24-hour cache in and through the cloud

Introduction

As explained in http://wis.wmo.int/doc=3117, the current any to any traffic pattern for the GISCs to exchange data to be stored in the 24h global cache creates some administrative and technical difficulties. As a potential solution, this paper proposes to use a server in the cloud to facilitate the data exchange between the GISCs. Since October 2015, a pilot has started to assess, from a technical point of view, the suitability of such a solution. It was also decided to launch a questionnaire to check if from an administrative and legal point of view this solution was acceptable by the members. As a data provider for the WIS (as NC or DCPC) or as a GISC maintaining the 24-hour cache, we need your opinion on the matter to assess whether this solution of cache in and through the cloud should become an operational service.

* 1. Please enter your name.

* 2. Please enter the name of your organisation.

* 3. Please enter your country name.

* 4. Please enter the WMO Regional Association of your country

- ◯ I
- ◯ II
- ◯ III
- ◯ IV
- ◯ V
- ◯ VI

* 5. Please enter the status of your organisation (select all applicable)

- ☐ GISC
- ☐ DCPC
- ☐ NC

24-hour cache in and through the cloud

Architecture and security of the solution

In the document http://wis.wmo.int/doc=3117, it is suggested that the architecture should include a group of cloud servers that are reachable via the RMDCN, and a separate group of servers that are reachable via the Internet. It is considered that this design could offer redundancy and independance for the RMDCN provider (if the servers on the Internet side are not provided by the same provider as the RMDCN).

6. Assuming that in term of performance, two servers are required in each environment (two servers in a virtual data centre on the RMDCN, two servers in a virtual data centre on the Internet) and based on your experience regarding the reliability of such cloud service, do you think that such a design will provide the required operational 24/7 service

- ◯ Yes
- ◯ No
- ◯ I don't know / No opinion

Please detail your answer, if needed

7. Again, assuming that we would use two different providers, do you think that one of the provider should be the provider of the RMDCN as it can provide unlimited access to the RMDCN?

- ◯ Yes
- ◯ No
- ◯ I don't know / No opinion

Please detail your answer, if needed

8. In addition of the redundancy explained above, it is paramount to guarantee the IT security of the cloud servers environment. The security of the servers will include at a bare minimum a "state of the art" patching strategy and configuring firewall rules (using iptables or equivalent) on the servers themselves. The firewall rules will allow only connection to and from the designated hosts from the GISCs as well as allowing ssh access for the managing entity. Do you consider this configuration secure enough?

- ◯ Yes
- ◯ No
- ◯ I don't know / No opinion

In addition, what would you suggest?

9. In the Attachment II.15, WMO has defined the use of socket or FTP to send and receive data on the WIS. On private networks (RMDCN or leased lines), these protocols can be considered as acceptable in term of security. However, data exchange from and to the cache in the cloud would use the Internet (probably in addition of the RMDCN). In such a case, it is important to use more secure protocol such as SFTP or HTTPS.
Assuming that the use of encrypted protocols (SFTP, HTTPS) is added as a standard protocol in the Attachment II.15, what is your opinion on the use of encrypted protocols compare to FTP.

- ◯ Nowadays transfering data over unencrypted channels is not suitable anymore and SFTP or HTTPS must be used.
- ◯ SFTP and HTTPS are more secure protocol, however our production systems are not yet ready. So FTP should be used.
- ◯ A mix and match solution is acceptable. Centres would be strongly advised to use SFTP or HTTPS, but FTP is still accepted.
- ◯ I don't know / No opinion

In addition, what would you suggest?

10. The document (http://wis.wmo.int/File=1859) presented during the last TT-GISC meeting gives an update on the current test and also consider some of the features required by the software on the cloud servesr.
For the test, we are using a software called AFD (http://www.dwd.de/AFD/). Following exchanges with various GISCs during the tests, AFD has been configured to support rather advanced features (such as packing/unpacking bulletins into files...).
On the one hand, the software on the cloud server could be as simple as possible (receiving files and pushing back) or, on the other hand, could handle the complexity of WIS data exchanges (eg. managing de-duplication of data, filtering data not intended for "GlobalExchange"...).
In the paper http://wis.wmo.int/File=1859, this is what is called "WMO awareness".

