**What is CDF?**

| **Slide #** | **Audio/Narrator** | **Graphics** |
| --- | --- | --- |
| 1 | Title page |  |
| 2 | Welcome to module 4, “What is CDF?”  In this module you will learn:  What CDF is?  How is CDF used?  Where to go to for more information on CDF?  Why use CDF?  How is CDF file organized? | I would imagine we would have the same opening for every lesson. |
| 3 | What is CDF?  CDF is an Acronym for NASA’s Common Data Format. | C <- ommon  D <- ata  F <- ormat |
| 4 | CDF, is basically a conceptual data object for storing, accessing, and manipulating multidimensional data sets. We refer to CDF as a data object because we describe the form of the data sets and the means (interface) by which the sets may | Start with a single cube and build |
| 5 | Note: the use of a data object in no way inhibits access to physical data. It merely provides a way of generalizing the data model and makes possible a uniform method for constructing and manipulating a data set. The data provides for conceptual simplicity while running on any machine or device. |  |
| 6 | The contents of a CDF fall into two categories: Metadata and Data Objects. The first is a set of attribute entries (metadata) describing the CDF in global terms or specifically for a single variable. The second is a series of Data Objects is a collection of variables consisting of scalars, vectors, and n-dimensional arrays. Both the metadata (attributes) and the data objects (variables) are combined into an integrated data set. | Metadata  CDF 3.x (attributes)  model  serial number  session start  Data Objects (variables)  Go with weather theme |
| 7 | How Is CDF used?  CDF (data) dates back to the development of the NASA Climate Data System at the National Space Science Data Center (NSSDC). As such, it had three main requirements.  1. Combine the Metadata and Data Objects into a single file.  2. Use standard common (metadata) to describe the data sets.  3. Develop higher-level applications |  |
| 8 | There are two classes of users for CDF: Producers and Consumers. Producers perform primarily data acquisition and are mainly involved in designing CDF files and the associated CBM metadata. Consumers build high-level applications which interact with CDF files at the programming level and at the toolkit level. | Producers  warehouse  Consumers  Java MatLab |
| 9 | The CDF toolkit provides a suite of utilities for creating, browsing, and modifying CDF files as well as exporting or importing CDF data to/from a regular text file or an eXtensible Markup Language (XML) file. The browsing tools allow a quick look at CDF data sets and aid in CDF validation. | Re-use 2a  CDFedit  CDFexport  CDFdump  CDFmerge  CDFvalidate  And others |
| 10 | The CDF library comes with C, Java and Fortran interfaces and provide the framework on which graphical and data analysis applications can be created.  The CDF library allows developers of CDFbased systems to easily create applications which permit users to slice data across multidimensional arrays, access entire data structures, perform data subsampling, and access one data element independently of each other. | ? catalog ? |
| 11 | CDF data sets are portable on any of the CDF-supported platforms. These currently include:  Solaris  IBM RS6000 series  HP 9000 series,  PC  Macintosh  And others |  |
| 12 | CDF is supported by commercial and open source data analysis/visualization software such as IDL, MATLAB, and IBM’s Data Explorer (XP). For those familiar with IDL or MATLAB, a user can easily create sophisticated plots from CDF files instead of writing lengthy programs. | MatLab  2a |
| 13 | Where to go to for more information on CDF?  (Hopefully, the answers to the first two questions have provided a basis for answering this question.) If you still have questions or would like to learn more about CDF, please refer to the CDF Frequently Asked Questions (FAQ) page (http://cdf.gsfc.nasa.gov/html/FAQ.html) for more detailed information about CDF. (It is important to understand that CDF has been designed to solve a number of data management and storage problems and has shown itself to be quite flexible in storing a wide variety of data sets.) | NASA CDF homepage background |
| 14 | Why use CDF?  Well one important reason is that the Department of the Army, Deputy Chief of Staff, G-4 includes CDF in the CBM+ Sustainment Implementation Guide dated 8 October 2010. In support of that guidance, CDF has been included in the DoD Information Technology Standards Registry (DISR). | Fly in G-4 document and shrink |
| 15 | When people first hear the term CDF, they intuitively think of data formats in the traditional sense of the word (i.e., messy/convoluted storage of data on disk or tape).  CDF is more than just a format. CDF is a "self-describing" format for managing data. In addition to the actual data being stored, CDF also stores user-supplied descriptions of the data, known as metadata. This self-describing property allows CDF to be a generic, data-independent format that can store data from a wide variety of disciplines. | Show typical file and records and then show the ‘don’t’ symbol.  Metadata  CDF 3.x (attributes)  model  serial number  session start  Data Objects (variables)  Go with weather theme |
| 16 | In addition to being a self-describing data format, CDF is also a software library. The library allows a user to randomly access and manage data, and metadata without regard to their physical storage. This completely relieves the user of low-level I/O operations allowing more time for data analysis. The actual format used to store the data and metadata is completely transparent.  If an application is written in Java, it can be executed without any modifications on any of the Java supported platforms. | List of tools  Maybe a action for each of the major tools |
