fast

Technical Support

BEST PRACTICES FOR FAST INTSALLATIONS (4.X.X)

Version: 0.6

Date: 2006-05-04

Author: Sangita Skipitaris

INTERNAL INFORMATION

1 Objectives

The main objective is to work proactively to prevent common critical issues by sharing operational knowledge from the Techinical Support organization. This should ulmiately improve the product experience of our customers as well as reducing support load both on FAST technical support organization and customer system admistration or engineering teams.

1.1 Target Audiences

The target audiences for this document are FAST SCS engineers, FAST project managers, Partners, System Integrators, Resellers and selected OEM customers.

1.2 WHAT TO DO OR NOT TO DO

The following information has been collected by analyzing Severity 1 and 2 problems reported by FAST production customers (includes OEMs, Partners and Direct Sale customers) over the past 4-5 months. Below is a compiled list of suggestions you can follow as best practices when deploying FAST Data Search (FDS) or Instream 4.x.x on client sites. We hope this information will provide you more operational insight and help you to avoid some of common mistakes when deploying FAST products on production systems - Enjoy!

1. Always use a **cleaned installed OS system** for FAST products, and always use a separate larger disk partition for FAST file location than the system partition.
2. Do not use **unsupported OS versions and libraries** - always follow the installation guide and read the release note prior to installing a new system. Issues reported from unsupported platforms will be rejected by R&D, hence **no maintenance support is guaranteed**.
3. Do not go on production without installing the **latest available hotfixes and patches**; follow the recommended patch level as indicated in Installation guide or Release Notes. Always install hotfixes in increasing order and follow the readme.txt during the hotfix installation. Upgrading patches requires system shutdown which is not always feasible for production systems.
4. On Windows do not run **Anti-Virus scanner** on the %FASTSEARCH% folder. The binary index may contain strange patterns that the anti-virus software could see as a real virus. In this case, it will quarantine the file just as it happened on your system. Besides, Windows does not allow more than one program to access a file. If an anti-virus is scanning a file, a process related to FAST may have problems with opening the file. Finally, scanning of files introduces more I/O. Since search and indexing processes are I/O bound, the performance of the cluster may suffer. That is why we strongly discourage using anti-virus software.
5. Do not run any kind of **network security scanners** on a system running indexer and search components; network security scanner or vulnerability scanners locks processes which might create intermittent indexing and search issues.
6. Do not run any **backup utility** while the indexer is running or feeding content; locked index process will cause indexing issues.
7. Do not allow any **3rd party monitoring tools** on a production system without testing on a staging system first; the consequence of running 3rd part applications is unknown.
8. In order to avoid issues related to the **Daylight Saving Time change**, set the server Time Zone to either GMT or UTC and always uncheck “Automatically adjust clock for daylight saving changes”. The release note says:

   > FAST search servers require that the clock is kept in sync and not abruptly corrected forwards or backwards. Abrupt clock-changes may lead to internal processes (such as the qrserver and fdispatch processes) believing that there are communication problems or processes that have been silent for too long and hence dead. This will in

turn lead to warnings in the logs and processes being automatically restarted. Avoid manual clock adjustments and consider using professional software for keeping clocks adjusted and in sync.

