SAOUG
CONFERENCE 2026

FROM CLOUD TO COMPETITIVE EDGE

GETTING MORE BUSINESS VALUE FROM OCI

BRETT PRIMO

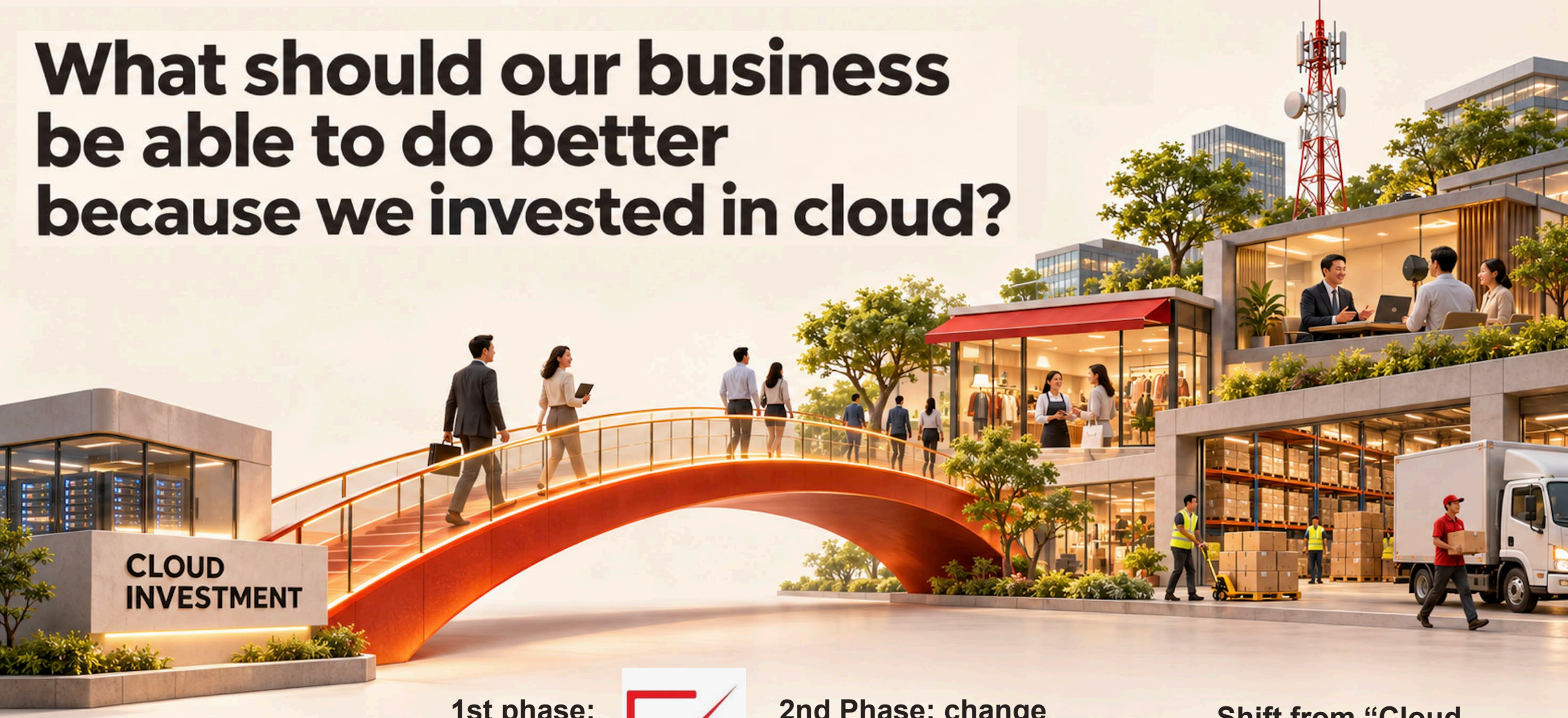
CLOUD ENGINEERING LEADER, ZA

HELLO AFRICA



UBUNTU
moving forward
TOGETHER.