What is your opinion regarding the "WMO awareness" of the software running on the cloud servers?

- ◯ The software (and its configuration) used on the cloud server should be as simple as possible. It will still be the responsibility of the GISCs to handle all the advanced featured (such as packing/unpacking bulletin into files, packing/unpacking files into compressed archives, managing duplication of data...). This way, the management of the software on the cloud server will be much easier.
- ◯ The aim of the "cache in and through the cloud" is to facilitate the data exchange between the GISCs. So, it is quite consistent with this approach to consider that the software on the cloud servers should handle as much as possible the complexity of data exchange between the GISCs. Having a more complex configuration on the software running on the cloud servers is an acceptable drawback.
- ◯ I don't know / No opinion

In addition, what would you suggest? Considering the features required by the software on the cloud server, do you know any "off the shelf" solution that could be used?

11. If you have any suggestion regarding the architecture, the software to be used and the security of the solution to be considered please add your comments and ideas below.

24-hour cache in and through the cloud

Policy for "GlobalExchange" data

The following questions consider the data that the National Centres and Production Centres are producing and that should be part of the "GlobalExchange" cache.

12. In such a design, the data intended for GlobalExchange will be sent by the GISCs to the cloud servers. These servers will then forward to data to all other GISCs. The cloud servers will act as a "relay". It is assumed here, that for this "relay" function data is not stored on disk on the cloud server. In this configuration, the "GlobalCache" will not be available in the cloud, but only kept by the GISCs. As a data producer (either in your National Center capacity or your Production Centre capacity), is this function compliant with your data policy?

- ◯ Yes
- ◯ No

Please detail your answer (if needed)

13. Again, for this "relay" function, is there any restriction regarding the physical location (country, continent...) of the cloud servers?

- ◯ Yes
- ◯ No

Please detail your answer, if needed

14. In such a design, the data intended for GlobalExchange may also be ***stored*** onto the cloud servers. The cloud servers will act as a "global cache" in addition to the cache already maintained by the GISCs. As a data producer (either in your National Center capacity or your Production Centre capacity), is this function compliant with your data policy?

- ◯ Yes
- ◯ No

Please detail your answer, if needed.

15. For this "global cache" function, is there any restriction regarding the physical location (country, continent...) of the cloud servers?

- ◯ Yes
- ◯ No

Please detail your answer, if needed

24-hour cache in and through the cloud

Policy for other data

The primary intent is to use this "cache in and through the cloud" only for "GlobalExchange" data (GISC must forward "GlobalExchange" data to all other GISCs so One GISC to All GISCs). However, this solution could potentially be extended to data with a large user base where your GISC would need to send the data to many GISCs. The following questions refer specifically to this type of data. For the sake of clarity we will designate this data as "WideExchange".

The following questions consider the data that the National Centres and Production Centres are producing and that can be considered as "WideExchange" data.

16. In such a design, the WideExchange data will be sent by the GISCs to the cloud servers. These servers will then forward to data to many other GISCs. The cloud servers will act as a "relay". It is assumed here, that for this "relay" function data is not stored on disk on the cloud server. As a data producer (either in your National Center capacity or your Production Centre capacity), is this function compliant with your data policy?

- ◯ Yes
- ◯ No
- ◯ I don't know / No opinion

Please detail your answer, if needed

17. Again, for this "relay" function, is there any restriction regarding the physical location (country, continent...) of the cloud servers?

- ◯ Yes
- ◯ No
- ◯ I don't know / No opinion

Please detail your answer, if needed

18. In such a design, the data intended for WideExchange data may also be stored onto the cloud servers. The cloud servers will act as a "cache". As a data producer (either in your National Center capacity or your Production Centre capacity), is this function compliant with your data policy?

