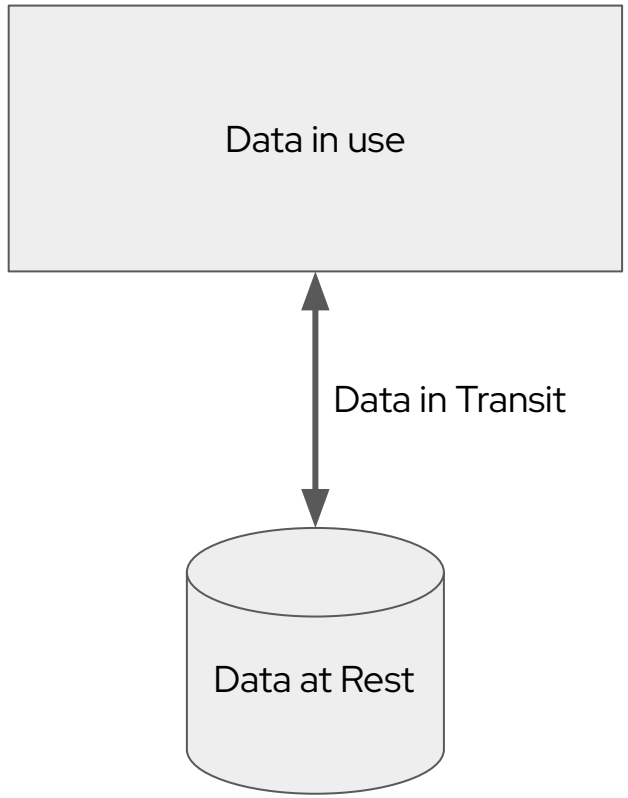
Sichere Nutzung der Cloud mit Hilfe von Confidential Computing

Red Hat | intel.

States of data at the core of security concerns.

- Move to the cloud -> Access to Data needs to be restricted.
  - See DORA (https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020PC0595 Chapter II Section II Article 8 Paragraph 2)
- 3 states of data are available.



