Red Hat OpenShift® on IBM zSystems® –

A Performance Tuning Case Study on the Example of MongoDB® & ODF 4.10

Dr. Manuel Gotin
manuel.gotin@ibm.com
Linux on IBM zSystems Performance Specialist

Motivation

Cloud Platforms are an emerging and disruptive technology changing the Software Landscape

IBM® enables with **OpenShift® Platform (OCP)** Hybrid-Cloud Strategies for Customers and provides with IBM zSystems® a powerful infrastructure

Customers operating Cloud Application Workloads on OCP want to know:

- How to tune my OCP on IBM zSystems® stack for a better performance of my workload?
- How does my tuning impact my infrastructure sizing?

In this talk, we present a Case Study with **OCP 4.10** and **MongoDB®** using a Storage provided by **OpenShift® Data Foundation (ODF)**

Background – OpenShift® Data Foundation

ODF (Red Hat **O**penShift® **D**ata **F**oundation) is a software-defined storage of containers with 3 main components:

- Ceph®: a software defined storage – https://ceph.io/
- Rook: OCP operators to deploy and maintain ODF – https://rook.io/
- noobaa™: Multi-Cloud Gateway for Object Stores – https://noobaa.io/

ODF is deployed on compute nodes of a OCP cluster:

- Seamless integration into OCP
- Provides file-, block- and object-storage interfaces for applications

ODF replicates data across multiple storage devices:

- Cluster is able to rebalance and recover
- Cluster maintains data integrity

Background – MongoDB®

„MongoDB® is a document database with the scalability and flexibility that you want with the querying and indexing that you need” – [*https://www.mongodb.com/*](https://www.mongodb.com/)

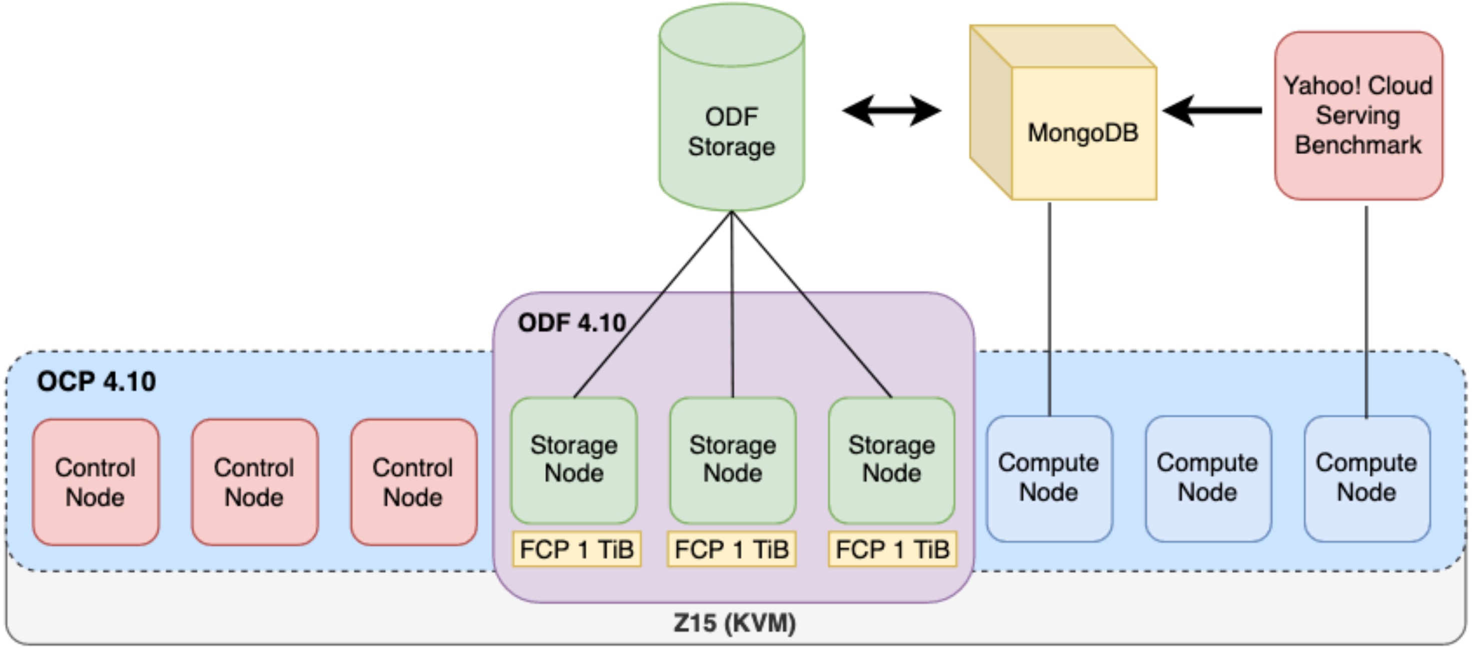
MongoDB® is a scalable high-performance document-orientated database:

- NoSQL Database: Data is not stored in relational tables
- Data is stored as a collection of key value in JSON structure

Popular for developing scalable internet applications with evolving data schemas, e.g. finance, e-commerce

Case Study

Case Study – Overview



Case Study Scenario:

- Cloud service workload with MongoDB®
- Storage managed and provided by ODF 4.10

Case Study System:

- KVM-based OCP 4.10 on a z15™ with 16 IFLs
- ODF 4.10 is using 3x 1 TiB FCP Disks
- MongoDB® backed by a 100G CephRBD disk

Case Study – YCSB Workload

The Yahoo!® Cloud Serving Benchmark (YCSB) [1] enables performance comparisons of the new generation of cloud data serving systems.