- ◯ Yes
- ◯ No
- ◯ I don't know / No opinion

Please detail your answer, if needed

19. For this "cache" function, is there any restriction regarding the physical location (country, continent...) of the cloud servers?

- ◯ Yes
- ◯ No
- ◯ I don't know / No opinion

Please detail your answer, if needed

24-hour cache in and through the cloud

Experience regarding the use of cloud services

20. The cloud services exist in many "flavours". A standard method to describe the service offered by Cloud provider is to distinguish between the following options:
- IaaS (Infrastructure as a Service): In this option, the provider is providing physical servers or virtual machines in its data center and the entire management of the server is done by the client
- PaaS (Platform as a Service): In this option, in addition to the above, the providers propose additional services (such as security patches of the servers. Amazon Web Service is a typical PaaS Provider.
- SaaS (Software as a Service): In the option, in addition of the above, the provider gives access to the application configured and tuned. The client uses the application as is.
In the cache in and through the Cloud, it seems quite unlikely that a commercial SaaS provider will be able to deliver the service we are looking for. The management of the platform could be done by the provider, but more as a bespoke service than a standard offer.
Considering what is expected by the servers and the providers in the "cache in and through the cloud", which solution(s) should we investigate further.

- [ ] IaaS
- [ ] PaaS
- [ ] SaaS
- [ ] Other
- [ ] I don't know / No opinion

If you have any experience, can you please provide details (providers, type of usage...)?

21. Do you have any experience regarding the use of "servers in the cloud", either as Infrastructure as a Service (IaaS) or as Software as a Service (Saas) or as Platform as a Service (PaaS)?

- ( ) Yes
- ( ) No

If you have any experience, can you please provide details (providers, type of usage...)?

22. From a contractual point of view, and if you have experience on the matter, does the provider(s) that you are using are offering SLA (Service Level Agreement)?

- ◯ Yes, our provider offers SLA
- ◯ No, our provider doesn't offer SLA
- ◯ No. We do not have such a contract.

Please detail your answer, if needed

23. Please provide any information related to your experience as a user of cloud services that might be helpful

24-hour cache in and through the cloud

Managing the service

Typical IaaS (Infrastructure as a Service) providers offers an interface to create/delete the virtual servers or dedicated servers in the cloud and a very basic monitoring function of the server environment.
Typical PaaS (Platform as a Service) providers offers additional services (improved management, better monitoring).
Typical SaaS (Software as a Service) providers offers additional services including application management.
All the tasks required to install and maintain the service will be called "management".
For the "cache in and through the cloud", the "connected" countries will be the GISCs. They will be the users of this service.
For its management, the situation with the RMDCN can be used as an example. You can also consider alternatives.
In typical IaaS or PaaS services (called XaaS below), the provider will not provide the entire support on the configuration of the servers and he will not manage the applications that are required to run the "cache" service... The "management" of the service has to be done by a separate entity.

24. Among the following options, which options do you consider appropriate for this service (select all applicable).

- [ ] The service should be managed by a commercial provider (probably not the provider of the XaaS) and one GISC should be the interface between the commercial provider of the management function and all other GISCs.
- [ ] The service should be managed by a commercial provider (probably not the provider of the XaaS) and ECMWF should be the interface between the commercial provider of the management function and all GISCs. Please note that ECMWF hasn't not been consulted on the matter.
- [ ] One or two members could work as a virtual team and manage the service for the connected GISCs (considering again that the XaaS provider is only responsible for providing the servers, not their full management)
- [ ] ECMWF, with its experience with the RMDCN, should manage the service (including application configuration). Please note that ECMWF hasn't not been consulted on the matter.
- [ ] I don't know / No opinion
- [ ] Other

Please explain how the service should be managed?

