MADIS-T, Satellite Radiance and Sounding Thread using Virtualization

Abhishek Agarwal[1], Shahram Tehranian[1], James McNitt[2], Joseph Facundo[2], Anand Swaroop[1]

[1]Nortel Government Solutions (NGS)
[2]National Weather Service (NWS), National Oceanic and Atmospheric Administration (NOAA)

Abstract**—** Meteorological Assimilation Data Ingest System (MADIS) is a complex data assimilation and distribution system developed by the National Oceanic and Atmospheric (NOAA) Office of Atmospheric Research to ingest, process quality control and distribute NOAA and non-NOAA surface observational data. Sources of data include networks for surface observations, upper air soundings, aircraft, remote sensing observations from both ground-based and satellite systems and profilers. MADIS has reached a point where it can now be transitioned from research into an operational environment under NOAA Policy on Transition of Research to Operations (NAO 216-105, dated May 31, 2005). NOAA's National Weather Service (NWS) plans to transition MADIS to the NWS's Telecommunications Gateway (NWSTG) in Silver Spring, Maryland. This transitioned system will be called MADIS Transition (MADIS-T). In this paper we present one of the subsystems within MADIS-T which processes satellite radiance and sounding data. MADIS-T satellite radiance and sounding data processing thread receives NOAA Polar Orbital Environmental Satellites (POES) data through NWSTG. The MADIS-T satellite dataset consists of multiple products from different providers which are integrated into a single dataset. In this paper we also discuss the use of virtualization technology for the migration of the current code base from the existing hardware to new blade architecture. Virtualization eases the migration of the legacy code into new architecture with minimal changes. Virtual machine technology also allows a single hardware platform to host multiple operating systems and applications, reducing the number of physical servers required while maximizing the utilization of each server. This results in fewer wasted CPU cycles, and a reduction in processors bringing power savings, management simplification, as well as enhanced security. Thus, we demonstrate the use of virtualization to migrate the current implementation of the satellite radiance and sounding subsystem from a 32-bit environment to the 64-bit blade architecture.

Index Terms **— Data Assimilation, MADIS, Data Dissemination**

I. Introduction

MADIS-T is an advanced data acquisition and distribution system developed by NOAA's Office of Atmospheric Research (OAR) with many innovative functions including: data ingest, decoding, integration, processing and transmission of NOAA and non-NOAA meteorological observations across the country and ocean-based, many types of upper air soundings. MADIS-T will process a subset of observations which are currently processed by MADIS. Observational datasets which will be processed and made available by MADIS-T system include:

- Radiosonde soundings
- Automated aircraft reports
- NOAA and non-NOAA wind profilers
- Non-NOAA experimental microwave radiometer observations
- Operational and experimental satellite observations and products
- Several types of surface datasets, including 20,000 Mesonet stations from local, state, and federal agencies
- Snow network data

The MADIS-T system will receive raw data from many different data providers in a variety of formats with observations reported in various meteorological units from stations reporting from different time zones. The MADIS-T system will read these data files, integrates them with other NOAA and non-NOAA observations. MADIS-T system will also provide encoding of data into a uniform format and convert all observations to standard observation units and time stamps. MADIS-T data files will be made available in uniform formats with uniform quality control (QC) data in NetCDF is compatible with the National Weather Service (NWS) Advanced Weather Interactive Processing System (AWIPS) systems. A series of flags will be included indicating the quality of the observation from a variety of perspectives (e.g. temporal consistency and spatial consistency), or more precisely, a series of flags indicating the results of various QC checks. MADIS-T users can design their application to then inspect these flags and decide whether or not to use the observation.

II. Concept of Data Processing Threads and Virtualization

The MADIS-T system will collect, organizes and perform quality check on the data, and then distribute the data for use by service providers and other members of the network. Figure 1 shows the high level functional components of MADIS-T middleware, including the data processing thread, Meta data information bases, the local data manager (LDM) function, storage, and the NWS Telecommunications Gateway (NWSTG). MADIS-T is conceived to expand into a number of data processing threads – each one tailored to include end-to-end processing for each individual dataset.

