Atmospheric and Climate Sciences, 2017, 7, 337-351
http://www.scirp.org/journal/acs
ISSN Online: 2160-0422
ISSN Print: 2160-0414

Workflow-Based Platform Design and Implementation for Numerical Weather Prediction Models and Meteorological Data Service

Xiaoxia Chen[1], Min Wei[1,2], Jing Sun[1]

[1]National Meteorological Information Center, China Meteorological Administration, Beijing, China
[2]Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, and Joint Center for Global Change Studies (JCGCS), Tsinghua University, Beijing, China
Email: weim@cma.gov.cn

How to cite this paper: Chen, X.X., Wei, M. and Sun, J. (2017) Workflow-Based Platform Design and Implementation for Numerical Weather Prediction Models and Meteorological Data Service. *Atmospheric and Climate Sciences*, **7**, 337-351.
http://dx.doi.org/10.4236/acs.2017.73025

Received: June 13, 2017
Accepted: July 23, 2017
Published: July 26, 2017

Abstract

In this paper, we present a set of best practices for workflow design and implementation for numerical weather prediction models and meteorological data service, which have been in operation in China Meteorological Administration (CMA) for years and have been proven effective in reliably managing the complexities of large-scale meteorological related workflows. Based on the previous work on the platforms, we argue that a minimum set of guidelines including workflow scheme, module design, implementation standards and maintenance consideration during the whole establishment of the platform are highly recommended, serving to reduce the need for future maintenance and adjustment. A significant gain in performance can be achieved through the workflow-based projects. We believe that a good workflow system plays an important role in the weather forecast service, providing a useful tool for monitoring the whole process, fixing the errors, repairing a workflow, or redesigning an equivalent workflow pattern with new components.

Keywords

Workflow, SMS, ecFlow, S2S, Numerical Model, Meteorological Data Service

1. High Performance Computing System in CMA

China Meteorological Administration (CMA) has been utilizing high performance computing systems (HPC) since the 1980s. The current HPC system was introduced in 2013, with a peak performance of 1054.2TFlops. There are two identical subsystems, one for production, and another for research (**Figure 1**) [1].

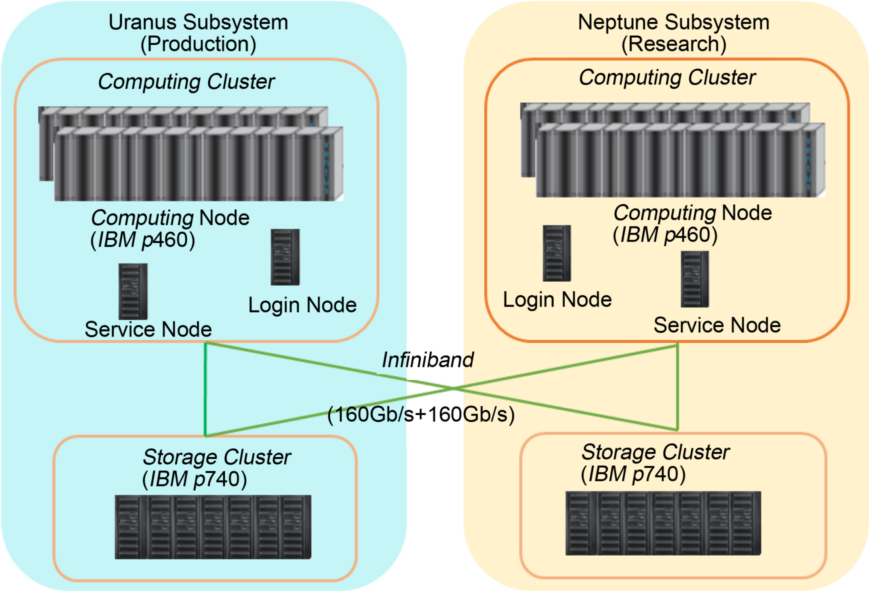


Figure 1. High performance computing system in CMA.

The computing cluster consists of 560 nodes, which is interconnected with the storage cluster via Infiniband network. There are login nodes which are used for users to log on the system and submit the jobs. Those service nodes are used for managing the jobs dispatch and workload balance. Numerical Weather Prediction Models such as Global and Regional Assimilation and Prediction System (GRAPES) are the top application, accounting for 40% of the total resources. CMA has a unique set of challenges since we require highly available, accurate, monitorable, and flexible systems, while simultaneously supporting the business and critical research development efforts.

2. Meteorological Workflow

Workflow management is a fast evolving technology which is increasingly being exploited by businesses in a variety of industries. Its primary characteristic is the automation of processes including a collection of tasks, whose execution is initiated by humans or machine-based activities. Workflows represent a repeatable and structured collection of tasks designed to achieve a desired goal.

A workflow specification spans at least three perspectives: control-flow, data-flow, and authorization (also called the resource perspective) [2]. Control-flow specifies the execution order of the tasks in sequential, parallel, or alternative execution methods; the data-flow defines the various input data objects or output data produced by these tasks; and the authorization constrains different executors responsible for the execution of the tasks in the form of authorization policies and constraints. These three dimensions are interconnected, as each one of them influences the others. The set of behaviors allowed by the control-flow is further constrained by conditions on the data or the status of its previous tasks.

Business Process Management includes the identification, execution, monitoring and improvement of business over time [3]. A scientific workflow system is a specialized form of a workflow management system designed specifically to compose and execute a series of computational or data manipulation steps, or workflow, in a scientific application [4].

When it comes to the meteorological field, there is a wide variety of workflow technologies which can support complex applications. CMA has adopted two technologies for HPC workflows: SMS (Supervisor Monitor Scheduler) and ecFlow from the European Center for Medium-Range Weather Forecasts (ECMWF) to support its needs in production, research and related services. Both have been proven effective in reliably managing the complexities of large-scale weather-related workflows. The SMS system has been replaced by an ECFLOW system which is the one that drives the integration in 2011 [5].

