JUPITER PRIME

Consulting • Training • Advisory

جوبيتر برايم للاستشارات والتدريب

COURSE BROCHURE

Advanced Cluster Computing and Platform Management Training Course

jupiterprime.uk

info@jupiterprime.uk

Registered in the United Kingdom and Saudi Arabia

Advanced Cluster Computing and Platform Management Training Course

Cluster Computing and Platform Management is a 2-week training course for IT managers and technical leads who oversee shared compute platforms, and it ends with each participant building a cluster platform management plan. As simulation, analytics and model training place heavier and less predictable demands on shared nodes, organisations need leaders who can balance throughput, fairness and spend across many user groups. Jupiter Prime Training delivers this course for leaders of cluster computing.

Why Cluster Computing and Platform Management Matters Now

Three forces shape cluster platforms today: accelerator-heavy workloads that change node design and power budgets, research and engineering users who expect queue times measured in minutes, and finance teams that ask what each core-hour returns. A well-run cluster turns scarce nodes into dependable capacity for every group that shares it; a poorly run one breeds idle hardware, long queues, shadow purchases and disputes over priority. The course rests on the IT Infrastructure Library ITIL service view of capacity and demand, together with established parallel performance models such as Amdahl's law and the roofline model.

Who Gains Most from Cluster Computing and Platform Management

- Leads responsible for the architecture, refresh and expansion of shared compute clusters
- Managers who set scheduling policy, fair-share rules and priority for competing user groups
- Technical leads who own parallel application performance and user enablement
- Platform owners accountable for capacity forecasts, utilisation reporting and chargeback
- Heads of research or engineering computing who negotiate service levels with user communities

Platform teams that send an architecture lead, a scheduling owner and a capacity owner together gain most, because they leave with one shared view of the cluster and its trade-offs.

Cluster Computing and Platform Management Capabilities

Participants move from node specifications to decisions about who gets which cycles, when and at what cost.

- Assess node, interconnect and storage options against workload profiles — the hardware refresh decision
- Design scheduling and fair-share policies that balance throughput and fairness — user community leads
- Calculate parallel speed-up and scaling limits for key applications — the core-count allocation decision
- Build capacity forecasts from queue, utilisation and demand signals — the expansion business case
- Specify chargeback or showback models for core-hours and accelerator time — finance partners
- Communicate platform service levels and change plans clearly — senior sponsors and user groups

Key Cluster Computing and Platform Management Terms

- Compute node — a single server in a cluster that runs jobs, with its own processors, memory and often accelerators.
- Interconnect — the high-bandwidth, low-latency network linking nodes so parallel jobs can exchange data quickly.
- Batch scheduler — software that queues submitted jobs and assigns them to nodes according to policy and available resources.
- Fair-share — a scheduling policy that adjusts job priority so each group receives its agreed portion of cluster capacity over time.
- Strong scaling — how run time for a fixed-size problem falls as more processors are added to the job.
- Core-hour — one processor core used for one hour, the common unit for measuring and charging cluster consumption.

Cluster Computing and Platform Management Day by Day

Day 1: The Cluster Landscape and Its Users

- Workload Taxonomy for Simulation, Analytics and Model Training — Platform Scope
- Tightly Coupled Versus High-Throughput Jobs — Architecture Priorities
- User Community Mapping by Group and Need — Service Expectations
- On-Premises, Cloud Burst and Hybrid Cluster Options — Sourcing Direction
- Cluster Maturity Self-Assessment — Gaps the Plan Must Close

Day 2: Node and Accelerator Architecture

- Processor, Memory Bandwidth and Core Count Trade-Offs — Node Specification
- Accelerator Node Design for Parallel Workloads — Partition Mix
- Power, Cooling and Rack Density Constraints — Facility Readiness
- Benchmark Selection Using Representative Applications — Procurement Evaluation Criteria
- Total Cost of Ownership Model for Nodes — Refresh Timing

Day 3: Interconnects and Parallel Storage

- Interconnect Topologies and Bandwidth Planning — Scaling Headroom
- Latency Sensitivity of Message-Passing Applications — Network Investment
- Parallel File System Layout and Metadata Load — Storage Tiering
- Scratch, Project and Archive Storage Policies — Data Lifecycle Rules
- Input and Output Bottleneck Diagnosis — Remediation Priorities