Within the conceptual architecture of the MADIS-T system,

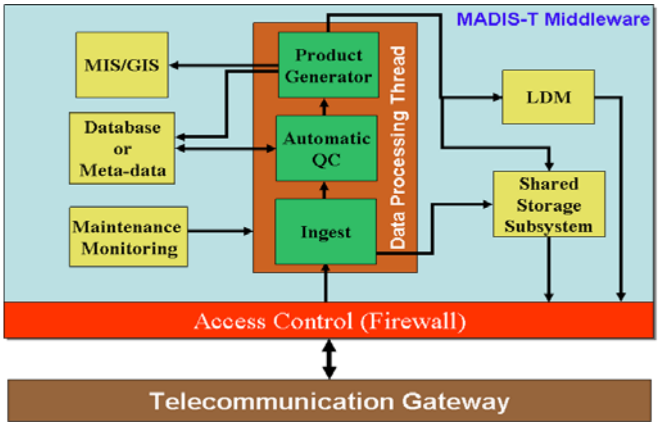

Fig. 1. High level functional components of MADIS-T System

each subsystem is represented as a *"data processing thread"* and includes end-to-end processing of a specific dataset. Each data processing thread consists of preprocessor, ingest, decoding, data processing, quality control (QC), and data delivery modules. Each thread can utilize multiple virtual machines (VM) and hardware can be added as required to host these VM's. This provides flexibility to increase number of stations processed for each data thread as new stations are installed. Figure 2 shows the various components of a data processing thread.

All observations will be received in the NWSTG. Once the data is received within the MADIS-T system, the data is ingested and sorted to a common exchange format. This data is then decoded using the decoder functions. Once the data is decoded, the thread creates standard NetCDF files for each observation. These input NetCDF files are made available immediately for delivery to customers. These files also serve as input to the QC processing. Based on the number of datasets, there can be two levels of QC processing categorized as static or dynamic QC processing. Once the data is quality controlled, the data is made available to the data delivery subsystem so that it can be sent out for distribution through NWSTG.

Virtualization is a broad term that refers to the abstraction of computer resources. This includes making a single physical resource, such as a server, an operating system, an application, or storage device appear to function as multiple logical resources; or it can include making multiple physical resources, such as storage devices or servers appear as a single logical resource. Virtualization is performed on a given hardware platform by a host software which creates a simulated computer environment, a virtual machine for its guest software. The guest software, which is often itself a complete operating system (OS) runs just as if it were installed on a stand-alone hardware platform. Virtualization provides isolation and encapsulation. Virtual machines are completely isolated from the host machine and other virtual machines. If a virtual machine crashes, all others are unaffected. Data does not leak across virtual machines and applications can only communicate over configured network connections. The concept of virtualization can be used for

- **Server Consolidation** where virtual machines are used to consolidate many physical servers into fewer servers, which in turn host virtual machines.
- **Storage Clustering and Virtualization** which gives the ability to add additional compute resources that

