A High Performance Virtual Workspace for Management and Monitoring of Scientific Applications

Bruno B. F. Leonor[a1], Walter A. dos Santos [a] and Reinaldo R. Rosa[a]

[a]National Institute for Space Research, São José dos Campos, SP, Brazil

Abstract

Several scientific areas benefits from the research methods empowered by computers in an e-Science paradigm. One of the major issues in this context regards to the underlying infrastructure interoperability. This work provides an e-Science environment named Sentinel to support scientists in their research in a cloud environment using any enabled device to access the web. The web portal allows parameterization, initialization and monitoring of experiments transparently from net infrastructure. As an illustration, techniques as such as Detrended Fluctuation Analysis (DFA) and Gradient Pattern Analysis (GPA) will be discussed in the context of a time series analysis application to understand seemingly unpredictable behaviors as extreme events in space physics. The EMBRACE (Brazilian Space Weather Program) program was chosen as a DFA case study for the knowledge and prediction of different phenomena that directly affect human activities, such as solar activity, magnetic storms and ionospheric storms, they have great importance for the well being of society. Furthermore, new applications can be uploaded to Sentinel easily empowering researchers to manage their scientific workflows. It is hoped that this work may contribute for a friendlier environment for e-Science users.

Keywords: Time Series Analysis, Web Portal, High-Performance Computing.

1. Introduction

The web has been causing a growing impact on scientific research, technological development and society as a whole with its growth over the past few years and has brought a different style of computing involving new concepts as: *cloud computing*, *e-Science*, *big data* and *data science* [1, 2].

Many important dynamical phenomena in science can only be investigated using methods for analyzing the correspondent complex underlying process. Usually, such process are represented as high resolution data collection so that meaningful statistics and other major data characteristics. Due to the huge amount of heterogenoues scientific datasets produced daily, the sinergy among those areas are being benefited with the use of new advanced scientific computing environments [3].

[1]E-mail Corresponding Author: brunobfl@gmail.com

This has contributed to the creation of a powerful compound ecosystem of humans and machines that together are able to solve problems that could not be solved alone before. Therefore, this paper presents an innovative e-Science environment named Sentinel to support management and monitoring scientific applications in a cloud environment where scientists can add, parameterize, initialize and monitor their applications transparently using any web-enabled device.

As a case study, the prototyped technique namely Detrended Fluctuation Analysis (DFA) [4, 5] is addressed for analysis of the time series obtained by Brazilian Space Weather Program (EMBRACE) [6], which has been useful for Space Weather forecast [7].

The Space Weather was chosen as a complex phenomena case study for the knowledge, monitoring and prediction of different solar geoeffective process that directly involve a huge variety of data. The term Space Weather refers to conditions on the Sun and in the solar wind, magnetosphere, ionosphere, and thermosphere that can influence the performance and reliability of space-borne and ground-based technological systems and that can affect human life and health [8]. Observation of Space Weather is done both for scientific research and for practical applications [9]. Scientific observation has evolved with the state of knowledge, while application-related observation expanded with the ability to exploit the analysis of such big real-time data and the variety of services potentially offered to the Space Weather area.

Sentinel uses the container's concept to manage the experiments in an isolated environment. A container is the grouping of an application along with its dependencies, which share the host operating system kernel, ie, the machine (virtual or physical) which is running.

Section 2 presents the concept of e-Science. Previous works are shown in Section 3. Section 4 contains a detailed description of the Sentinel framework. In Section 5, we briefly describe the case study data base and shown the practical results. Finally is presented the main concluding remarks.

2. Potencial Application on Scientific Workflows

e-Science mainstream are commonly used to describe the development of software services infrastructure to support scientific workflows in which the Sentinel web framework fit in [10]. This enables access to remote facilities, distributed computing resources, information storage in dedicated databases, dissemination and sharing of data, results and knowledge.

Research in e-Science requires joint and multidisciplinary efforts where computer scientists assist researchers from other fields of knowledge to develop their research more quickly and efficiently and, in this partnership, the computer solutions tend to be made innovative and transformatively [11].

Basically, the e-Science research can be described by the following characteristics [11]: (1) involves collaboration of researchers computing with researchers from other fields of knowledge, and (2) requires the creation of sophisticated computational methods to handle large volumes of data (big data) and/or to perform simulations and programs that require complex computational systems.

The data base from Space Weather observations was chosen as a complex data warehousing for the knowledge, monitoring and prediction of different solar geo-effective process that directly involve a huge variety of data (time series and digital images). The volume of data must be processed several times by different users on different platforms. The workflow of usage involving different applications (analytical tools, such as DFA) should be made optimal use of the cloud and the algorithm (which may distribute threads among the cores of a hybrid HPC architecture).