Day 4: Parallel Processing Models

- Message Passing and Shared Memory Models Compared — Application Support Strategy
- Amdahl's Law and Gustafson's Law in Practice — Realistic Job Sizing
- Roofline Model for Compute Versus Memory Limits — Optimisation Focus
- Strong and Weak Scaling Studies — Core Allocation Limits
- Profiling and Performance Counters Review — User Coaching Needs

Day 5: Workload Scheduling Policy

- Queue and Partition Design by Job Class — Wait Time Targets
- Fair-Share Weights and Decay Settings — Group Entitlements
- Backfill and Reservation Strategies — Utilisation Versus Responsiveness
- Pre-Emption and Priority Escalation Rules — Urgent Workload Handling
- Job Size and Wall-Time Limits — Policy Enforcement

Day 6: Capacity Planning and Demand Forecasting

- Utilisation, Queue Depth and Wait Time Metrics — Capacity Health Signals
- Demand Forecasting From Project Pipelines — Expansion Timing
- Queueing Theory Basics for Shared Clusters — Saturation Thresholds
- Cloud Burst Triggers and Thresholds — Peak Demand Response
- Capacity Scenario Modelling — Investment Options

Day 7: Cost Governance and Chargeback

- Cost Pools for Hardware, Energy and Staff — Unit Cost Baseline
- Core-Hour and Accelerator-Hour Pricing Models — Chargeback Design
- Showback Dashboards for User Groups — Consumption Accountability
- Allocation Committees and Proposal Review — Fair Award of Capacity
- Energy Efficiency and Power Capping Measures — Running Cost Control

Day 8: Operations, Reliability and Security

- Node Health Checks and Automated Draining — Availability Targets
- Configuration Management and Image Control — Change Risk
- Checkpoint and Restart Practices — Job Loss Tolerance
- Access Control and Multi-Tenant Isolation — Data Protection Duties
- Incident and Problem Records for Cluster Outages — Service Improvement

Day 9: Service Levels and Stakeholder Leadership

- Service Catalogue for Cluster Offerings — User Expectations
- Service Level Objectives for Wait Time and Uptime — Performance Commitments
- User Training and Onboarding Paths — Efficient Job Submission
- Platform Team Skills Matrix — Staffing and Succession
- Steering Group Reporting Pack — Sponsor Decisions

Day 10: Practising the Cluster Platform Management Plan

- Architecture Roadmap Drawn From Node and Storage Choices — Refresh Sequence
- Scheduling Policy Set Tested on Sample Workloads — Fairness Check
- Capacity and Cost Model Combined — Expansion Business Case
- Peer Challenge of Service Level Proposals — Stakeholder Readiness
- Cluster Platform Management Plan Finalised — Sponsor Approval

Applying Cluster Computing and Platform Management After the Course

- Run a fair-share and queue review using one month of scheduler records, and propose revised weights and partitions to the user community.
- Build a core-hour unit cost baseline and publish a first showback report to the heads of the largest user groups.

Cluster Computing and Platform Management Questions FAQ

What is cluster computing and platform management, and why does it matter now?

It is the leadership of shared compute clusters, covering architecture, scheduling, capacity and cost. It matters now because accelerator workloads, impatient users and closer cost scrutiny make every node and every queue decision more visible and more contested.

Which roles gain most from cluster computing and platform management training?

Platform owners, scheduling policy owners, architecture leads and heads of research or engineering computing gain most. Each sets rules or makes investments that decide how fairly and efficiently many groups share the same nodes.

How does cluster computing and platform management differ from a cluster administration course?

It focuses on decisions rather than commands. An administration course teaches installation and daily operation, while this course covers architecture choices, policy design, capacity forecasting and cost governance that leaders must defend to sponsors and users.

What do you bring back from cluster computing and platform management training?

You bring back a finished plan for your own cluster. It combines an architecture roadmap, a tested scheduling policy, a capacity forecast and a cost model, ready to present to your sponsors and user groups.

Cluster Computing and Platform Management: The Way Forward

The cluster platform management plan first informs the next refresh or expansion decision, giving sponsors evidence on nodes, storage and cost per core-hour. Platform owners, scheduling leads and finance partners use it to agree priorities with user communities. After its first use, the fair-share weights and capacity thresholds deserve a fresh look, since real queue behaviour rarely matches the first forecast.