IBM® InfoSphere Streams® v4.0 Performance Best Practices

Abstract

Streams® v4.0 introduces powerful high availability features. Leveraging these requires careful consideration of performance related configuration options.

Henry May
hjmay@us.ibm.com
David Engebretsen
engebret@us.ibm.com
Walt Madden
wrmadden@us.ibm.com

1 Introduction

IBM InfoSphere Streams v4.0 is a major new release with significant advances in high availability and ease of use. This release includes many new features which make InfoSphere Streams simpler to manage and more resilient. InfoSphere Streams v4.0 features "always on" recovery, support for redundant services, automatic failover and restart of management services, and automatic recovery from host failures.

A significant infrastructure change adds Apache ZooKeeper to store the configuration and state information required for enhanced high availability functions thus removing IBM DB2 as the repository for recovery information. Users upgrading from a prior release who used DB2 recovery typically will not experience a significant difference in system load, however users upgrading from the prior release's default configuration with no recovery enabled should expect additional overhead due to recovery being enabled by default in V4.

Due to the critical role ZooKeeper plays in Streams v4.0, the performance of ZooKeeper will directly influence the performance of Streams management commands. A low latency storage subsystem is absolutely critical to ZooKeeper performance. Configurations without low latency storage on the ZooKeeper hosts might experience significantly longer latencies on operations such as Job Submission.

Streams V4 also implements two new optional features that have performance considerations for people who choose to use them. The first is Consistent Regions, which guarantee that all tuples are processed at least once by periodically establishing checkpoints. The second is IBM InfoSphere Streams for Microsoft Excel, which allows real time data display in a Microsoft Excel spreadsheet.

The tests used as the basis for the recommendations in this document were run on environments ranging from a single host running application and services to a cluster of hosts with dedicated roles running on well-configured, development-level computer systems dedicated to the tests being performed. Emphasis was placed on operations that were impacted by the V4 changes. All performance benchmark values are provided "AS IS" and no warranties or guarantees are expressed or implied by IBM. Actual performance of InfoSphere Streams applications and management functions may vary and is dependent upon many factors, including, but not limited to:

- Application design;
- Application complexity;
- Server hardware, including number of cores, CPU speed, and memory size;
- Server cluster communications bandwidth;
- Number of InfoSphere Streams jobs running simultaneously.

Furthermore, some measurements may have been estimated through extrapolation. Since actual results may vary, buyers should consult other sources of information to evaluate the performance of systems they are considering buying, including conducting application oriented testing.

2 ZooKeeper Host Storage Requirements

Support for job recovery functions implemented in V4 uses Apache ZooKeeper to maintain state associated with PEs and connections where each state transition is written synchronously to disk. Consequently, low latency storage in any host running a ZooKeeper process is absolutely critical for acceptable Streams infrastructure performance. The *Sample Configurations* Section in this document

shows options with the ZooKeeper service running on the same node as the Streams Management Services.

Generally this storage requirement applies to application management functions such as job submission and will not impact the computational performance of a running Streams application. Exceptions to this general rule exist, for example when a Processing Element (PE) goes through a recovery operation. Disk performance can limit the PE recovery time, which in turn can impact the processing rate of related PEs in the application.

Figure 1: Storage Latency Impacts on Job Submission Performance shows the dramatic effect the storage subsystem can have on performance of the submitjob operation as PEs transition to a Healthy state. Here we show the time required for a job with 301 PEs to transition to the Healthy state following job submission. Two configurations were measured, one using a single direct attached 7200 RPM SATA drive dedicated to the ZooKeeper log per best practices recommendations while the second used an XIV Fibre Channel (FC) connected storage server with a non-volatile write cache.

The direct attached SATA configuration required 305 seconds to reach a Healthy state, with 260 of those seconds spent waiting for ZooKeeper disk writes to complete. The FC disk configuration required only 56 seconds to complete the same operation.

The blue "*submitjob*" component of this measurement was the amount of time taken between issuing the *submitjob* command and return to the command prompt. The orange "*PEs Healthy*" component was the additional time required for all PEs to become healthy. Similar results can be expected on any function that requires state to be written to ZooKeeper; e.g., *canceljob* or the establishment of dynamic connections.

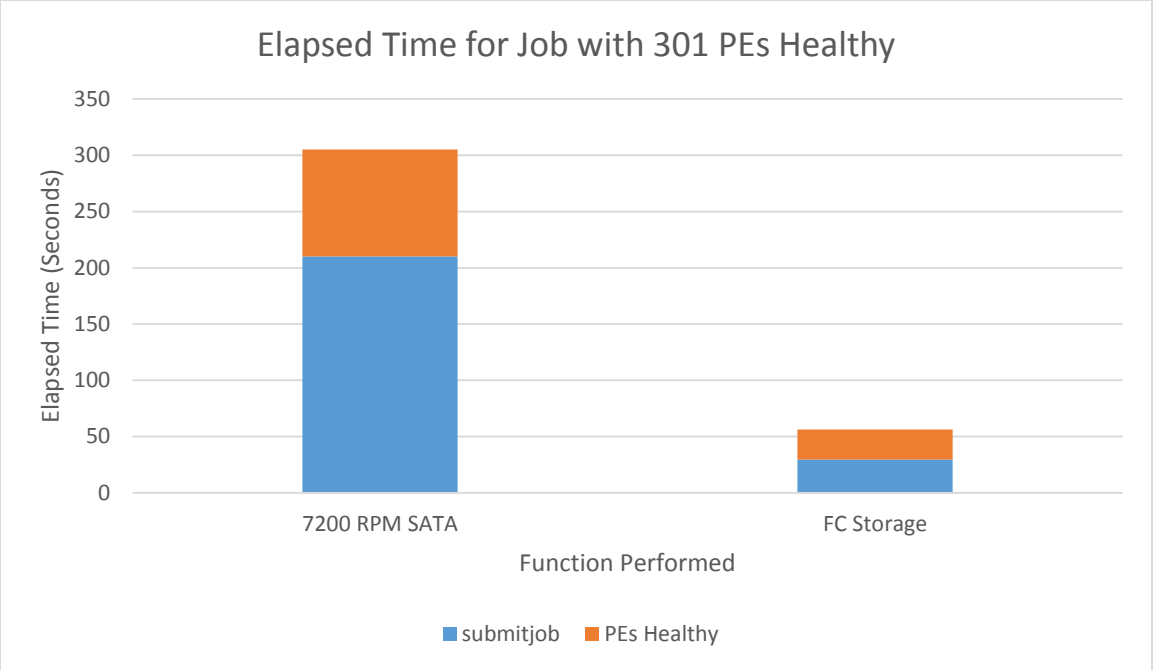

Figure 1: Storage Latency Impacts on Job Submission Performance

While these results demonstrate superior performance with an FC attached storage server with non-volatile write cache, similar results can be expected with any low latency storage subsystem. In general, solutions from any of the following categories will provide adequate performance.

