Automating usability of ATLAS Distributed Computing resources

S A Tupputi[1], A Di Girolamo[2], T Kouba[3] and J Schovancová[4] for the ATLAS Collaboration

[1] INFN - CNAF, viale B. Pichat, 6/2, Bologna 40127, Italy
[2] CERN, CH - 1211 Geneva 23, Switzerland
[3] Institute of Physics, Academy of Sciences of the Czech Republic, Na Slovance 2, CZ-18221 Prague 8, Czech Republic
[4] Brookhaven National Laboratory, Upton NY, USA

E-mail: salvatore.tupputi@cern.ch

Abstract. The automation of ATLAS Distributed Computing (ADC) operations is essential to reduce manpower costs and allow performance-enhancing actions, which improve the reliability of the system. In this perspective a crucial case is the automatic handling of outages of ATLAS computing sites storage resources, which are continuously exploited at the edge of their capabilities. It is challenging to adopt unambiguous decision criteria for storage resources of non-homogeneous types, sizes and roles. The recently developed Storage Area Automatic Blacklisting (SAAB) tool has provided a suitable solution, by employing an inference algorithm which processes history of storage monitoring tests outcome. SAAB accomplishes both the tasks of providing global monitoring as well as automatic operations on single sites. The implementation of the SAAB tool has been the first step in a comprehensive review of the storage areas monitoring and central management at all levels. Such review has involved the reordering and optimization of SAM tests deployment and the inclusion of SAAB results in the ATLAS Site Status Board with both dedicated metrics and views. The resulting structure allows monitoring the storage resources status with fine time-granularity and automatic actions to be taken in foreseen cases, like automatic outage handling and notifications to sites. Hence, the human actions are restricted to reporting and following up problems, where and when needed. In this work we show SAAB working principles and features. We present also the decrease of human interactions achieved within the ATLAS Computing Operation team. The automation results in a prompt reaction to failures, which leads to the optimization of resource exploitation.

1. Introduction

The Large Hadron Collider (LHC) at CERN has delivered colliding beams at the centre-of-mass-energy of 7 TeV since March 2010 and at the center-of-mass-energy of 8 TeV since April up to December 2012. The ATLAS Experiment [1], one of the LHC general purpose detectors, has collected over 30 PB of RAW data in the same period. ATLAS benefits of the World-wide LHC Computing Grid (WLCG) to process data and simulations.

The ATLAS grid resources (CPU resources, storage systems, network links) are spread worldwide over more than 120 computing centres to make up the ATLAS Distributed Computing (ADC) infrastructure [2]. Such variety entails different flavours of storage systems and heterogeneous CPU resources which have to be fully exploited in order to provide a good quality service to the ATLAS Collaboration. Thus it is crucial for the ADC Operations teams to be able to identify and promptly address any issue within the ADC infrastructure.

The task of monitoring ADC activities (Data transfers, Data processing, Distributed analysis) is addressed by the ADC Monitoring team [3]. Each activity is supported by a shift team, backed up by two ADC Experts, while site issues are addressed by site administrators. The shifters report the problem to expert or activity requester or to the site team whenever an issue within the ADC infrastructure is identified. Over 8800 problem reports have been issued for ADC sites since 1st January 2010, averaging to 7 reports per day. Such rate involves a huge manual effort: shifter investigating on issues, creating the report for site administrators or activity requesters to address the issue, then setting up functional testing of the reported service, and eventually including the service back into production.

The manual repetitive work for the reported issues involves dealing with many occurrences of similar-looking cases. Moreover human errors (for instance, color-blindness or fatigue due to long time spent in iterative operations) may occur or actions may result depending on the rules as interpreted by the person currently in charge. Hence the need for non-ambiguous and correct actions calls for automation on well known issues. In order to meet such needs several tools have been implemented so far both for CPU (e.g.: Switcher, PanDA site exclusion) and storage resources management [4]. Among these, an automatic tool able to provide performance-wise tests of any resource type has been developed and has been named Storage Area Automatic Blacklisting (SAAB), since it has been first applied on storage resources.

