

Just Streams: Real-Time Data Pipelines in OCI

With a Live Demo Twist

Oracle AI Data Platform
Powered by



Sandi Holub

Principal Consultant



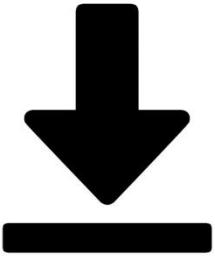
Žiga Vaupot

Director & Lead Analytics Consultant

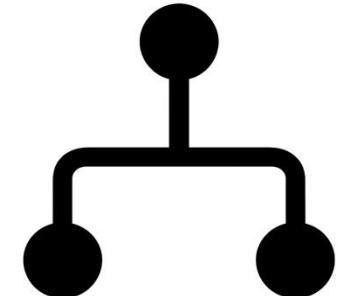
We master Data. Every kind of Data.



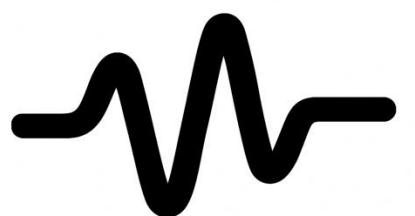
source
data



bulk
load



incremental
refresh



real-time
& streaming



data platform

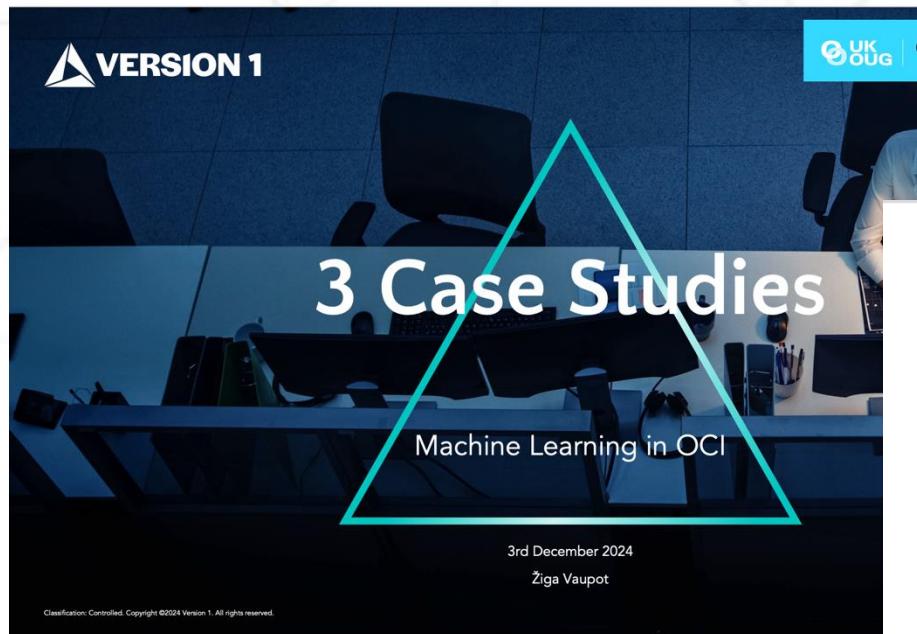


analytics



machine learning
& artificial
intelligence

In my presentation at last year's UKOUG event...