25. If one or two centres manage this service, do you think it is acceptable for those centres to ask for a "management fee" to the connected parties

- ◯ Yes
- ◯ No
- ◯ I don't know / No opinion

Please explain how the service should be managed?

24-hour cache in and through the cloud

Paying for the service

26. One of the challenge for establishing the RMDCN was to find a suitable way to procure and to pay for the service. Following a memorandum of understanding between WMO and ECMWF, ECMWF is in charge of this and is signing the contracts on behalf of all connected members and then each member pays for its own connection either directly to the provider or through ECMWF (for the Member States). Such an arrangement is not really applicable to this service. There isn't one connection per connected member but instead a ressource (the servers in the cloud) that are shared by all the members.
In your opinion, how should we organise the contract?

- [ ] One Member is signing the contract and the connected parties will pay that Member for the service.
- [ ] WMO is signing the contract (or using the RMDCN contract with the RMDCN operator) and the paying parties will contribute by using a "trust fund" that WMO will put in place. In this case, it is likely that WMO will charge a management fee of approx. 13%
- [ ] ECMWF is signing the contract, in addition of the networking function of the RMDCN (in this case, it is possible that ECMWF will charge a management fee). Please note that ECMWF hasn't not been consulted on the matter.
- [ ] Each paying party should pay directly the provider(s) after signing an individual contract.
- [ ] No opinion / I don't know
- [ ] Other

Please detail your answer, if needed.

27. In the current design, in order to maintain the 24h global cache, the GISCs have to send and receive the data to and from all other GISCs. As a consequence, a high speed connection to the RMDCN is therefore required. Using the "cache in and through the cloud" would limit the need of high speed RMDCN access. In that respect, the GISCs would save some money on their infrastructre. As a consequence, and considering that the servers in the cloud would be another way to do what the GISCs already agreed to do, only the GISCs should have to pay for that service.

- ( ) Yes
- ( ) No
- ( ) No opinion / I don't know

Please detail your answer, if needed.

24-hour cache in and through the cloud

Potential future use

This method to exchange data between the GISCs through the cache in and through the cloud can be extended to potential other uses.
In this first use case, the data is sent by one GISC and redistributed to all the others.
Among other potential use cases, one might consider:
- data sent by one member (it could be a DCPC or a NC) to be redistributed to many members (some GISCs, other DCPC/NC...)
- data provided by one member and accessible via a pull service to other authorised parties
- ...

Most of the data exchange between WMO members is based on a store and forward mechanism. Data is received and forwarded to interested parties based on routing (GTS or WIS) or subscription (WIS) strategies.
Cloud servers and pull solutions (webservices...) are currently hardly used.

This cache in and through the cloud could be the first step toward a complete redesign of our data exchange strategy.

The following questions assume that a member will be able to enforce the appropriate data policy to the data stored in the cloud.

28. In your capacity of data producer, either as a National Centre or a Production Centre, would you be interested in using a similar design (pushing data in the cloud) to make your data available to other members?

- ◯ Yes
- ◯ No
- ◯ I don't know / No opinion

Please detail your answer, if needed.

29. In your capacity of data producer, either as a National Centre or a Production Centre, and assuming that you would use the cloud service for data exceeding the "GlobalExchange" category, would you be in a position to contribute financially to the service?

- ◯ Yes
- ◯ No
- ◯ I don't know / No opinion

Please detail your answer, if needed.

30. When you receive data either from a National Centre or a Production Centre, would you be interested in using a similar design (data pushed in the cloud) to get the required data via this extended cache?

- ◯ Yes
- ◯ No
- ◯ I don't know / No opinion

Please detail your answer, if needed.

31. Do you consider that the data flow between WMO members should evolve toward accessing data through pull mechanisms in addition (or even replacement) to the current data push architecture?

- ◯ Yes
- ◯ No
- ◯ I don't know / No opinion

Please detail your answer, if needed.

32. Please provide any additional information that you consider as relevant on this topic of cloud servers.