



The Server Labs G-Cloud 15 Service Description

G-Cloud 15 Service Definition –
AI Infrastructure and Enablement

Ref: TSL/GCLOUD15/SERVICE/AI

Issue Number
1.0

Date
Jan 2026



AI Readiness & Infrastructure Assessment for AWS

Overview

With over **20 years of AWS expertise** and as one of the **first UK AWS Partners**, The Server Labs helps organisations establish whether their cloud environments are genuinely capable of supporting AI workloads in production.

Many organisations explore AI tools and models without first confirming that their platforms can support secure inference services, controlled model access, scalable workloads, and reliable operational ownership. This assessment provides a structured, engineering-led view of how ready the existing AWS environment is for real AI deployment, and what must change to make it viable.

Benefits

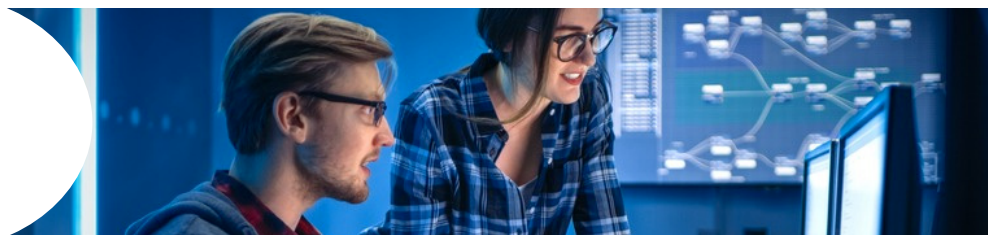
- Clear view of whether your AWS platform can support AI workloads
- Early identification of architectural and security constraints
- Practical guidance on enabling AI execution environments
- Focus on production readiness rather than experimentation
- Suitable for regulated and high-assurance environments

What this service provides

AI adoption most often fails at the platform layer rather than the model layer.

Organisations discover too late that their cloud environments were not designed to support secure inference services, controlled access to models, reliable integration with enterprise systems, or the operational governance required for AI workloads in production.

This service examines how the current AWS environment behaves in practice, focusing on the structural, security, and operational characteristics that determine whether AI services can be deployed safely and sustainably. The assessment looks beyond theoretical capability and concentrates on real-world readiness, including how workloads are deployed, accessed, monitored, and governed. The outcome is a practical, evidence-based view of what must change before AI services can operate reliably in production.



Who is this for

This service is suited to organisations that are considering AI deployment on AWS and need confidence that their existing platforms can support secure, reliable, and well-governed AI workloads.

It is particularly relevant for organisations operating in regulated or high-assurance environments, where AI services must meet the same standards of security, resilience, and operational control as other critical systems.

What's included

- AI infrastructure readiness assessment for inference and MCP workloads
- Review of compute, scaling, and capacity suitability
- MCP platform compatibility analysis
- Identity, access, and network boundary assessment
- Secure model integration pathway review
- CI/CD and deployment readiness evaluation
- Observability and operational maturity review
- Constraint, dependency, and risk mapping
- Practical AI enablement backlog and roadmap

Delivery Approach

The assessment is delivered through a structured, evidence-led engagement that examines how the AWS environment is designed, operated, and governed in practice.

Work begins with a focused review of the existing platform architecture, security model, and operational processes to establish a factual baseline. This is followed by targeted analysis of how AI workloads would interact with the environment, including compute behaviour, access controls, network boundaries, and deployment pathways.

Findings are then consolidated into a prioritised view of constraints and opportunities, with clear recommendations on what must change to support secure and reliable AI execution. The emphasis is on producing an actionable roadmap that internal teams can use to plan and deliver the necessary platform improvements.



Key Deliverables

- AI Infrastructure Readiness Report covering platform, security, and operational capability
- MCP and inference compatibility assessment with identified constraints
- Security and integration risk summary for AI workloads
- Platform maturity baseline for AI execution readiness
- Prioritised AI enablement backlog
- Phased execution roadmap for platform improvements
- Executive summary suitable for technical and leadership audiences

AWS Services & Framework Alignment

This service is aligned with the AWS Well-Architected Framework, with particular focus on the Security, Reliability, and Performance Efficiency pillars as they relate to AI workloads and inference platforms.