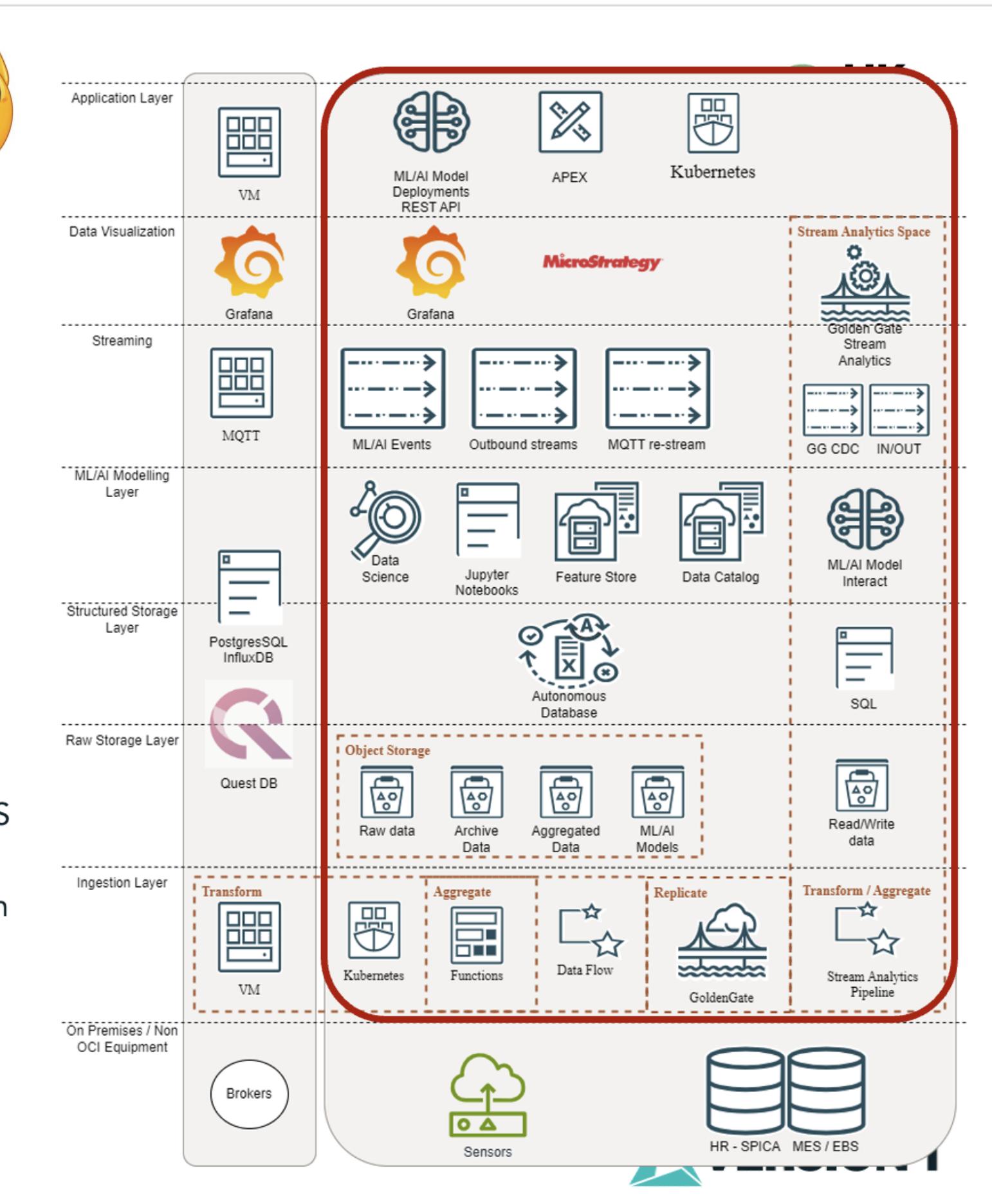


... and after Solution Design

- It's not just some OCI products, but we are talking about full OCI Data Lakehouse implementation that includes several services such as:
 - Oracle Autonomous Database
 - OCI Compute
 - OCI Object Storage
 - OCI Streaming
 - OCI Data Flow
 - OCI GoldenGate (replications)
 - OCI GoldenGate Stream Analytics
 - OCI Data Science
 - OCI Functions
 - Kubernetes
 - (unfortunately, no Oracle Analytics Cloud)
- Services are organised into logical layers which are planned to support MES and sensor data real-time streaming and ML/AI analysis.
- The 1st phase of the project is about data management platform rather than data analysis, which is planned for the 2nd phase.