- RAID adapter with non-volatile write cache. See http://www.redbooks.ibm.com/abstracts/tips0054.html for a selection of RAID adapters.
- Enterprise class SSDs. In this case a dedicated physical drive should be allocated to the log file. If multiple ZooKeeper instances are running on the same host, each log file should employ a dedicated physical SSD.
- Fibre Channel SAN Storage Servers. See http://www-03.ibm.com/systems/storage/storwize/ and http://www-03.ibm.com/systems/storage/disk/xiv/index.html for examples.

2.1 Additional Storage Considerations for ZooKeeper

2.1.1 Dedicated Device for the Transaction Log File

The key to ZooKeeper performance is provisioning a storage subsystem that will consistently provide fast write response times to the ZooKeeper transaction log. ZooKeeper documents the best practice of a dedicated physical disk drive for this purpose. Our testing demonstrates that while this helps with consistency of performance, it does not guarantee adequate performance; low latency storage as described above is the preferred approach. In fact, the 7200 RPM SATA measurements shown conformed to this best practice of dedicating a drive for logging. Rotational drives without a non-volatile write cache are inherently incapable of that level of performance. Note that many rotational drives incorporate a write cache. This is almost always a volatile cache; i.e., data is lost when the power is removed. Because ZooKeeper must flush each disk write through to the media, volatile caches do not help and may actually degrade performance.

The FC Storage Server used for this measurement was highly virtualized was being accessed by multiple machines running workloads unrelated to our Streams workload. Even with this additional load, the performance of a disk subsystem with a non-volatile write cache is clearly superior.

As with any performance recommendation, there are scenarios where specific requirements may change the configuration requirements. For example, if performance of functions such as job submission are not a primary concern, a rotational disk without non-volatile cache may be adequate. In that situation, dedicating a drive to the ZooKeeper logs may be desirable to maintain consistency of performance.

3 Management Host Memory Requirements

InfoSphere Streams V4 includes significant updates to its management services to improve resiliency and system management. These changes and new features result in requirements for additional system memory. *Figure 2: Management Host Memory Consumption* illustrates the relative memory consumption for bringing up and using the infrastructure necessary to run Streams jobs in the V3.2.1 and V4.0.0 releases. Management Host memory consumption was measured using the same hardware and a typical set of job management functions; e.g., submitjob, capturestate, canceljob. V3.2.1 was measured both with and without DB2 as a state repository for recovery data and V4.0.0 was measured with all Streams Management Services and ZooKeeper running on the same host. No application processes were allowed to run on the Management Host. We have concluded that users migrating from V3.2.1 installations that did not use DB2 as a state repository (the default configuration) to V4.0.0 with recovery automatically enabled, should expect to see memory requirements for Streams Services increase by 50% to 100%. For those situations where customers were running V3.2.1 and DB2 based

recovery enabled, they should see a reduction in the memory requirements for Streams Management Services.

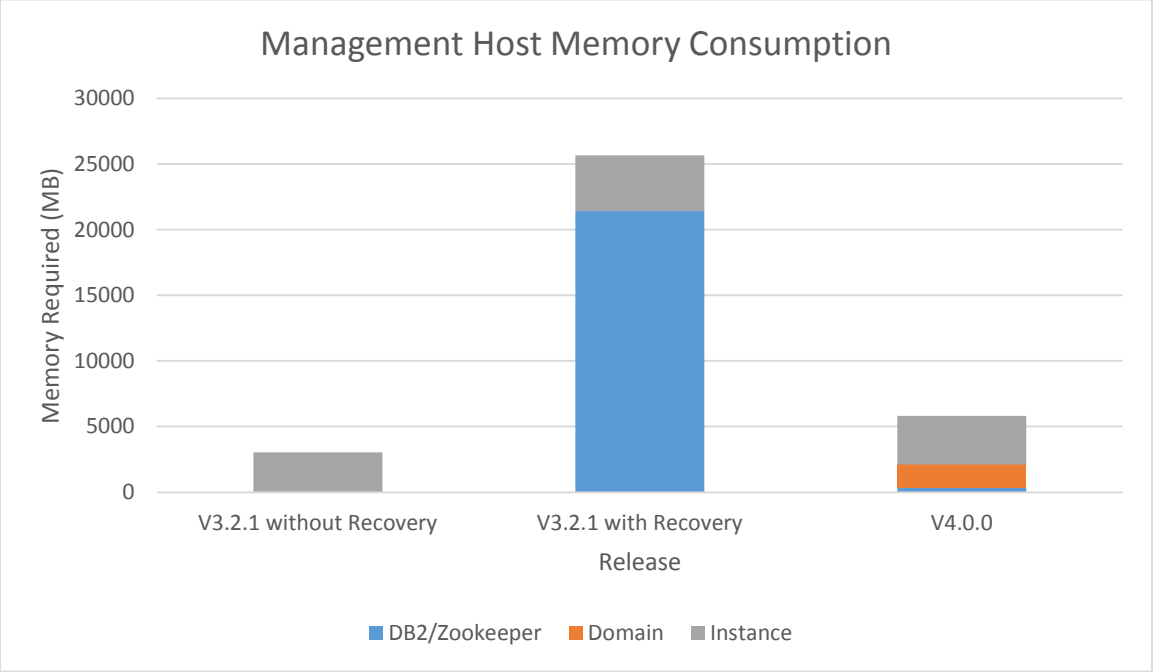

Figure 2: Management Host Memory Consumption

4 Management Host Processing Requirements

As with memory, the automatic recovery mechanisms implemented in InfoSphere Streams V4 will require additional processing resources in some scenarios. As illustrated by *Figure 3: Elapsed Time and CPU Consumption for canceljob*, we have not seen this CPU resource overhead significantly degrade elapsed time as long as sufficient processing resources are available on the system to absorb the additional processing required. The job measured in this example contains 301 PEs and was cancelled with Streams Management Services and ZooKeeper running on a single host while the application itself was running on four separate hosts.