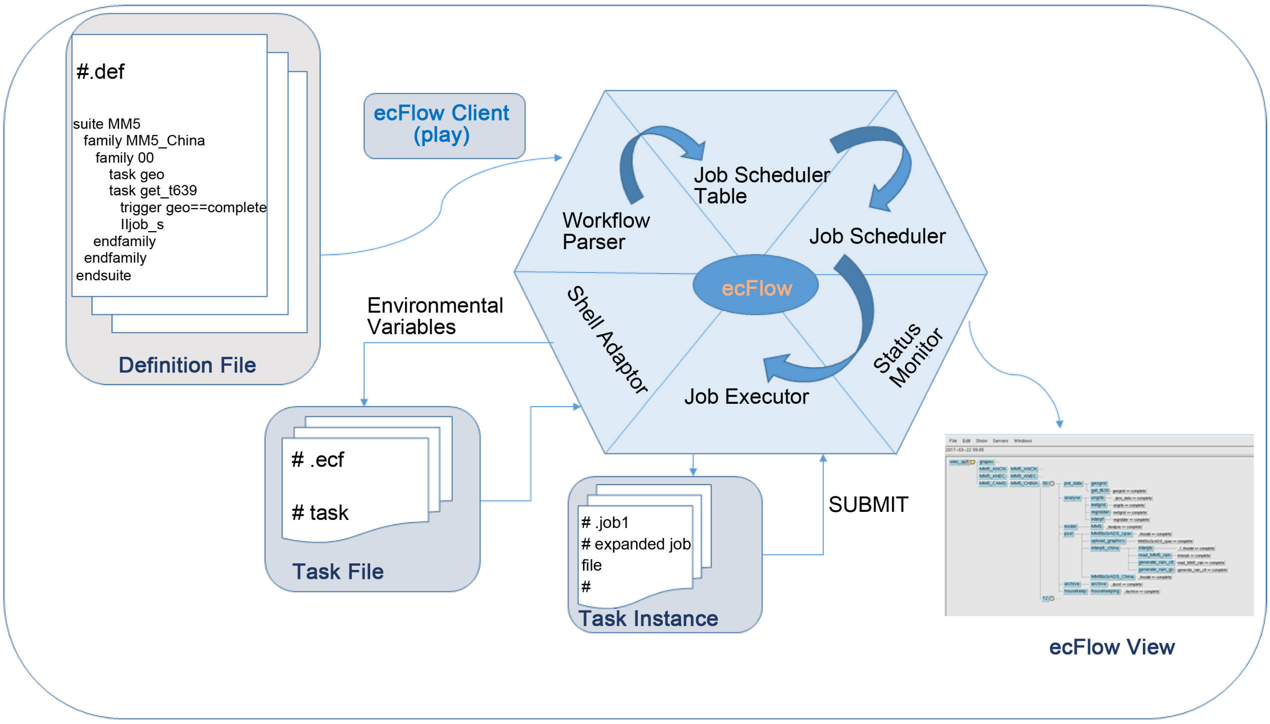
The whole process of the model including the dependencies is defined in a text definition file (.def) which is then loaded to the ecFlow server. ecFlow is a workflow manager with an intuitive GUI that is used to handle dependencies, schedule jobs, and monitor the production suite. In the production environment, all jobs are scheduled and submitted to the CMA Uranus resource manager, Loadleveler, by ecFlow. Each job in ecFlow is associated with an ecFlow script which acts like a Loadleveler submission script, setting up the llsubmit parameters and much of the execution environment and calling the job file to execute the job. All jobs must be submitted to Loadleveler via llsubmit. It is at the ecFLow or submission script level where certain environment specific variables must be set. ecFlow executes tasks (jobs) and receives acknowledgements from the associated job when it changes status or when it sends events. It does this using child commands embedded in the scripts. ecFlow stores the relationship between tasks and is able to submit tasks dependent on triggers. ecFlow is complemented by ecflow_ui, its graphical interface that allows users to have immediate knowledge, using color coding, of the status of the various programs or processes handled by the scheduler (**Figure 2**).

3. Best Practices for Workflow Design

There is no standard workflow design in meteorological applications. Based on the previous work on the platforms, we argue that a minimum set of guidelines at the workflow design stage are highly recommended, serving to preventing the need for maintenance and adjustment in the future.

In order to enable operational stability and efficient troubleshooting, we define four steps to follow to allow the creation of high quality workflows (**Figure 3**).

1) Workflow Scheme: After clarifying the whole process of the operational model, make a reasonable workflow. Consult with the developer if possible to better understand the flow chart. A good workflow will potentially increase its reproducibility and reuse. Create dependencies between nodes in a suite to guarantee the flow and improve the efficiency by time/date, or the status of other events. To design appropriate modes of failure is essential. An executable task should not terminate abnormally with a segmentation or memory fault for errors



```
#.def

suite MM5
  family MM5_China
    family 00
      task geo
      task get_t639
        trigger geo==complete
        lljob_s
    endfamily
  endfamily
endsuite
```

Figure 2. Architecture of the ecFlow system.

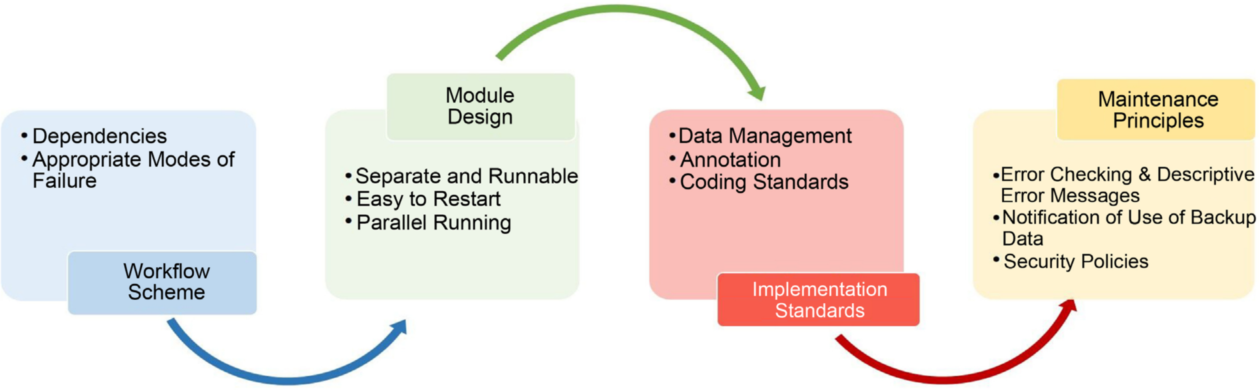


Figure 3. Workflow design guidelines.

that are discoverable or trappable. For example, lacking of input data should be handled in the script before the executable runs. Diagnosing failures quickly is a necessary component of maintaining a suite of products that boasts a great on-time delivery rate.