Key value proposition for the client

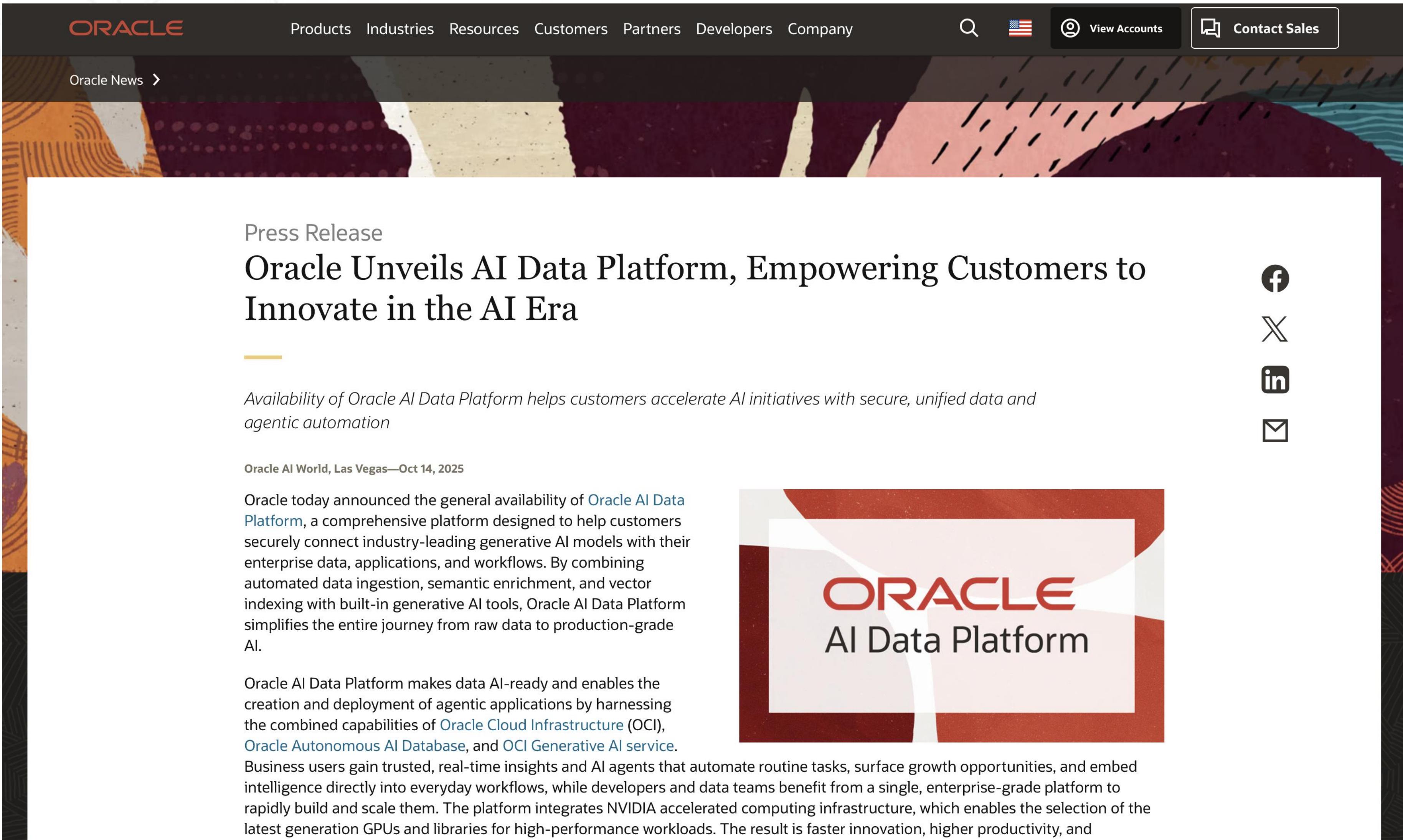
- They already use EBS in OCI, this project is natural extension of their OCI usage.
- Enterprise-level services in one place.



... and this was basically our idea for this years' presentation too.



...but last month, this was introduced...



The screenshot shows a press release from Oracle. At the top, there's a navigation bar with links for Products, Industries, Resources, Customers, Partners, Developers, Company, and a search bar. Below the navigation is a banner with the text "Oracle News >". The main content area starts with a "Press Release" heading and the title "Oracle Unveils AI Data Platform, Empowering Customers to Innovate in the AI Era". A sub-headline below the title reads: "Availability of Oracle AI Data Platform helps customers accelerate AI initiatives with secure, unified data and agentic automation". The text continues with details about the platform's general availability at Oracle AI World, Las Vegas—Oct 14, 2025. It describes Oracle AI Data Platform as a comprehensive platform designed to help customers securely connect industry-leading generative AI models with their enterprise data, applications, and workflows. The platform integrates NVIDIA accelerated computing infrastructure, which enables the selection of the latest generation GPUs and libraries for high-performance workloads. The result is faster innovation, higher productivity, and business users gain trusted, real-time insights and AI agents that automate routine tasks, surface growth opportunities, and embed intelligence directly into everyday workflows. The text also mentions the combined capabilities of Oracle Cloud Infrastructure (OCI), Oracle Autonomous AI Database, and OCI Generative AI service. To the right of the text, there's a graphic for "ORACLE AI Data Platform" featuring the Oracle logo and the text "AI Data Platform". Below the graphic, there are social media sharing icons for Facebook, X (Twitter), LinkedIn, and Email.