It defines a set of workloads of a Web Service, e.g. session storing or threaded conversations.

We use **Workload A – Update Heavy**, which represents storing recent actions in a user session.
It consists of 50% read and 50% update operations on the records.

We conduct a total of 8500k operations over 7000k records.

[1] https://courses.cs.duke.edu/fall13/compsci590.4/838-CloudPapers/ycsb.pdf

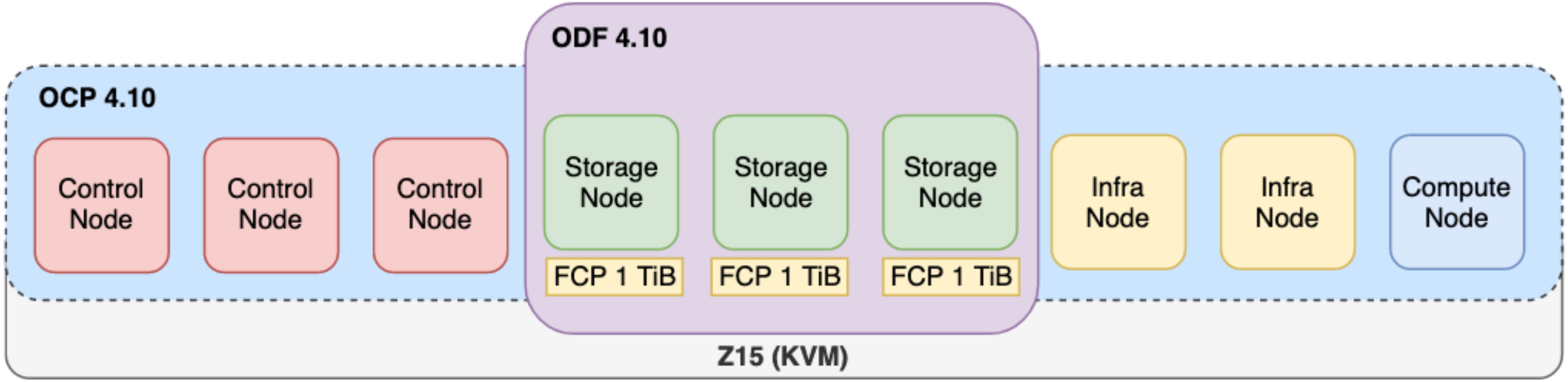
Case Study – Tuning Scenarios

Scenario	Tuning	Description	Reference
A - Baseline	-	Default installation of OCP and ODF	-
B - Infra-Nodes	Deployment	Separate OCP infrastructure workload from others	**IBM Blog Post** Boosting Performance by Infrastructure Nodes [1]
C - RFS Tuning	Network	Tune the Network configuration of the Cluster with Receive Flow Steering (RFS)	**IBM Developer Tutorial** Tune the network performance by RFS [2]
D - KVM-Tuning	Network, Disk, Memory	Set of Tuning Practices tailored to KVM-based OCP installations, including RFS, Memory, Disks	**Tuning Playbook** KVM-IPI-Automation Playbooks [3]

[1] https://www.linkedin.com/pulse/boosting-performance-using-infrastructure-nodes-your-cluster-miranda/
[2] https://developer.ibm.com/tutorials/red-hat-openshift-on-ibm-z-tune-your-network-performance-with-rfs/
[3] https://github.com/ibm-s390-cloud/ocp-kvm-ipi-automation