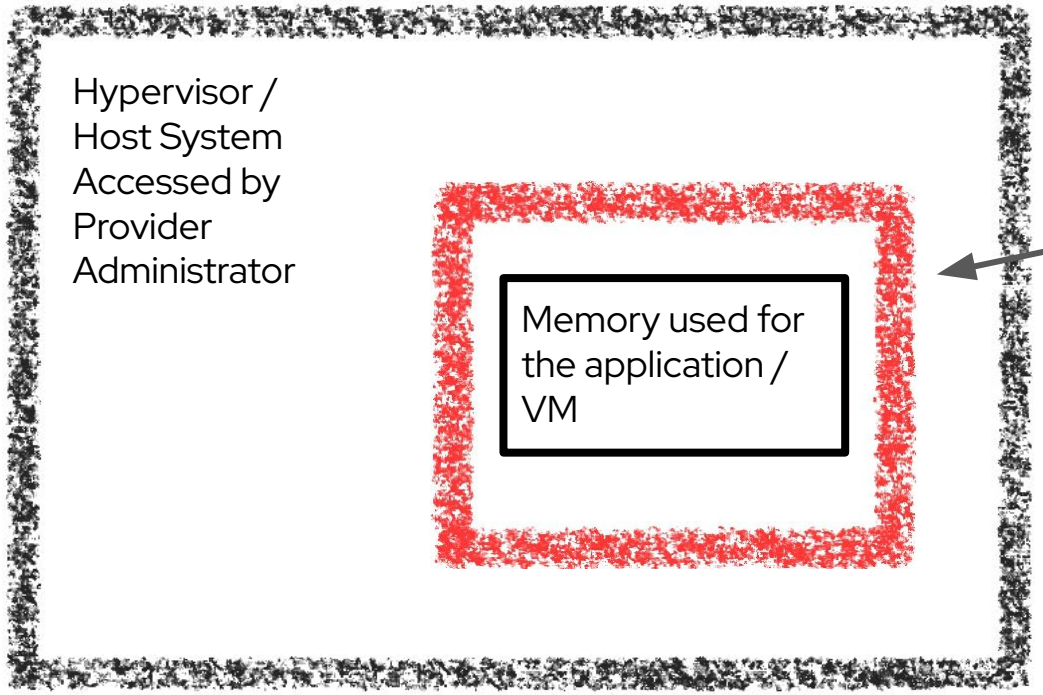
Red Hat

What is Confidential Computing

- Confidential Computing (CC) is the protection of **data in use** by performing the computation in a **hardware-based, attested Trusted Execution Environment**, according to the Confidential Computing Consortium's definition. The three primary attributes of a Trusted Execution Environment are **data integrity, data confidentiality, and code integrity**.

Data in use

- How does Confidential Computing work?



Red Hat

2 security concepts are used here:

- Remote Attestation

  Is my application really running in a confidential environment?

- Separation of responsibilities

  Distribution of responsibilities between different roles in the overall process of Confidential Computing

Hardware implementations

- So this confidentiality needs to be implemented at the hardware level.
  - AMD SEV SNP: Confidentiality on a Core based VM and on the Memory

    https://www.amd.com/en/processors/epyc-confidential-computing-cloud
  - Intel TDX (VM)

    https://www.intel.com/content/www/us/en/products/docs/processors/xeon-accelerated/security-accelerators-product-brief.html
  - IBM z HyperProtect + Secure Execution

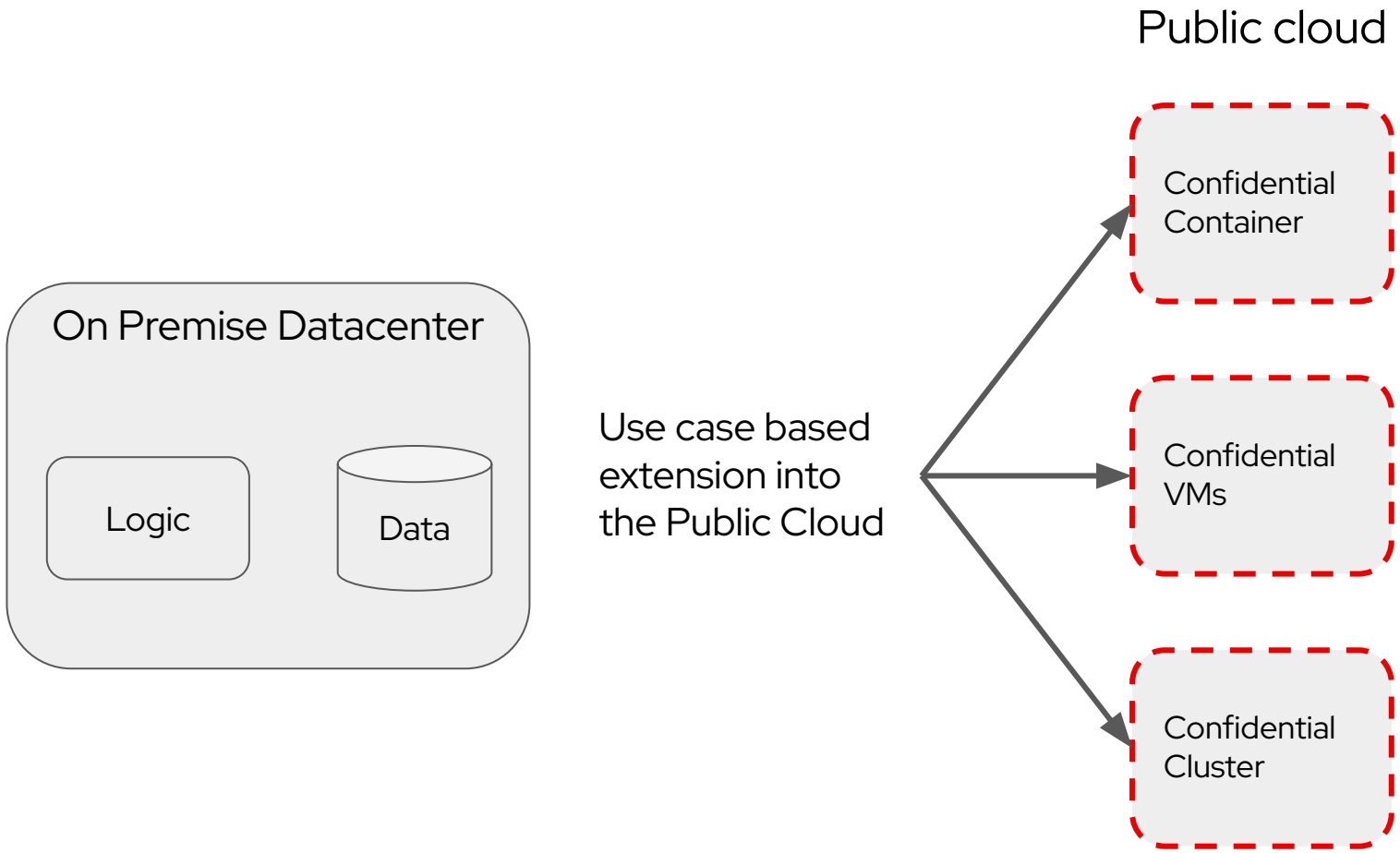
    HSM modules available to enable confidential computing environments
  - ARM CCA