ADC storage resources are operated by the Distributed Data Management (DDM) system [5] which performs operations on and monitoring of the sites storage elements. The elemental ATLAS storage unit is called DDM endpoint (DDM EP) and all storage elements comprise at least one or more DDM endpoints. The DDM system keeps track of DDM endpoints status: each DDM endpoint can be in *on* or *off* status, meaning no automatic action upon them is allowed. Only DDM endpoints set in *auto* status can be acted upon and have their accessibility and availability changed. In this latter case, only declared downtimes and space shortage conditions have been considered in recent past to change DDM endpoints status. Thus, automatic actions based on performance-related metrics fill a gap in the DDM endpoints management.

In Section 2 the SAAB tool general principles and implementation features are described. In Section 3 it is outlined SAAB operational experience in assessing ADC sites storage resources performance and consequently acting on their status. Finally, in Section 4 conclusions about such experience are drawn and the future developing perspectives of the tool are briefly discussed.

2. SAAB

In grid computing for large experiments, it is pivotal to monitor the resources status for each grid site by means of a suited metric. To this purpose, several suitable metrics have been developed and set to constantly test sites facilities, thus assessing the status of monitored resources.

2.1. Principles

Historical information on performance provided by such metrics is useful in the scenario of automating repetitive actions for known problems. Indeed such information can be suitably used to implement decisions-enabling automatic tools for real time action; the SAAB tool has been drawn out of this perspective and its designed workflow can be described as follows:

(i) Consider a time series of test results for a resource yielded by a given metric at a constant frequency; a fixed-width time window is defined over such series, so that a given number of test results is comprised. Such time window establishes the SAAB tool elemental input unit for single decisions (top of fig. 1).

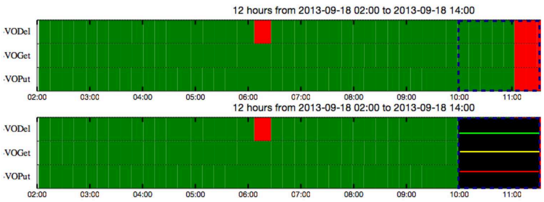


Figure 1. Metric results in time window over which thresholds are applied. Tests execution time is on X axis, test type is on Y axis.

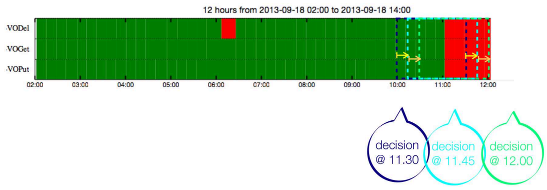


Figure 2. At each algorithm execution a decision is taken.

(ii) A suitable algorithm for processing input data and providing an unique interpretation of the overall resource status is designed. The most straightforward approach consists in introducing a set of thresholds which the input data are compared against (bottom of fig. 1).

(iii) Each possible outcome of the algorithm is associated to an action decision to be taken.

(iv) The algorithm is executed at a steady frequency with a different set of input data, so that action decisions are drawn out at regular time intervals (fig. 2).

Such approach can be represented by a time window sliding over the metric history by means of which automatic exclusion/re-inclusion of resources is inferred in compliancy with a common policy.

2.2. Implementation

The SAAB tool is coded in python and at each execution collects information on ADC sites from AGIS (ATLAS Grid Information System) [6] and on the metric results. The results are evaluated according to the positive or negative tests outcome. SAAB outputs a set of operational instructions, according to which a resource is excluded from usage or re-included (in the following referred to as blacklisting and whitelisting, respectively). SAAB output is also used for resources status monitoring. Real-case details are provided in Section 3.

The SAAB algorithm features two adjustable thresholds, a “lower” (l) and an “upper” (u) one, which test results are compared against. Three percentage intervals are identified by applying these thresholds to the percentage of tests with a positive outcome in the time window. For a given percentage p of tests number with positive outcome, the following flags are associated to each interval (fig. 3):

- RED if $p \leq l$.
- YELLOW if $l < p \leq u$.
- GREEN if $p > u$.

Figure 3. Example of percentage thresholds applied to positive outcome tests.

In addition to that, the number of tests to be processed need to be above a critical threshold for the algorithm to ensure meaningful results. Indeed the metric results may be unexpectedly unavailable at times for a given reason. Hence a minimum amount of tests is required, which is set to be greater than a given percentage m of the expected total number of tests falling within the fixed time interval. If such requirement is not met, the flag GRAY is used. The SAAB flags are coded as strings and used as input for exclusion actions purposes, while are used as colors for monitoring-only purposes. The color choice follows the one in use for ADC monitoring tools.