Tuning Scenario B – Infrastructure Nodes

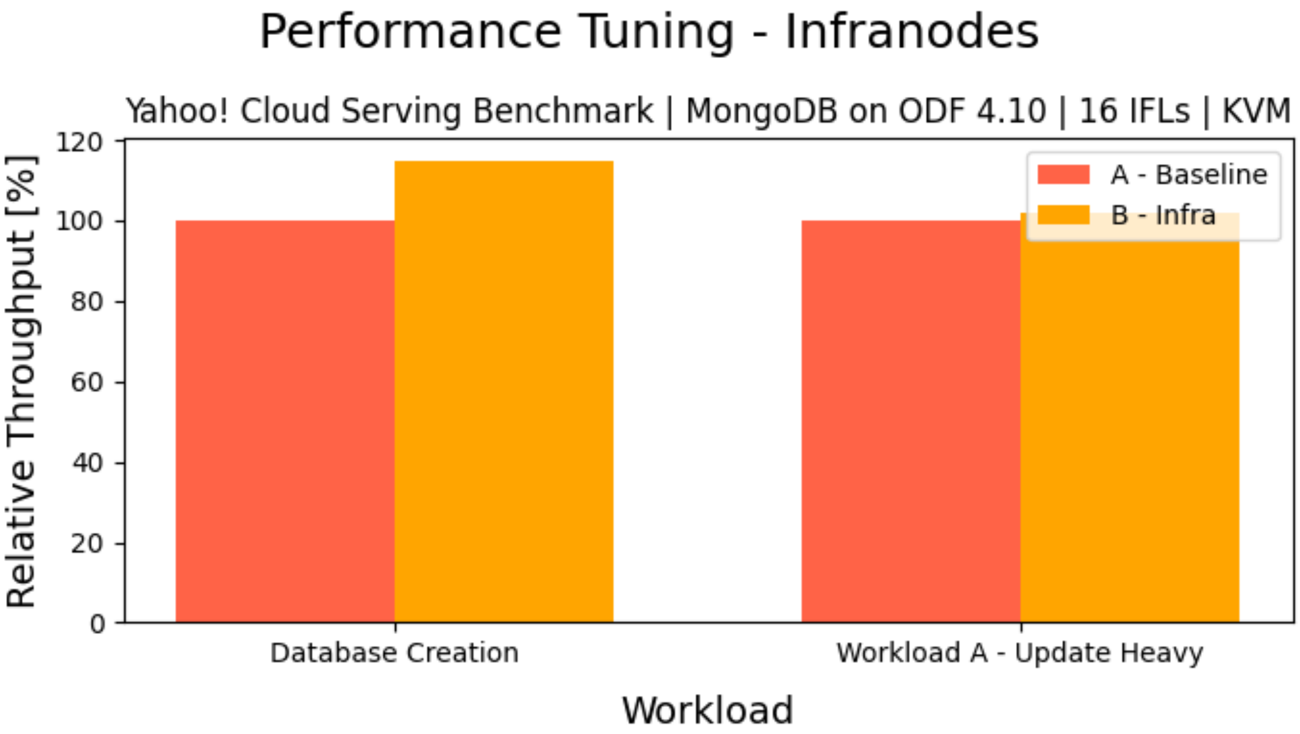
Infranode Tuning



Infrastructure nodes allow to isolate infrastructure workloads, e.g. of openshift-monitoring, openshift-dns, ...

- Prevents that other workloads are scheduled on infrastructure nodes
- Done by labelling compute nodes with the infra node-role

