Abstract

Our energy production increasingly depends on renewable energy sources, which impose new challenges for distributed and decentralized systems. One problem is that the availability of renewable energy sources such as wind and solar is not continuous as it is affected by meteorological factors. The challenge is to develop forecast methods capable of determining the level of power generation (e.g., through solar power) in near real-time in order to control solar power plants for optimal energy production. Another scenario is the identification of optimal locations for such power plants. In our collaborative project, these tasks are investigated in the domain of energy meteorology. For that purpose large data sources from many different sensors (e.g., satellites and ground stations) are the base for complex computations. The idea is to parallelize these computations in order to obtain significant speedup. This paper reports on an ongoing project employing Grid technologies in that context. Our approach to processing large data sets from a variety of heterogeneous data sources as well as ideas for parallel and distributed computing in energy meteorology are presented. Preliminary experience with several Grid middleware systems in our application scenario is discussed.
1 Introduction

The importance of renewable energy sources for our future energy supply is steadily increasing; thus, our dependency on these sources will simultaneously increase. However, the availability of these sources is highly influenced by meteorological factors. Thus efficient forecast methods for simulations are required to provide a near real-time estimation of the power generation (e.g., through wind power). Another scenario is the identification of applicable locations for building such power plants. For example an analysis of (archived) solar irradiation data combined with geographical information (lakes, rivers, etc.) and financial information (costs, etc.) can be used to identify optimal places for solar power plants. Such simulations or analyses are based on large heterogeneous data sets, which originate from various satellites, earth stations or other sources. Through the large amount of data the calculation time on a single computer becomes unacceptable. Next generation satellites with higher resolutions will further increase that amount of data. Today, approximately one terabyte of new data per month is received and archived in the context of our energy meteorology project WISENT (http://wisent.d-grid.de).

A solution to speedup the simulation and analysis lies in exploiting parallel and distributed computing. Therefore, the challenges of handling large amounts of data as well as the parallelization of simulations and analyses are addressed in WISENT.

For distributed processing of large data sets, the Grid infrastructure has been developed during the last decade. Grid aims at the reliable interconnection of various computing resources to enable e-Science applications. We intend to use the Grid infrastructure for data handling and distributed execution in the energy meteorology community. Various Grid middleware platforms are available for deploying a Grid infrastructure. To gain experience for our application domain, we are investigating Grid platforms such as Globus Toolkit 4, Condor, Unicore, and gLite.

Section 2 first introduces energy meteorology as an interdisciplinary application domain for e-Science before our project consortium and project structure are presented in Sections 3 and 4, respectively. In Section 5, our preliminary experience with Grid technology for e-Science in energy meteorology is discussed. We conclude and discuss future work in Section 6.

2 Energy meteorology

Energy meteorology is an interdisciplinary and application-oriented field of research. Besides physical and meteorological methods an extensive know-how in the field of wind and solar energy conversion (physics, engineering), energy supply structures (electrical engineering, economics), control techniques (computer science), and environmental science is necessary.

Basically, energy meteorology investigates the influence of weather and climate on transformation, transport, and utilization of energy. In recent years, the increasing relevance of renewable energy sources (biogas, solar, wind, hydro) in global energy supply, and therefore the rising dependence on meteorological conditions, made energy meteorology an essential part of energy research. Meteorological information has become an enormous economical factor in the energy business. [2]

Linking meteorological, physical, and technical topics is a characteristic attribute of energy meteorology. Examples of meteorological questions in energy industry are:

- Influence of weather on energy consumption, especially in power supply systems
- Determination of wind, solar, bio, and hydro energy resources
- Methods for power output prediction of solar and wind power plants
- Interaction of turbulent wind flow and wind farms
- Spatiotemporal statistics of fluctuating power production for better grid integration
- Influence of long-term weather patterns and climate changes on the power production
- Consequences of extreme weather events on power plants and the electrical grid
- Optimized measurement procedures for solar and wind power plants

The retrieval of application specific meteorological information is an important prerequisite. Usually, information provided by meteorological institutes is not directly usable (spatial resolution, accuracy) and requires suitable processing methods for amelioration.