What should our business be able to do better because we invested in cloud?



Cloud is no longer the competitive advantage

1st phase: Cloud Migration

2nd Phase: change "HOW" business performs

Shift from "Cloud Adoption" to **Business Performance**

From Cloud to Competitive Edge

SAOUG CONFERENCE 2026

Cloud investment is not the same as business value

Many organizations have modernized the platform—but still measure success in technical terms.

What is typically measured

- Workloads migrated
- Compute and storage consumed
- Availability and uptime
- Infrastructure cost per month

What executive KPI's measure

- Revenue and margin growth
- Cost-to-serve
- Faster time-to-market
- Customer and citizen outcomes

The missing link

A clear line of sight from cloud capability to business KPIs—with owners, baselines and measurable benefits.

A cloud implementation program is successful when the business is measurably better – not simply when the migration is complete.

SAOUG CONFERENCE 2026

What makes OCI into a Business Performance Enabler?

01	**Infrastructure Choice → Outcomes**	Revenue, service quality, agility and time-to-market
02	**Cloud Adoption → AI-powered work**	Connect enterprise data, models, applications and agents
03	**Lock-in → cloud choice**	Hybrid, multicloud and workload economics
04	**Security Controls → Trusted AI**	Identity, data access, governance and sovereignty
05	**Cloud Spend → Return on Investment**	Optimize total value - not just unit cost

SAOUG CONFERENCE 2026

Infrastructure Choice: The right workload deserves the right economics

OCI is expanding compute and networking options for demanding AI workloads.

NVIDIA Vera Rubin

Vera Rubin NVL72 systems and Vera CPUs announced as generally available on OCI.

Google TPUs

Oracle and Google plan to bring TPUs to OCI starting in 2027—adding another accelerator architecture.

AMD, Intel & networking

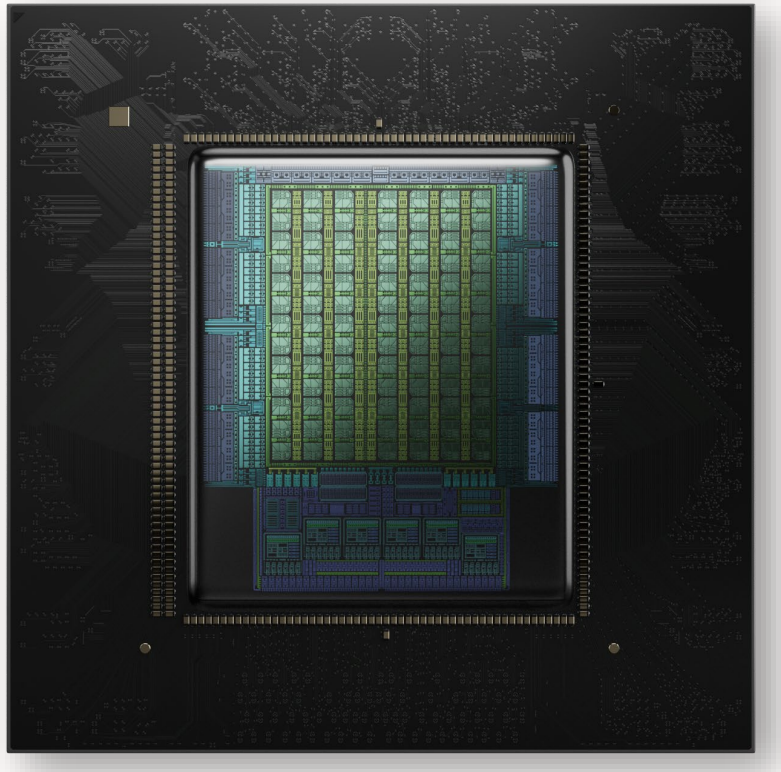
Continued expansion of CPU options, accelerated networking and high-performance infrastructure.

The business question: Which architecture delivers the best cost, throughput, latency and energy profile for this workload?

Important: Google TPU availability on OCI was announced for 2027, not as a generally available service at AI World.