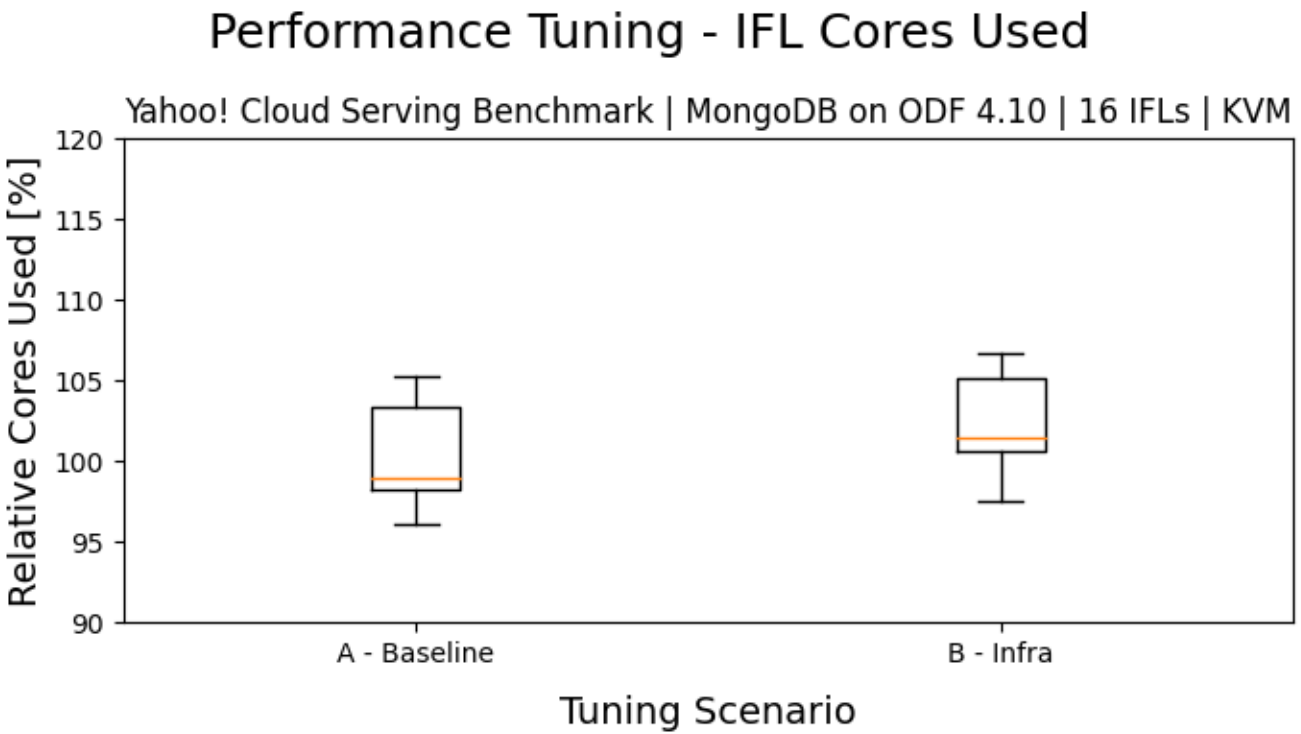
Results - Infranode Tuning



Throughput – Workload A
Infra has a minor impact on Workload A of **+2%**

Throughput – DB Creation
Infra has a noticeable impact on database creation of **+15%**

Results - Infranode Tuning

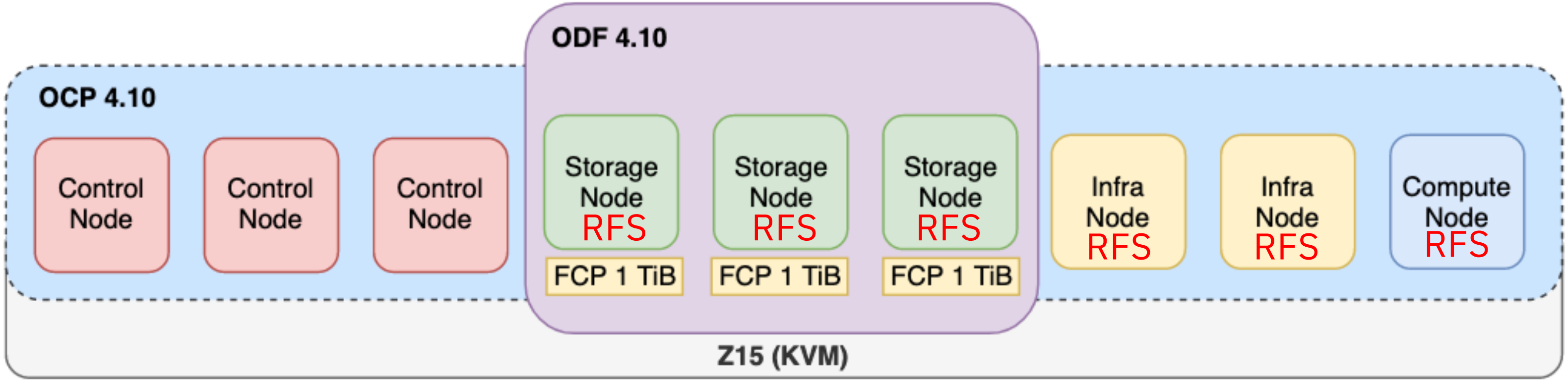


Core Utilization – Workload A
Minor impact on utilized cores of **+2%**

Discussion
Gain in core utilization is as high as the gain in throughput

Tuning Scenario C –
Receive Flow Steering (RFS)