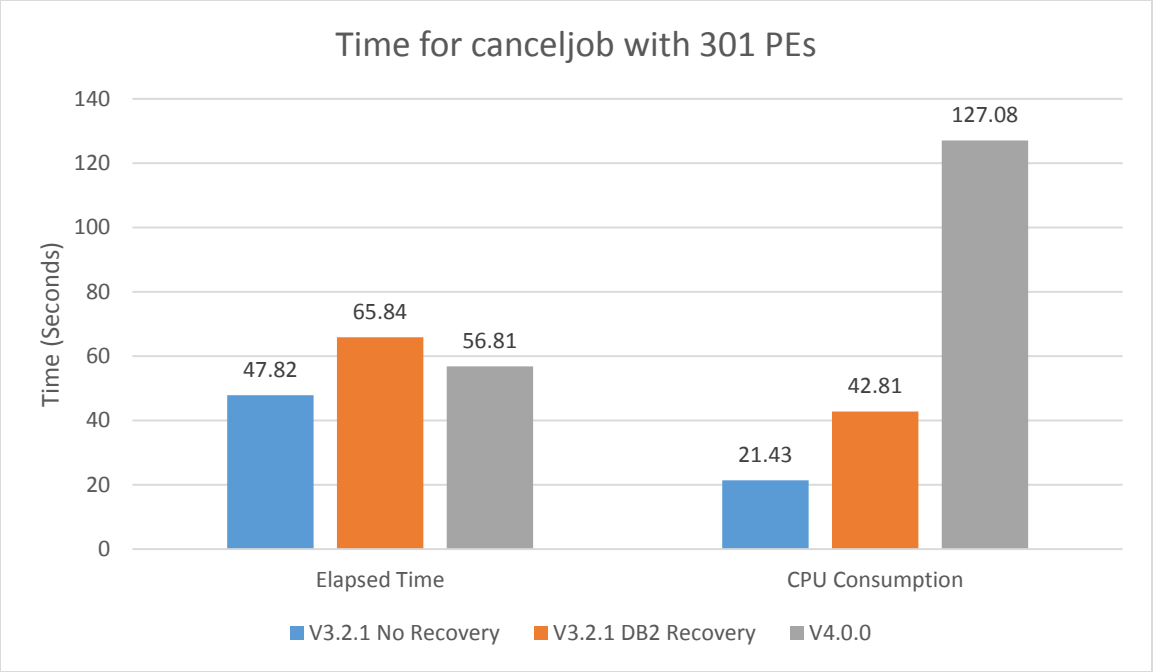

Figure 3: Elapsed Time and CPU Consumption for canceljob

5 Service Placement

5.1 Streams Management Services

As in prior releases, contention for processing resources between the application processes and Streams services must be eliminated for optimal performance of management functions. This is best achieved by providing dedicated hosts for management services as illustrated in *Figure 6: Non-Redundant Streams Configuration* and *Figure 5: Redundant Streams Configuration*. Streams can also be run in a single host configuration as shown in *Figure 7: Single Host Configuration*. This configuration should be deployed with caution, since a running application may consume resources required by the Streams Services and could result in slower response times for the functions performed by those services; e.g., job submission or instance graph display. Descriptions of services and details on tagging mechanisms used for their placement can be found in the following link:

http://www-01.ibm.com/support/knowledgecenter/SSCRJU_4.0.0/com.ibm.streams.welcome.doc/doc/services.html?lang=en.

It is generally acceptable to run all Streams Management Services on the same host, but there are scenarios where consideration should be given to dedicating a separate host for a particular service. For example, IBM InfoSphere Streams for Microsoft Excel can place significant demands on processing resources as the number of clients increases so heavy users of this function may want to consider dedicating resources to support the service.

5.2 ZooKeeper Service

Note that ZooKeeper is not a Streams service. While an embedded version may be deployed in a basic configuration (see http://www-

01.ibm.com/support/knowledgecenter/SSCRJU_4.0.0/com.ibm.streams.cfg.doc/doc/creating-basic-domain-and-instance.html?lang=en), ZooKeeper is generally a standalone application that is deployed separately from the Streams installation. Our testing demonstrates that the ZooKeeper service may be run on the same set of hosts as the Streams Management Services although configuring a separate set of hosts as a dedicated ZooKeeper ensemble is also an acceptable practice. In any event, if performance is a consideration it is critical to follow the guidelines in ZooKeeper Host Storage Requirements. More details on configuring ZooKeeper may be found in the following link:

http://www-01.ibm.com/support/knowledgecenter/SSCRJU_4.0.0/com.ibm.streams.cfg.doc/doc/configuring-external-zookeeper.html?lang=en.

6 Application Performance

Our testing demonstrates that application performance in InfoSphere Streams V4.0.0 is comparable to the previous V3.2.1 release. *Figure 4: Application Performance Comparison* shows a comparison between the releases across a set of applications that focus on basic Streams operators. All tests were performed on a host with 28 Intel Xeon model E5-2697 v3 cores running at 2.60GHz and Hyper Threading enabled. Note that Streams Management Services and Zookeeper processes were running on the application host, but measurements were performed at a point where there was minimal use of system resources by those processes.

