



CHECKLIST

AWS Spot Instances Virtualization Cost Optimization Checklist

A practical checklist to help virtualization teams decide whether AWS Spot Instances fit a workload, validate interruption tolerance, and confirm the operational controls needed before production use.

AWS Spot Instances Virtualization Cost Optimization Checklist

Use this checklist to decide whether Spot Instances are a good fit for a virtualization workload and to validate the controls needed before production use.

1) Workload fit

- ☐ The workload can tolerate interruption without causing customer impact.
- ☐ The VM is disposable, replaceable, or short-lived by design.
- ☐ The workload is batch-based, ephemeral, or horizontally scalable.
- ☐ Long-lived user sessions are not dependent on a single VM.
- ☐ The workload does not require continuous host residency.
- ☐ The application can restart cleanly from image, config, or automation.

2) State and data handling

- ☐ Critical data is stored outside the instance root disk.
- ☐ Logs, artifacts, and queue state are exported to durable storage.



- ☐ Local disk usage is limited to cache or temporary data.
- ☐ If attached storage is used, recovery behavior is documented.
- ☐ Backup and snapshot procedures are tested and current.
- ☐ Encryption requirements for disks and backups are confirmed.
- ☐ Recovery time for persistent data is acceptable.

3) Interruption readiness

- ☐ The workload has graceful shutdown handling.
- ☐ Any pre-stop or termination notice is consumed automatically.
- ☐ Jobs can checkpoint progress before interruption.
- ☐ Retries are safe and do not create duplicate side effects.
- ☐ Replacement capacity can be launched automatically.
- ☐ Health checks detect and remove failed or reclaimed nodes.
- ☐ The team knows what happens during capacity reclamation.

4) Automation and orchestration

- ☐ Images are standardized and version-controlled.
- ☐ Configuration is externalized and reproducible.
- ☐ Node provisioning is automated end-to-end.
- ☐ Cluster or scheduler reschedules work without manual intervention.
- ☐ Build, scan, or batch jobs can resume or rerun safely.
- ☐ Manual rebuild steps are not required for normal recovery.

5) Capacity and placement strategy

- ☐ The workload can run on multiple instance sizes or families.



- ☐ The design can tolerate capacity variability across Availability Zones.
- ☐ A fallback capacity type is defined if Spot is unavailable.
- ☐ Placement constraints are not overly strict.
- ☐ Surge handling has been tested with replacement nodes.
- ☐ The workload does not depend on one fixed host identity.

6) Cost validation

- ☐ The expected Spot discount is documented.
- ☐ The cost of interruption handling is estimated.
- ☐ The cost of recovery time is estimated.
- ☐ The cost of extra storage, logging, or data movement is included.
- ☐ Operational labor required for recovery has been considered.
- ☐ Savings still hold after overhead is added.

7) Security and control checks

- ☐ Instance roles follow least privilege.
- ☐ Network boundaries are defined and tested.
- ☐ Access paths are documented for transient hosts.
- ☐ Sensitive workloads have appropriate isolation controls.
- ☐ Data protection requirements remain intact on Spot nodes.
- ☐ Monitoring and alerting cover interruption and recovery events.

8) Pilot and rollout

- ☐ A non-critical pilot pool has been defined.
- ☐ Pilot success criteria are documented.



- ☐ Interruption rate, queue drain time, and recovery time are measured.
- ☐ Failure scenarios were tested before broader rollout.
- ☐ The team reviewed whether savings persist after operational overhead.
- ☐ The workload was not moved to Spot until the pilot passed.

9) Red flags: do not use Spot yet if any of these are true

- ☐ The workload is stateful and cannot quickly rehydrate.
- ☐ The VM holds unique transactional state.
- ☐ Interactive users require stable sessions.
- ☐ Local disks contain data that cannot be lost.
- ☐ Manual intervention is required for every interruption.
- ☐ Recovery time would break service targets.
- ☐ The workload is latency-sensitive and cannot absorb restart delays.

10) Decision summary

- ☐ Spot is appropriate for this workload.
- ☐ Spot is appropriate only for part of the stack.
- ☐ Spot is not appropriate for this workload.
- ☐ A fallback capacity model has been defined.
- ☐ Owner approval has been recorded.
- ☐ Production rollout is approved only after pilot validation.

Notes



- If the workload treats a VM as an interchangeable execution unit, Spot is often a strong fit.
- If the workload treats a VM as a durable server, Spot is usually a poor fit unless the architecture is redesigned.
- The goal is not only lower compute cost, but lower total cost after interruption, recovery, and data protection are included.