Receive Flow Steering

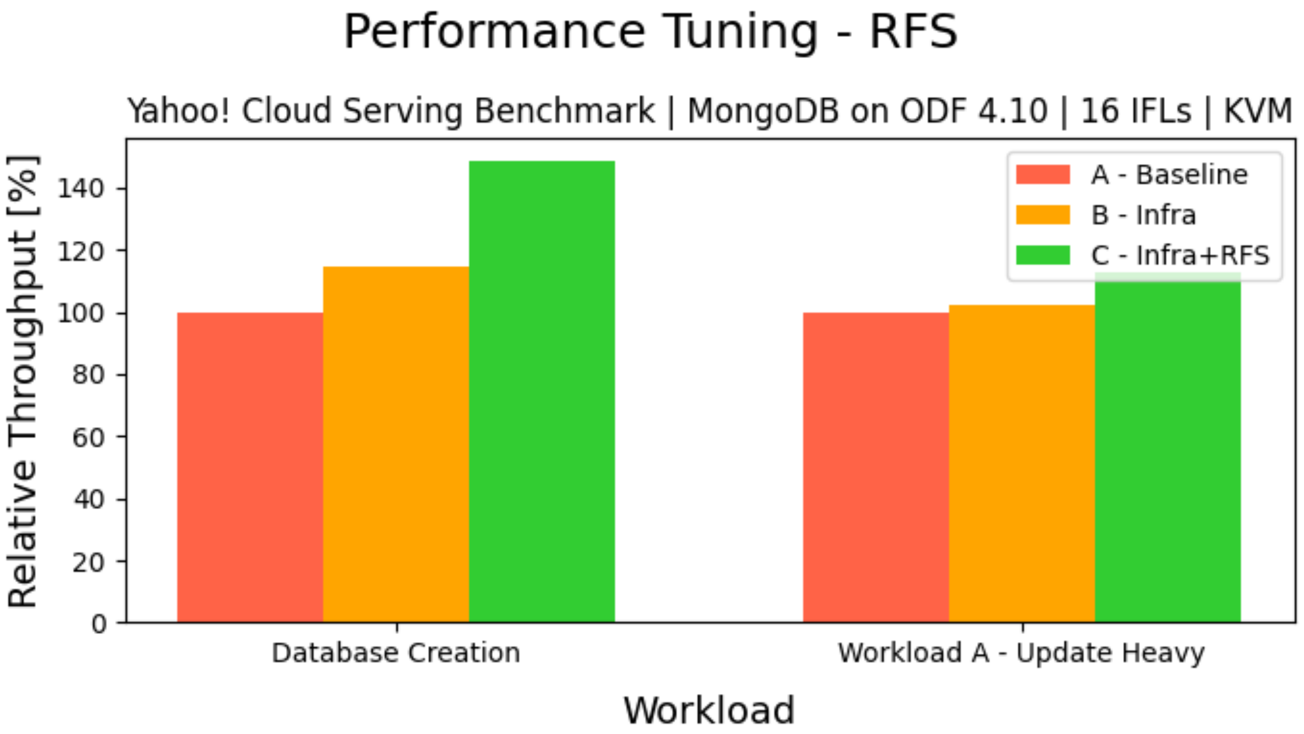


Receive Flow Steering (RFS) calculates the most appropriate CPU to forward network packets to increases CPU cache hit ratio. It has shown to reduce network latency for OCP workloads [2].

We activate it for the **Compute, Storage** and **Infra Nodes** in the cluster

[2] https://developer.ibm.com/tutorials/red-hat-openshift-on-ibm-z-tune-your-network-performance-with-rfs/

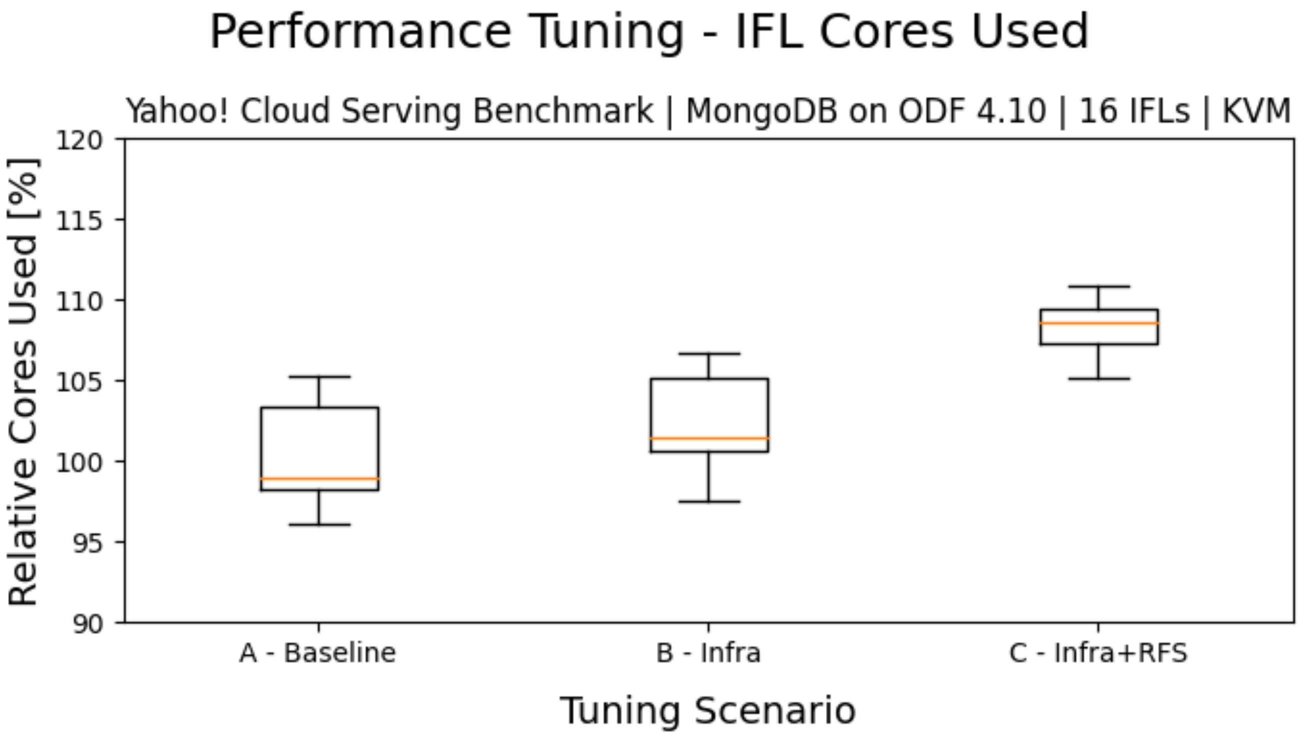
Results – Receive-Flow Steering



Throughput – Workload A
RFS has a noticeable impact of **+12%**

Throughput – DB Creation
RFS has a strong impact on database creation of **+48%**

Results – Receive-Flow Steering



Core Utilization – Workload A
Noticeable impact on utilized cores of **+8%**

Discussion
RFS results in a decent throughput gain on the expense of core utilization.