2) Module Design: All executable components are implemented as separate, runnable workflows. There is the possibility to continue the job which was interrupted or failed-in places where restarts can be applied to save time when recovering from a failure. Long running jobs that have multiple executable calls might be a good candidate to break into two smaller jobs so that if a failure occurs, only the problem part needs to re-run and the time to completion is shorter. This helps to minimize the time it takes to re-run a failed job. Also consider the possibility of parallel running. It is very typical to run the post-processing in

parallel method [6] [7].

3) Implementation Standards: Focus on data management, functional annotation and coding standards.

Firstly, data management is one of the essential parts. Consider both input and output data. Understand well the location of the input and output data on the resource to analyze the job when the job fails. Make a list or table for output of each module to get a clear idea of the data usages. Output can be used as input to another workflow, stored in a database, or be presented to the end user. The format of the output should be considered too, in the form of graph, table, text, GRIB file, binary file or NetCDF (Network Common Data Form) [8].

Secondly, make high-level functional annotation. Careful annotation of a workflow helps to record all steps and assumptions hidden in the workflow. There is no accepted standard for annotating a workflow. We propose to choose meaningful names for the workflow title, inputs, outputs, and for the processes that constitute the workflow as well as for the interconnections between the components [8], so the annotations are not only a collection of static tags but capture the dynamics of the workflow. Besides, source code and scripts should be annotated with information that may help staff remedy a problem if something goes awry. In some cases, too much information is as bad as none at all. Use the best judgment to include information that will be of the most help in troubleshooting potential issues.

Thirdly, standard environment variables, file name conventions, production utilities, date utilities, GRIB utilities, general application standards, compiled code, interpreted code (bash, ksh or perl scripts), and directory structures should be used to improve the troubleshooting process. For example, file names should indicate the name of the model run, the cycle, the type of data the file contains, the resolution of the data, other data related elements, the three-digit forecast hour the data represents, and the file type. Do not specify absolute paths to executables, libraries, or any other products inside the make file. All components of an application to be implemented into the production environment are required to be in vertical structure, where, with the exception of system or standard production libraries and input data, all of the files required to completely build and run the jobs are contained in an application-specific directory.

4) Maintenance Principles: Focus on the descriptive error messages, notification of backup data, and security policies.

Firstly, it is imperative that all production code and scripts broadly employ error checking to catch and recover from errors as quickly as possible. Use descriptive error messages. Fatal errors should print a descriptive message beginning with "FATAL ERROR". Warnings or non-fatal error messages should be prefaced with "WARNING". Failures should not be allowed to propagate downstream of the point where the problem can first be detected.

Secondly, notification of use of backup data. For scripts that have a secondary data source to be used when the primary data is not available, the script should include a message that indicates the primary data is not available and backup data is being used. The application cannot fail when this data is missing. Appropriate notification of use of backup data should be made and the job should continue with other operationally supported input data.

Thirdly, enforce security policies in the form of access control, specifying which users can execute which tasks, and authorization constraints, such as staff from 7 × 24 maintenance duty team can resubmit the tasks and developers can edit the script for further debugging.

4. Use Case 1: NWP Models

Operational numerical weather prediction models (NWP) have been applied in CMA for decades and for provincial meteorological bureaus. Our NWP Operational System consists of real-time observation data, data processing module, model run, post processing and archive module. Currently, there are more than 30 operational NWP models including T639 global, GRAPES (Global and Regional Assimilation and Prediction System), haze, ocean sea fog, and climate prediction models such as BCCCSM (Beijing Climate Center Climate System Model) which all employ SMS to achieve real-time monitoring.

4.1. GRAPES_CAMS Model in Weather Modification Center, CMA

While for the new centers, such as the Weather Modification Center, the NWP models run in shell, without the ability of being monitored and tracked in real-time. Weather Modification Model (Version 2013) includes GRAPES_CAMS mode (Global and Regional Assimilation and Prediction System, Chinese Academy of Meteorological Sciences) and MM5_CAMS mode (Meso-scale model 5, Chinese Academy of Meteorological Sciences). GRAPES_CAMS mode covers the whole country with the aim to predict the large scale layered cloud system artificial precipitation, with a horizontal resolution of 25 km. MM5_CAMS model provides the accurate forecasting service for a mixed cloud system and artificial precipitation of convective clouds for the area of focus, including the Northeast region and the North China region, with a horizontal resolution of 15 km and a second grid level resolution of 5 km.

We followed the workflow design guidelines to establish the GRAPES_CAMS models. To begin with, we clarify the whole process of the models and re-design the modules. Then we take the implementation standards to write the modules. Lastly, through the maintenance process, we update or rewrite the scripts to improve the efficiency.

From Figure 4, we can see the whole process as below. The Data Acquisition (pre_data) module is designed to receive and process the T639 global forecasting data with all T639 data being downloaded in parallel. The data analysis (analyse) module is the pre-process of the model run, which is divided into two steps (make a namelist and run si.exe). After generating the namelist file, the initial field data and the lateral boundary data are formed. The main program (model) module is for forecasting, by using the initial field data and the lateral boundary data. The Data post-processing (post) module is used to generate the rainfall data.