ORACLE

Products Industries Resources Customers Partners Developers Company

Oracle News >

Press Release

Oracle Unveils AI Data Platform, Empowering Customers to Innovate in the AI Era

Availability of Oracle AI Data Platform helps customers accelerate AI initiatives with secure, unified data and agentic automation

Oracle AI World, Las Vegas—Oct 14, 2025

Oracle today announced the general availability of [Oracle AI Data Platform](#), a comprehensive platform designed to help customers securely connect industry-leading generative AI models with their enterprise data, applications, and workflows. By combining automated data ingestion, semantic enrichment, and vector indexing with built-in generative AI tools, Oracle AI Data Platform simplifies the entire journey from raw data to production-grade AI.

Oracle AI Data Platform makes data AI-ready and enables the creation and deployment of agentic applications by harnessing the combined capabilities of [Oracle Cloud Infrastructure \(OCI\)](#), [Oracle Autonomous AI Database](#), and [OCI Generative AI service](#). Business users gain trusted, real-time insights and AI agents that automate routine tasks, surface growth opportunities, and embed intelligence directly into everyday workflows, while developers and data teams benefit from a single, enterprise-grade platform to rapidly build and scale them. The platform integrates NVIDIA accelerated computing infrastructure, which enables the selection of the latest generation GPUs and libraries for high-performance workloads. The result is faster innovation, higher productivity, and

View Accounts

Contact Sales

f

X

in

✉

ORACLE

AI Data Platform



Let's rewrite everything to run on top of the OCI AI Data Platform — it'll be fun!

AI Data Platform - SMARTQ_AIDP

Search

AD

+ Create

- Home
- Master catalog
- Workspace
- SMARTQ
 - Workflow
 - Compute
- Data Sharing
- Auto populate catalog
- Notifications 0
- Roles
- Audit logs

Activity

- tfl_bronze_layer.job
- tfl_silver_layer.job
- tfl_stream_producer....
- tfl_arrivals_poller....
- spark_cluster_defaul...
- bronze_util.ipynb
- tfl_arrivals_poller....

Recent

Type	Name
code/tfl-bronze/tfl_arrivals_poller.i...	
code/tfl-bronze/bronze_util.ipynb	
spark_cluster_default	
tfl_arrivals_poller.job	
code/streaming_producers/tfl_stre...	
tfl_silver_layer.job	
tfl_bronze_layer.job	

AI Data Platform

All 7 Tasks 5 Resources 2 What's new 0



Welcome to AI Data Platform

See an overview of what you can do with AI Data Platform.