    https://www.arm.com/architecture/security-features/arm-confidential-compute-architecture
  - AWS Nitro

    https://docs.aws.amazon.com/whitepapers/latest/security-design-of-aws-nitro-system
  - RISC-V (plan)

    https://github.com/riscv-non-isa/riscv-ap-tee/blob/main/specification/riscv-cove.pdf

Confidential Computing and its outlook



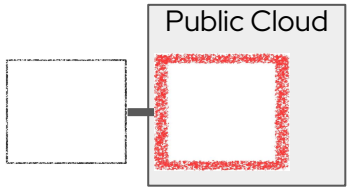
Categorization of Use cases

What use cases are our customers telling us about?

Partner Interaction

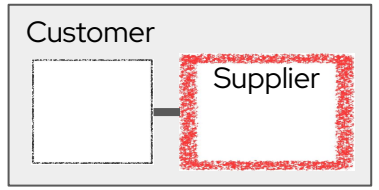
2 protected datasets interacting in confidential container

Secure Cloudburst



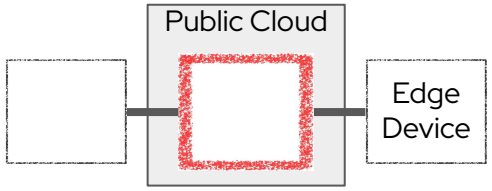
Using the public cloud to for peak workload or shared resources

IP Protection/Integrity



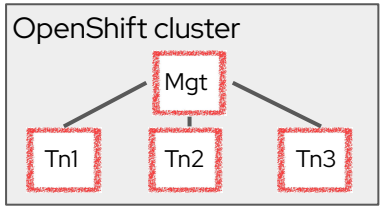
Protection of supplier data and business logic in customer environments

Edge use case



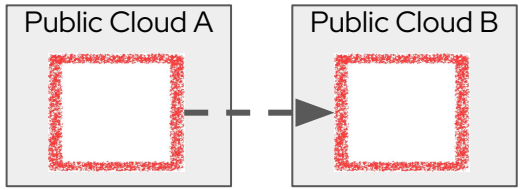
Protecting Edge device data in the public cloud for aggregation

Total Tenant Isolation



Isolating OpenShift Tenants

Digital Sovereignty



Encapsulating and moving workload from one provider to the next.