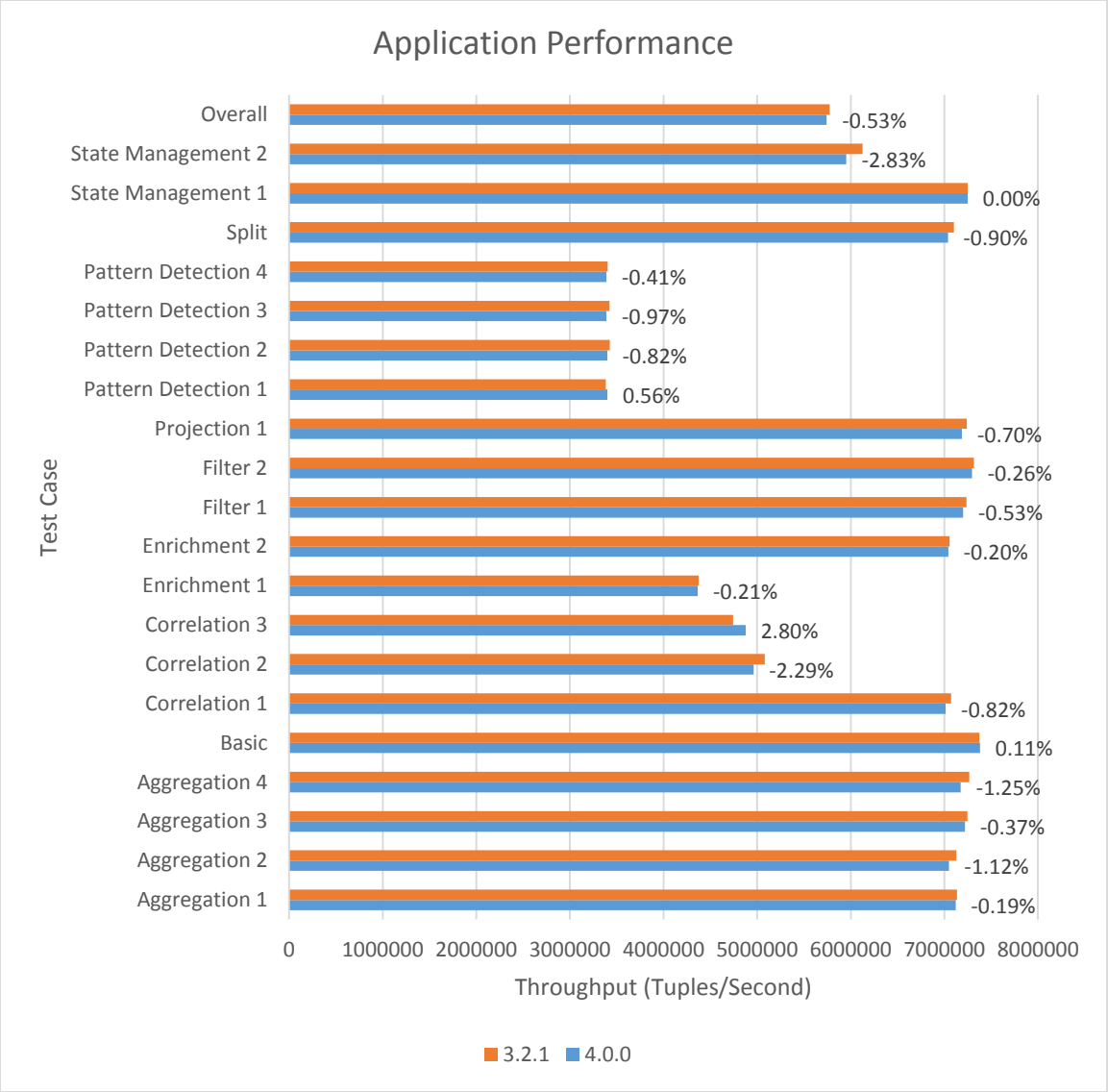

Figure 4: Application Performance Comparison

7 Sample Configurations

InfoSphere Streams V4 supports enterprise class high availability with redundant hosts playing specialized roles as discussed in *Redundant Management Services Example*. For small applications not requiring high availability, single host configurations as shown in *Single Host Configuration* are also supported. Along this entire range of configurations, resource requirements for Application, Streams Management, and ZooKeeper functions must be maintained. This section provides three examples as guidelines for provisioning Application, Management, and ZooKeeper Hosts. Application Host configuration is highly dependent on the specifics of the application, but generally speaking Application Host performance will depend on appropriate processing, memory, and network capacity. Management Hosts will generally demand fewer processing and memory resources than Application Hosts. ZooKeeper hosts require low latency storage, as documented in *Section 2: ZooKeeper Host Storage Requirements*.

7.1 Redundant Management Services Example

Figure 5: Redundant Streams Configuration shows a highly available solution with separation of management and application functions. In this example, **instance.highAvailabilityCount=3**, which allows complete failure of two Management hosts at the same time; e.g., an unexpected failure during an outage for planned maintenance. There are three hosts running Streams Management Services and ZooKeeper. The two optional hosts running ZooKeeper only would be required for ZooKeeper to have the same level of protection as Streams Services to tolerate two failures. This reflects a difference in the redundancy mechanisms used by the two management subsystems. ZooKeeper relies on a quorum of operational instances to maintain the service, while Streams management services only require a single operational master. Therefore ZooKeeper requires five hosts to tolerate two ZooKeeper hosts failing. Note that there are alternatives to improve ZooKeeper availability in the event of software faults such as by provisioning two separate ZooKeeper processes on the same physical host. In any event, it is critical to provision an adequate storage subsystem as described in *ZooKeeper Host Storage Requirements*. For illustration in this example, five hosts were selected for application processes but of course the actual number will depend on specific application requirements.