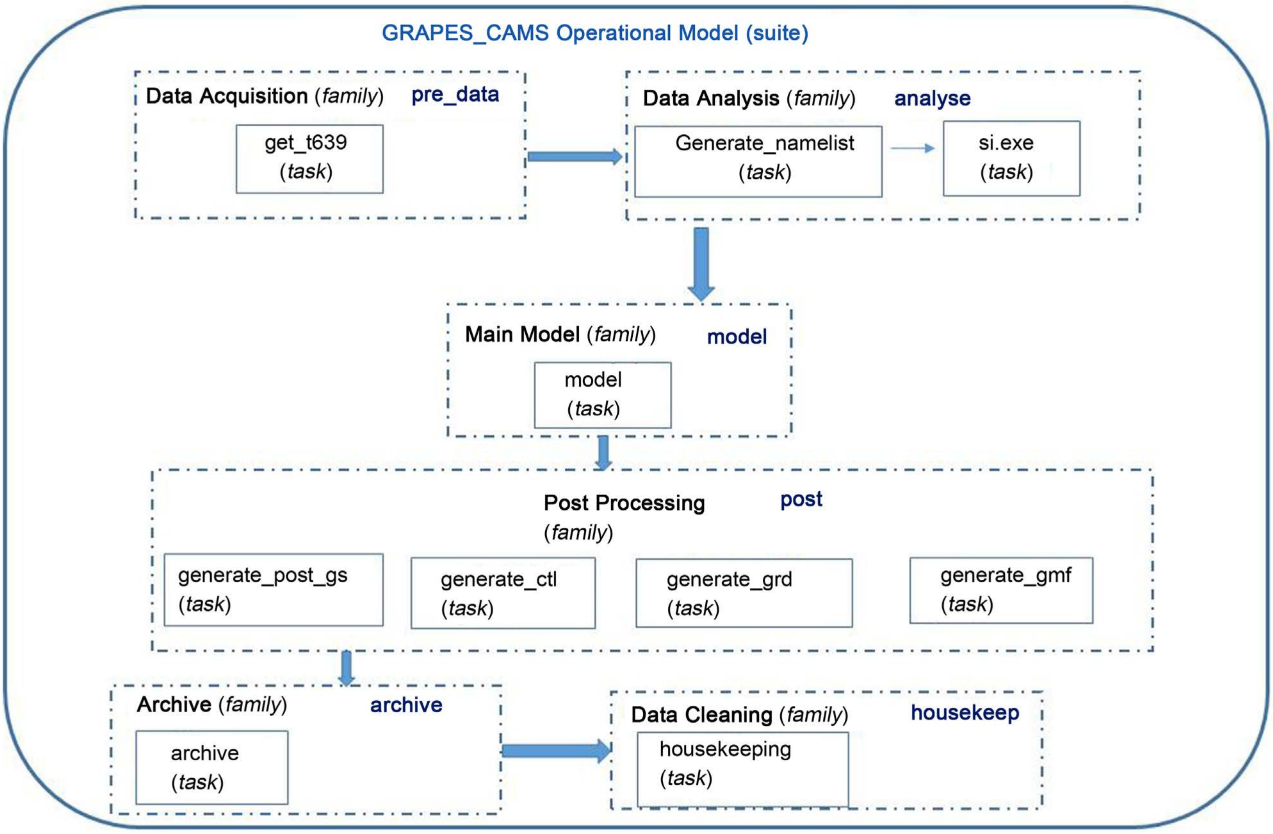


Figure 4. Re-organized GRAPES_CAMS operational process.

The Data Archiving (archive) module is for forecast product archive, which is served to transmit data to three different FTP servers. The data Cleanup (housekeeping) module is used for the deletion of the predicted results, mainly to remove intermediate data and outdated pattern data generated during the whole mode operation periodically.

The Figure 5 shows the final monitoring interface of the GRAPES_CAMS model. From the system, we can easily check where the process is, whenever there is an alarm or error, it is easy to track and to analyze the cause of the failure through the output files. MM5_CAMS model has three systems based on different coverage. After deploying the MM5_CAMS model for the Northeast region, it is very effective to develop the other two for the simple reason that they share a lot in common except some time and input data variables. Most of the modules are reusable. In conclusion, a good module design is a prerequisite for code reuse.

4.2. Regional NWP Models' Backup System in CMA

There are two identical subsystems located in Beijing and another seven regional HPC centers which are installed all over the country: Guangdong, Liaoning, Shanghai, Hubei, Sichuan, Gansu and Xinjiang, with the names of Mercury, Venus, Saturn, Mars, Jupiter, Dipper and Polaris respectively. The performance of each subsystem is listed in the Figure 6. Pluto, located in Guangdong, is for