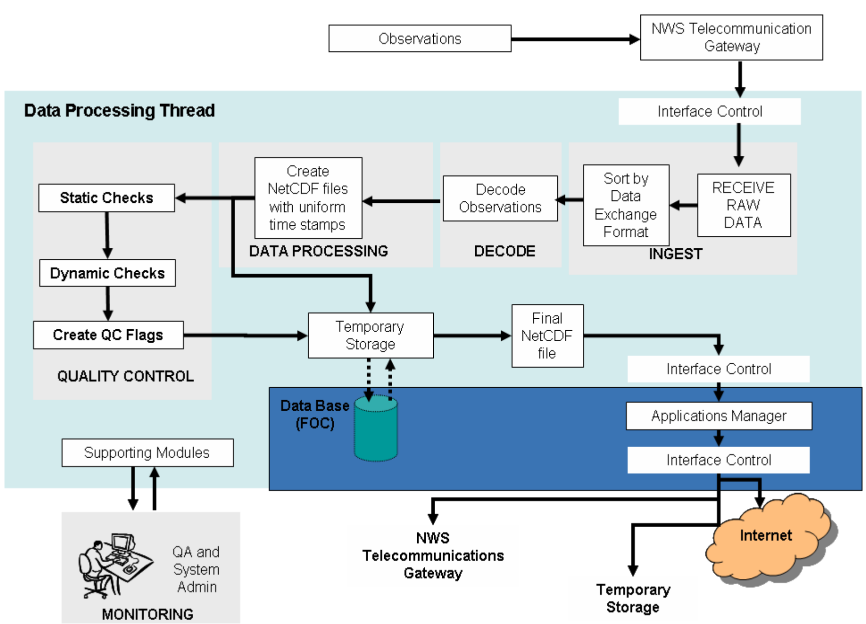

Fig. 2. MADIS-T data processing thread system functions

share the same data and enables organizations to more easily scale-out their environment to meet changing needs. It also enables an organization to provide an opportunity to build a more cost-effective, highly available, application environment.

- **Network Virtualization** which entails load balancing TCP and UDP traffic to and from applications to provide high availability, increased performance, and transparent failover.
- **Disaster Recovery** where virtual machines can be used as *hot standby* environments for physical production servers. Backup images that can *boot* into live virtual machines, capable of taking over workload for a production server experiencing an outage.
- **Development and Test** to create archived libraries of virtual machines, use virtual network segments or virtual switches to network virtual machines together for more efficient and lower cost software development and testing.

MADIS-T will use blade server and virtualization technology to migrate the current system functionality to the NWSTG environment. Figure 3 shows an illustration of MADIS-T virtualized architecture within a blade clustering environment. This also demonstrates how we can create a pool of resources and use as required by the application.

This architecture provides additional advantages by using virtualization software for replication and migration of virtual machines. These benefits address scalability and manageability needs, which are critical for data centers faced with the requirement to grow. Seamless scalability is a primary benefit of implementing virtualization on blade systems. As additional server blades are added to the chassis, virtual machines can be moved to newly added server blades, thereby enabling dynamic provisioning of workloads.

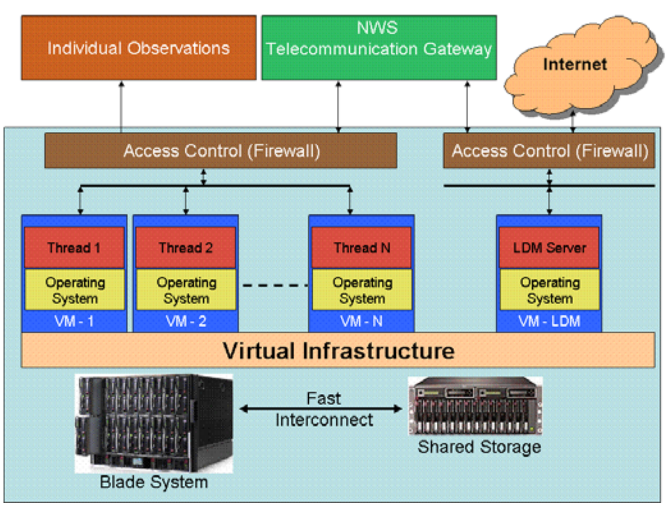

Fig. 3. Illustration of the MADIS-T virtualized architecture using Blade System

III. MADIS-T SATELLITE RADIANCE AND SOUNDING DATA PROCESSING THREAD

The MADIS-T satellite radiance and sounding dataset consists of multiple products from different "providers" that are integrated into a single dataset. Currently, these products consist of radiances (also called brightness temperatures) from the NOAA Polar Orbital Environmental Satellites (POES). Through the use of various mathematical techniques the radiances can be used to calculate vertical profiles of temperature and moisture. MADIS-T satellite radiance and sounding data processing thread also provide users with soundings (i.e. temperature and moisture profiles) generated by ATOVS (Advanced TIROS Operational Vertical Sounder) processing system

The POES system offers the advantage of daily global coverage, by making nearly polar orbits roughly 14.1 times daily. Since the number of orbits per day is not an integer the suborbital tracks do not repeat on a daily basis, although the local solar time of each satellite's passage is essentially unchanged for any latitude. NOAA maintains a four satellite configuration (morning -- METOP-A primary, NOAA-15 secondary, NOAA-17 backup, and afternoon -- NOAA-18 primary, NOAA-16 secondary), allowing NOAA to obtain data five times daily in any one location.