Red Hat

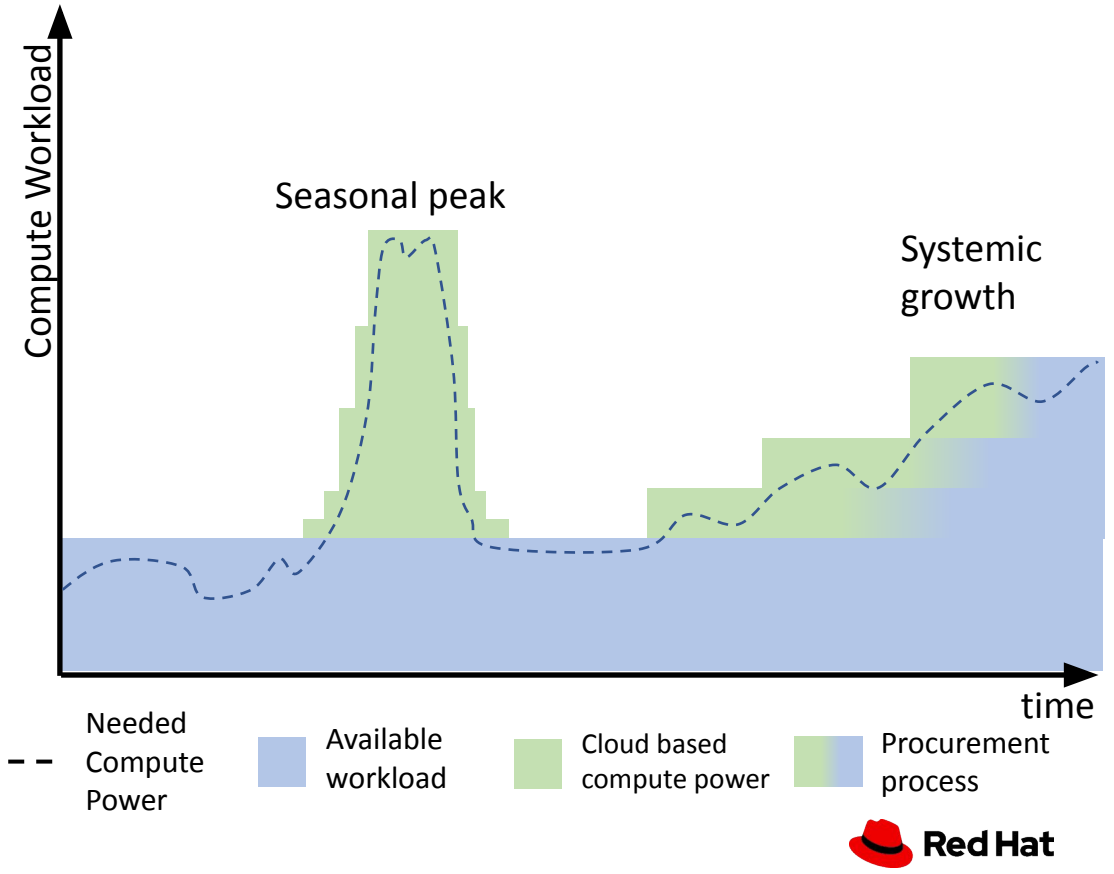
When and how to use the Public Cloud securely?

- Main services will always use the On Premise DataCenter.
- The procurement process to add computational Nodes takes a considerable time.
- Data can made available in the Cloud (partitioning/duplication/preemptive deployment)

- Seasonal Peak
  - Have enough unused compute power available to grow incurs a lot of cost.
  - Not necessary to add Compute nodes.
  - Temporary compute power can be used in the cloud

- Systemic growth
  - The average on premise load rises.
  - To be able to provide enough compute power

The question remains how to secure your business logic and data in the cloud?