Moreover SAAB tool also provides functionalities such as: automatic mail alerts and notifications, filtering and redirection of produced information to distinct log files for different purposes and a configuration file for tuning algorithm parameters, setting logging verbosity or debugging mode and enabling or disabling actual actions.

3. Operational experience of SAAB applied on ADC storage resources

The storage resources metrics whose results are used as input for SAAB are the Site Availability Monitoring (SAM) tests [7, 8]. SAM is the framework used by LHC experiments to test resources status and availability. It is deployed as a Nagios-based [9] service and, as such, yields three possible outcomes: OK, WARNING, CRITICAL. Among the SAM tests probing the sites storage resources, the most relevant ones are the *Put*, *Get*, *Del* tests, which perform basic I/O operations on storage resources. In the current ATLAS SAM configuration, tests are executed every 13 minutes: this value has been chosen as the best trade-off, given the Nagios machine hardware, providing a tests frequency reasonable for the experiment needs. This parameter has to be considered as a constant for SAAB purposes. The SAM tests outcome for each tested DDM endpoint at each site is stored in a database. From such database both latest results and historical data are exposed through a RESTful interface. As a first stage, only the results from *Put* SAM test have been used as SAAB input; the *Put* test probes the status of uploading and writing permissions on DDM endpoints. Hence the blacklisting or whitelisting actions performed by SAAB on a DDM endpoint translate to disabling or re-enabling the uploading and writing permissions on DDM endpoints. In any case SAM tests keep being executed on DDM blacklisted endpoints.

3.1. SAAB as automation tool

The SAAB time window has been set to 90 minutes. This is a trade-off between needing a window narrow enough to appreciate the information from the most recent tests and avoiding fluctuations in the number of tests falling in the window. Such fluctuations may arise from test results which may be missing for various reasons (delays in writing the results in the database, tests stuck in some storage resources for resource problems, etc.). Moreover, given the SAM tests frequency of 13 minutes, SAAB runs every 15 minutes. This ensures that, on average, at least one new SAM test result is included with respect to the previous execution. These settings translate to an average of 6 expected SAM tests for each DDM endpoint at each SAAB execution. With reference to Section 2, the actions carried out by SAAB at each execution break down as follows:

- Each RED-flagged DDM endpoint is set to be blacklisted for 90 minutes, regardless its current status.
- Each GREEN-flagged DDM endpoint currently blacklisted is set to be whitelisted.
- No action is taken in any other case (YELLOW, GRAY)

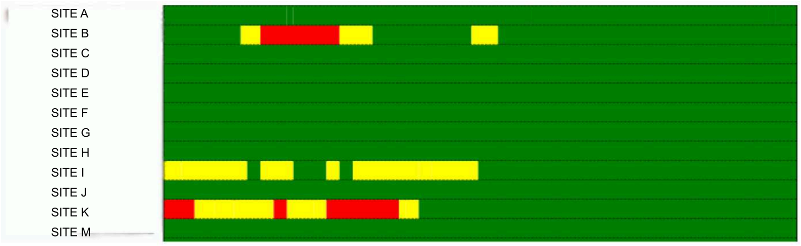


Figure 4. SAAB as Site-grained SSB metric. SAAB execution time is on X axis. Y axis displays ADC sites (Site names are placeholders).

In this context, SAAB is not the means by which the actual blacklisting is executed: SAAB makes us of APIs of DQ2, the DDM tool designated for such actions [5]. The l and u thresholds (see Section 2.2) have been set as $l = 0\%$ and $u = 60\%$, after being tested with different values. The SAAB observed behavior with tighter values ($l = 50\%$, $u = 80\%$) presents the benefit of prompt action. However it is too much sensitive to temporary system glitches (i.e. SAM machinery or storage resources or both): indeed it may blacklists sites for too short periods to be actually meaningful for the ADC operations. For the whitelisting action, on the other hand, cases have been observed where a recovered resource had to wait longer than necessary for being used again. Both cases may be referred to as false negatives. Hence a conservative approach has been preferred, which reduces the occurrence of false-negatives and results in a more stable resource management. A cross-check with other DDM blacklisting tools (for downtimes and disk-space) as well as SAM results history analysis confirms the consistency of the choice. Finally, the minimum SAM tests number m required for SAAB algorithm is set to $m = 30\%$. A tighter threshold ($m = 65\%$) has been observed to lead to too many indefinite cases on DDM endpoints status.

