Confidential - Oracle Restricted \Employees Only

From HPC to AI: An open compute fabric

Performance, openness and portability for financial services

James Calise
Oracle Cloud Infrastructure

ORACLE

Confidential - Oracle Restricted \Employees Only

Confidential - Oracle Restricted \Employees Only

ORACLE + FINOS

Oracle's progression to open compute

A familiar perception: "closed source"

The engineering story is broader

COMPUTE EVOLUTION

AI's roots: Dartmouth, 1956

1960s–70s

HPC takes shape

CDC 6600 (1964)
Cray-1 (1976)

1990s–2000s

Grid computing

Shared resources
across organizations

2012

Deep learning (AI)

AlexNet demonstrates
GPU-powered learning

2022

Generative AI

ChatGPT brings
AI to a broad audience

ORACLE'S RESPONSE

1977

Oracle begins

Ellison, Miner, Oates
Three engineers

1998–2010

Open source

1998: commitment
2010: Sun joins

2018

Gen 2 cloud

Architecture reset
HPC foundations

2026

FSI + FINOS

Compute + AI access
AI Governance

NOW → 2028

Compute squeeze

AI growth + power limits
325–580 TWh / year*

Common access and governance across compute and AI services

ORACLE

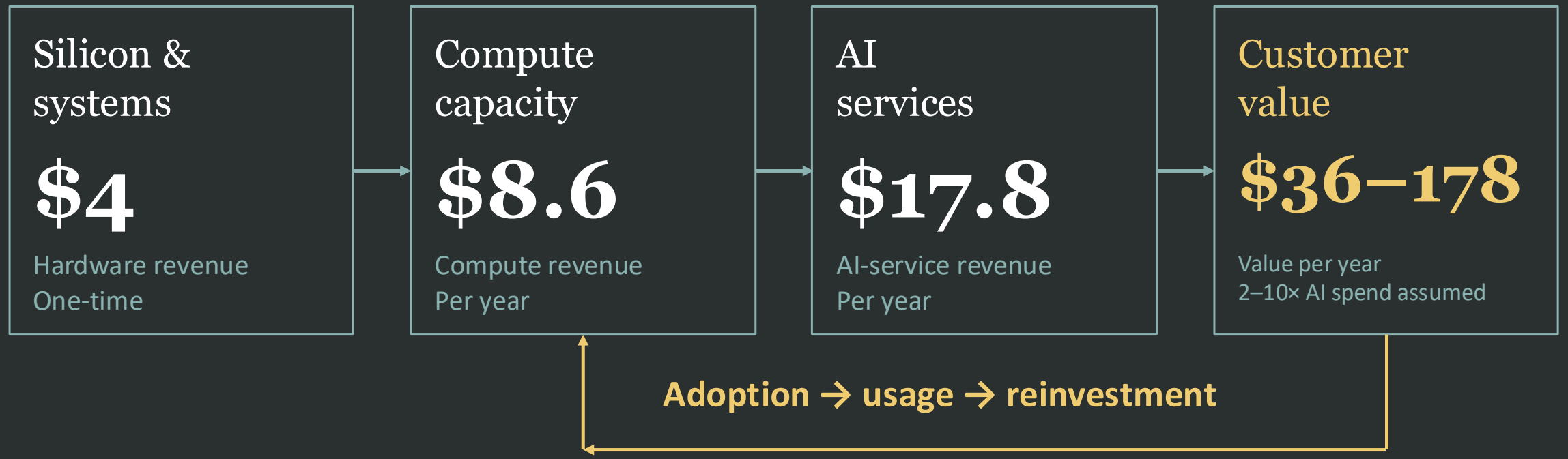
*2028 US data-center electricity projection. DOE / Berkeley Lab, 2024. Milestones not to scale.