Red Hat

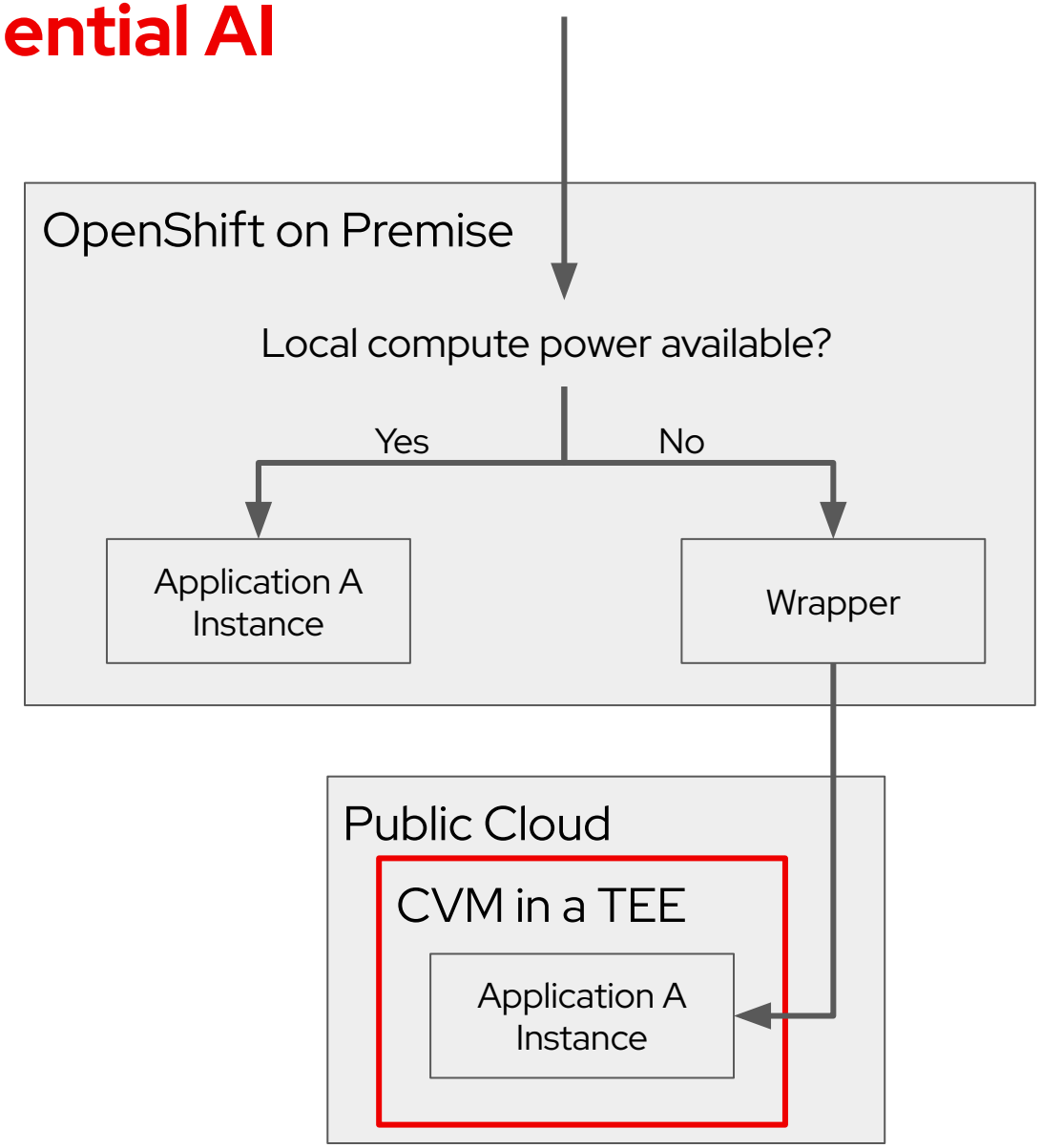
Secure Cloud Bursting with Confidential AI

- Why are we looking at this?
  - Requirements are mostly coming from FSI and regulated industry.
  - Regulations impose need to secure access to logic and data running in the public from external players.

- What are the requirements?
  - When on premise compute power reaches its limits the public cloud should be used.
  - When using the cloud data and business logic need to be protected so no one else can access it or be forced to access it.
  - Regulations enforce security for Data at Rest, Data in Transit and Data in Use.
  - This also needs to include AI related workload.