3. Previous Works

Currently there are several scientific portals around the world, which enables the sharing of knowledge generated by projects in their research areas. The vast majority of scientific portals were created to solve specific scientific problems, so have a few and restricted applications.

In Brazil we can highlight the scientific portals offered by SINAPAD (National System of High Performance Processing) that allow the execution of scientific applications on clusters that make up the national system. Currently available gateways are [12]: (1) ACES3, (2) BRAMS, (3) MAC-GRID, (4) DANCE, (5) DockThor, (6) Gaussian, (7) GdfidL, (8) PrimTest, (9) Profrager, (10) SPiNMe and (11) TrueRNG.

Other research areas explored by scientific portals are: big data, astrophysics, climate and Earth science, biology and chemistry, among others.

4. The Sentinel Virtual Workspace

Sentinel e-Science environment described in this paper allows each researcher to develop their scientific research through any web-enabled device. In their scientific research, each researcher will dispose of a set of tools created by your own as well as take advantage of tools shared by other researchers in your workspace.

There are two ways to researchers to access the portal, either as a ”researcher” or ”guest”. Figure 1 shows the main features offers by Sentinel.

The Sentinel portal allows researchers to concentrate all necessary applications in order to develop their research in a single environment with no knowledge of the computational environment infrastructure. Each researcher will have your own

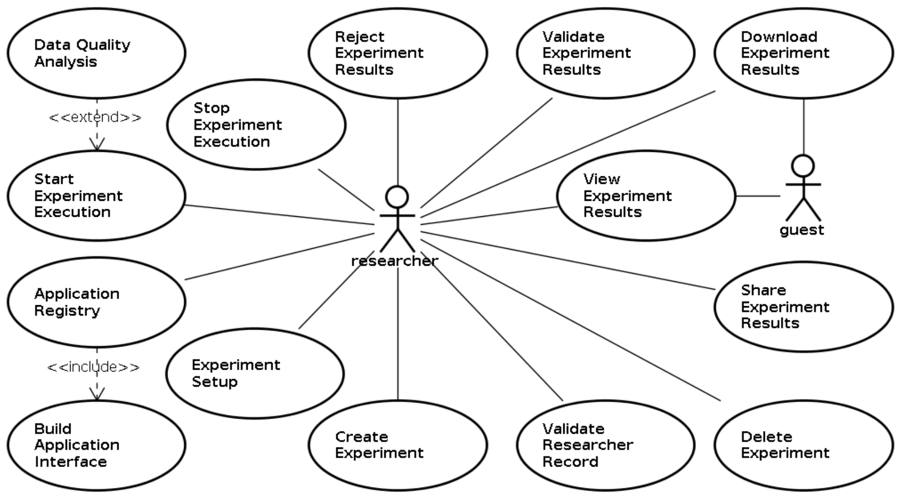

Figure 1 Use case diagram - Main Sentinel's functionalities.

desktop with your set of tools that will be built from the Application Registry interface.

The Sentinel implementation is currently being done using Java [13–15]. The authors believe that this may facilitate future maintenance and improvements in the environment due to the large community of developers.

All data necessary to experiment's run are stored at the MongoBD, a document database that provides high performance, high availability, and automatic scaling [16].

In order to automate the deployment of new applications each one is performed in an isolated and secure environment through the Docker containers [17]. By using containers, resources can be isolated, services restricted, and processes provisioned to have an almost completely private view of the operating system [18].

The Sentinel's workflow is divided in three simple steps as shown in Figure 2.

Figure 2 A typical Sentinel's workflow.

Step 1 - Create Application: In order to have its own set of tools available in your workspace, researchers must first incorporate them into the environment so they can perform their experiments. Therefore, in Application Registry interface the researcher may add all the information necessary for the implementation of application, ranging from the programming language used in the application setting up parameters needed for its execution. This informations are grouped into the tabs `File`, `Variables`, `Lib` and `Compiler`.

Each registered application may have their visibility set by the researcher to be either public or private. If set as public, other researchers will have access to use it. The `File` tab is intended for uploading the application source code. The input variables are defined for the application parameterization in the `Variables` tab. If necessary, the inclusion of some special library may be added to the environment by `Lib`. The last tab, `Compiler`, is used to define the application's build parameters within the Sentinel portal environment where the researcher can enter a script to compile and run the application.

