properties, but tool operations on virtual datasets will be essentially transparent to the user.

Virtual Dataset Design

The Virtual Dataset design document below describes feature requirements, how the feature works, and why design choices were made.

RFC: HDF5 Virtual Dataset (PDF)	<p>This document describes requirements that guided development of the Virtual Dataset (VDS) feature, feature constraints, several use cases, the VDS programming model, and some details of the implementation. This document contains useful illustrations that provide an intuitive understanding of virtual datasets.</p> <p>Note: The current version reflects the design, strategies, and general approach employed in the VDS feature, but the API implementation had to be modified from the specification. An expected update will correct this divergence.</p>
---	--