Typical AWS services and frameworks referenced during the assessment include:

- AWS Well-Architected Tool
- Amazon EC2 and GPU instance families
- Amazon EKS and ECS
- AWS Identity and Access Management (IAM)
- AWS Identity Center
- Amazon VPC and network security services
- AWS CloudWatch and X-Ray
- AWS Control Tower and Organisations
- AWS CI/CD tooling

The exact services reviewed and referenced will vary depending on the client's existing platform architecture, operating model, and regulatory context.

Pricing & Commercial Model

-Engagement Model: Professional Services

-Pricing: Fixed-price or Time & Materials (depending on complexity)

-Procurement: Available via G-Cloud 15 and AWS Marketplace



AI Platform Architecture & Integration Design for AWS

Overview

With over twenty years of AWS platform engineering experience, The Server Labs designs production-grade AI platform architectures that integrate securely and predictably into complex enterprise environments.

Many organisations reach a point where AI models or services exist, but there is no clear, defensible architecture for how those capabilities should be deployed, secured, integrated, and operated alongside existing applications. This service defines how AI execution platforms should function in practice, ensuring they align with enterprise security models, operational expectations, and long-term platform strategy.

Benefits

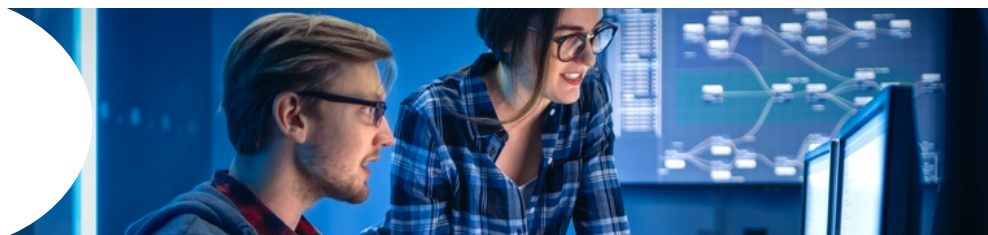
- Clear architectural blueprint for running AI services in production
- Secure integration of inference platforms and MCP services
- Explicit treatment of identity, network, and trust boundaries
- Reduced delivery and operational risk through disciplined design
- Suitable for regulated and high-assurance environments

What this service provides

AI services cannot be introduced as isolated components. They must operate as part of an existing platform ecosystem, subject to the same security controls, governance structures, and operational disciplines as other production services.

This service focuses on defining how AI execution platforms should be structured and integrated within the client's AWS environment. It addresses inference service topology, MCP design, identity and access boundaries, network segmentation, application integration, and operational ownership. The emphasis is on architectures that support sustained production use, rather than experimentation or proof-of-concept deployments.

Design decisions are grounded in how systems are actually built, deployed, and operated, with attention given to resilience, scaling behaviour, observability, and long-term maintainability. **4**



Who is this for

This service is suited to organisations that have confirmed AI platform readiness and now require a clear architectural foundation before committing to implementation.

It is particularly relevant for organisations operating in regulated, mission-critical, or high-assurance environments where AI services must integrate safely with existing platforms, data flows, and operational models.

What's included

- AI execution platform architecture covering inference and MCP services
- MCP topology design reflecting trust zones and service separation
- Secure integration patterns for enterprise applications
- Identity and access architecture aligned to security models
- Network segmentation and traffic control design
- Deployment and release architecture aligned to existing CI/CD practices
- Observability and operational control design for AI services
- Architecture governance artefacts for assurance and delivery

Delivery Approach

The engagement is delivered through a structured architectural design process focused on translating business, security, and operational requirements into an implementable AI platform architecture.

Work begins by establishing architectural constraints and integration requirements based on the existing AWS environment and operating model. The AI execution platform design is then developed with explicit consideration of service boundaries, trust relationships, deployment pathways, and operational ownership. Designs are refined through review and validation to ensure they are practical, defensible, and ready for implementation by internal teams or delivery partners.