From notifications standpoint, a log file containing output from each SAAB execution and another one only featuring SAAB actual actions are accessible via web. Mail alerts are sent to SAAB developers in case of issues with SAM results retrieval and, in addition to that, sites are informed whenever one of their DDM endpoints is acted upon by SAAB. However, if a problem or a blacklisting action for a site keeps continuously occurring, no further mails are sent since the system keeps memory of previously taken alert actions.

3.2. SAAB as monitoring metric tool

As a monitoring tool, SAAB suitably arranges the produced information to be exposed through the ATLAS Site Status Board (SSB) [10]. The SSB is a framework which allows to monitor the ADC computing activities at distributed sites and to evaluate site performance. It is used in ADC distributed shifts, for estimating data processing and data transfer efficiencies at a particular site, to provide real-time aggregated monitoring views and for rendering the history of the monitoring metrics. Based on this history, usability of a site from the ATLAS perspective can be calculated.

SAAB results history is exposed as both site-grained and DDM endpoint-grained SSB metrics. In the first case (fig. 4), the site status time-bar reflects the OR of the site's single DDM

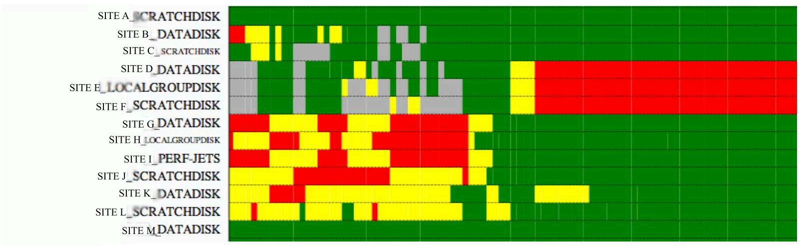
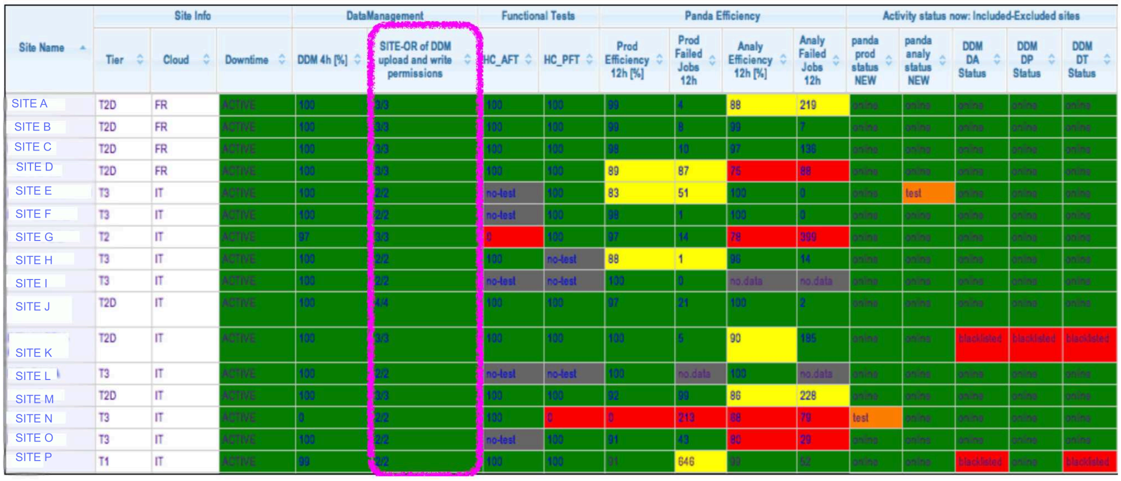


Figure 5. SAAB as DDM endpoint-grained SSB metric. Y axis displays the ADC DDM endpoints (Site names are placeholders).

endpoints statuses. Clicking on the site's bar redirects to the DDM endpoint-grained SSB metric (fig. 5), which features the status history for each DDM endpoint. From the DDM endpoint metric, clicking on the DDM endpoint's bar redirects the history of SAM tests results for the site DDM endpoints belong to. An instance of the latter is employed in the figures 1 and 2.