Step 2 - Set up Experiment: After registering by the researcher in the environment, the application is ready to run. The "Experiment Setup" interface is created automatically by the framework based on the information entered in the Application Registry interface - `Variables` tab. The user needs to set the values that will serve as input for the execution of the application.

All the settings created to perform this step are generically called as an experiment.

Step 3 - Experiment Management and Monitoring: The Sentinel Dashboard interface, shown in Figure 3, is where the user can track all experiments.

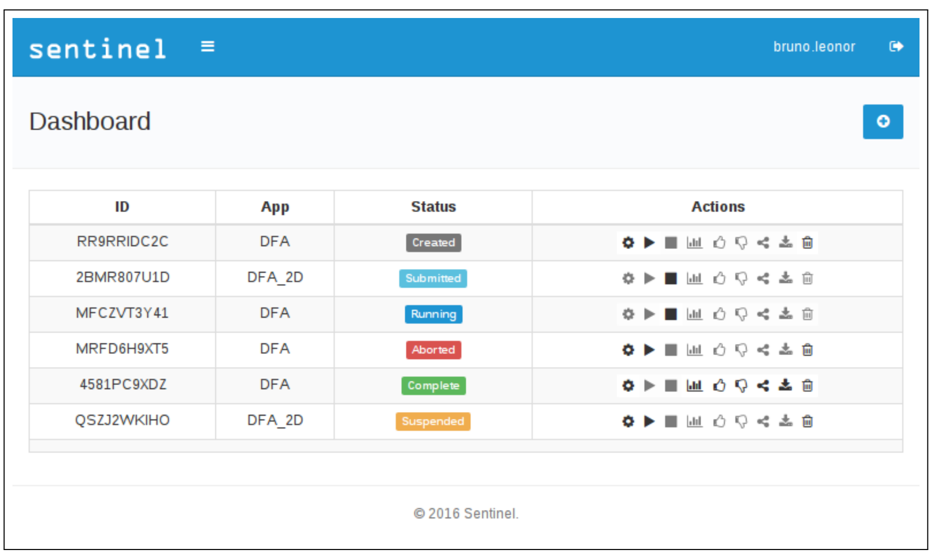

Figure 3 The Sentinel Dashboard Interface.

In order to help a researcher to monitor the experiments, Sentinel offers a set of information regarding each experiment, namely: `ID` (experiment identifier), `App` (chosen application), `Creation Date`[2] (creation date of the experiment), `Submission Date`[2] (request date for execution), `Start Execution`[2] (initial date of execution), `End Execution`[2] (end of the execution date) and `Status` (current situation of the experiment).

Through the `Actions` column the researcher can manage each one of experiments with the following commands: Set up Experiment (⚙), Start Execution (▶), Stop Execution (■), View Results (📊), Validate Results (👍), Reject Results (👎), Share Results (<), Download Results (⤓) and Delete an Experiment (🗑) and its results.

5. Case Study on Space Weather Monitoring

A case study from Solar Physics is hereafter presented as an example of using the system. To this study, it was used the Total Solar Irradiance (TSI) data from VIRGO/SOHO mission [19] extracted from the dataset file named composite_d41_62_1204.dat [20].

The time series, shown in Figure 4, was cutted in windows representing the observations made to each year (from 1978 to 2012), totalizing 35 windows.

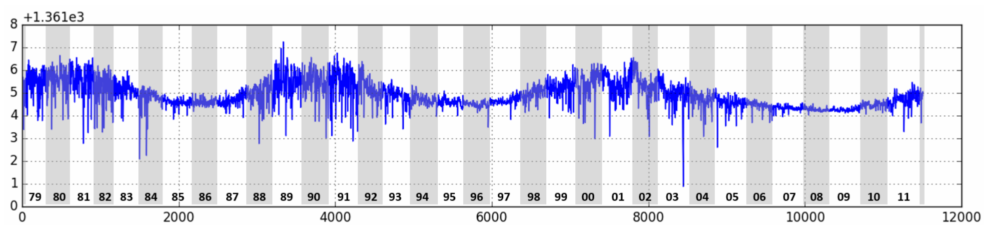

Figure 4 Total Solar Irradiance time series cutted into 35 windows.

For analysis of the time series obtained by EMBRACE [6], this work has selected only one well accepted technique as a case study, namely DFA.

The Brazilian Space Weather Program was created in 2007 with the mission to carry out the monitoring, modeling and dissemination of Space Weather information regarding to phenomena research and forecasting of significant effects on the space and on the surface of the Brazilian territory, including impacts on space and ground technological systems [6].