The project partners in WISENT provide several resources that should be shared within the Grid infrastructure. The different types of these resources and their usage are illustrated in Figure 1. Meteorological data is received from many heterogeneous sources such as satellites and earth stations. Based on this data, various complex computations are performed. For these computations, we intend to exploit resources of the German Grid initiative D-Grid (www.d-grid.de). The results of these computations are used for multiple purposes, for example to control solar power plants for optimal energy production.
The project WISENT builds the knowledge network for energy meteorology. It started in October 2005. The University of Oldenburg, three institutes of the German Aerospace Center (DLR) and the company meteocontrol are collaborating with OFFIS as project coordinator. The project is funded by the German Federal Ministry of Education and Research (BMBF, http://www.bmbf.de) for a period of three years. The University of Oldenburg and each project partner of the DLR are already collaborating in the Virtual Institute of Energy Meteorology (vIEM, http://www.viem.de). Figure 2 illustrates the expertise areas contributed to WISENT by the individual project partners, who are briefly introduced in the following subsections.

3.1 OFFIS

OFFIS (http://www.offis.de), the Oldenburg Institute for Computer Science, is a technology transfer institute associated with the Department of Computer Science at the University of Oldenburg, Germany (http://www.informatik.uni-oldenburg.de). The mission of OFFIS is to transfer innovative methods developed in academic contexts into industrial practice. OFFIS is divided into six research and development units, each with its specific technology focus: safety-critical systems, embedded hardware/software systems, medical information systems, multimedia and Internet services, microsystems and nanorobotics, and business information management. The Business Information Management unit, whose research concentrates on new software engineering methods and enterprise information systems, coordinates the project WISENT. Currently, a team of about 20 people is working in this research unit (from a staff of 200 employees in total).

3.2 University of Oldenburg

The Energy and Semiconductor Research Laboratory (EHF, http://ehf.uni-oldenburg.de/) is a major focus of the Institute of Physics’ research at the University of Oldenburg which concentrates on both basic and applied research in semiconductor physics, photovoltaics, hydrodynamics, energy meteorology and wind power research. Within this spectrum, energy meteorology (http://www.energy-meteorology.de) has recently developed as a major field of applied energy research. Currently, a team of about 20 people is working in this field.

3.3 German Aerospace Center (DLR)

The German Aerospace Center (DLR, http://www.dlr.de) was formed in 1969 as a result of the merger of several former research organizations. With a staff of approximately 4000 people DLR today is the largest research establishment for engineering sciences in Germany. In recent years DLR has extended its activities to related areas, such as traffic and transport, communications, non-nuclear energy technology and environmental research, where the knowhow acquired can be made available to new, but technologically related fields of application. Within the scope of its research and development capacities, the DLR also supports technology transfer companies.
3.3.1 German Remote Sensing Data Center (DLR-DFD)

The German Remote Sensing Data Center (DFD, http://www.caf.dlr.de/caf/institut/dfd/) as the national center for remote sensing data has over the last 10 years built up an infrastructure on behalf of ESA and other commercial partners to fulfill the needs of the EO (Earth Observation) market. Together with its partners, DFD has provided data of major EO missions, such as Landsat, ENVISAT, MSG, Ikonos, SRTM, and other mainly commercially oriented satellite missions. Additionally, DFD with its partners has provided operational services, for example data acquisition and air quality monitoring.

3.3.2 Institute of Atmospheric Physics (DLR-PA)

The Institute of Atmospheric Physics (PA, http://www.dlr.de/ipa) is conducting research in physics and chemistry of the troposphere and lower stratosphere. In particular, the department of Remote Sensing of the atmosphere involved in the WISENT project is developing remote sensing methods for the characterization of ice and water clouds based on satellite observations (e.g., Meteosat first and second Generation, AHVRR, ATSR, MODIS). The developed methods are used in various applications such as climate or solar energy research.

3.3.3 Institute of Technical Thermodynamics (DLR-TT)

The department of systems analysis and technology assessment of the Institute of Technical Thermodynamics (TT, http://www.dlr.de/tt) has a long standing record in assessing the available solar resources using earth observation data. This data is processed in geographical information systems (GIS) to assess potentials of solar energy use or to select the best sites for project development. This requires the handling of very heterogeneous data sets. This is done in close cooperation with the other involved DLR institutes of Atmospheric Physics (DLR-PA) and the German Remote Sensing Data Center (DLR-DFD), see above. The focus of the GIS work has been on the support of solar technologies. It was recently widened to the integration of all renewable energy sources into the electricity grid, which will increase the amount of data to be handled.

3.4 meteocontrol GmbH

meteocontrol GmbH is the market leader for industry-specific energy and meteorological services. Numerous innovative products and services combine with energy-related know-how and meteorological competence. meteocontrol relies on state-of-the-art information technology and can draw on many years of experience in monitoring renewable energy systems. The service packages offered by the Augsburg based company with branches in Bochum and Bremerhaven are tailored to the needs of the insurance and utility industry, commerce and trade. meteocontrol arose from the IST consortium founded in 1976 and is now a subsidiary of S.A.G. Solarstrom AG.