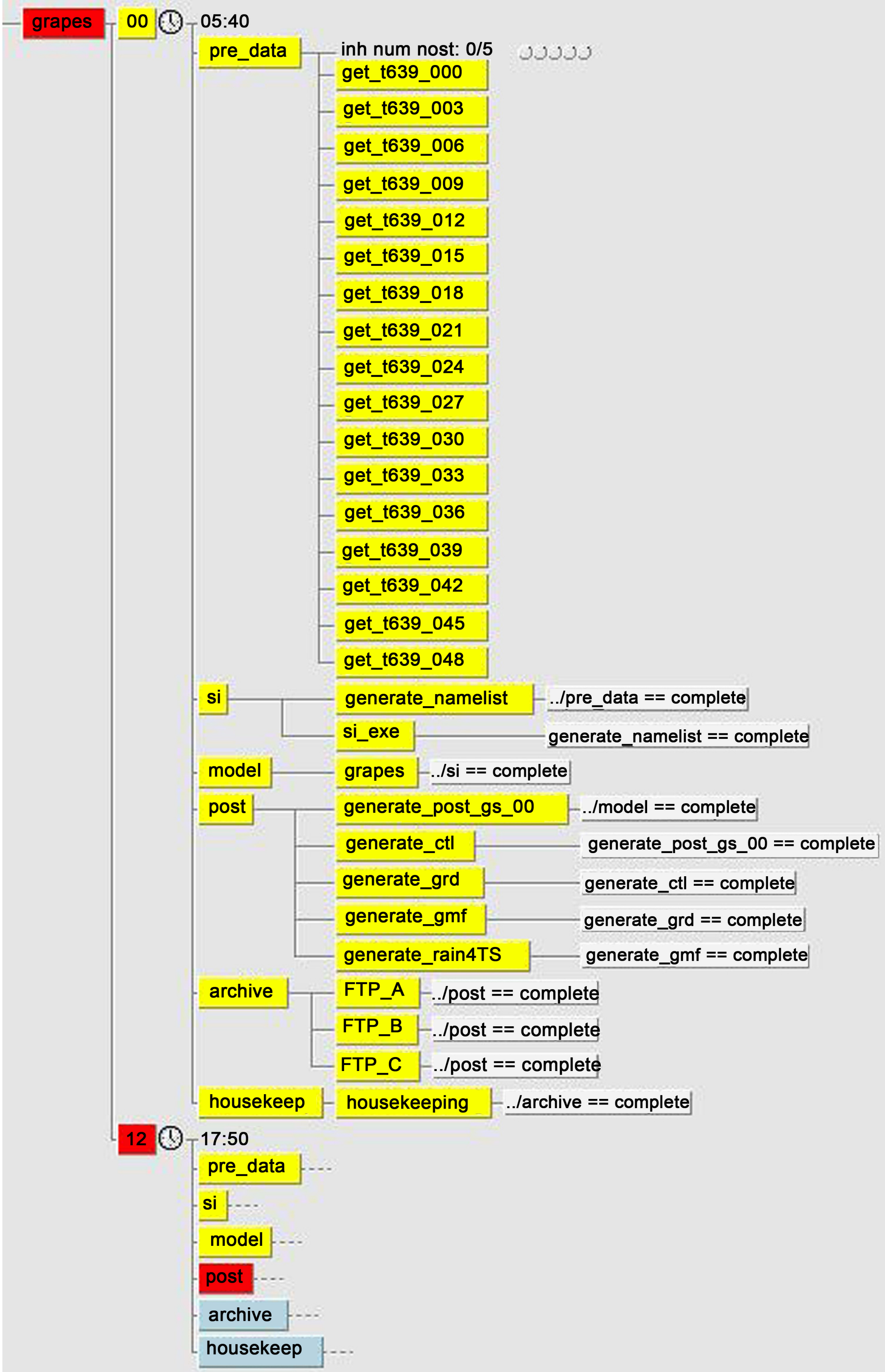
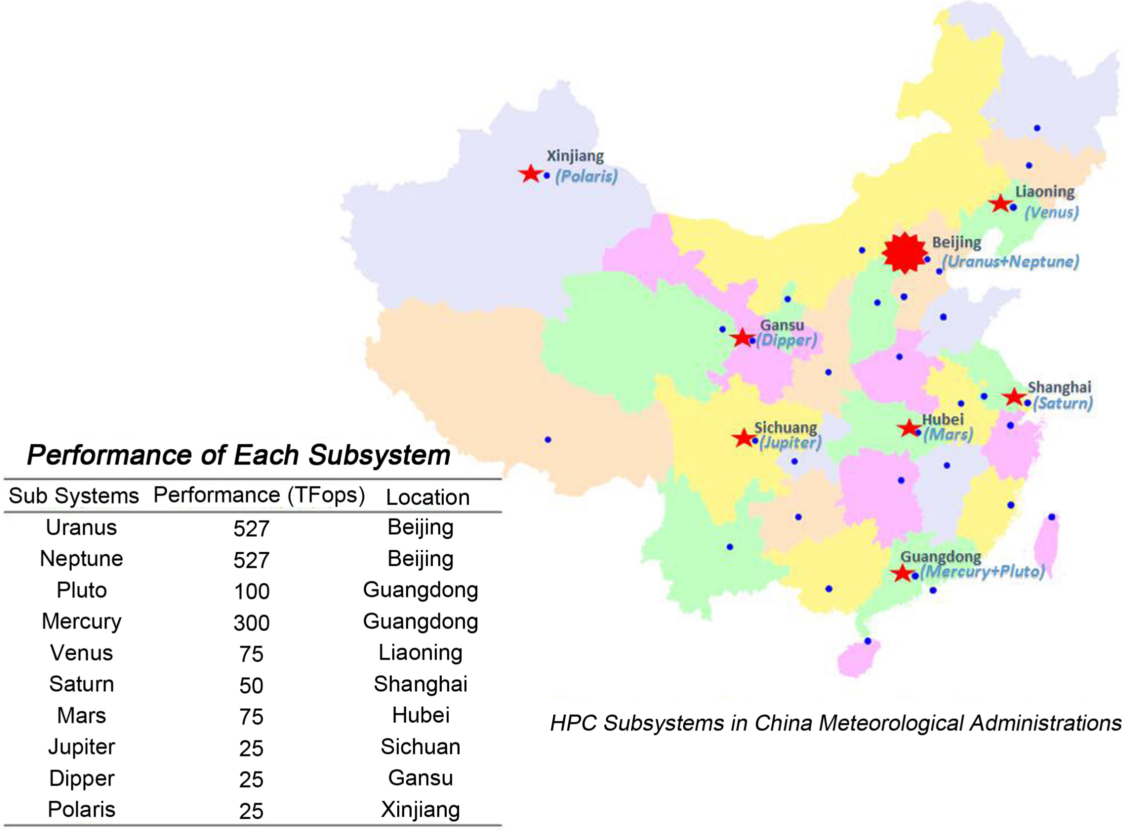


Figure 5. The Interface of the GRAPES_CAMS model.



Performance of Each Subsystem

Sub Systems	Performance (TFops)	Location
Uranus	527	Beijing
Neptune	527	Beijing
Pluto	100	Guangdong
Mercury	300	Guangdong
Venus	75	Liaoning
Saturn	50	Shanghai
Mars	75	Hubei
Jupiter	25	Sichuan
Dipper	25	Gansu
Polaris	25	Xinjiang

Figure 6. The distribution of HPC systems in meteorological field.

CMA users in Beijing.

In regional centers, numerical weather prediction models have been established and monitored via SMS. Operational models require a disaster recovery plan and backup system in order to minimize the duration of disruption in the event of failure of HPC systems. For the aim of building the off-site backup systems, all regional operational models are designed to be able to run on CMA HPC systems immediately when the regional HPC systems are down. The key issue of the development is the data management. Normally, the input data includes radar, observations, satellites and automatic observation data shared by provinces. We firstly establish all the data that are needed for the system and the products that are to be disseminated to the regional centers. Secondly, we establish a workflow to get and distribute the data. Thirdly, considering the replacement of SMS, we use ecFlow to establish the system.

During the development of the operational NWP models' backup system, we focus on the process, the data input and output, and make the right directories. Take the WRF (The Weather Research and Forecasting) running on Venus in Liaoning for example. Data is the key design of the whole system. When the model runs in local area, the input data comes from CMACast and shared regional systems. So we apply for CMACast to directly feed the data to the HPC systems (**Figure 7**). To get the necessary data and process them, there is a separate workflow-based system to do the data acquisition and pre-processing part (**Figure 8**).