The Solution: NVIDIA Vera CPU

OCI



Unmatched Single Thread Speed: Up to 80% faster with 1.8× higher single-thread performance.

Energy Efficiency: Up to 2× better performance per watt for lower infra TCO.

Faster AI Training: Up to 10% faster time-to-quality through faster RL evaluations.

High-Performance Custom Olympus Cores: 88 custom Arm cores optimized for low-latency execution.

Spatial Multithreading: 176 threads with predictable performance under load.

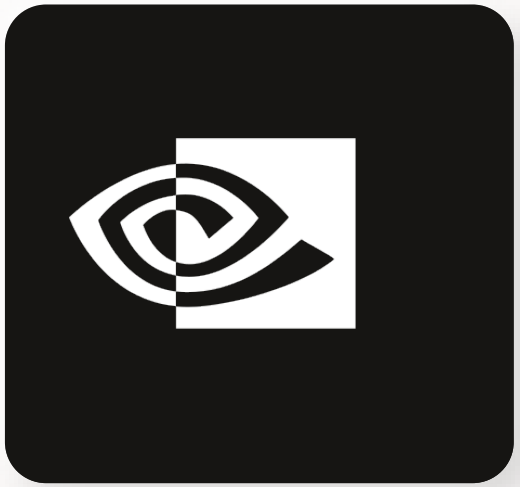
High LPDDR5X Memory Bandwidth: Up to 1.2 TB/s bandwidth with 2× bandwidth at half the memory power.

Fast On-Chip Fabric Fabric: 3.4 TB/s on-chip fabric for low-latency data movement.

Confidential – © 2026 Oracle Restricted

Vera CPU Performance Benchmarks

OCI



1.61x

Vera vs. Grace
overall geomean:
Phoronix

1.54x

Vera vs. Xeon 6980P
overall geomean:
Phoronix

1.30x

Vera vs. AMD Zen 5
baseline:
Prime Intellect

Confidential – © 2026 Oracle Restricted

SAOUG
CONFERENCE 2026

Recommended Vera Workloads

01
Agentic Orchestration & Tool Serving

02
Sandbox Execution & RL Environments

OCI

03
GPU-Fleet Control Plan

04
Vector Search & Retrieval

05
In-Memory Databases & Analytics

06
Operational Apps & General Services

Confidential – © 2026 Oracle Restricted

OpenAI

"NVIDIA Vera's high memory bandwidth and per-core performance on OCI make it a promising platform for emerging agentic AI workloads."

Uday Ruddarraju
CTO of Compute, OpenAI

Oracle Brings Google TPUs to OCI

OCI will be the first Hyperscaler outside Google to offer Google TPUs

Why It Matters:

- Giving customers more choice to match AI training and inference infrastructure to their workload, performance requirements and cost targets
- Choose TPUs based on model architecture, memory needs, software compatibility, performance and cost
- Build and serve demanding AI models on OCI, combining TPU architecture with infrastructure designed for large deployments



Oracle Acceleron:
Networking designed to run demanding AI workloads

New Acceleron capabilities improve performance and reliability to support demanding AI infrastructure requirements

Why It Matters:

- **Acceleron High-Performance Fabric** - Extends RDMA connectivity to OCI front-end compute to support inference, storage, and HPC workloads, enabling more efficient data movement and greater performance
- **Front-End Multiplanar Networking** - Introduces independent network planes that provide transparent path selection, automatic failover, and congestion control for front-end RDMA traffic, helping customers maintain responsive, resilient AI infrastructure
- **Collective Communications Library (OCCL)** - Distributes collective traffic across OCI's backend Multiplanar Network based on congestion and path health, enabling instantaneous recovery from network failures and helping keep AI workloads running



