From: Eva Dafonte Perez
To: Jamie Shiers; Maria Girone;
Subject: Streams problem on the ATLAS setup during the last weekend
Date: Thursday, October 23, 2008 12:41:03
Attachments: SR Details_ORA600droppingPropagation.pdf

Hi,

First I would like to mention:
- NDGF is running production on a test system: single instance with only 6 disks. This system is very limited.
- Tier1 sites and Streams setups are only covered during working hours.
- We have already pointed to ATLAS that the scalability/stress tests they are running must be supervised by us and the Tier1s database administrators in order to identify the bottlenecks and try to fix them.

However, the test which was originating the whole problem started on Friday 17.10.2008 at 18:00.

Description of the problem affecting the ATLAS Streams setup during the weekend:
On Saturday (18.10.2008) morning, around 09:30 a.m. the high load at NDGF caused the database to be totally unresponsive. The propagation job connection to NDGF got blocked due to this situation and could not deliver the LCRs. When the LCRs are not consumed by all the destination databases, they are maintained in the capture queue (spilled LCRs). The capture process and all the propagation jobs to the other 9 Tier1 sites were still working fine. The replication was completely stopped on Sunday (19.10.2008) morning, around 07:30 a.m. when the large number of messages build up in the capture process's buffered queue kicked up the flow control state causing the capture process to temporarily "pause" capturing any new messages until some messages are removed from the queue. On Monday morning (20.10.2008), NDGF had to reboot the host in order to fix the problem with the database. When NDGF was back, LCRs started to be consumed (consuming spilled LCRs is slower than consuming LCRs in memory, this explains why the apply at NDGF was running slower than during normal activity). Once the LCRs were consumed and removed from the queue, capture process started capturing changes again and all the Tier1 sites recover from the accumulated backlog during the afternoon.

This problem could only have being fixed by the Tier1 site intervention (but Tier1 sites do not intervene during weekends) or by performing the split of the streams setup in order to separate NDGF from the main streams replication (but Streams support is restricted to working hours only and the split method, which requires a lot of manual work due to an Oracle bug (I have attached the SR), is not included in

our procedures outside working hours).

We are currently investigating the possibility of increasing the memory dedicated to the capture queue. This will allow the capture process to capture LCRs during a longer period of time (and possibly cover the weekends) when one destination site is not consuming the LCRs and they are being spilled.

Please let me know if you have any question.
Cheers,
Eva