4 Project Structure

Figure 3 gives an overview of the work packages in WISENT which will be briefly sketched in the following subsections.

4.1 WP 1 Use of Grid technologies in virtual organizations

This WP focuses on the collaboration in virtual organizations (VOs). The Virtual Institute of Energy Meteorology is an existing VO established by several project partners prior to WISENT, thus our project builds upon the given technical and organizational structures in vIEM to extend and enhance the collaboration. On the technical level, this collaboration is based on the common use of distributed resources (CPU power, data), which are connected within a Grid infrastructure utilizing appropriate Grid technologies. The collaboration on the organizational level mainly depends on software tools for Computer Supported Cooperative Work (CSCW). Therefore, we are evaluating and establishing several software tools to build up the structure of networked knowledge and information management.
4.2 WP 2 Data transfer and integration

In WP 2, we are primarily concerned with the automation, simplification and monitoring of data transfers in the operational and research context of the project partners. The partners exchange data within so-called process chains, which cover the transport of near-real-time and long-term data, used for both scientific and commercial purposes.

We are substituting the currently deployed basic communication protocols such as e-mail and FTP by Grid technologies to deal with the quality of service aspects of data transfer and integration. A uniform, standards-based data interchange platform provides several benefits. First, the administrative burden of monitoring existing transfers and dealing with failures is decreased. Second, best practices for building new process chains are established. Apart from introducing Grid technologies, improved data exchange interfaces are defined, which contribute to a smoother integration of physically and organizationally distributed applications of the project partners. Finally, we address organizational aspects which affect service quality, particularly documentation and application development practices.

4.3 WP 3 Domain-specific standards and semantic interoperability

In this WP we transfer the knowledge gained in WP 2 to energy meteorologic applications in WISENT. The focus is on standard data formats and interfaces to access heterogeneous data sources. With regard to standard data formats, we evaluate appropriate standards for the data itself and for metadata to describe the quality of data. At the end we obtain standard data formats to which different data formats can be transformed before they are further transferred or processed. Today, as mentioned before, the data sources in WISENT are distributed and accessed in different ways. Thus, we develop standard interfaces, using the standard data formats, which can be used remotely to access these data sources. Process chains to compute data sets from distributed heterogeneous data sources can be more easily developed, because the use of standardized data formats and interfaces encapsulate those details from the energy meteorologic application.

4.4 WP 4 Efficient interaction with large data objects

This WP addresses the problem of visualizing large distributed data sets over networks with low bandwidth. First, we develop new types of visualization methods within the scope of the energy meteorologic community. Second, preemptive data distribution techniques are applied within the Grid infrastructure in order to improve access to data sets by choosing data sources with the (current) network bandwidth. The results of these two approaches will be compared to each other to achieve the most applicable combination.

4.5 WP 5 Distributed processing of energy meteorology data

WP 5 encompasses the application of the Grid technology for the parallelization of computations in context of energy meteorology. Through its implementation, scaling effects of the existing systems are taken advantage of and general knowledge is gathered that will enable the energy meteorology community to use external Grid resources. These topics will be further elaborated in Section 5.

4.6 WP 6 Cooperation with other Grid communities

During the whole project we are cooperating with other Grid projects and Grid initiatives. The exchange of experiences with other national and international Grid projects is essential for the transfer of our knowledge related to Grid technologies from the energy meteorologic community to other communities and vice versa. This prevents redundant work in research of similar problems. Furthermore, the integration into the German Grid initiative D-Grid has been established (the scientific coordinator of WISENT is member of the D-Grid steering committee). D-Grid coordinates many Grid projects in different communities with the aim to integrate their Grid resources and Grid infrastructure, and to build a Grid infrastructure with a national span. This integration process itself is coordinated by the D-Grid integration project (DGI). Thus, the cooperation with D-Grid projects plays an important role in supporting the integration process. For example, Grid technologies used to build up such an infrastructure are selected based on the requirements of each scientific project. A cooperation with the European e-Science project Enabling Grids for E-Science (EGEE) is planned as well.

5 Deploying Grid-technologies for energy meteorology

Grid infrastructures are typically classified into Intra-, Extra-, and Inter-Grids (Figure 4). The term Intra-Grid refers to a Grid set up within a single organization. Several Intra-grids of different organizations are connected to an Extra-Grid, which requires stronger security policies, such as virtual private networks (VPN). An Extra-Grid often provides access to a Grid infrastructure for a certain user group with established working relationships, thus it is typical for community grids. The final extension is Inter-Grid, a global
Grid infrastructure for a wide range of independent users. At each of these three levels different Grid technologies can be utilized. WISENT belongs to Extra-Grid level, but because our project partners currently lack an existing Intra-Grids infrastructure, we are deploying Grid technologies at both levels.