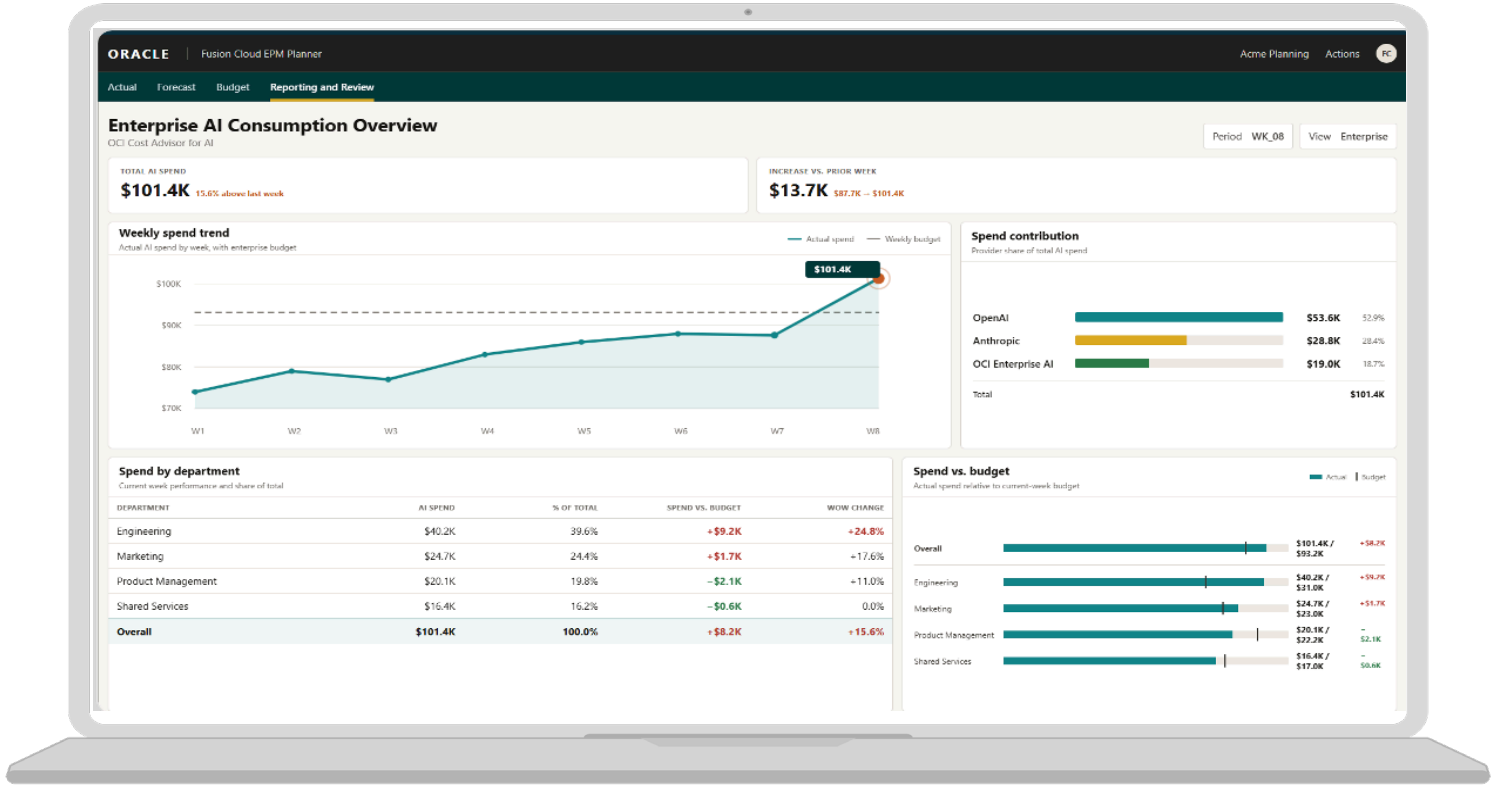
SAOUG CONFERENCE 2026

Oracle Cost Advisor for AI

New Oracle Cost Advisor for AI helps companies see all their AI spend, connect it to budgets and owners, and get recommendations on how to reduce costs

Why It Matters:

- Consolidates cost and usage data across OCI Enterprise AI, OpenAI, and Anthropic, all within existing Fusion EPM workflows.
- Sets budgets and approvals by cost center, model, or tool; flags anomalies, waste, and unmapped spend.
- Compares costs by application, workflow, or owner to identify cost drivers and where closer review is needed.
- Recommends lower-cost model alternatives for comparable tasks, with estimated savings.