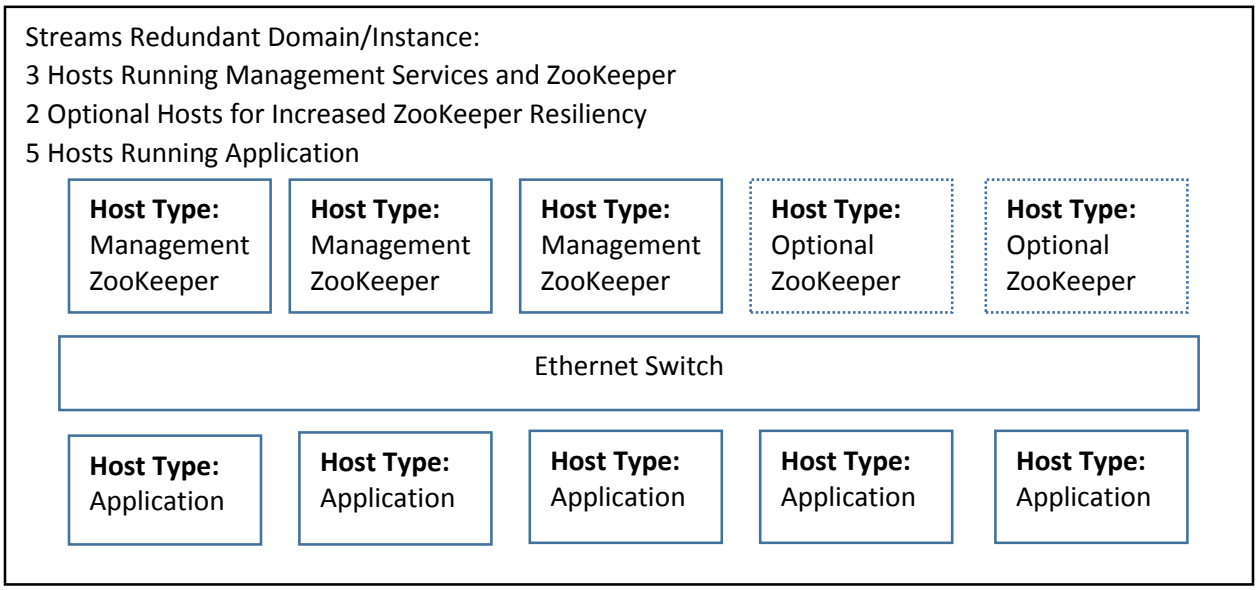

Figure 5: Redundant Streams Configuration

7.2 Non-Redundant Management Services with Multiple Application Hosts

Figure 6: Non-Redundant Streams Configuration shows an example of a simple Streams configuration that does not support redundant services where recovery is supported via service restart. All Management Services, as well as a standalone ZooKeeper, are running on a single host. In this example, there are 4 hosts dedicated for application processes, but the actual number will vary according to the requirements of the specific application.

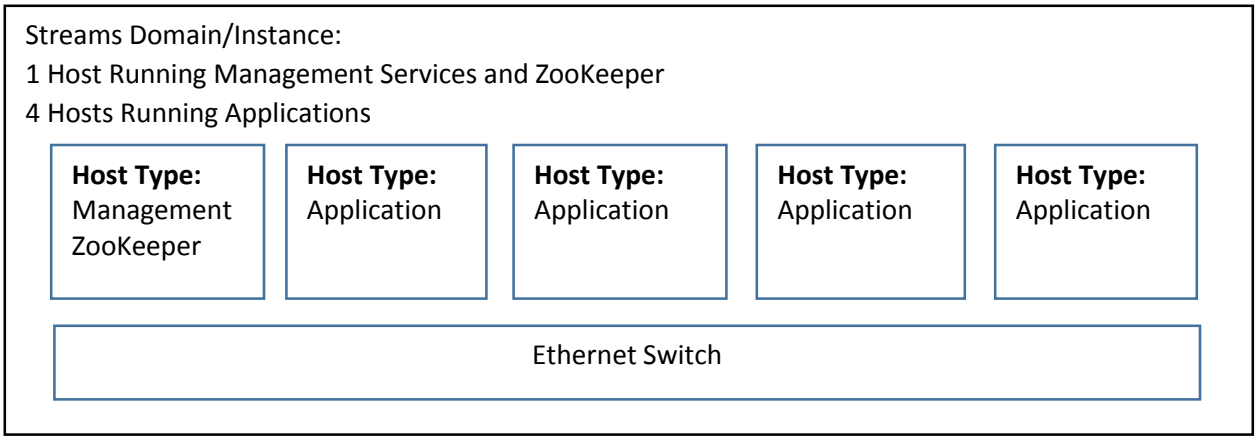

Figure 6: Non-Redundant Streams Configuration

7.3 Single Host Configuration

For development environments and small production application deployments which do not require high availability, Streams can be run in a single host configuration. Because Streams services and the application are competing for resources, consistent performance requires careful configuration of the system. Primary considerations include:

- **Ensure sufficient CPU and memory resources are allocated to both services and applications.** We recommend that at least 6 GB of memory be available for Streams Services and ZooKeeper. Processing resources will vary with installation requirements. Configurations supporting a single application which is submitted once and runs for long periods without the need for monitoring will require minimal processing capacity beyond the application requirements. If multiple applications will be submitted while other applications are already running, or dynamic connections between jobs are heavily used, we recommend that 15% of the processing capability of the host be reserved for Streams Services and ZooKeeper.
- **Provide dedicated or high performance disk resources for ZooKeeper.** The ZooKeeper best practice recommendation is to put the transaction log on a dedicated device, however this may not be sufficient. See *Dedicated Device for the Transaction Log File* for more information. Use of a high performance disk subsystem for the ZooKeeper log is the preferred solution for environments where performance is a consideration. Options include controllers with a non-volatile write cache, SSD drives, or fibre channel devices. See the *ZooKeeper Host Storage Requirements* section for more details.

Failure to follow the recommendations listed above will result in slow response times to Streams services such as job management and monitoring operations, and may cause application failures due to insufficient resources and timeouts.

It is important to remember that when running on a single host, failure of certain key system components (hardware platform, operating system, Streams services, etc) may result in an application outage. This minimal system configuration should only be used in situations when loss of application availability can be tolerated.