Site Name	Site Info			DataManagement		Functional Tests		Panda Efficiency				Activity status now: Included-Excluded sites				
	Tier	Cloud	Downtime	DDM 4h [%]	SITE-OR of DDM upload and write permissions	HC_AFT	HC_PFT	Prod Efficiency 12h [%]	Prod Failed Jobs 12h	Analy Efficiency 12h [%]	Analy Failed Jobs 12h	panda prod status NEW	panda analy status NEW	DDM DA Status	DDM DP Status	DDM DT Status
SITE A	T2D	FR	ACTIVE	100	3/3	100	100	99	4	88	219	online	online	online	online	online
SITE B	T2D	FR	ACTIVE	100	3/3	100	100	99	8	99	7	online	online	online	online	online
SITE C	T2D	FR	ACTIVE	100	3/3	100	100	98	10	97	136	online	online	online	online	online
SITE D	T2D	FR	ACTIVE	100	3/3	100	100	89	87	75	88	online	online	online	online	online
SITE E	T3	IT	ACTIVE	100	2/2	no-test	100	83	51	100	0	online	test	online	online	online
SITE F	T3	IT	ACTIVE	100	2/2	no-test	100	98	1	100	0	online	online	online	online	online
SITE G	T2	IT	ACTIVE	97	3/3	0	100	97	14	78	399	online	online	online	online	online
SITE H	T3	IT	ACTIVE	100	2/2	100	no-test	88	1	96	14	online	online	online	online	online
SITE I	T3	IT	ACTIVE	100	2/2	no-test	no-test	100	0	no.data	no.data	online	online	online	online	online
SITE J	T2D	IT	ACTIVE	100	4/4	100	100	97	21	100	2	online	online	online	online	online
SITE K	T2D	IT	ACTIVE	100	3/3	100	100	100	5	90	185	online	online	blacklisted	blacklisted	blacklisted
SITE L	T3	IT	ACTIVE	100	2/2	no-test	no-test	100	no.data	100	no.data	online	online	online	online	online
SITE M	T2D	IT	ACTIVE	100	3/3	100	100	92	99	86	228	online	online	online	online	online
SITE N	T3	IT	ACTIVE	0	2/2	100	0	0	213	68	70	test	online	online	online	online
SITE O	T3	IT	ACTIVE	100	2/2	no-test	100	91	43	80	29	online	online	online	online	online
SITE P	T1	IT	ACTIVE	99	2/2	100	100	91	646	93	52	online	online	blacklisted	online	blacklisted

Figure 6. Shifter SSB view where the highlighted column features the SAAB-collected results (Site names are placeholders).

SAAB-produced information is also displayed in the general SSB view, which is used by ADC shifters as the overall sites status dashboard, as sketched in fig. 6.