Figure 4. Intra-, Extra-, and Inter-Grid (Source: IBM)

Today, several Grid technologies and corresponding Grid middleware platforms are available. We gained initial experience with Condor at Intra-Grid level and with Globus Toolkit 4 at the Extra-Grid level. These experiences are reported in the following sections.

5.1 Condor at the Intra-Grid level

The WP 5 in WISENT focuses on improving computational performance through parallelization (see Section 4.5). One promising approach in that context is provided by Condor [10], a freely available software package developed at the University of Wisconsin-Madison. Condor is a software package that has been designed with the aim to support “cycle scavenging”, that is, taking advantage of computational resources during idle periods without interfering with their primary users.

A typical Condor installation consists of a central coordinating machine and one or more machines acting as compute and/or job submission nodes. Both classes of nodes communicate with the coordinator to announce the type and availability of resources or the availability of jobs to be computed. A coordinator machine with its compute and submission nodes is also called a Condor pool. Thus, the architecture of a Condor pool is a traditional master/worker architecture. Furthermore, Condor provides explicit support for monitoring of a desktop PC user’s activities, enabling unobtrusive background execution of jobs. Condor also offers the possibility to connect several Condor pools within one or across several organizations. The latter scenario belongs to the Extra-grid level. However, we intend to use Condor only at the Intra-Grid level, that is, to connect several machines within one company. The connection at the Extra-Grid level will be accomplished by other Grid-middleware platforms such as Globus Toolkit, further described in Section 5.2.

At DLR, two scenarios, discussed next, are parallelized utilizing Condor in the domain of remote sensing methods. Their testbed consists of a Linux cluster, several servers and Linux desktop workstations.

5.1.1 Simulation of cloud scenes

For cloud scenes based on data from current weather prediction models two kinds of simulations will be carried out. The measurement of current meteorological satellite sensors will be simulated with the SIMSAT algorithm that has been developed at DLR-PA (Figure 5). The resulting data sets of synthetic observations and solar irradiance at the surface under known cloud conditions are used to validate and improve existing methods for the derivation of solar irradiance from satellite measurements. In a second step, even more realistic simulations of full 3-dimensional radiative transport for sections of the cloud scene will be conducted with the radiative transfer model MYSTIC that has also been developed at DLR-PA. Both, SIMSAT and in particular 3-dimensional MYSTIC simulations consume a lot of computational resources and can therefore benefit greatly from grid computing.

Figure 5. Meteosat 7 image (left) and SIMSAT simulation of the scene based on meteorological data from weather prediction models (right)

5.1.2 Analyzing locations for solar plants

Figure 6 illustrates the analysis process of identifying appropriate locations for solar plants [1]. The displayed area is North-West Africa (including Morocco). The analysis of solar irradiance potentials and the selection of sites involves the processing of highly heterogeneous data. It starts
from zero-dimensional data which describe some country-specific key figures such as labor cost. It includes many two-dimensional GIS data sets such as land cover, land use, available infrastructure, location-dependent insurance costs (natural risks as hail, sand storms, earthquakes). 3-dimensional data (2-dimensional in space and one time dimension) on the meteorological parameters such as solar radiation, temperatures and wind speeds are also needed. All these data come from very different sources, which additionally increases our challenges.

Figure 6. STEPS: Evaluation System for Solar Thermal Power Stations [1]

Within WISENT, strategies for an optimized handling, archiving and processing of the data will be developed at DLR TT. This will make the applications easier to use and easier to extend to new energy technologies. Distributed processing of the data will decrease computing time allowing more detailed analysis, for example by processing more parameter variations.

5.2 Globus Toolkit 4 at the Extra-Grid level

Globus Toolkit 4 (GT4) [5] was the first Grid middleware platform we started to evaluate. This decision was made because of the comprehensive available documentation [5, 8], the long development history since Globus Toolkit 1 (released in 1996), and the implementation of the Grid standard Web Services Resource Framework (WSRF) [11]. With WSRF, each resource is offered as a Grid service, that is a stateful web service implementing interfaces specified in WSRF. The development of Globus Toolkit 4 is supported by the Globus Alliance [6], a community of organizations and individuals developing Grid technologies. Ian Foster and Carl Kesselman, the initiators of the development of Globus Toolkit and inventors of the term Grid, are members of this community.