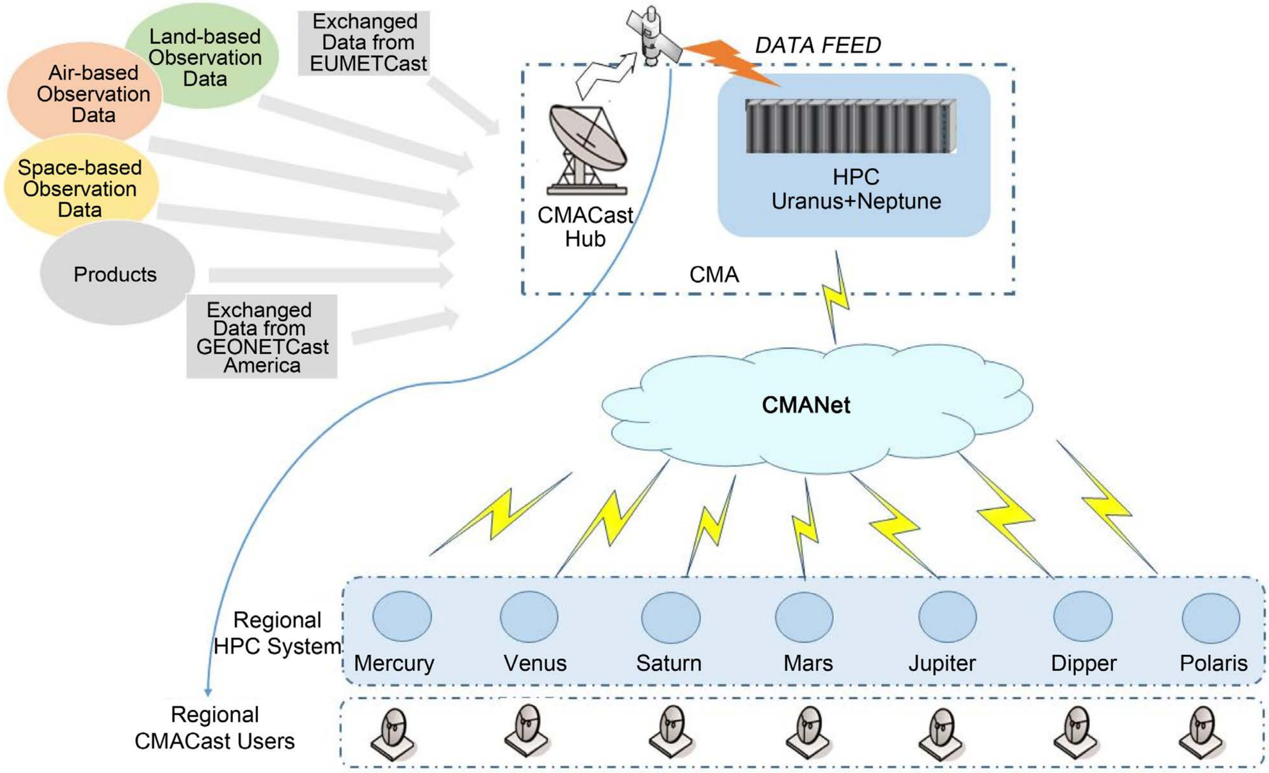
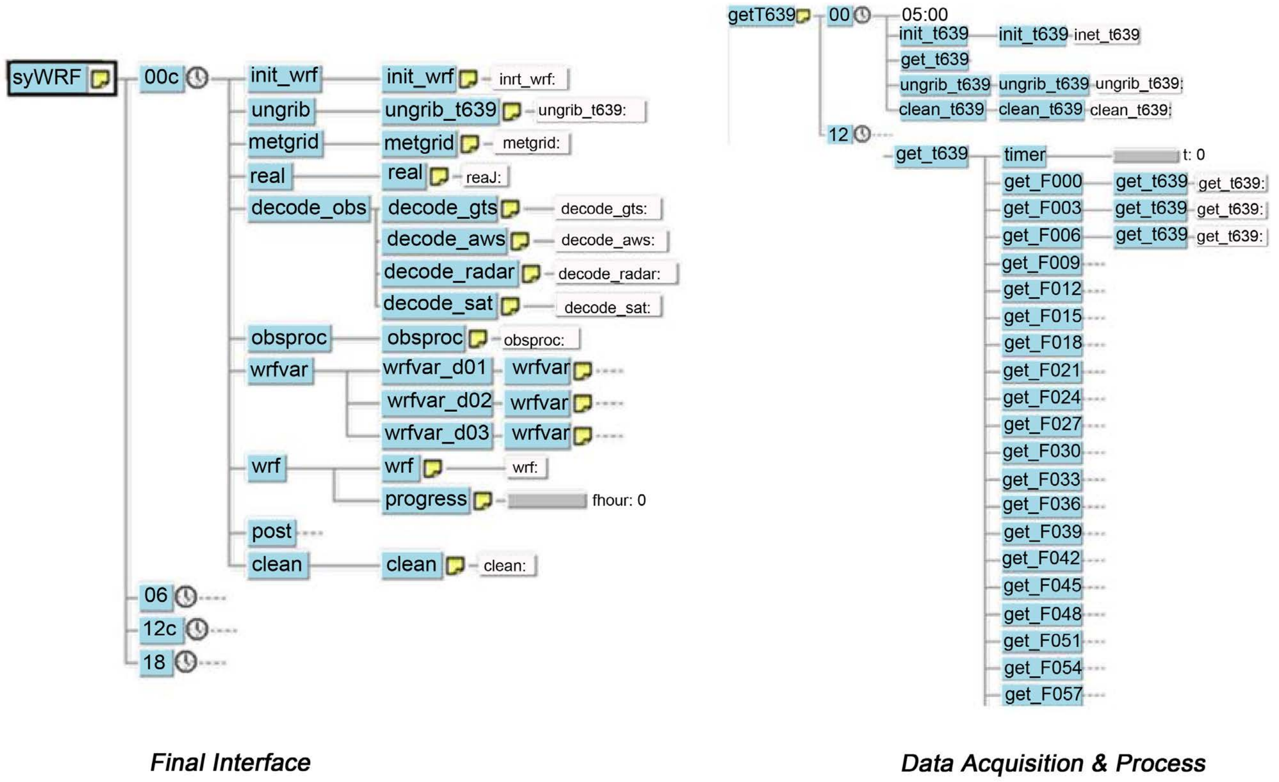

Land-based Observation Data
Air-based Observation Data
Exchanged Data from EUMETCast
Space-based Observation Data
Products
Exchanged Data from GEONETCast America
DATA FEED
CMACast Hub
HPC Uranus+Neptune
CMA
CMANet
Regional HPC System
Mercury
Venus
Saturn
Mars
Jupiter
Dipper
Polaris
Regional CMACast Users

Figure 7. HPC & data between CMA and regional centers.



Final Interface

Data Acquisition & Process

Figure 8. Backup WRF system running on Venus in CMA.

CMACast is one of the three broadcasting system of WMO, taking responsible for the data broadcast and distribution for Asia-Pacific users. The data includes ground, air, sea and other global exchange data, numerical weather prediction Products and satellite data such as FY-2E/F/G, FY-3A/C, Meteosat-7/8/10, Metop-A/B and GOES-13/1. Currently, CMACast distributes 250 GB data for the users on a daily basis. For the backup system establishment there are data feeds from CMACast to CMA HPC system, receiving the radar, surface, observation, and satellite data.

