

Cloud Linked High Performance Computing (HPC) Deployment on AWS

Professional Deployment and Industry Relevance Guide

Component	Version
Operating System	Ubuntu 22.04.5 LTS
Workload Manager	SLURM
Parallel File System	BeeGFS
Cluster Manager	Pacemaker, Corosync
Monitoring	ELK Stack, Zabbix
Cloud Platform	AWS EC2

Cloud HPC Overview

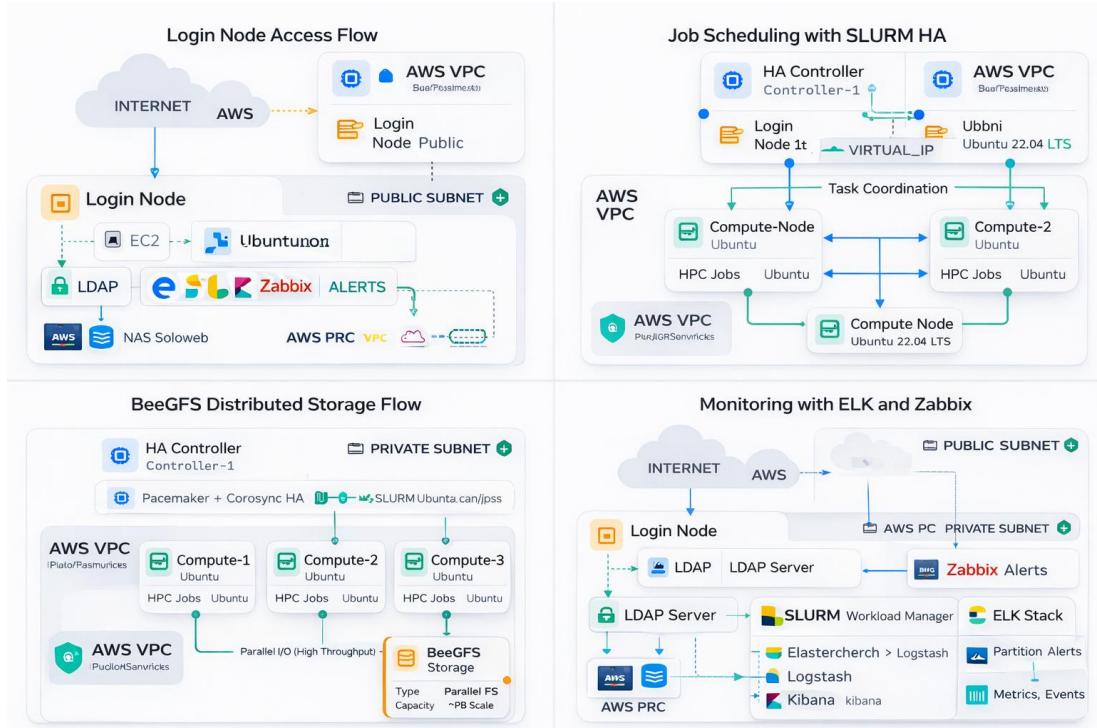
What it is: Cloud HPC is a high-performance computing cluster deployed on cloud infrastructure like AWS instead of physical hardware.

Why it is Important: It provides scalable computing power on demand without requiring physical servers, reducing infrastructure cost and increasing flexibility.

Industry Relevance: Cloud HPC is used in AI training, scientific simulations, financial modeling, weather forecasting, and genome analysis.

Problem it Solves / Disadvantages Removed: Traditional HPC requires expensive hardware, limited scalability, and long deployment time. Cloud HPC removes these limitations.

Real Industry Example: Companies like NVIDIA and OpenAI use cloud HPC clusters to train AI models that require thousands of CPUs and GPUs.



Network Architecture (VPC, Public and Private Subnet)