The DFA method was first proposed by [4] to analysis of DNA sequences. DFA was basically designed to reveal long range correlation in non stationary processes [5]. However, DFA can also be regarded as a suitable method to investigate both long-range and short range correlation in non-stationary and stationary systems.

[2]Hidden element in Figure 3 to fit the page layout according to the device resolution used by the researcher.

The behavior from DFA analysis can be useful in a framework of operational data analysis providing monitoring, alerts and forecasts of solar flares and, possibly, of geomagnetic activity [7].

This case study uses an environment based on AMD A10-4655M APU @ 2.0GHz processor with 6 GB of memory for DFA's execution under two comparative cases. In the first one, the application was executed into a Debian GNU/Linux 8.3 (jessie) x86_64 OS. The second represents the application running in a docker container using Docker version 1.12.1 (build 23cf638) and the same OS of case 1. In case 2 are necessary two extra steps: go up and down the container.

Table 1 shows some Elapsed Time and Computing Throughput (data points processed by seconds) in both case.

Table 1: Elapsed Time of execution and Computing Throughput

Window	**Data Points**	**Elapsed Time**		**Computing Throughput**	
		(Case 1)	(Case 2)	(Case 1)	(Case 2)
78	34	0m1.286s	0m1.151s	26,44	29,54
79	262	0m1.318s	0m1.377s	198,79	190,27
80	331	0m1.357s	0m1.547s	243,92	213,96
81	274	0m1.305s	0m1.257s	209,96	217,98
82	272	0m1.299s	0m1.718s	209,39	158,32
83	301	0m1.299s	0m1.188s	231,72	253,37
84	347	0m1.321s	0m1.385s	262,68	250,54
85	348	0m1.302s	0m1.430s	267,28	243,36
86	338	0m1.294s	0m1.461s	261,21	231,35
87	340	0m1.305s	0m1.292s	260,54	263,16
88	362	0m1.327s	0m1.475s	272,80	245,42
89	352	0m1.275s	0m1.319s	276,08	266,87
...	...	...	...	...	...
02	362	0m1.516s	0m1.016s	238,79	356,30
03	363	0m1.512s	0m1.267s	240,08	286,50
04	358	0m1.516s	0m1.631s	236,15	219,50
05	365	0m1.511s	0m1.297s	241,56	281,42
06	365	0m1.493s	0m1.409s	244,47	259,05
07	365	0m1.488s	0m1.561s	245,30	233,82
08	366	0m1.504s	0m1.460s	243,35	250,68
09	365	0m1.513s	0m1.341s	241,24	272,18
10	365	0m1.512s	0m1.407s	241,40	259,42
11	360	0m1.602s	0m1.542s	224,72	233,46
12	88	0m1.465s	0m1.521s	60,07	57,86

Elapsed times of execution for all windows of the time series were plotted in a line chart, Figure 5, to both cases.

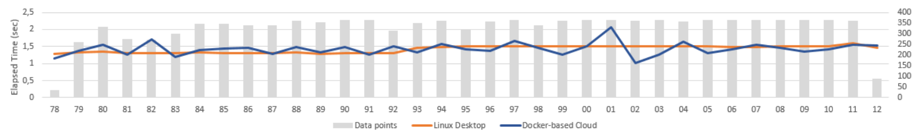

Figure 5 Comparative DFA execution time under a Linux desktop vs. a Docker-based cloud environment.

Concluding Remarks

This work presented a web framework, named Sentinel, dedicated and suitable for managing e-Science applications for Space Weather purposes. In particular, the EMBRACE Space Weather program was chosen as a domain for the case study for the knowledge and prediction of different phenomena that directly affect human activities, such as solar activity, magnetic storms and ionospheric fluctuations. These topics have great importance for the well-being of society. As a prototype application, the Sentinel framework has been applied for the solar irradiance time series monitoring using DFA.

Since the cloud computing structure is flexible, new techniques and algorithms can be incorporated into the Sentinel portal by scientists and scientific teams. It is important to mention that a module of paramount importance in the system concerns the data quality. This module will be incorporated and discussed in a future paper. The authors expect that this work may contribute to a friendlier and safer environment to e-Science users, where they can gather all their applications used in the development of their research and information technology services.

Future work will incorporate more challenging experiments and algorithms. Some example of applications that can be incorporated to the Sentinel for mutidimensional time series analysis considering big data files and digital images are: PSD (Power Spectral Density) [21], Singularity Spectrum [22], DFA2D (Detrended Fluctuation Analysis 2D) [23], GPA [24], among others.

Acknowledgments.