Confidential - Oracle Restricted \Employees Only

Confidential - Oracle Restricted \Employees Only

AI's value chain feeds compute demand

ILLUSTRATIVE MODEL: PER $1 OF NVIDIA COST OF GOODS SOLD



Open access lets more innovators participate in the growth engine

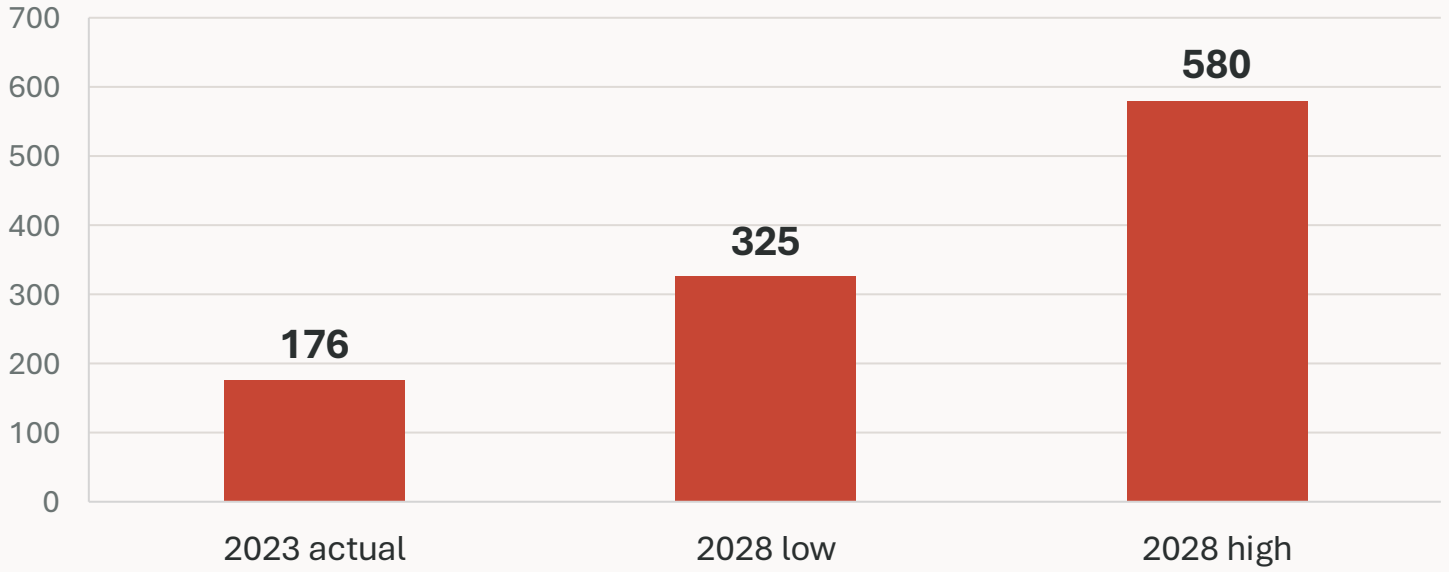
Source: supplied illustration. Assumptions, not verified results. Amounts are not additive.

Confidential - Oracle Restricted \Employees Only

ORACLE + FINOS

The compute squeeze is physical

US data-center electricity demand



TWh per year. 2028 values are projection bounds.

Power

Where capacity can grow

Chip supply

Leading AI chip fabrication is concentrated at TSMC

Confidential - Oracle Restricted \Employees Only

ORACLE + FINOS

Capacity across every environment

Customers are sourcing capacity wherever they can

On-premises

Existing grids and
private infrastructure

Public clouds

Hyperscale compute
and managed services

Neo-clouds

Specialist providers
and accelerated compute

Choose by capability, data location, policy and available capacity

ORACLE

Confidential - Oracle Restricted \Employees Only

The future of compute is heterogeneous

CPUs

Vera and general-purpose CPUs

Data preparation
Tools and orchestration

GPUs

Parallel accelerators

Model training
AI inference

TPUs

Google ML ASICs

ML workloads matched to TPU software

Wafer-scale

Cerebras

AI workloads matched to its architecture

Choose by workload, software fit, data movement and measured outcome

ORACLE

Confidential - Oracle Restricted \Employees Only

Confidential - Oracle Restricted \Employees Only

Common interfaces for compute and AI

Access	Jobs, model endpoints and tool access
Governance	Identity, data policy and use-case approval
Workflow	Tasks, agents, decisions and human review
Evidence	Evaluation, provenance, cost and audit records

Open interfaces reduce the cost of changing providers

ORACLE

Confidential - Oracle Restricted \Employees Only

ORACLE + FINOS

Our FINOS work spans compute and AI

OpenGRIS	Execute distributed Python work OCI worker managers
Open Resource Broker	Request and release compute OCI provider merged upstream
Fluxnova	Coordinate workflows and decisions OCI evaluation work
AI Governance Framework	Apply shared AI risks and mitigations Our team is participating

Common foundations for capacity, AI services and governed workflows

Confidential - Oracle Restricted \Employees Only

Confidential - Oracle Restricted \Employees Only

Bring a workload or AI use case

Help build the common path

Choose a project. Test a boundary. Contribute the evidence.

github.com/finos/opengris github.com/finos/open-resource-broker
fluxnova.finos.org air-governance-framework.finos.org

ORACLE

Confidential - Oracle Restricted \Employees Only