High Availability using Pacemaker and Corosync

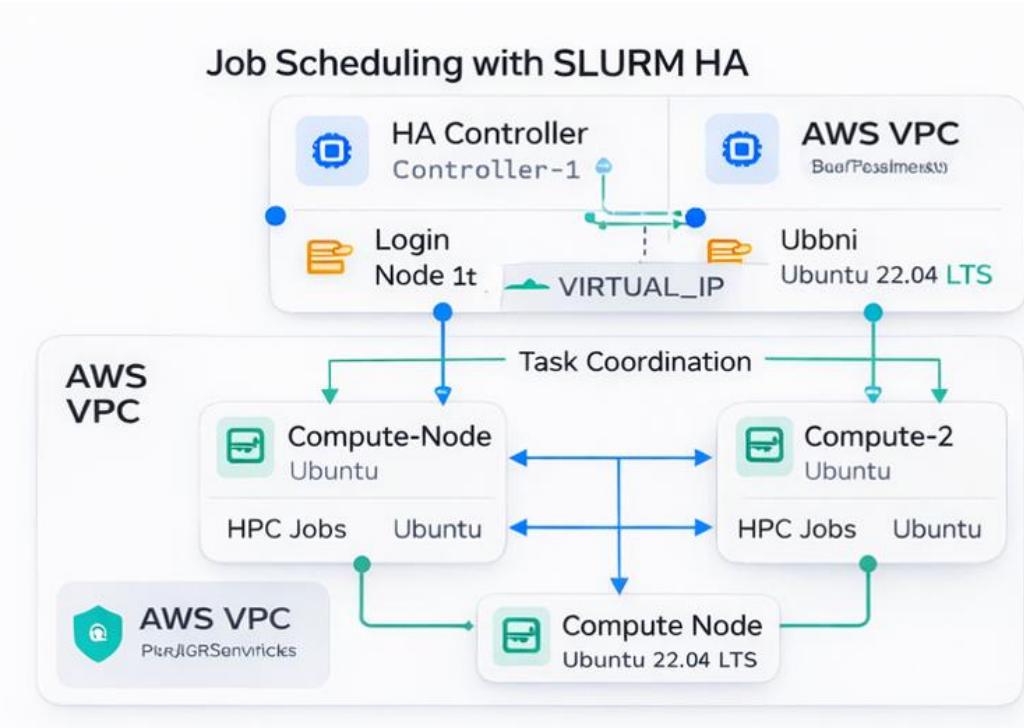
What it is: High availability ensures system continues working even if one controller fails.

Why it is Important: It prevents system downtime and ensures continuous job scheduling.

Industry Relevance: Used in banking systems, cloud platforms, and enterprise servers requiring 24/7 uptime.

Problem it Solves / Disadvantages Removed: Without HA, controller failure would stop all workloads, causing downtime and productivity loss.

Real Industry Example: Stock exchange systems use HA clusters to ensure trading continues even if one server fails.



SLURM Workload Manager

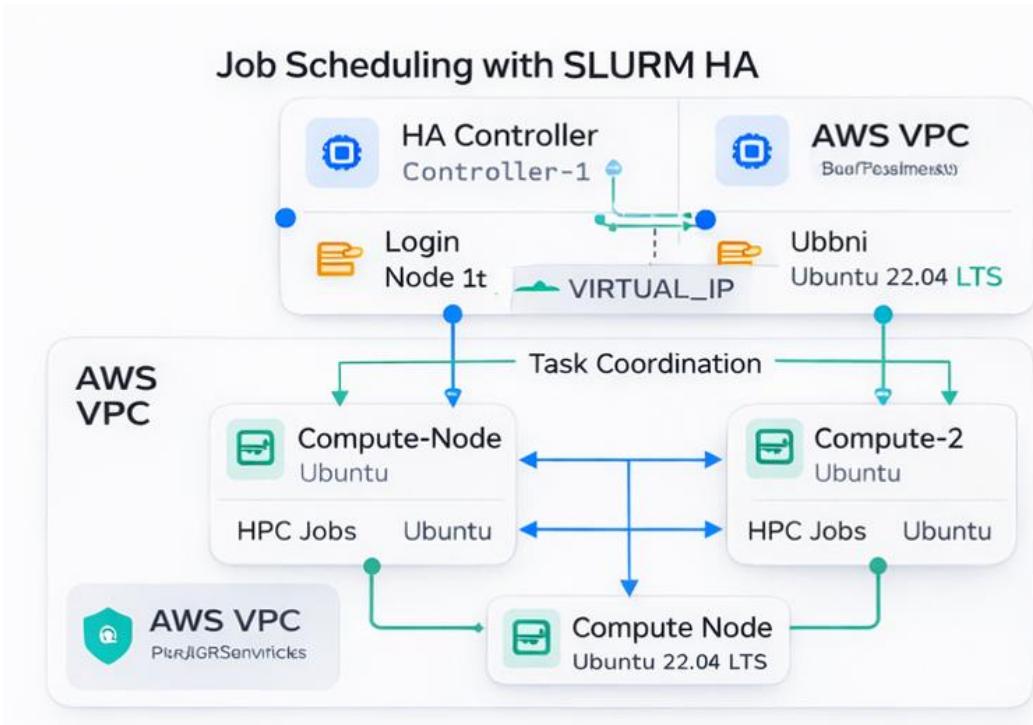
What it is: SLURM schedules jobs across compute nodes efficiently.