In this section we will now describe various components of MADIS-T Satellite radiance and sounding data processing thread. Currently TOC does not receive all satellite radiance and sounding data for the POES constellation and hence we have ingested these files directly from a NESDIS source.

A. Data Ingest and Decode

The Ingest module receives the input data and sorts these files in a standard data format. The output of the ingest module is fed into the decoding module where the received data is decoded and stored in memory for further processing. The decoder software has been designed currently to decode data received from NESDIS. Once the system is moved to NWSTG the decoding software will be modified to decode the NWSTG data format. This decoded data is then written out to temporary NetCDF files which are then made available for immediate delivery.

B. Quality Control Function

Two categories of QC checks, static and dynamic are implemented for each observation type. The static QC checks are single-station, single-time checks which, as such, are unaware of the previous and current meteorological or hydrologic situation described by other observations and grids. Checks falling into this category include validity, internal consistency, and vertical consistency. Although useful for locating extreme outliers in the observational database, the static checks can have difficulty with statistically reasonable, but invalid data. To address these difficulties, MADIS-T also implements dynamic checks which refine the QC information by taking advantage of other available hydrometeorological information. MADIS-T satellite radiance and sounding data processing thread only contains static checks. This quality control information is made available to data users though the output NetCDF files which includes a QC applied bitmap indicating which QC checks were applied to each observation, and a QC results bitmap indicating the results of the various QC checks. MADIS-T also describes single character *data descriptors* for each observation, which provide an overall opinion of the quality of the observation by combining the information from the various QC checks. Algorithms used to compute the data descriptor are a function of the types of QC checks applied to the observation, and the sophistication of those checks. Level-1 QC checks are considered the least sophisticated, whereas level-3 the most sophisticated. Below follows a complete list of the data descriptors:

- No QC Available
  - Z – Preliminary, No QC
- Automated QC Checks
  - Level-1, validity, position consistency checks
  - Level-2, internal, temporal Consistency, time-height consistency, hydrostatic, super-adiabatic lapse rate, and wind shear checks
  - Level-3, spatial consistency checks
  - C –Course, passed level-1
  - S – Screened, passed level-1 and level-2
  - V – Verified, passed level-, level-2 and level-3
  - X – Rejected, failed level-1
  - Q – Questioned, passed level-1, failed level-2 and level-3
- Subjective Intervention
  - G – Subjective good
  - B – Subjective bad
- Interpolated / Corrected Observations
  - I – Interpolated
  - W – Raw data was wrong, but has been corrected using estimates from automated QC checks
  - T – Virtual temperature could not be calculated; air temperature passing all QC checks has been returned

Table I shows a list of variables and the associated QC checks applied to MADIS-T satellite sounding data. Level-1 QC checks are considered the least sophisticated while Level 2 checks are the more sophisticated checks. Note that the observed water vapor mixing ratio is converted to dew point, after which the quality checks are applied.

TABLE I
List of satellite sounding variables quality controlled within the MADIS System

Code	**Name**	**Max. QC Level**	**Level 1**	**Level 2**
T	*Temperature*	*1*	*X*	
TV	*Virtual Temperature*	*1*	*X*	
WVMR	*Water Vapor Mixing Ratio*	*2*	*X*	*X*
TD	*Dew point Temperature*	*2*	*X*	*X*
RH	*Relative Humidity*	*2*	*X*	*X*
Q	*Specific Humidity*	*2*	*X*	*X*
DPD	*Dew point Depression*	*2*	*X*	*X*
AH	*Absolute Humidity*	*2*	*X*	*X*