Tuning Scenario D – KVM-Tuning Playbook

KVM Tuning Playbook

The KVM-Tuning Playbook is part of the OCP KVM-IPI-Automation [1] collected and maintained by Dirk Haubenreißer [2]

It provides an Ansible® playbook for automated tuning of KVM-based OCP installations.

The collection configures the following (non-exhaustive):

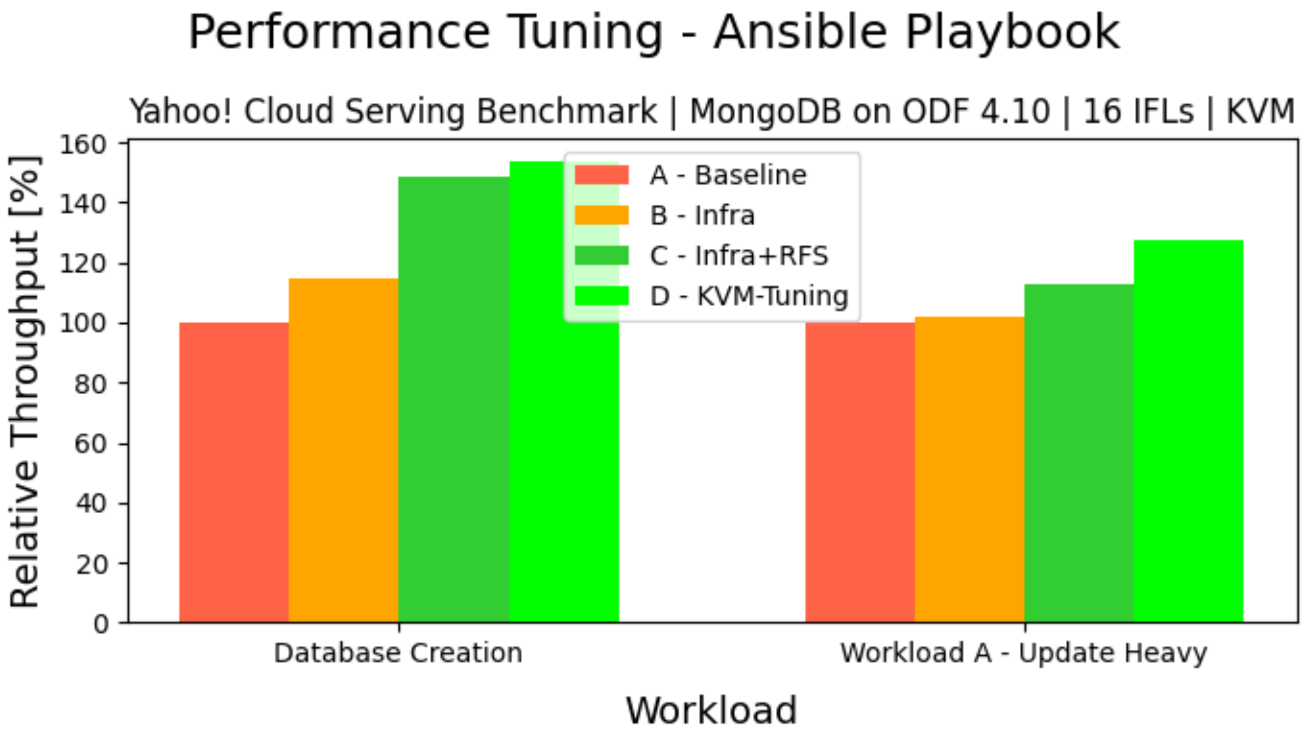
- Dedicated I/O Threads for disk access
- Increases CPU weight for cluster guests
- Memory tuning (transparent huge pages)
- Network tuning (RFS, libvirt)

It extends the previous Infra+RFS scenario by disk and memory tunings.

Please note, that RFS and transparent huge pages are not limited to KVM-based OCP installations.