We intend to use GT4 at the Extra-Grid level because of the following reasons. GT4 offers comprehensive services such as GridFTP, Reliable File Transfer (RFT), Replica Location Service (RLS), and Data Replication Service (DRS) for data transfers. Most data transfers in WISENT are performed between our project partners and can precede computational jobs. Furthermore, GT4 offers the possibility to utilize Condor as a scheduler, thus GT4 can act as a proxy between the Extra-Grid and Intra-Grid level by submitting computational jobs to Condor. In combination with the good support of data transfers, GT4 is more applicable at the Extra-Grid level than Condor, whose services for data transfers are not as comprehensive as in GT4. Another reason is the use of the standard WSRF which is based on web services, an established technology for using distributed services across the Internet.

With GridFTP, RFT, RLS, and DRS GT4 offers a layered architecture of services for data transfers. The GridFTP service is the only non-WSRF-compliant service, which is running as a standalone server and belongs to the lowest layer. It uses an enhanced FTP protocol with features such as the support of encryption and authentication based on X.509-certificates and the option of using several parallel data channels. GridFTP provides only the basic service to transfer files from a source to a destination. The Grid service named Reliable File Transfers (RFT) builds upon this capability. It accepts a list of files that are to be transported, whereby each data transfer is processed in a reliable manner. That means, for example, that the RFT service is monitoring each transfer and retries interrupted transports up to a limited number of times before reporting an error. Another service is the Replica Location Service (RLS). The RLS supports a 1:N mapping between logical file names and physical files. Thus a physical file can have several identical copies at different locations. Finally, the Data Replication Service (DRS) is built upon both RLS and RFT. This service ensures that redundant storage servers are automatically synchronized. We will deploy GT4 to support data transfer (see WP 2 in Section 4).

Toward Grid-based data transfer in WISENT

Currently, the implemented data transfers are all point-to-point connections realized using standard technologies such as FTP, e-mail and custom web services. However, these technologies are suboptimal from the quality-of-service point of view, including reliability and maintenance.

DLR DFD, for example, is involved in most data trans-
fers. The execution of a data transfer used for delivery of a data product may be performed in two ways. Either DLR DFD offers the desired data product for download on their public FTP server or the data product is transferred directly to an FTP server of a project partner or another external facility. This FTP-based approach has some disadvantages, such as unnecessary manual overheads. For example, a transfer initiated by a cronjob might fail due to a temporary network timeout and may thus require a user intervention. Another potential problem results from the lack of synchronization between file uploads and downloads. The required data products are stored together with raw satellite data in a multi-terabyte archive called Digital Information and Management System (DIMS), which is secured behind a firewall. Thus, the data product is periodically transferred to the public FTP server, which is located in a DMZ. The recipient is also periodically looking for changes on this FTP server and downloads the desired data products. This procedure does not include checks on whether the required data products are completely available. If the series of uploads to the public FTP server is interrupted, an incomplete product may be downloaded, which is discovered late.

Our approach is a migration of each transfer toward GT4 services starting with GridFTP and RFT. Furthermore, we evaluate the introduction of the RLS and DRS. The full integration of the DIMS archive into the Grid infrastructure is planned as well.

6 Conclusions and Future Work

The use of Grid infrastructure offers many benefits for parallel processing and handling of large data sets. But this approach creates new challenges related to efficiently utilizing distributed and parallel computing resources. This paper presents the project WISENT where these challenges are addressed in a meteorological context, including data access, exchange of large heterogeneous data sets and parallelization of existing applications in energy meteorology.

After gaining first experience with Condor [10] and the Globus Toolkit 4 [4], we are experimenting with additional Grid middleware platforms such as Unicore [9] and gLite [3]. It is very likely that several Grid middleware platforms will be deployed for different purposes such that they interoperate to build our WISENT community network for energy meteorology. For instance, we intend to assess the Unicore client’s utility for modeling and executing process chains.

In WISENT, we are building an Extra-Grid whereby security issues are one central concern, besides other quality-of-service attributes. A critical success factor for a Grid infrastructure lies in achieving trustworthiness [7]. The project partners need to be sure that the Grid infrastructure is secure, and they intend to control their own participating systems. Therefore, services for monitoring, access control and logging must be supported. The Grid infrastructure must be reliable and robust in case of failures. Finally, the usability of the new technology needs to be improved. Easy access to the Grid infrastructure (e.g., Single Sign-On), fast installation and deployment of new Grid nodes and simple creation and signing of certificates are important concerns.

References