Currently, CMACast pushes the data needed to HPC systems every day in case of the start of the backup system, and regular data housekeeping is applied to save the space. Whenever there is failover of the regional NWP models, the backup system will be working right away.

5. Use Case 2: Meteorological Data Service

5.1. The Sub-Seasonal to Seasonal (S2S) Data

S2S Project (Sub-Seasonal to Seasonal Prediction project) is a joint initiative of the World Weather Research Program (WWRP) and the World Climate Research Program (WCRP) for observational systems research and predictability testing (THORPEX), aiming to enhance the ability to forecast technology and services to seasonal forecasts [9]. CMA, serving as a second archiving center, has launched a data portal for sub-seasonal to seasonal (S2S) weather forecasts to help researchers study predictability on time-scales of up to 60 days.

CMA established the whole process including data collection, quality inspection, archiving (MARS & Online Disk) and portal service modules (**Figure 9**). And the operational process system of data collection has been established based on shell and ecFLOW.

Data collection module is used to receive the data and historical data from all 11 centers including European Center for Medium Range Weather Forecasts (ECMWF), the US National Centers for Environmental Prediction (NCEP), Australia's Bureau of Meteorology (BoM).

Quality inspection module is the second step following the data collection

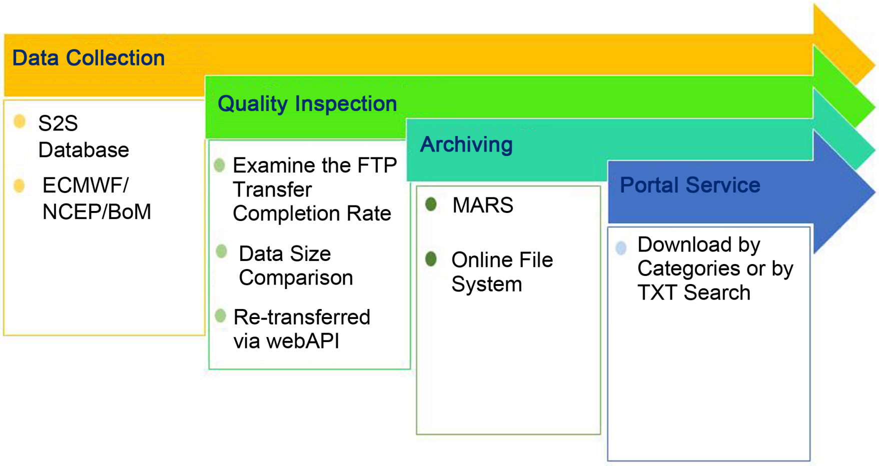


Figure 9. The whole process of the S2S service.

module. Firstly, decompress the received packet, and then check the file to see whether it meets S2S data definition based on GRIB_API definition. Data integrity will be checked by two ways, one is to examine the FTP transfer completion rate; the other is to compare the data size between the local file downloaded and the original file on the ECMWF FTP server. If the reception is incomplete, limited number of transfers will be firstly initiated to ensure the data integrity. Once this method fails, there would be a notification record, and the data will be re-transferred based on the web API from ECMWF.

Archiving module consists of two parts, one is to archive the data to MARS; the other one is to process the data to a specific file and then upload to the online file system. CMA optimize the MARS retrieval system, adjust the database structure of archiving data in MARS, clean all the S2S data which has already been archived, and archive the data based on the new rule. Online data storage system transfers data to portal sharing servers based on the new S2S data according to the single center, single date, single variable and GRIB_API programming.

The CMA data portal, like the ECMWF data portal, provides descriptions of the models from the different centers and S2S data parameters, in addition to the data download service. Two ways of searching and accessing the data are supported: free text search and faceted search.

After the development of the workflow for S2S project, it changed the previous situation without monitor alert and direct manual interference. Real time running has improved the efficiency of data processing, archiving and service processing, and at the same time, data accuracy is guaranteed.

5.2. Meteorological Data from the Website

CMA HPC provides an array of data storage areas, each designed with a particular purpose in mind. Storage areas are broadly divided into two categories: those intended for user data and those intended for project data. Within each of the two categories, we provide different sub-areas, each with an indented purpose. CMA HPC users have a number of file systems available for their use. Each user has a home directory. Mass storage space is intended to hold important files that are too large to be stored in users' home directories. CMA HPC provides the meteorological data including the NCEP reanalysis data, Global Forecast System data and Global Data Assimilation System data, NCEP FNL (Final) Operational Global Analysis data, NCEP Global Ocean Data Assimilation System (GODAS) and other data for our users on a shared file system. The advantage of providing shared data space is so that users don't bother to download the public data by themselves which saves a lot of individual space.

This data is available on the NOAA websites or ftp servers. Considering the security issue, we firstly download the public data to Demilitarized Zone (DMZ), and then download to CMA HPC system. The whole process can be monitored through the workflow-based platform with alarms.

Each module consists of two parts, one is to download the data to the server which locates in DMZ area, the other one is to download the data in DMZ to HPC server GPFS (Figure 10).