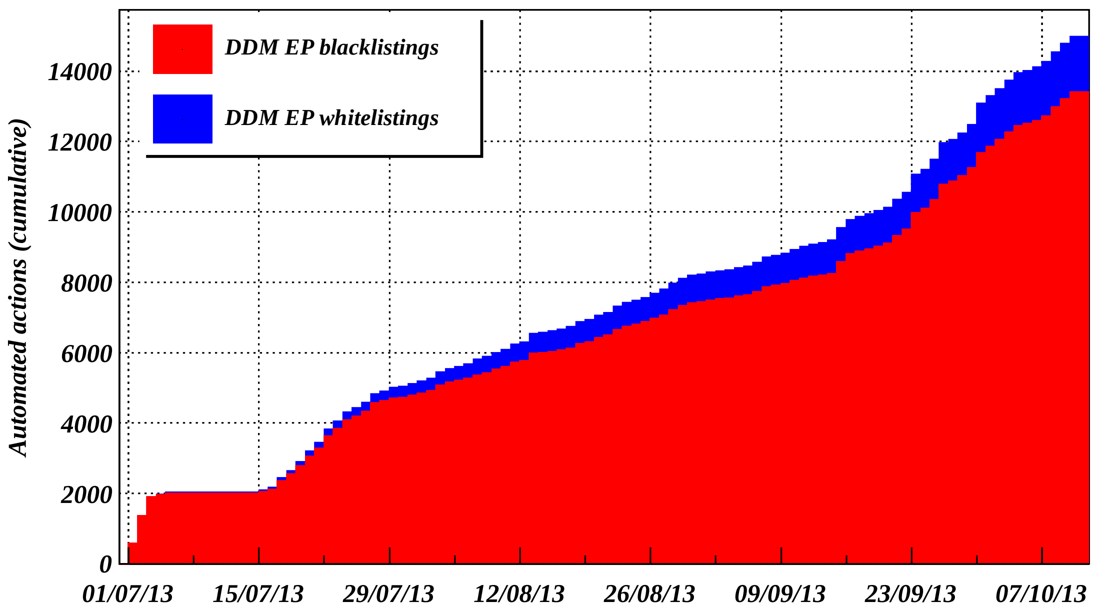


Figure 7. Cumulative plot with the amount of automatic actions taken by SAAB.

Finally, in fig. 7 it is possible to appreciate the amount of automatic actions taken by SAAB starting from July 2013, when it has been put in production, up to about the beginning of October 2013: it directly translates to more than 14.000 spared manual actions in the span of little more than three months. The asymmetry between automatic blacklisting and whitelisting actions is due to whitelisting occurring only when a DDM endpoint gets the GREEN flag before the blacklisting time expiration (90 minutes).

4. Conclusions and further perspectives

4.1. SAAB production experience

The SAAB tool principles have been implemented for achieving automatic management of storage resources based upon their performance. After the first production operations it is possible to conclude that they have proved successful in enhancing automated ADC activities. A first improvement achieved consists in replacing human educated-guesses with unambiguous and reproducible decision criteria. Such criteria lead to a uniform performance-wise management of the heterogeneous ADC storage resources. An operational improvement is achieved by enabling automatic actions to decrease the amount of manpower needed for repetitive operations.

In addition, its clear and minimal design principles provide SAAB with the potential to act as a management tool for different types of grid resources, thus envisioning the possibility of deploying automatic, performance-based actions by means of a single, uniform and clear approach.

4.2. Extending SAAB

Regarding the SAAB storage resources implementation, the most forthcoming advance in SAAB is to use as input also the results from *Get* and *Del* SAM tests, which probe fetching, reading and deleting permissions on DDM endpoints.

Besides, it has been observed that, in most cases, when a site's DDM endpoint gets blacklisted following its poor performance, the other DDM endpoints shortly get blacklisted as well. This suggests that when a DDM endpoint is blacklisted it may be enough to observe a few more test failing on the other DDM endpoints to blacklist those too, instead of waiting for the 0% positive tests threshold to be reached: the implementation of this feature is currently being pondered over as a next-to-follow new functionality.

On a different perspective, one of the branch of future work on SAAB, as already mentioned in 4.1, concerns the application of SAAB tool (regardless of the acronym meaning) to metrics testing other resources such as, for instance, networks and data transfer.

Further on, SAAB does not feature any core dependence on a particular experiment environment, so it may also be applied as a general control decision tool.

References

[1] The ATLAS Collaboration 2008 *JINST* **3** S08003
[2] Jézéquel S and Stewart G 2012 *J. Phys..: Conf. Series* **396** 032058
[3] Schovancová J for the ATLAS Collaboration 2012 *J. Phys.: Conf. Series* **396** 032095
[4] Schovancová J et al for the ATLAS Collaboration 2012 to appear in *Proc. 5th Int. Conf. Distr. Comp. and Grid-Techn. in Science and Educ. (Dubna, 16-20 July 2012)* JINR, Dubna
[5] Branco M et al for the ATLAS Collaboration 2008 *J. Phys.: Conf. Series* **119** 062017
[6] Anisenkov A et al for the ATLAS Collaboration 2012 *J. Phys.: Conf. Series* **396** 032006
[7] Babik M et al 2012 *J. Phys.: Conf. Series* **396** 032008
[8] Andreeva J et al for the ATLAS Collaboration 2012 *J. Phys.: Conf. Series* **396** 032100
[9] Nagios *www.nagios.org*
[10] Andreeva J et al for the ATLAS Collaboration 2012 *J. Phys.: Conf. Series* **396** 032072