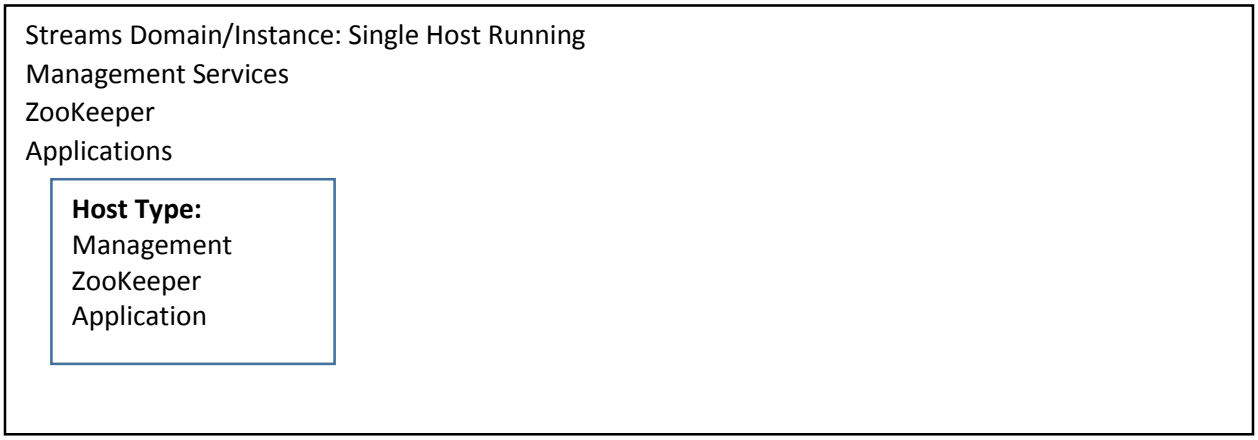

Figure 7: Single Host Configuration

7.4 Colocating Streams Services on Application Hosts

As shown in Single Host Configuration,Streams may be configured with application, management, and Zookeeper processes running on the same node. This concept may be applied to provide redundancy with services running on application hosts as shown in Figure 8: Combined Host Configuration, but consideration must be given to scenarios where contention for resources will prevent optimal performance on these configurations.

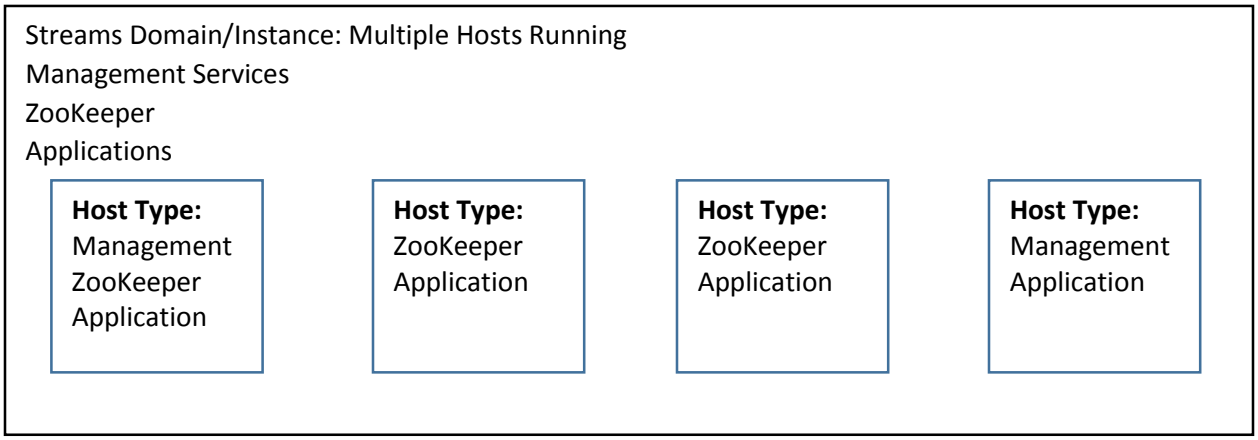

Figure 8: Combined Host Configuration

7.4.1 CPU Resource Utilization by Streams Services

Figure 9: CPU Consumption by Streams Management Services below shows system level CPU consumption over time by Streams Management Service and Zookeeper process during job submission and the subsequent process of PEs transitioning to the healthy state. This measurement was performed on a 16 core system that was hosting all of the streams management services and Zookeeper. There were no application processes running on the system. There are intervals where the management services consume significant amounts of CPU. Performance would clearly be degraded in the presence of an application consuming large quantities of CPU.

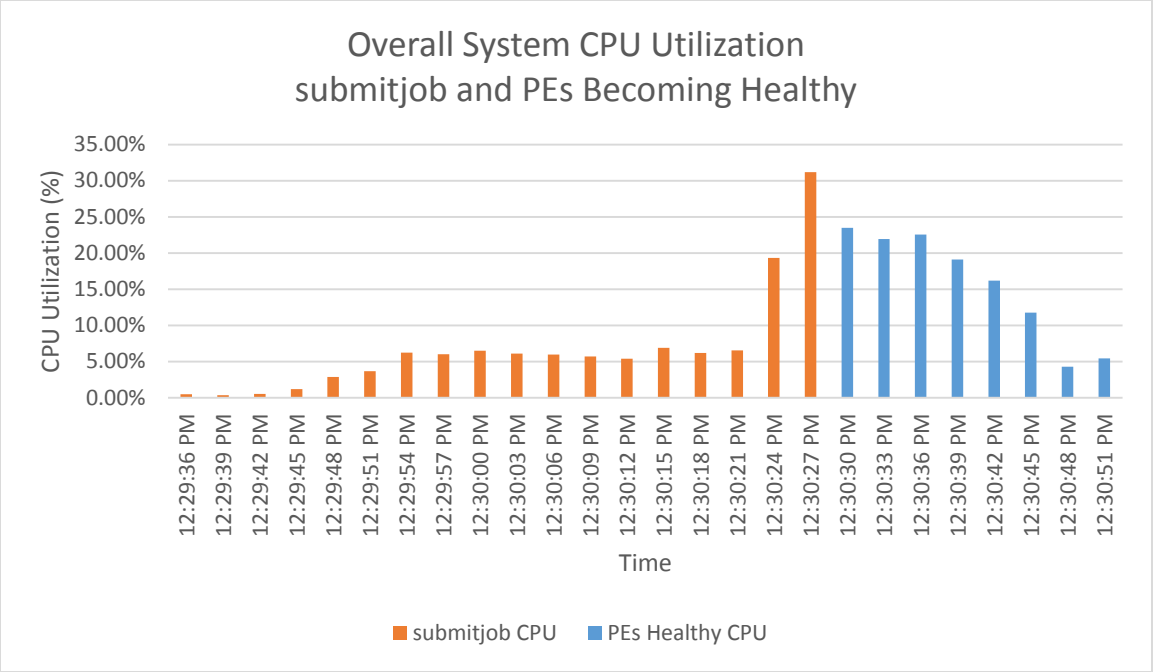

Figure 9: CPU Consumption by Streams Management Services

Another scenario that may require large amounts of CPU resource is shown in IBM InfoSphere Streams for Microsoft Excel. Other functions that may be degraded by the presence of a running application are collection of metrics and viewing the instance graph from Streams Studio or the web based console. Installations that do not require optimal performance for these scenarios may find Colocating Streams Services on Application Hosts a cost effective solution.