Key Deliverables

- AI Execution Platform Architecture Pack describing deployment, integration, and operation
- MCP and inference topology design with defined trust and ownership boundaries
- Secure integration architecture for identity, networking, and application access
- Deployment and release model aligned to existing operational practices
- Observability blueprint supporting monitoring, reliability, and supportability
- Architecture governance summary suitable for internal approval and delivery planning

AWS Services & Framework Alignment

This service is aligned with the AWS Well-Architected Framework, with particular emphasis on the Security, Reliability, and Performance Efficiency pillars as they apply to AI execution platforms

Architectural designs typically reference the following AWS services and frameworks:

- AWS Well-Architected Framework and Tool
- Amazon EKS and Amazon ECS
- Amazon EC2, including GPU instance families
- AWS Identity and Access Management (IAM)
- AWS Identity Center
- Amazon VPC and network security services
- Amazon API Gateway and Application Load Balancers
- Amazon CloudWatch and AWS X-Ray
- AWS CI/CD tooling

The specific services and patterns used will vary depending on the client's existing AWS architecture, operational maturity, and regulatory context.

Pricing & Commercial Model

- Engagement Model:** Professional Services
- Pricing:** Fixed-price or Time & Materials (depending on complexity)
- Procurement:** Available via G-Cloud 15 and AWS Marketplace



AI Infrastructure Implementation & Enablement for AWS

Overview

With over twenty years of AWS platform engineering experience, The Server Labs builds secure, production-ready AI execution environments that operate reliably within complex enterprise estates.

Many organisations reach a point where AI platform designs exist, but the underlying infrastructure has not yet been implemented in a way that supports secure integration, operational ownership, and long-term stability. This service delivers the practical engineering required to turn approved AI platform architectures into functioning, well-governed production environments.

Benefits

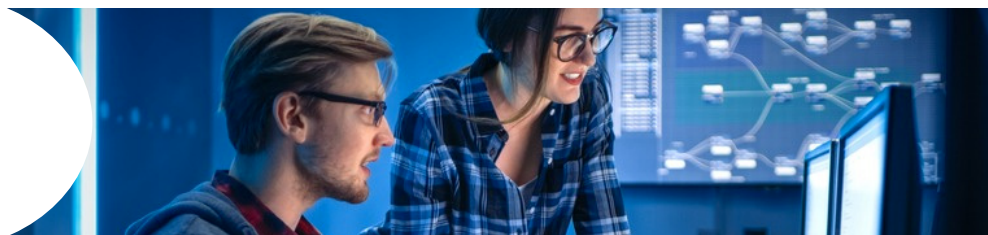
- Delivers a working AI execution platform on AWS
- Implements MCP and inference services securely
- Integrates AI services with existing environments
- Embeds identity, network, and access controls
- Supports operational handover and sustainability

What this service provides

An AI platform only becomes useful when it is built, secured, integrated, and operated as part of the wider enterprise environment.

This service focuses on implementing the agreed AI execution platform architecture within the client's AWS estate. It covers the deployment of MCP components, inference services, and supporting infrastructure, with careful attention to identity management, network segmentation, secure service exposure, and integration with existing applications.

The work is carried out in a way that aligns with established delivery pipelines and operational practices, ensuring that AI services are introduced as a natural extension of the existing platform rather than as a separate or experimental capability. The emphasis is on reliability, security, and operational readiness from the outset.



Who is this for

This service is suited to organisations that have approved AI platform designs and now require practical engineering delivery to establish secure, production-ready AI infrastructure.

It is particularly relevant for organisations operating in regulated, mission-critical, or high-assurance environments where AI services must meet the same standards of security, reliability, and operational control as other production systems.

What's included

- Deployment of AI execution platform components
- Implementation of MCP and inference services
- Identity and access control configuration
- Network segmentation and security controls
- Integration with existing enterprise applications
- CI/CD alignment for AI service deployment
- Observability and monitoring configuration
- Operational runbook development
- Knowledge transfer to internal teams

Delivery Approach

The engagement is delivered through hands-on engineering work focused on implementing the approved AI platform architecture within the client's AWS environment.

Work begins with the establishment of the core AI execution infrastructure, followed by the configuration of security controls, identity boundaries, and network segmentation. Integration with existing applications and delivery pipelines is then completed to ensure AI services can be deployed and operated in line with established practices. Documentation, runbooks, and knowledge transfer are provided so that internal teams can confidently support and evolve the platform



Key Deliverables

- Deployed AI execution platform with MCP and inference services
- Secure identity, access, and network configurations
- Integrated AI service interfaces for enterprise applications
- CI/CD enablement for AI platform components
- Observability dashboards and monitoring configuration
- Operational runbooks for support and maintenance
- Handover and enablement pack for internal teams

AWS Services & Framework Alignment

This service is aligned with the AWS Well-Architected Framework, with particular emphasis on the Security, Reliability, and Operational Excellence pillars as they apply to AI execution platforms.