9. Make sure that the **system clocks are in sync** between the Active Directory server and FDS production servers. If not, on windows system the AD server will fail to deliver the DNS information to the FAST server, which will first of all cause windows login problems. Secondly, possible DNS problems will cause communication problems between the feeder and other components within FAST, which will typically cause symptoms like no activities even though the search may work. Users will report login issues as the system is freeze condition, no other options but rebooting the system.
10. Do not **store or backup any files or folder in the $fastsearch/var/file_export folder**. This causes two problems:
    a. When the search-1 (searchctrl) process starts it copies any content from to the $fastsearch folder; large archive will fill the disk.
    b. Later the searchctrl will parse every single file located in the **file_export** before it can start properly. For smaller amount of files this problem might not be noticeable however for large amount of data (we had one issue with 40GB data querylogs archived in that folder) the low level search will not be responsive until parsing has been completed. The typical symptom is search does not work after a restart and you will find some innocent Info messages in the rtsearch log like “transferfile: Sending file rtsearch\rtsearch.log.9 to plum.boston.fast.no:15604”. In this case, we reproduced this behavior by copying the rtsearch log folder into the **$fastsearch/var/file_export** folder.
11. Do not run **VMWare** for production systems. VMware as a virtual machine technology is subject to resource sharing which implies resource **bottlenecks**. As a properly configured FAST installation utilizes all available memory, disk IO bandwidth and CPU there is not much to gain in implementing on top of VMware except for a few special cases such as: Functionality testing, ex QA, as VMware gives operational features such as filesystem rollback, simplified install and resource sharing – lowering the cost of functionality platform. Hot standby failover but this would severely degrade performance when fault occurs
12. Do not use “**Veritas Cluster Server”** (provides disaster recovery plan, which also includes replication and backup utilities) which uses virtual hostnames and prevents FAST indexer and search processes to come up properly. Instream 4.x supports “Veritas File System (The ``basic'' VxFS filesystem product is included on all UnixWare 7 systems”) – check the Installation Guide and Release Notes for supported platforms information.
13. Our experience is placing **firewall** between separate FDS components cause communication problems in the 4.x.x versions. By default, in FDS Configuration Server, Content Distributor, Status Server, and Name Servers are not tied to specific server ports. Especially CORBA opens several ranges of ports dynamically and when the requirement is content push from the outside of the firewall this creates feeding problems. In this case, it’s safer to open all 4000 ports, from the baseport. Check the 4.x product guide for information on port ranges when using Search front-end from outside or having admin GUI within the firewall.
14. Provide customers guidelines on **recovery procedure** before signing off the project and ask the customer system admin team to practice it before going live (rebuilding from FIXML, rebuilding from another row, disconnecting a faulty node etc.). The feedback we get is customers want to know the basic on disaster recovery procedure. Check the Operations Guide delivered with the product distribution for more information.
15. A **properly sized system** should handle a system restart without hitting the resource limit. Always be mindful of the operational impact of pushing the limits of server hardware when initially scaling systems. Issues of type sizing/scaling problems reported from a production system requires significant engineering resources from the customer, R&D, and Technical Support and finally these issues will in most cases be escalated back to the

implementation team. In addition to running necessary load testing, we recommend that the customer exercises stop and start the search/index nodes or rebuilding the index from the FIXML at least once. During this process monitor the system resource like memory usage, page files swaping and recovery time. Check the Deployment Guide delivered with the product distribution for more information.

