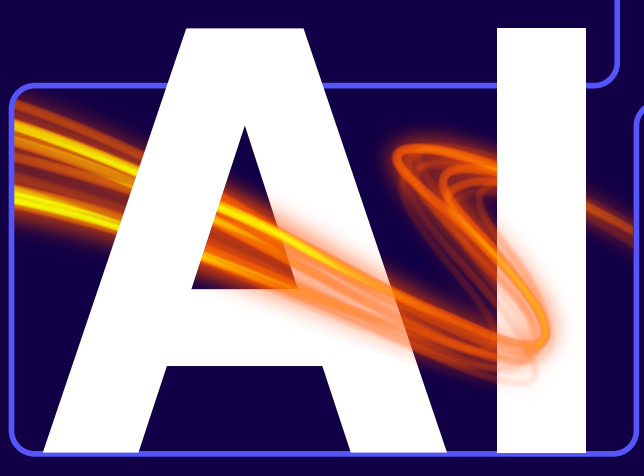
CLOUDERA

The New Rules of Scale: How Enterprises Are Adapting to AI's Operational Reality

AI is rewriting the rules of enterprise operations. As organizations move workloads into production and expand AI's role across the business, they face the realities of scale: rising costs, production risk, governance requirements, and changing workload needs.

Now, business and IT leaders must sustain AI at scale while maintaining control over both their data and the flexibility to adapt as requirements change.



Scale Is Changing the Economics of AI

As AI usage grows, cost is beginning to influence which use cases remain viable and where workloads run. Enterprises are already redesigning applications and changing workload placement in response.

89%

of organizations have limited or redesigned AI use cases because of rising token consumption.

How enterprises are responding to higher or more volatile AI costs

45%
moved a workload to another cloud, private environment, or on-premises platform

40%
reduced processing frequency, model size, context size, or output volume

24%
paused or canceled an AI use case

Production Is Where AI is Put to the Test

Once AI reaches production, enterprises get a clearer picture of where it delivers and where it needs adjustment. Pullbacks are common, but they are happening alongside meaningful productivity gains.

98%

of organizations have pulled back an AI deployment because of failure, hallucinations, or unexpected behavior.

But despite pullbacks, enterprises are also starting to see some productivity gains. However, rising costs continue to influence how AI's value is measured.

43%
report greater-than-expected productivity

23%
report lower-than-expected productivity

41%
encountered higher-than-expected costs

As AI Grows, Control Becomes More Complex

Scaling AI also means extending governance across a larger and more complex footprint. While enterprises report confidence in applying controls, governance remains the most frequently cited constraint on scale and its costs are growing.

75%

of organizations said governance and compliance is a constraint on scaling AI.

How enterprises are responding to higher or more volatile AI costs

54%
say they have struck an effective balance between AI adoption and appropriate governance

23%
say they are moving too cautiously

23%
say AI adoption is outpacing governance and controls

Where AI Runs Can't Be a Fixed Decision

Scaling AI means operating in an environment where production needs, governance requirements, and costs are constantly changing. That stresses increasing importance on enterprises' ability to place and move AI workloads according to their requirements.

84%

of organizations say data residency, sovereignty, or jurisdiction-specific requirements influence where AI workloads are placed.

Why AI needs more than one environment

42%
Resilience

38%
Sovereignty and residency requirements

38%
Cost reduction

37%
Proximity to data

58%

of organizations say relocating an AI workload requires moderate or major redesign.

What's Next

Scaling AI has uncovered gaps that leave many rethinking their approach. If they could start over, **49%** would strengthen governance from the beginning, **46%** would invest in data quality earlier, and **45%** would modernize infrastructure sooner.

As models, economics, and workload requirements continue to change, enterprises need the flexibility to adjust where and how AI runs while maintaining control.

Read the Full Report

Contact us today for more information about **Cloudera**
US: +1 888 789 1488 | Outside the US: +1 650 362 0488

CLOUDERA