[1] https://github.com/ibm-s390-cloud/ocp-kvm-ipi-automation/tree/main/ansible/roles/tuning
[2] haubenr@de.ibm.com

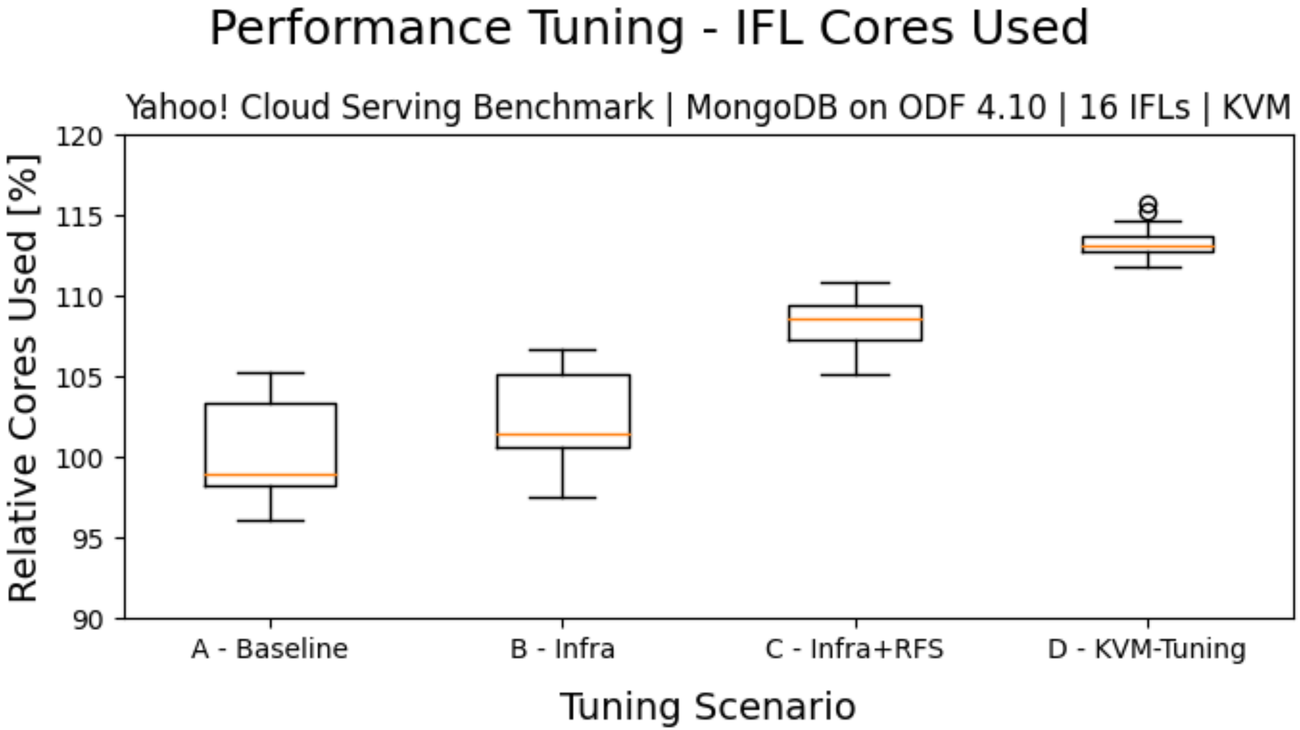
Results – KVM-Tuning Playbook



Throughput – Workload A
The Tuning has a strong impact on Workload A of **+27%**

Throughput – DB Creation
RFS has a strong impact on database creation of **+53%**

Results – KVM-Tuning Playbook



Core Utilization – Workload A
Strong impact on utilized cores of **+14%**

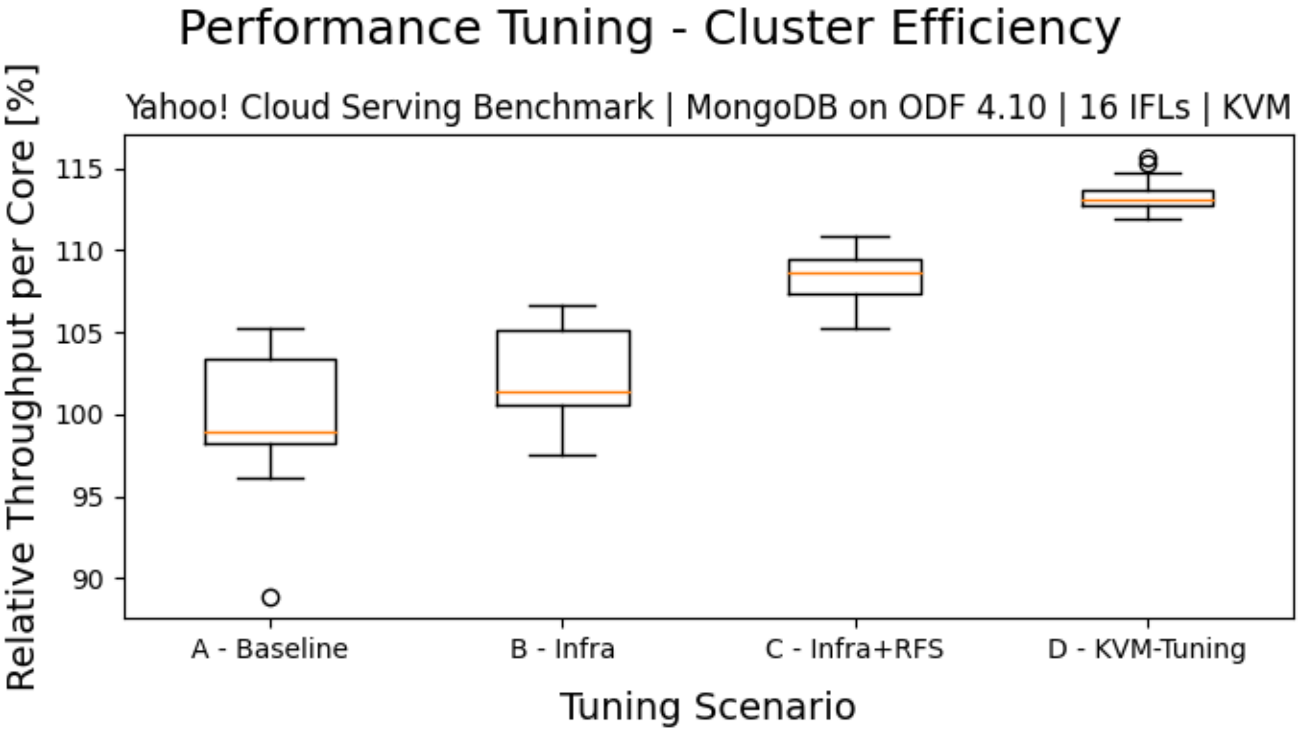
Discussion
Tuning Scenario results in the highest throughput gain and core utilization increase with **+27%** and **+14%**

