Methodology for evaluating the exploitation of distributed data providers

Evgenii Viazilov, All-Russian Research Institute for Hydrometeorological Information – World Data Centre (Russia), vjaz@meteo.ru

Nick Mikhailov, All-Russian Research Institute for Hydrometeorological Information – World Data Centre (Russia), nodc@meteo.ru

Denis Melnikov, All-Russian Research Institute for Hydrometeorological Information – World Data Centre (Russia), melnikov@meteo.ru

Introduction.

To assess the exploitation of distributed, heterogeneous data integration systems, IT infrastructure monitoring systems are used. In these systems the metrics are used that help to understand that data providers of integration system has reached the required indicator values. When monitoring the IT infrastructure are used threshold values of metrics. For example, the time for automatic data upload to the integrated database (IDB) should not exceed 60 minutes. The relevance of the integrated data (updating the data in accordance with the established a regulations in the metadata) should be at least 90% from the number of information resources (IR) supplied to the IDB; or the reliability of the IT infrastructure must be >98%. These indicators allow it to take the necessary actions. For example, this allows loading IR, when the computational resources are available. To increase the reliability of the IT infrastructure in the event of an incident, an automatic software reboot of one or more software components is implementing.

In addition to operational using of the results of IT infrastructure monitoring, it is necessary to monitor long-term trends in the change of various indicators, to compare the work of data providers. Moreover, here already separate indicators are not enough, it is necessary to conduct an integrated assessment of data providers work and to calculate of rating of data providers.

Proposed indicators of integrated systems exploitation.

To calculate the rating of data providers, a methodology are developed for the integrated evaluation of the exploitation of distributed data providers. The purpose of developing such a methodology is to compare the exploitation level of data providers. Such an assessment allows data providers to see bottlenecks in the exploitation of the IT infrastructure.

For assessment of the exploitation of the integrated system IT infrastructure, the following groups of indicators defined:

1) reliability of the IT infrastructure of each data provider (the number of data providers in the system; reliability of the IT infrastructure; the downtime of IT infrastructure);
2) the relevance of IR provided by data providers (total number of IR integrated into the system; number of new IR; number of relevance IR; number of free IR);
3) normative accessibility of IR (percentage opened IR from general number of IR);
4) the level of information service of users (total number of views or IR downloads; average number of IR downloads; the specific weight of IR downloads in the total number of resources; the total number of IR transferred by subscription);
5) ensuring rights for access to IR - the number of assigned roles for access to IR;
6) feedback from users of the system - the number of complaints to the data.

The most important groups of indicators from the system's point of view are the reliability of the IT infrastructure, the relevance of IR and the level of information services. These indicators have the

greatest contribution to an integrated assessment of the performance of data providers. In the future, when the reliability of the work and the relevance of IR will reach the planned values, their contribution can be reducing, and the contribution of a group of indicators on information services to users will increased. These contributions can install, when you configure an application that implements this method of assessing IT infrastructure. The rating of data providers is equal to the sum of the scores obtained for the first five groups of indicators. The last indicator "feedback from users" subtracts from the sum of the scores.

Most important indicators of integrated systems exploitation.

Let's consider some of the important indicators (the work reliability of the data providers and the relevance of IR) in more detail.

The reliability of the data providers is the percentage of time that the data provider software were on-line. The IT infrastructure downtime is equal to the time of the problem on any of the software components of the data provider. The total time of downtime is sum of time bands of idle. The monitoring system are controlling a portal, a geographic information system, an integration server, IDB, etc. For each software component, objects such as a virtual machine, network, and application are monitoring. If the network is unavailable, than the problem of network is fixing, since determination of the work other components is physically impossible.

Indicators of the relevance of IR provided by data providers are their number and relevance. The number of IRs is the quantity described resources in the form of structured files, or database tables, or object files on the every data provider. The actual date and time of updating the IR should be in accordance with the data provider's obligation to supply the data declared in their metadata. The relevance indicator is the specific weight of relevant resources in the total number IR of the data provider. He is calculate as the average ratio of the number of regularly updated resources to the total number of resources for a reporting period. This indicator evaluation based on a daily-automated verification of the composition of IR ready for work.

Integral evaluation of the exploitation of data providers is normalizing by calculating the ratio of the values of each indicator to the average value of this indicator as a whole for the entire integrated system.

Conclusions.

The methodology has developed for determining the integral indicator for evaluating the exploitation of data providers in the form of their rating, which makes it possible to identify good and poorly working data providers. For the first time in this work, a set of indicators is proposed to assess the performance of data providers, some of which were previously considered only separately or not at all in other systems (for example, normative accessibility and regulation of rights to access to IR).

The considered methodology use for calculate the rating of data providers of the Russian national system "Unified State Information System on the World Ocean" (http://esimo.ru) and showed the significance of the selected indicators. This methodology can apply in the SeaDataCloud project to evaluate the exploitation of the Download Managers, which installed in more than 100 marine centers of European countries, and others system of data integration.