SAOUG CONFERENCE 2026

The Shift: Cloud Adoption → AI-powered work

The strongest announcements are not isolated features—they change how people, data and systems work together.

OpenAI collaboration

AI Workspace, App Composer and AI Security Mesh bring frontier AI closer to enterprise workflows.

Deep Trust AI Architecture

Trust, privacy, security and business rules embedded across the enterprise stack.

More infrastructure & data choice

Optionally manage your entire AI Agent Ecosystem within OCI – deploying the relevant components you see fit, across vendors.

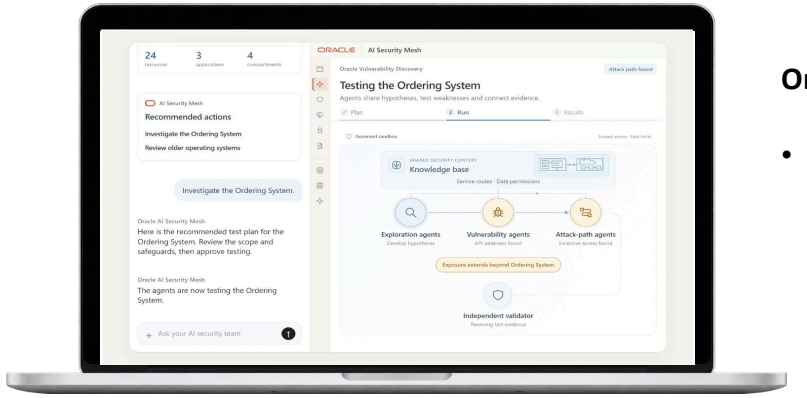
The strategic direction: make AI useful in real business processes—with enterprise context, control and choice.

Announcements: 26–27 October 2026 | Check availability and regional rollout before committing to delivery dates.

SAOUG CONFERENCE 2026

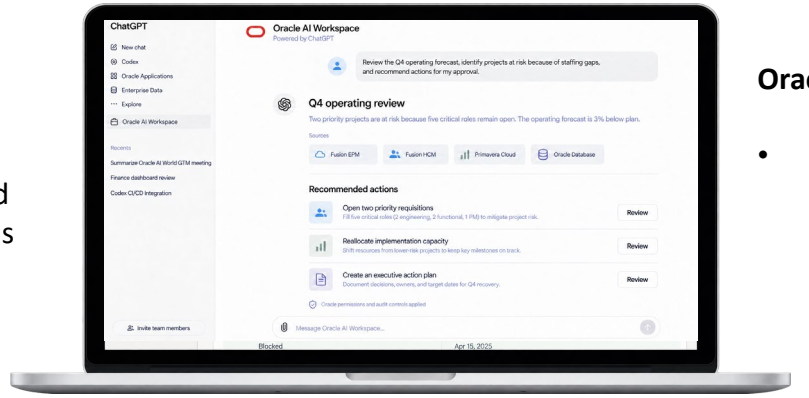
Oracle | OpenAI: Put frontier AI to work for your business

Oracle and OpenAI are bringing ChatGPT, Codex, and advanced AI models together with Oracle's applications, data, and cloud infrastructure.