Results – Summary

Scenario	Tuning	δtps [DB Creation]	δtps [Workload A]	δUtilized IFL Cores
Baseline	-	-	-	-
Infra-Nodes	Deployment	+ 15 %	+ 2 %	+ 2 %
RFS Tuning	Deployment, Network	+ 48 %	+ 12 %	+ 8 %
KVM-Tuning	Deployment, Network, Disk, Memory	+ 53 %	+ 27 %	+ 14 %

Sizing Considerations

Tuning Scenarios & Cluster Efficiency



Cluster Efficiency
Throughput per utilized IFL Core

Results – Workload A
In terms of cluster efficiency:

- **B –** Infra is on pair with the baseline
- **C –** RFS increases the cluster efficiency by **+3%**
- **D –** KVM-Tuning further manages to improve it by **+12%**

Discussion
The tunings enable to use the cores more efficiently

Takeaways & Outlook

Takeaways & Outlook

The presented tuning scenarios for a cloud service workload with MongoDB® & ODF 4.10 ...

- ... increased the throughput of the cloud service workload
- ... increased the core utilization during workload peaks
- ... were able to utilize the cores more efficiently – which is awesome!

The case study demonstrated the capability of the OCP on IBM zSystems® stack for tuning the performance of workloads in many ways, e.g. topology, network, hypervisor, ...

Outlook of the Case Study:

- **Deployment:** Deploy *ycsb* and *mongodb* pods on one Node (pod-2-pod communication)
- **Network:** Configure Multus for ODF to isolate storage traffic from application traffic

Notices and Disclaimer (1/2)

Information in these presentations (including information relating to products that have not yet been announced by IBM) has been reviewed for accuracy as of the date of initial publication and could include unintentional technical or typographical errors. IBM shall have no responsibility to update this information. **This document is distributed "as is" without any warranty, either express or implied. In no event, shall IBM be liable for any damage arising from the use of this information, including but not limited to, loss of data, business interruption, loss of profit or loss of opportunity**. IBM products and services are warranted per the terms and conditions of the agreements under which they are provided.

IBM products are manufactured from new parts or new and used parts. In some cases, a product may not be new and may have been previously installed. Regardless, our warranty terms apply.

Any statements regarding IBM's future direction, intent or product plans are subject to change or withdrawal without notice.

Performance data contained herein was generally obtained in a controlled, isolated environment. Customer examples are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual performance, cost, savings or other results in other operating environments may vary.

References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business.

Notices and Disclaimer (2/2)

Workshops, sessions and associated materials may have been prepared by independent session speakers, and do not necessarily reflect the views of IBM. All materials and discussions are provided for informational purposes only, and are neither intended to, nor shall constitute legal or other guidance or advice to any individual participant or their specific situation.

It is the customer's responsibility to insure its own compliance with legal requirements and to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer's business and any actions the customer may need to take to comply with such laws. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the customer follows any law.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products about this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products. IBM does not warrant the quality of any third-party products, or the ability of any such third-party products to interoperate with IBM's products. **IBM expressly disclaims all warranties, expressed or implied, including but not limited to, the implied warranties of merchantability and fitness for a purpose.**

IBM, the IBM logo, and ibm.com are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at: www.ibm.com/legal/copytrade

The registered trademark Linux® is used pursuant to a sublicense from the Linux Foundation, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis.

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents, copyrights, trademarks or other intellectual property right. Refer to www.ibm.com/legal for further legal information.