Through the platform, we can re-download the data which are not available at

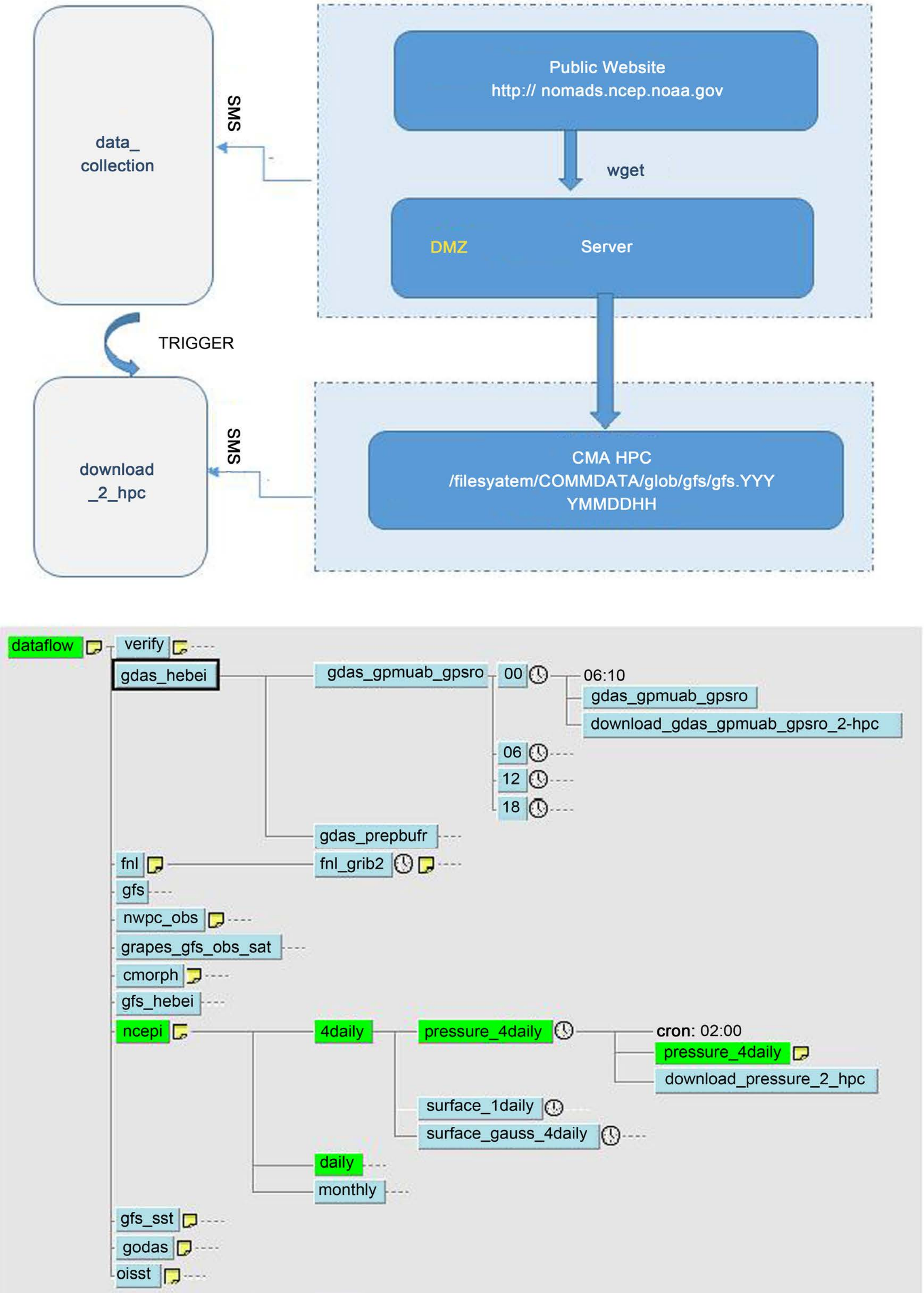


Figure 10. The architecture of the meteorological data service.

the required time. There are some strategies which are applied to improve the efficiency: Retry within a limited number times when the data is not available; check the integrity of the data; Make jobs easy to resubmit.

6. Performance of the Workflow Systems

In this section, we present the results of the systems. All the projects mentioned in this thesis have been operational systems. These show that a significant gain in performance can be achieved through the workflow-based projects.

Highlights are shown as below:

1) High efficiency. After the establishment of the workflow-based system, it is easy to check the whole process in real-time. Error handling becomes more effective. All the errors are good material for the establishment of the knowledge base for future statistics and analysis. Resubmission of jobs becomes possible without human interference. Without this system, we needed to read the output file to check the errors, which was very difficult and time-consuming.

2) High scalability. The developer can adjust the process through the definition file which will be displayed clearly through the interface and can be debugged more effectively. To combine with the technology of the instant messaging application, take the popular app WeChat for example, we can push the error information to the cell phone which will notify the operator on duty immediately.

3) Great computing and storage resources saver. With the great parallel design, the jobs can run simultaneously which fully utilizes the computing cores. Data cleaning can be done when the previous related jobs have been completed.

4) Friendly user interface for monitoring and control. Different colors represent different status of each task which makes it easy to check the status. When a task aborts, the notification can be sent automatically to the operator containing the error message.

7. Concluding Remarks

When following the four basic best practices for workflow design, we find that it is much easier to design and implement a new application for processing and monitoring the system. The lack of a standard in the Meteorological field makes it even more important to follow the basic principles. We propose that in the future there will be a meteorological standard for workflow design and development.

Acknowledgements

The projects presented in this paper are supported by the National Key Research and Development Program of China (Grant No. 2016YFA0602102) and China Special Program for Meteorological Research in the Public Interest (Grant No. GYHY201306062).

References

[1] Zhao, L.C., Shen, W.H., Xiao, H.D., *et al.* (2016) The Application of High Perfor-

mance Computing Technology in Meteorological Field. *Journal of Applied Meteorological Science*, **27**, 550-558.

[2] dos Santos, D.R. (2017) Automatic Techniques for the Synthesis and Assisted Deployment of Security Policies in Workflow-based Applications, Ph.D. Thesis, University of Trento, Trento.

[3] Yudin, Y., Krasikova, T., Dorozhko, Y. and Currle-Linde, N. (2013) Modular Workflow System for HPC Applications, World Academy of Science, Engineering and Tchnology. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, **7**, No. 2.

[4] Płóciennik, M., Winczewski, S. and Ciecieląg, P. (2014) Tools, Methods and Services Enhancing the Usage of the Kepler-Based Scientific Workflow Framework. *Procedia Computer Science,* **29**, 1733-1744. https://doi.org/10.1016/j.procs.2014.05.158

[5] https://software.ecmwf.int/wiki/display/ECFLOW/Tutorial

[6] Zhao, J., Zhang, Li.-L. and Song, J.-Q. (2010) Job-Level Parallel Implementation Method of Numerical Weather Prediction Post-Processing System. *Journal of Computer Applcations*, **30**(Suppl.1), 241-242.

[7] Mahura, A., Petersen, C., Amstrup, B. and Sass, B.H. (2016) Post-Processing of NWP Models Forecasts: Case of Denmark and Greenland. *Geophysical Research Abstracts*, **18**, EGU2016-10130.

[8] Hettne, K., Wolstencroft, K.J., Belhajjame, K. and Roos, M. (2012) Best Practices for Workflow Design: How to Prevent Workflow Decay. *SWAT4LS*, January 2012.

[9] Sub-Seasonal to Seasonal Prediction Project. http://www.s2sprediction.net/