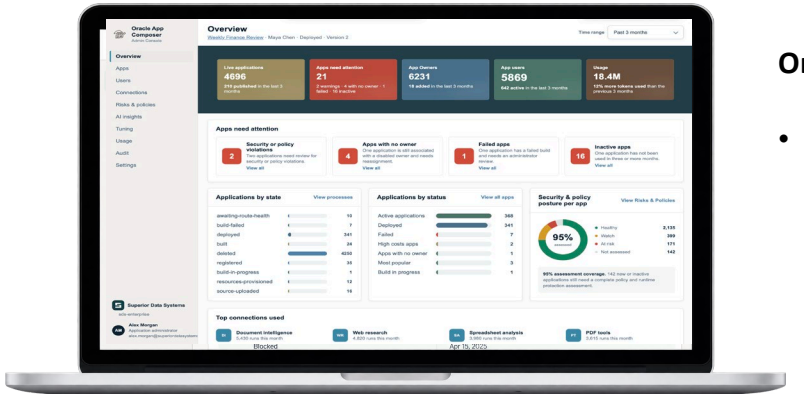
Oracle AI Security Mesh

- Harness advanced AI models for security. Your teams can find and fix vulnerabilities with safeguards that keep you in control.



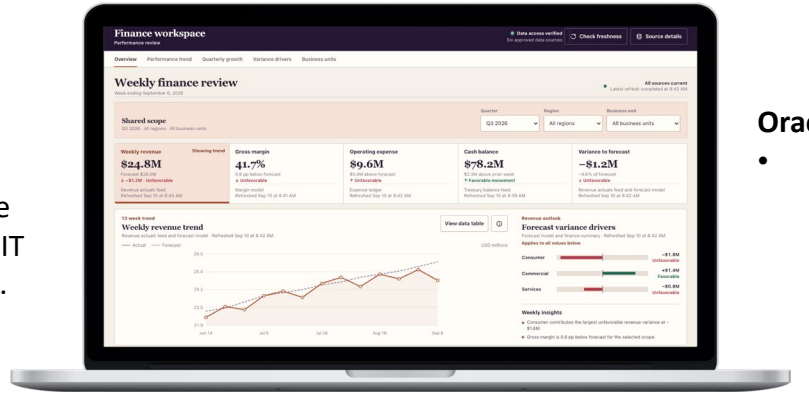
Oracle AI Workspace

- Connect apps, data, and AI through a single interface to find information and take approved actions.



Oracle App Composer

- Vibe code apps with AI, then share easily across teams. Oracle manages OCI deployment while IT oversees access, data, and costs.



Oracle AI SMB Bundle

- A free-to-start package for eligible, ***approved*** businesses, including OCI credits, NetSuite QuickStart, and Oracle AI tools.