7.5 Example Hardware

7.5.1 Management/ZooKeeper Host

For general management functions such as submitting, monitoring, and cancelling jobs an 8 core host with 64 GB of memory is generally adequate. This will be able to run all of the Streams services and an embedded or standalone ZooKeeper. Note that the bulk of our testing was done with jobs that had hundreds of processing elements. In addition, these requirements take into account the need to monitor jobs and moderate use of IBM InfoSphere Streams for Microsoft Excel. Installations that are submitting only small jobs and have minimal need for monitoring and data collection functions will achieve satisfactory performance with smaller configurations.

http://www-03.ibm.com/systems/x/hardware/rack/x3550m5/ and http://www-03.ibm.com/systems/power/hardware/linux.html?LNK=browse provide more information on hosts that provide these capabilities. See *ZooKeeper Host Storage Requirements* for examples of low latency storage solutions appropriate for use in Management Hosts. Appropriate networking adapters may be found in the following link:

http://www-03.ibm.com/systems/x/options/networking/adapters.html.

For the Redundant Management Services configuration discussed in *Section 7.1 Redundant Management Services Example*, multiple identical hosts should generally be used. There are cases, however, where specific services require more compute power or memory. For example, the services used for IBM InfoSphere Streams for Microsoft Excel require significant computational resources for large numbers of clients, and hosts providing those services should be sized accordingly.

7.5.2 Application Host

Application Hosts differ from Management Hosts in that they generally require more processor and memory resources, but do not require low latency storage. For example, a typical application host may have 24 cores and 512 GB of memory. These guidelines are highly dependent on the application characteristics.

7.5.3 Network Switch

Selection of a network switch depends on the number of hosts in the configuration and application bandwidth requirements. High bandwidth applications may require 10 Gb/sec Ethernet, but in many cases 1 Gb/sec connections are sufficient.

8 New Features

8.1 Consistent Regions

Consistent Regions enable applications to guarantee that all tuples are processed at least once by periodically establishing checkpoints. This does require additional system resource utilization. That utilization will vary by application, but is generally correlated to

1. Frequency of checkpoint, and
2. Checkpoint size.

Figure 10: Effects of Checkpoint Size on Application Throughput illustrates that as Checkpoint Size increases, application throughput decreases and Figure 11: Effects of Checkpoint Frequency on Application Throughput shows that increasing the interval between checkpoints results in increasing application throughput. Note that increasing the time between checkpoints will result in longer recovery times in the event of failure. The data shown is from a sample application; results are highly dependent on individual application characteristics. Finally, since each PE for which Consistent Regions is enabled will start a Java Virtual Machine, care should be taken to limit the number of PEs employing Consistent Regions. This can be accomplished by fusing PEs using Consistent Regions or simply avoiding unnecessary use of Consistent Regions.