The authors are greatful to the PMOD/WRC, Davos, Switzerland for the dataset provided (composite_d41_62_1204.txt) from the VIRGO Experiment on the cooperative ESA/NASA Mission SOHO. The authors would like to thank Vanessa C. Oliveira de Souza for providing the DFA code used in the case study presented. RRR acknowledge financial support from FAPESP through the grant number 2014/11156.

References

[1] Shadbolt, N., Hall, W., Hendler, J.A., Dutton, W.H., Web science: a new frontier, The Royal Society **371**(1987), 2013.

[2] Alisson E., eScience revoluciona a forma como se faz ciência, available at: http://agencia.fapesp.br/17279, accessed January 2016.

[3] Tolle, K., Tansley, S., Hey, T., O Quarto Paradigma: Descobertas Científicas na era da eSciente, Oficina de Texto, 2011.

[4] Peng, C.K., Buldyrev, S.V., Havlin, S., Simons, M., Stanley, M.H.E., Goldberger, A.L., Mosaic organization of DNA nucleotides, Phys. Rev. E **49**(2): 1685-1689, 1994.

[5] Morariu, V. V., Buimaga-Iarinca, L., Vamos, C., Soltuz, S., Detrended Fluctuation Analysis of Autoregressive Processes, Cornell University Library, 2007.

[6] INPE, O Programa Embrace, available at: http://www2.inpe.br/climaespacial/portal/, accessed January 2016.

[7] Veronese, T.B., Rosa, R.R., Bolzan, M.J.A., Fernandes, F.C.R., Sawant, H.S., Karlicky, M., Fluctuation analysis of solar radio bursts associated with geoeffective X-class flares, Journal of Atmospheric and Solar-Terrestrial Physics **73**(11-12): 1311-1316, 2011, available at: http://www.sciencedirect.com/science/article/pii/S1364682610002907.

[8] Schwenn, R., Space Weather: The Solar Perspective, Living Reviews in Solar Physics **3**(2), 2006, available at: http://www.livingreviews.org/lrsp-2006-2.

[9] Cade III, W.B., Chan-Park, C., The Origin of Space Weather, Space Weather **13**(2): 99-103, 2015.

[10] Palazzi, D., Silva, L., Mendes, L.F., Gaspar, W., Matos, E., Campos, F., Braga, R., Uso de ontologias em projetos de e-science, BDBComp, 2009.

[11] Fapesp, Programa FAPESP de Pesquisa em eScience, available at: http://www.fapesp.br/publicacoes/2015/folder_escience.pdf, accessed December 2015.

[12] LNCC, Portais Científicos, available at: https://www.lncc.br/sinapad/gateways.php?pg=gateways, accessed September 2016.

[13] Tiobe, TIOBE Index for February 2016, available at: http://www.tiobe.com/tiobe_index, accessed February 2016.

[14] RedMonk, The RedMonk Programming Language Rankings: January 2016, available at: https://redmonk.com/sogrady/category/programming-languages/, accessed February 2016.

[15] Spectrum, The 2015 Top Ten Programming Languages, available at: http://spectrum.ieee.org/computing/software/the-2015-top-ten-programming-languages, accessed February 2016.

[16] mongoDB, Introduction to MongoDB, available at: available at: https://docs.mongodb.com/manual/introduction/, accessed August 2016.

[17] Docker, What is Docker?, available at: https://www.docker.com/what-docker, accessed August 2016.

[18] Cirrus Tech Ltd, Docker: Reorganizing the Data Center, available at: http://www.cirrushosting.com/cloud-hosting-blog, accessed August 2016.

[19] Frohlich, C., Lean, J. The Sun's total irradiance: Cycles, trends and related climate change uncertainties since 1976, Geophysical Research Letters **25**(23): 4377-4380, 1998, available at: http://dx.doi.org/10.1029/1998GL900157.

[20] ACRIM, composite_d41_62_1204.dat, 2012, available at: acrim.com/TSI/composite_d41_62_1204.txt.

[21] Oppenheim, A.V., Verghese, G.C., Signals, Systems and Inference, Prentice-Hall signal processing series, Pearson Education, 2015.

[22] Bolzan, M.J.A., Tardelli, A., Pillat, V.G., Fagundes, P.R., Rosa, R.R.,

[23] Gu, G.F., Zhou, W.X., Detrended fluctuation analysis for fractals and multifractals in higher dimensions, Phys. Rev. E **74**(6), American Physical Society, 2006.

[24] Rosa, R.R., Sharma, A.S., Valdivia, J.A., Characterization of Asymmetric Fragmentation Patterns in Spatially Extended Systems, International Journal of Modern Physics C **10**(01): 147-163, World Scientific Publishing Company, 1999.