Implementation work typically references the following AWS services and frameworks:

- AWS Well-Architected Framework and Tool
- Amazon EKS and Amazon ECS
- Amazon EC2, including GPU instance families
- AWS Identity and Access Management (IAM)
- AWS Identity Center
- Amazon VPC and network security services
- Amazon API Gateway and Application Load Balancers
- Amazon CloudWatch
- AWS CI/CD tooling

The specific services and patterns used will vary depending on the client's existing AWS architecture, operational maturity, and regulatory context.

Pricing & Commercial Model

- Engagement Model:** Professional Services
- Pricing:** Fixed-price or Time & Materials (depending on complexity)
- Procurement:** Available via G-Cloud 15 and AWS Marketplace



AI Platform Optimisation & Managed Enablement for AWS

Overview

With over twenty years of AWS platform engineering experience, The Server Labs helps organisations keep AI execution platforms reliable, performant, and operationally sustainable over time.

As AI services move into regular use, workload demand increases, integrations expand, and operational complexity grows. This service provides ongoing architectural and engineering oversight to ensure AI platforms continue to operate securely, predictably, and at scale within enterprise environments.

Benefits

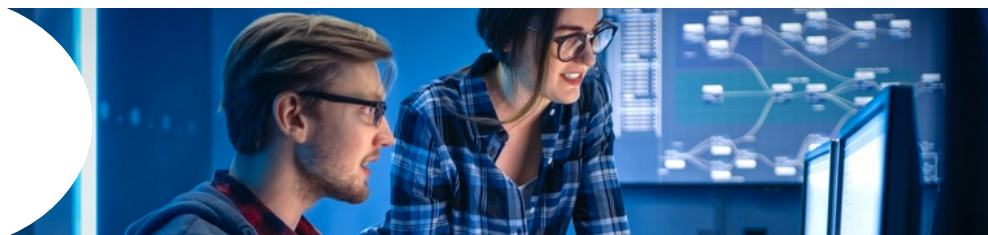
- Sustains performance and reliability of AI platforms
- Strengthens operational stability as usage grows
- Improves scaling behaviour and resilience
- Prevents security and configuration drift
- Supports long-term platform sustainability

What this service provides

AI platforms do not remain stable by default. As inference workloads increase, new integrations are introduced, and operational teams change, performance, security, and reliability can gradually degrade.

This service focuses on maintaining the quality of the AI execution platform over time. It provides architectural oversight, performance tuning, reliability improvements, and operational refinement to ensure that AI services continue to meet enterprise standards for security, availability, and control.

The emphasis is on sustaining production readiness and preventing avoidable degradation, rather than reacting to issues after they occur.



Who is this for

This service is suited to organisations that already operate AI execution platforms on AWS and require continued assurance that those platforms remain secure, reliable, and operationally effective.

It is particularly relevant for organisations operating in regulated, mission-critical, or high-assurance environments where AI services must consistently meet defined standards for availability, security, and operational governance.

What's Included

- Regular AI platform health and performance reviews
- Scaling and capacity optimisation guidance
- Reliability and resilience improvement recommendations
- Security posture and configuration drift checks
- Observability and monitoring refinement
- Operational runbook updates
- Ongoing architectural advisory support

Delivery Approach

The engagement is delivered through ongoing architectural and engineering oversight focused on how the AI platform behaves in real operational conditions.

Work begins with a structured review of platform performance, reliability, and security posture, followed by targeted recommendations to address emerging risks or constraints. Improvements are introduced in a controlled manner, aligned to existing operational practices, with continued advisory support to help internal teams maintain platform quality as demand and complexity evolve.



Key Deliverables

- AI platform health and performance assessment reports
- Scaling and optimisation recommendations for inference workloads
- Reliability and resilience improvement guidance
- Security and configuration drift review summaries
- Updated operational runbooks and support guidance
- Executive-level platform assurance summaries

AWS Services & Framework Alignment

This service is aligned with the AWS Well-Architected Framework, with particular emphasis on the Reliability, Security, and Operational Excellence pillars as they apply to AI execution platforms.