| 17 | The term "CDF" is also used to refer to the physical files that the CDF library generates. A data set stored using the CDF library is called a "CDF File". | Standard  tools  And files |
| 18 | How is CDF file organized?  An important feature of CDF is that it can handle data sets that are inherently multidimensional in addition to data sets that are scalar. To do this, CDF groups data by "variables" whose values are conceptually organized into arrays. CDF's "variable" is a generic name or an object that represents data, and it does not have any scientific context associated it. |  |
| 19 | For example, a variable can be data representing an independent variable, a dependent variable, time and date value, or whatever data might be (e.g. image, XML file, etc.). In other words, the variables contain no hidden meanings other than the data itself. |  |
| 20 | One may describe a variable or one variable's relationship with other variable(s) through "attributes" |  |
| 21 | How many dimensions the data occupies is determined by the user. For scalar data, as an example, the array of values would be 0-dimensional (i.e., a single value); whereas for image data the array would be 2-dimensional. Similarly, the array for volume data would be 3-dimensional. CDF allows users to specify arrays of up to ten dimensions. | Variable records |
| 22 | The array for a particular variable is called a "variable record." A collection of arrays, one for each variable, is referred to as a "CDF record." A CDF can, and usually does, contain multiple CDF records. This is useful for data with repeated observations over time. | CDF records |
| 23 | Two types of variables may exist in a CDF:  rVariables  zVariables  Every rVariable in a CDF must have the same number of dimensions and sizes.  zVariables may have a different number of dimensions and/or sizes.  Since zVariable is more efficient in terms of storage and offers more functionality than rVariable, use of zVariables is strongly recommended. | rVariable zVariable |
| 24 | The advantage of CDF is that the Data values can be organized in whatever way makes sense to a user.  It is important to note that there is no single "correct" way to store data in a CDF.  The user has complete control over how the data values are stored in the CDF depending on how the user views the data. | Producer 1 Produce 2 |
| 25 | CDF’s “attribute” is a mechanism for describing the CDF file and the individual CDF variables in the file. There are two types of attributes in CDF: global attribute and variable attribute. Global attribute is used for describing the platform in the CDF file and the variable attribute is used for describing CBM sensor data. | Modify for Army CBM theme |
| 26 | Examples of global attributes include such things as file creation date, file author, source of data, and data set documentation. | Metadata  CDF 3.x (attributes)  model  serial number  session start |
| 27 | Examples of variable attributes includes such things as a field name for the variable, the valid minimum and maximum, the units in which the variable data values are stored, the format in which the data values are to be displayed. | Data Objects (variables)  Go with weather theme |
| 28 | CBM File Features (include) :  Compatibility  Checksum  Compression  Data Encoding  Sparseness  Variable Data Methods  File Format Options |  |
| 29 | Compatibility with Previous CDF Releases  One of the CDF 3.0 requirements was an ability to create files bigger than 2G bytes. However, CDF 3.\* can read files that are created with any of the previous CDF releases. | CDF 3.x  > 2GB  CDF 2.7 don’t symbol 3.x  CDF 3.x read 2.7 and earlier |
| 30 | Checksum  To ensure the data integrity in a CDF file, the checksum option has been added. This is a form of redundancy check, a very simple measure for protecting the integrity of data by detecting error in data that is sent through space or time. | Show a hacker trying to change the CBM file in transit from producers to consumers. |
| 31 | Compression  File compression occurs transparently to the user. When a compressed CDF is opened, the CDF library automatically decompresses it. An application does not have to even know that a CDF is compressed. The user or application doesn’t even have to know a file is compressed.  When a compressed CDF is closed by an application, it is automatically recompressed as it is written back to disk. | Checked compression box  Larger file  shrinks  to small file |
| 32 | The CDF library supports several different compression algorithms. When compression is specified for a CDF or one of its variables, the compression algorithm to be used must be selected.  Run-Length Encoding  Huffman  Adaptive Huffman  GZIP |  |
| 33 | The individual variables of a CDF can also be compressed.  The CDF library handles the compression and decompression of the variable values transparently. The application does not have to know that the variable is compressed as it accesses the variable's values.  There are the compression algorithms: |  |
| 34 | The CDF format includes additional options which are outside the scope of this document.  Data Encoding  Sparseness  Variable Data Access Options  File Format Options | Unix versus PC |
| 35 | This completes module 4, “What is CDF?”  In this module you should have learned:  What CDF is?  How is CDF used?  Where to go to for more information on CDF?  Why use CDF?  How is CDF file organized?  You may either review parts of the information, or move on to the next module. |  |
| 36 | User Comments Page:  I could have any comments or questions please send them to  logsa.ccbmdw@conus.army.mil  referring the module name, version and date.  Thank you.  Click next to continue. |  |