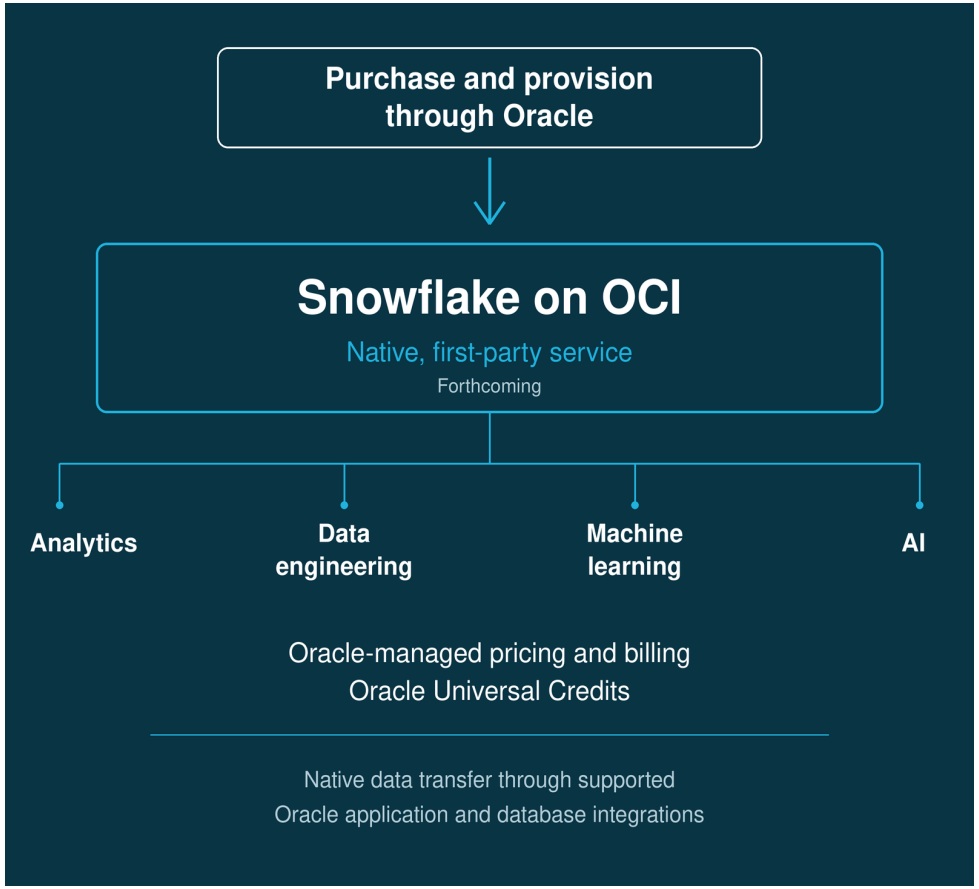
SAOUG CONFERENCE 2026

Oracle and Snowflake Announce a Native, First-Party OCI Service

Snowflake AI Data Cloud will be available as a native, first-party service on OCI, bringing analytics and AI closer to enterprise data. Customers will be able to purchase, provision, and consume the service directly through Oracle and use Oracle Universal Credits.

Why It Matters:

- Simplifies adoption and accelerates AI: native data transfer through supported Oracle integrations reduces the need for custom pipelines, helping customers use enterprise data for analytics and AI faster.
- Unifies the buying experience: Oracle-managed pricing, discounts, and billing provide a single commercial relationship, while Oracle Universal Credits let customers use existing cloud commitments.
- Maintains data security and customer control: customers can use Snowflake on OCI alongside their enterprise data while maintaining control over data access, use, and governance.



Use Cases: enterprise analytics and reporting, data engineering, machine learning and AI, and enterprise data integration through supported Oracle application and database integrations.

Mining: from running workloads to improving production

Illustrative use case—connect OCI capabilities to operational KPIs.



TRADITIONAL CLOUD MEASURES

- Workloads migrated
- Infrastructure availability
- Storage and compute consumed

COMPETITIVE-EDGE OUTCOMES

- Predict equipment failures
- Reduce unplanned downtime
- Optimise production and asset utilisation
- Improve safety and operating margin

The mine doesn't buy compute. It wants more production, less downtime and safer operations.

SAOUG CONFERENCE 2026

Four levers of OCI business value

Map every capability and investment to one or more measurable business outcomes.

GROW REVENUE

New digital services
Better customer experience
Faster innovation

IMPROVE EFFICIENCY

Automation & AI productivity
Lower cost-to-serve
Operational optimisation

REDUCE RISK

Security & resilience
Governance & compliance
Data sovereignty

OPTIMISE INVESTMENT

Consumption efficiency
Workload economics
Improved ROI

Competitive advantage = the combined effect of all four levers.

Replace cloud activity metrics with business value metrics

Define a baseline, target, accountable owner and review cadence.

REVENUE	Digital revenue contribution • conversion • time-to-market
EFFICIENCY	Cost-to-serve • process cycle time • hours returned to teams
RISK	Recovery time • control exceptions • security exposure
INVESTMENT	Cost per transaction • utilization • unit economics • forecast accuracy

Start with a visible base…

No baseline = no credible improvement claim. No business owner = no sustained value.

SAOUG CONFERENCE 2026

A 90-day value acceleration plan

Start with one business priority—not a broad technology transformation.

DAYS 1–30

DISCOVER

Identify 2–3 value pools
Baseline KPIs and costs
Select executive sponsor

DAYS 31–60

DESIGN

Map data and workload needs
Validate security and governance
Prioritise use cases by value

DAYS 61–90

PROVE

Launch one measurable pilot
Track adoption and outcomes
Agree scale / stop decision

Success at day 90: a validated business case, a working proof point and a decision backed by evidence.