Secure Cloud Bursting with Confidential AI

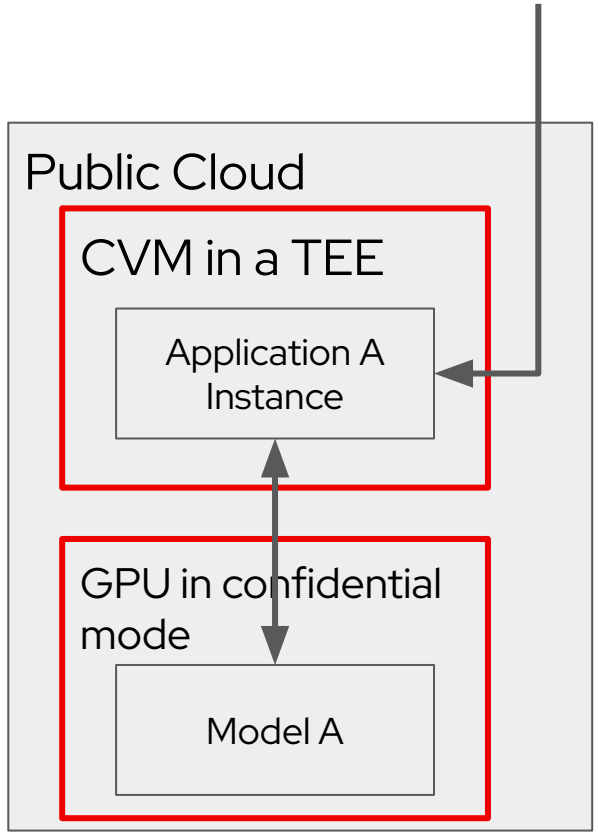
- What is Dynamic Secure Cloud Bursting
  - The on premise resources are used as long as there is compute power left to execute those request.
  - For each new request it is evaluated if the on premise infrastructure can handle the execution. When not, dynamically a cloud based instance is initiated and the request is diverted to the public cloud.
  - When the load of the request is reduced the cloud instances are removed.



Red Hat

Secure Cloud Bursting with Confidential AI

- Integrating the execution of the AI model into a Confidential Computing context.
  - The GPU is external resource outside of the CPU and therefor has its own confidential environment.
  - When an application is started which needs a GPU
    - a GPU is reserved.
    - the GPU is started in confidential mode.
    - the model needs to be loaded into the GPUs memory using a secured connection.
  - When the cloud instance is deleted also the GPU is released.



Red Hat

Secure Cloud Bursting with Confidential AI

- How is the scheduling in the On Premise OpenShift done?
  - KEDA is used in our demo use case.
  - KEDA is a lightweight, open-source Kubernetes event-driven autoscaler used by DevOps, SRE, and Ops teams to horizontally scale pods based on external events or triggers.
    https://docs.openshift.com/container-platform/4.11/nodes/cma/nodes-cma-autoscaling-custom-rn.html

  - For this demo a policy has been developed which can be customized to address the requirements of the target system.

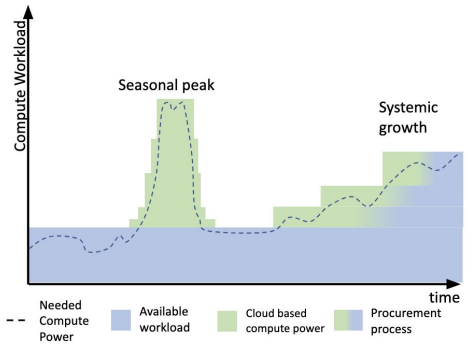
  - With KEDA we are evaluating the load on the on premise node. If space a local instance is created and the requests forwarded to that instance.

  - If there is not enough space then KEDA is instructed to create the local wrapper of the cloud instance which then initiates the cloud instance and manages the connection to the instance etc.

Red Hat

Secure Cloud Bursting with Confidential AI

- Secure Cloud Bursting
  - enables the public cloud acting as a consumption based flexible extension of the on premise Datacenter.
  - ensures the implementation of the highest security standards including securing Data in Use.
  - implementing confidentiality of the business logic and data. And also integrating consumption based GPUs and executing AI models in a confidential context.
  - enabling companies to use the Public Cloud in a secure way



Red Hat

The End