16. Do not allow **a separate IS team to manage the network** without understanding FAST requirements; operations like changing IPs or DNS, shutting down servers, run automated upgrades, running network intensive apps on same subnet should be strictly forbidden on FAST systems without consulting FAST system administrators.
17. Be mindful when feeding **too large documents**; the reason is during document processing the memory footprint will increase and doc procs will stop sending callbacks or become unresponsive which will cause errors from the ContentDistributor. Even if the processing passes the fixmlindex will be the next bottleneck. FIXML is parsed in memory first then written to disk in batches, and if the system is Low on memory fixmlindex will be unable to complete parsing and in the worst case the process will crash.
18. Do not add **more than 2 doc procs per physical CPU;** doc procs are CPU intensive, hogs the server for resource. If procservers and indexer run on the same node it will cause indexing or full API queue issues – in this case suspend the indexer during feeding and suspend the doc procs during the indexing. Consider adding separate physical servers for document processing server. Adding document processors on the index/search nodes might slow down the indexing and search performance.
19. Use separate nodes to run **medium/large crawlers** than indexers and search – both crawler and indexer is I/O intensive.
20. It is recommended that customers have a **staging system.** A staging system is helpful for patch verification especially for systems with high customization. Ask customers to keep the staging system up-to-date.
21. Do not run FAST on a system without **proper DNS setting.** Any DNS issue will cause communication problems between the FAST servers and within some of FAST components.
22. Do not ignore **intermittent indexer, search or callback issues** reported during the implementation phase; make sure to apply the latest available fixes or service packs before customer goes on production. If the issue persists open a problem report to R&D and escalate it to your manager.
23. Always **test and document your custom code** before delivering to customers. Provide only a functional overlay to technical support during the handover to support.
24. Do not leave your **apache server unattended** if it is accessible from “the internet” - hackers can take control of your server and damage the FAST system.
25. In 4.x, do not use **NAS** - not even to store the crawled data or copy of indexes or log files
26. In 4.x, do not use **separate disk drives** to store index partitions; the indexer will crash intermittently
27. Do not attempt using **index and search on separate nodes** on Windows with scp; scp is not supported on windows
28. Avoid having the **Windows Indexing Service** running on your $FASTSEARCH/data/data_index tree – one indexer is enough. The windows Indexing Service will index all the content on the server which can lock-up files in datasearch and cause serious problems
29. Do not use **multicasting** with the following RTS versions:
    a. FDS 4.1 RTS version 3.1.46 – 3.1.49 (inclusive)
    b. FDS 4.0 RTS version 3.0.120 – 3.0.123 (inclusive)
    c. InStream 4.x RTS version 3.0.120 – 3.0.123 (inclusive)

30. Make sure that any production system is (proven to handle) capable of handling the load (**stress testing**).
    a. Stress test of the disks prior to going live; too many query performance issues are reported related to slow disk I/O
    b. Stress test with comparable data stream and QPS load;
31. Do not run **unnecessary services on a production system**, such as IIS, domain controller, exchange, print server, DHCP server, SQL, etc... Obviously, these may be useful in some installations, but careful consideration should be taken to install these servers with FDS components that are not using the same disk, memory, or CPU resources.
32. Do not run **multiple instances of FDS on a single node** if the server will be used for production. Although it works sometimes, operational procedures are very complex and confusing. The performance will be slower.
33. If a **resetindex is suspended with rtsadmin** it will not continue with a resumeindex.
34. Do not recommend **updating index profile on production systems** without providing correct and tested procedure. Customer should backup their system prior to updating the index profile. In general, the FAST system admin should practice any maintenance procedure and get hands-on experience before going on production.
35. Do not **update index profile from the system admin GUI** especially on a multiple node system; use bliss in a controlled manner or else the system will hose, both index and search will break.
36. Provide customers customized **disaster recovery documentation** for disk failure or server crash issues; too many severity 1 issues fall in to this category
37. Do not install **FDS on a Domain Controller** – FAST processes will start as it should be but Admin Gui will be unable to add any collections with an **XMLRPC** error saying collection not found. The root cause is unknown.
38. Do not **reboot production servers** without stopping FDS processes first – in the worst case the index may get corrupted
39. Do not **restart FDS processes** (nctrl stop and killing processes randomly) on production systems without isolating the issue first or being sure that restart will resolve the issue
    a. **killing statusserver** will cause statusserver recovery which can take some time to come up; will depend on the number of documents
    b. **killing fixmlindex or findex** processes may corrupt the FIXML or the Gigabase DB or the indexer which may do larger damage to a production system
    c. Restart of **FAST processes without stopping processes** correctly will cause random issues – wait until all processes have stopped
40. Do not allow running **PRINT Server** on the FAST system and **turn of automated windows** or other application updates; recommend system updates in a controlled manner. Most system updates required reboot of the server.
41. Make sure that production systems have **production license installed** not an eval license; in order to get a more accurate picture of the QPS and index update performance, any acceptance testing should be performed with the production license installed

1.3 References

[1] Support tickets and product documentation