Ongoing optimisation and assurance activities typically reference the following AWS services and frameworks:

- AWS Well-Architected Framework and Tool
- Amazon CloudWatch
- Amazon EC2, including GPU instance families
- Amazon EKS and Amazon ECS
- AWS Identity and Access Management (IAM)
- Amazon VPC and network security services
- AWS Config and Security Hub

The specific services and optimisation focus will vary depending on the client's existing AWS architecture, operational maturity, and regulatory context.

Pricing & Commercial Model

- Engagement Model:** Professional Services
- Pricing:** Fixed-price or Time & Materials (depending on complexity)
- Procurement:** Available via G-Cloud 15 and AWS Marketplace



TSL Company Profile

Founded in 2004, TSL are cloud architects and integrators.

We offer a full range of lifecycle cloud solutions and services, from consultancy, software/application development, to cloud architecture design, and integration, to fully managed services.

Our specialisations are High Performance Computing and AI/ML deployed in highly secure cloud environments, with best practice built in from day one.

Our clients include private and public organisations, with an emphasis on those processing large amounts of data in mission critical environments, particularly in the Lifesciences, Utilities, Banking, Space, Metrology and Telecoms sectors.



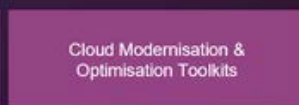
The Server Labs 3 Layer Service Map



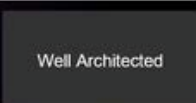
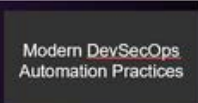
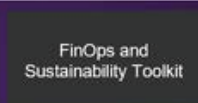
AI Readiness

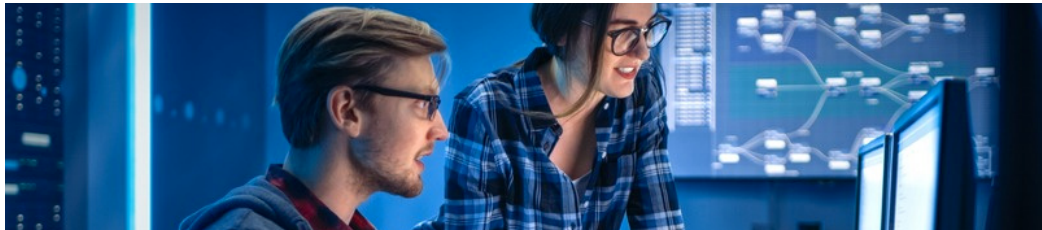


Cloud Maturity Accelerators



Trusted Foundations





The Company's mission is:

- To provide expert services in the field of cloud architectures, High Performance Computing, and advanced cloud software engineering
- To radically improve cloud software development processes
- To help organisations achieve better business results through the correct use of cloud technologies
- To create high quality innovative cloud software solutions, providing added value to our customers

The specific value, experience and expertise that The Server Labs can provide are:

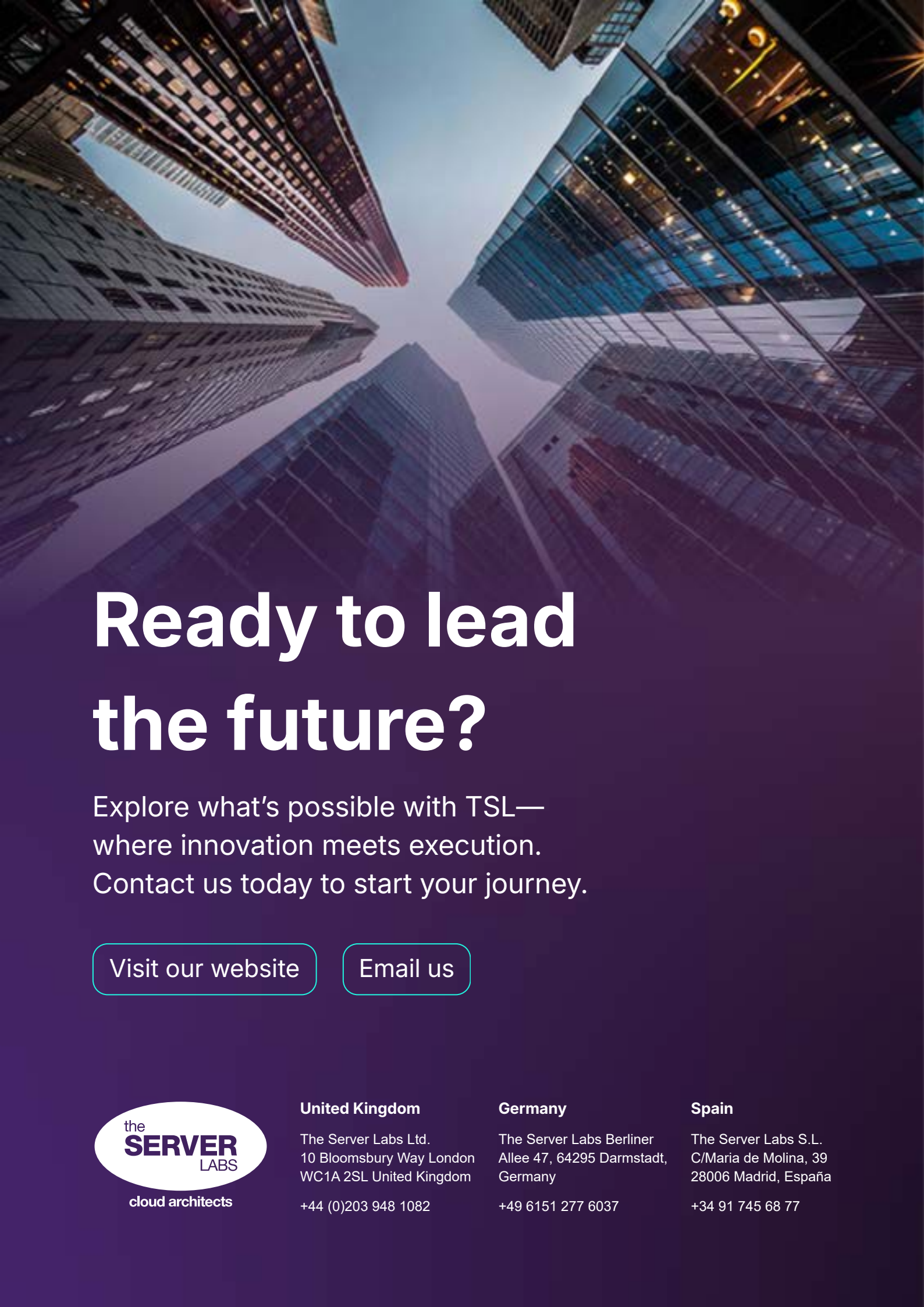
1. **Technical excellence** and capability to act as lead on architectural decisions and as technology expert in software and system subjects.
2. **Proven experience** in implementing HPC, FinOps, Cloud Security and Big Data Projects.
 - FinOps sustainable change management to enable innovation
 - Industry leading experts in high performance computing
 - Cloud security and sensitive, big data management and processing
3. **Cloud migration:** Proven track record in migrating large, complex and sensitive dataset from on-premise to cloud environments and from legacy to target cloud.
4. **Architecture experience** at software and system level.

AWS MAP and WAF certified

TSL have achieved AWS MAP competence, which is a rigorous and demanding standard to achieve and few boutique consultancies like TSL have achieved this.

Our architects and engineers are experts with an average of 10 years' experience in the planning, design and development of complex software systems. Our multinational team has been a pioneer in cloud native service, serverless computing, automation, infrastructure as code, and distributed architectures, and has the required hands-on experience in many state of the art technologies.

The Server Labs is a leader in cloud computing solutions and services, helping organisations to move to the cloud at all levels.



Ready to lead the future?

Explore what's possible with TSL—
where innovation meets execution.
Contact us today to start your journey.

[Visit our website](#)

[Email us](#)



United Kingdom

The Server Labs Ltd.
10 Bloomsbury Way London
WC1A 2SL United Kingdom

+44 (0)203 948 1082

Germany

The Server Labs Berliner
Allee 47, 64295 Darmstadt,
Germany

+49 6151 277 6037

Spain

The Server Labs S.L.
C/María de Molina, 39
28006 Madrid, España

+34 91 745 68 77