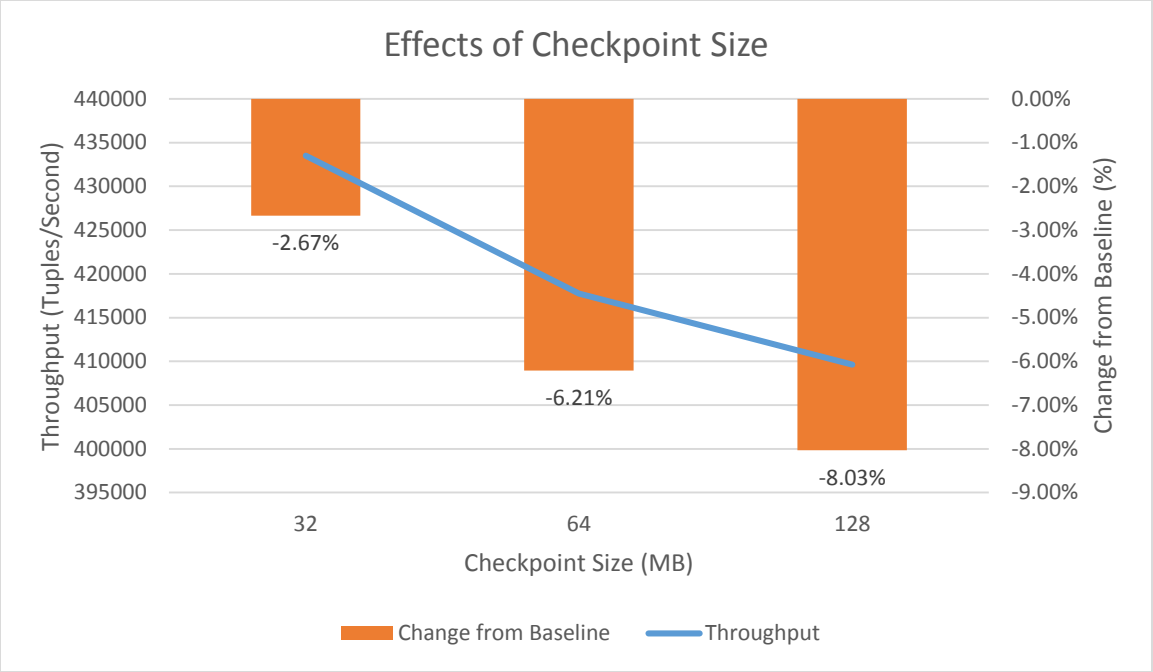

Figure 10: Effects of Checkpoint Size on Application Throughput

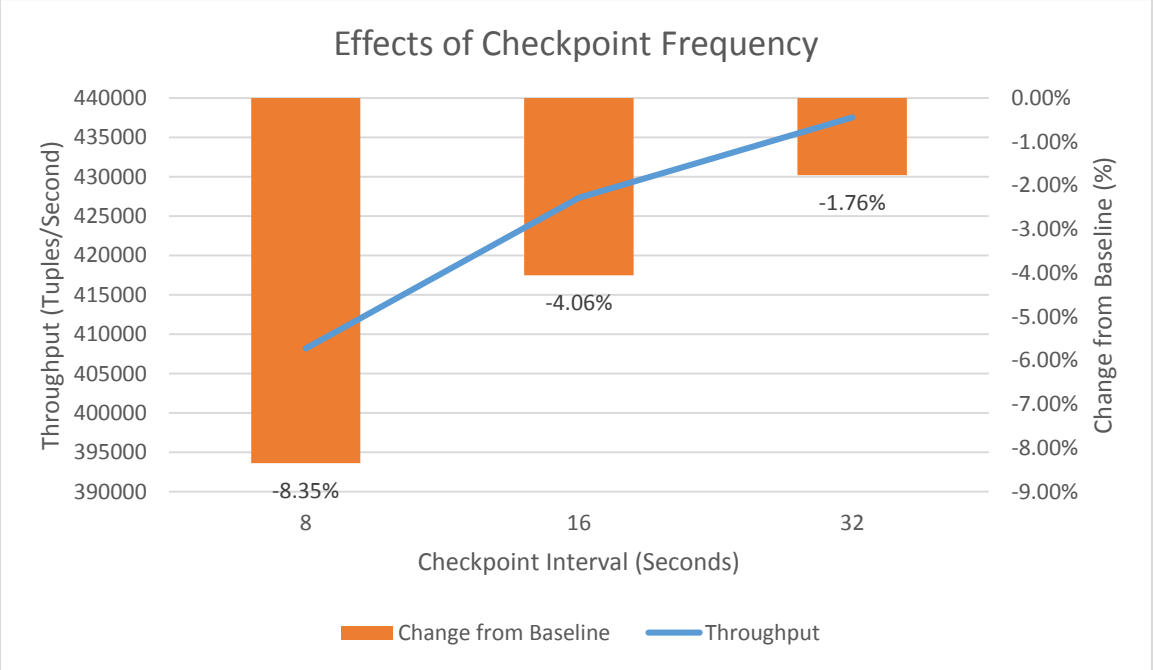

Figure 11: Effects of Checkpoint Frequency on Application Throughput

8.2 IBM InfoSphere Streams for Microsoft Excel

IBM InfoSphere Streams for Microsoft Excel provides a facility to import data from a running Streams application into a Microsoft Excel spreadsheet in real time. *Figure 12: Characteristics of IBM InfoSphere Streams for Microsoft Excel* shows how response time and CPU utilization increase as the number of

client threads accessing a view increases. Sub-second response time is maintained with up to 500 clients accessing the view. This test was executed on a host with 28 Intel Xeon model E5-2697 v3 cores running at 2.60GHz with Hyper Threading enabled. For planning purposes, we recommend that at least one core be provisioned for every 75 clients simultaneously using IBM InfoSphere Streams for Microsoft Excel, assuming a refresh rate of 3 seconds. With that in mind, heavy users of this function may want to provision larger hosts than recommended in *Management/ZooKeeper Host* for running the view services.

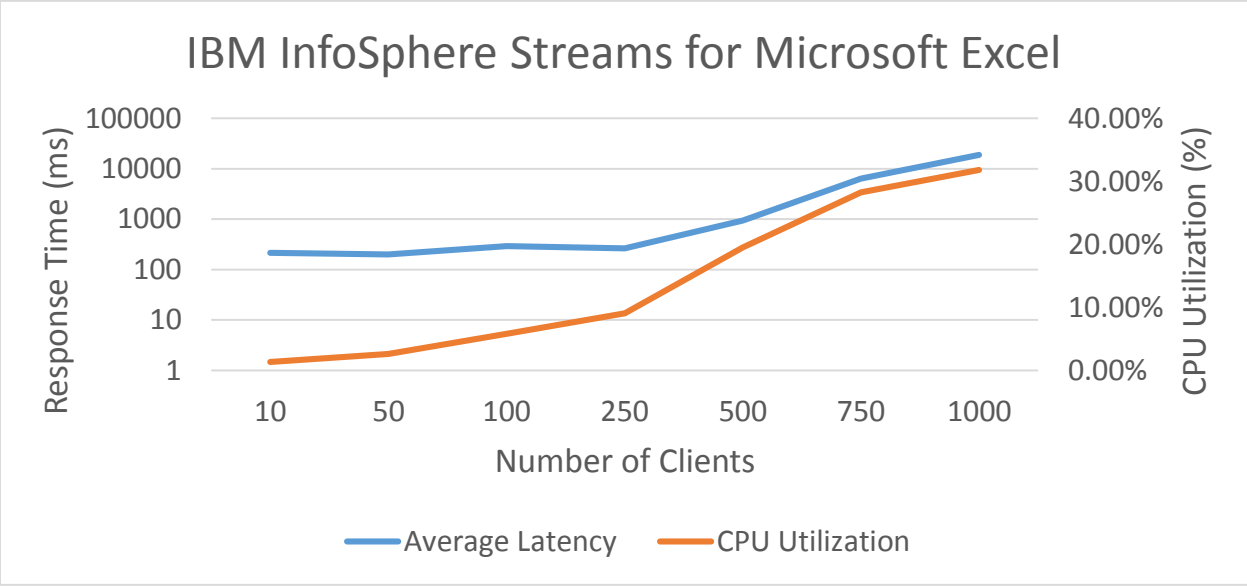

Figure 12: Characteristics of IBM InfoSphere Streams for Microsoft Excel

9 Summary

In this document we have included many best practices for successfully deploying IBM InfoSphere Streams V4.0.0. These include:

- Hosts running ZooKeeper require low latency storage.
- The robust high availability mechanisms implemented by V4 require additional memory on hosts running Streams Management Services.
- With careful planning, Streams Management Services may be run on the same host as application process.
- In configurations where the application must be spread across multiple hosts, it is best to dedicate separate hosts for the Management Services.
- It is generally acceptable to run all Streams Management Services and ZooKeeper on a single host. There are exceptions to this practice, particularly for heavy users of specific functions.