The Level-1 validity checks restrict each observation to falling within a predetermined tolerance limits. The tolerances are specified as a function of pressure level, and the radiometer heights are converted into pressure using the U.S. standard

TABLE II
List of satellite sounding variables quality controlled within the MADIS System

Tolerance Level (mb)	**Air and Dew point Temperature (C)**	
	Low	**High**
1000	*-65*	*60*
850	*-50*	*45*
700	*-50*	*30*
500	*-57*	*5*
400	*-66*	*-10*
300	*-72*	*-20*
250	*-76*	*-25*
200	*-78*	*-30*
150	*-85*	*-30*
100	*-95*	*-30*
70	*-95*	*-25*
50	*-95*	*-15*
30	*-95*	*-5*
20	*-95*	*5*
10	*-95*	*15*
<10	*-95*	*15*

atmosphere calculation. Observations not falling within the limits are flagged as failing the QC check. Table II lists the tolerance limits used in QC checks for the satellite sounding variables.

The Level-2 internal consistency checks enforce reasonable, meteorological relationships among observations measured at a single location. In this case, the dew point temperature must not exceed the temperature observation. If it does, the dew point is flagged as failing the internal consistency check. Similar checks are implemented for other variables.

C. Product Generation

System output functions are used to generate the integrated NetCDF files including the QC flags. This requires data rearrangement, timestamp adjustments, additional metadata as well as any subjective intervention provided by meteorologists. It also provides the data to the delivery system for users to receive data as soon as the data becomes available. Once QC has been applied, the data is checked against any subjective intervention provided by the meteorologists. Two text files, a *reject* list and an *accept* list provides the capability to subjectively override the results of the automated QC checks. The reject list is a list of satellites and associated input observations that are labeled as *bad*, regardless of the outcome of the QC checks; the accept list is the corresponding list of satellites that are labeled as *good*, regardless of the outcome of the QC checks. In both cases, observations associated with the satellites in the lists can be individually flagged. For example, moisture observations for a particular satellite may be added to the reject list, but not the temperature observations.

Once we have received the QC flags we create the QC *data descriptors*. A data descriptor is a single-character used to define an overall quality of each observation by combining the information from various QC checks. There are two word generated. The first is named as “applied word” which defines the QC’s which were applied on the observations. The second is the “result word” which defines the results of the QC’s when applied word is executed. The following chart provides a complete list of the data descriptors and the bitmap used for the satellite sounding quality control. The Master Check bit in the chart below is used to summarize all of the checks in a single bit. If a QC is applied to an observation the bit is set in applied word. Similarly if the observation failed any QC check, the bit for the observation is set in the QC results word.

IV. Results and Conclusions

In this paper we have presented an approach which can be utilized for the transition of the MADIS system from a research environment to the NWSTG with minimal impact/changes to the infrastructure at TOC. We have performed a risk reduction activity in which we have ported one of the data processing threads to a blade clustering environment using the virtual infrastructure. QC results were compared with output produced by the baseline GSD system and there was no difference in the output for the sampled input data. Since this work was performed using the virtual environment it can easily be ported to any other architecture as required. It was also demonstrated that virtualization not only helps in speeding up the process of putting the system in production but it also helps in increasing reliability, manageability, server consolidation and scalability.

V. Acknowledgements

The authors acknowledge the Earth System Research Laboratory Global Systems Division’s MADIS group for providing the data to compare the baseline results. We would also like to acknowledge Steve Schofield of Science Application International Corporation (SAIC) for providing the technical insight of the NWS environment to generate the System Requirements Specifications for MADIS-T system.

VI. References

[1] http://madis.noaa.gov

[2] NOAA’s National Weather Service, 2007: Meteorological Assimilation Data Ingest System Functional Requirements Document

[3] NOAA’s National Weather Service, 2007: Meteorological Assimilation Data Ingest System Transition System Requirements Specification.

[4] NOAA’s National Weather Service, 2007: Meteorological Assimilation Data Ingest System Transition Satellite Radiance and Sounding Test Plan and Procedures.

[5] http://vmware.com/virtualization