Why it is Important: It ensures proper resource utilization and prevents system overload.

Industry Relevance: Used in supercomputers worldwide including NASA, research labs, and AI companies.

Problem it Solves / Disadvantages Removed: Without workload manager, jobs would conflict, overload nodes, and reduce performance.

Real Industry Example: NASA uses SLURM to manage space simulation workloads across thousands of compute nodes.



BeeGFS Parallel Storage

What it is: BeeGFS provides distributed storage accessible by all compute nodes.

Why it is Important: It ensures fast data access required for parallel computing workloads.

Industry Relevance: Used in AI training, video rendering, and scientific computing.

Problem it Solves / Disadvantages Removed: Without parallel storage, storage becomes bottleneck, reducing performance.

Real Industry Example: AI model training requires fast parallel access to training datasets stored in distributed storage.

Monitoring using ELK and Zabbix

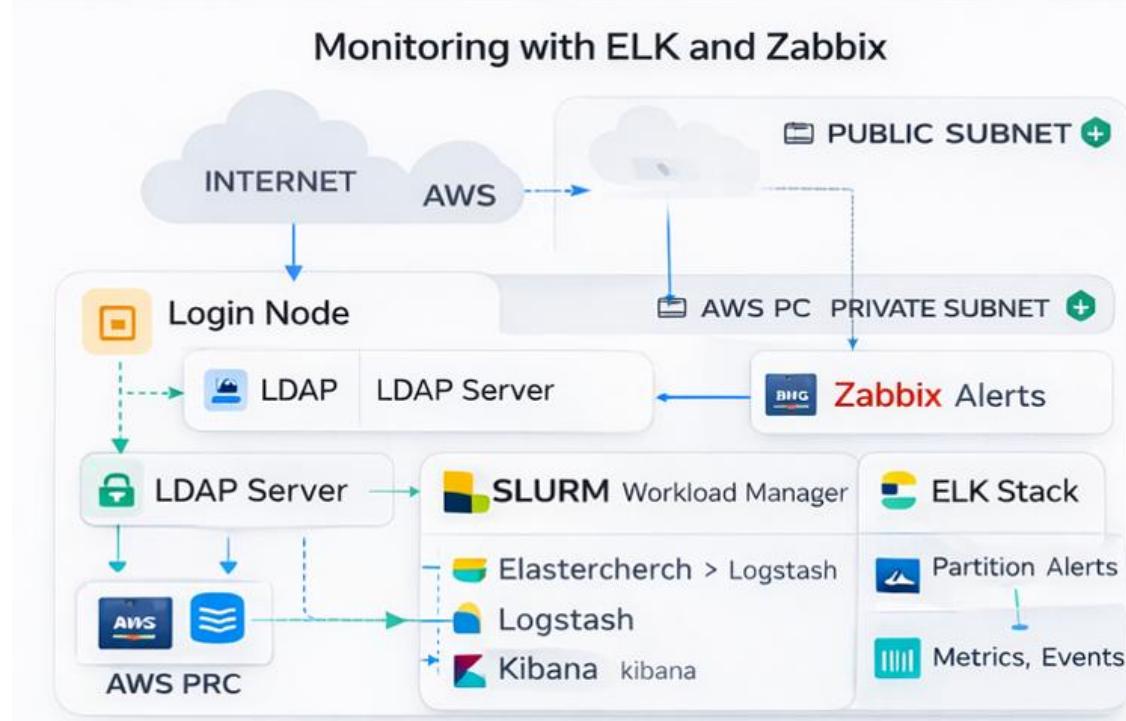
What it is: Monitoring tracks system health, performance, and failures.

Why it is Important: It ensures system reliability and allows early detection of problems.

Industry Relevance: Used in cloud infrastructure, enterprise servers, and production systems.

Problem it Solves / Disadvantages Removed: Without monitoring, failures go undetected causing downtime.

Real Industry Example: Google uses monitoring systems to detect failures instantly in data centers.



Security Implementation

What it is: Security ensures only authorized users can access HPC cluster.

Why it is Important: It protects system from cyber attacks and unauthorized access.

Industry Relevance: Used in enterprise cloud systems and government infrastructure.

Problem it Solves / Disadvantages Removed: Without security, attackers could access or damage compute infrastructure.

Real Industry Example: Financial institutions secure compute clusters for transaction processing.

Conclusion

This project demonstrates industry-grade HPC cluster deployment using cloud infrastructure. It reflects real-world enterprise deployment practices including scalability, high availability, monitoring, security, and automation.