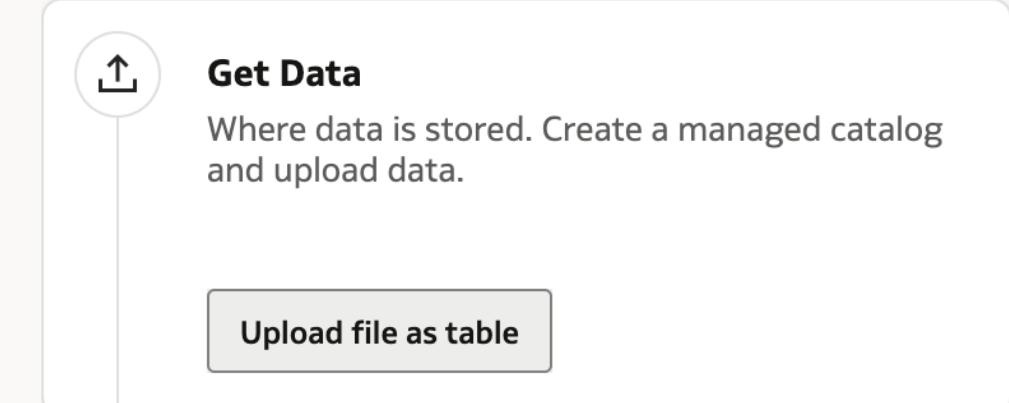
Help page



Download sample code

See examples of how you can use your AI Data Platform APIs

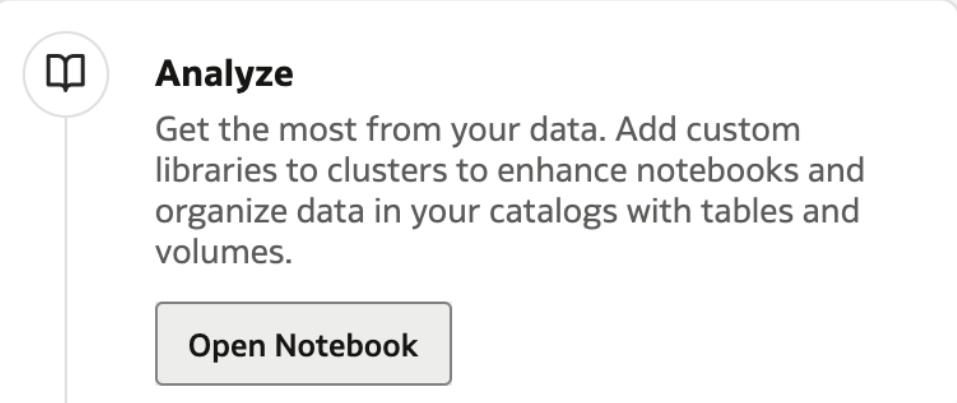
Download zip file



Get Data

Where data is stored. Create a managed catalog and upload data.

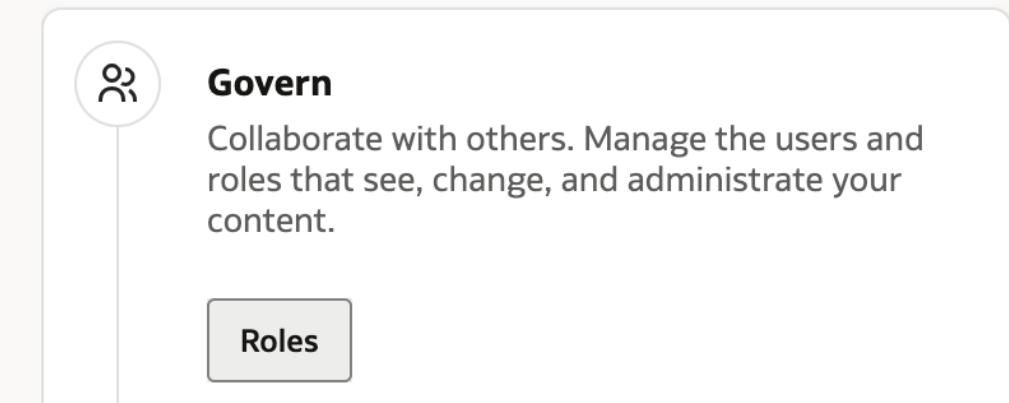
Upload file as table



Analyze

Get the most from your data. Add custom libraries to clusters to enhance notebooks and organize data in your catalogs with tables and volumes.

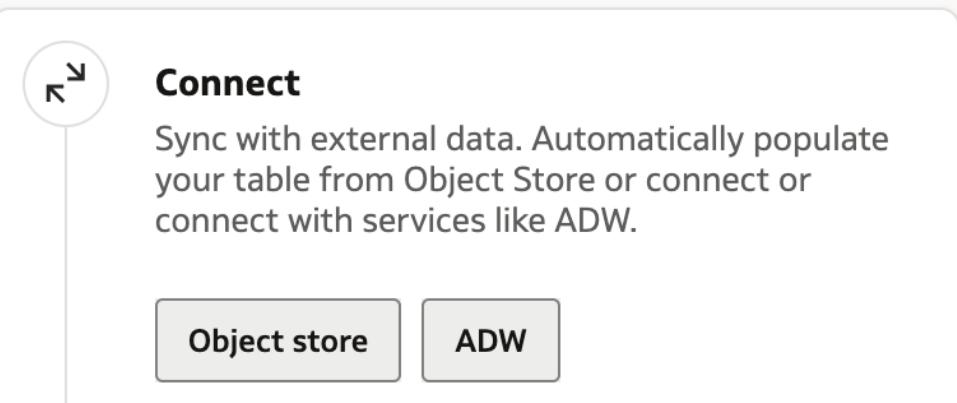
Open Notebook



Govern

Collaborate with others. Manage the users and roles that see, change, and administrate your content.

Roles



Connect

Sync with external data. Automatically populate your table from Object Store or connect or connect with services like ADW.

Object store ADW

Basic idea ...

... since we cannot use the client's data

- Let's identify a Public API that provides high-volume data streams.
- Ingest these data streams into the AI Data Platform.
- Transform the data to make it structured and query-ready.
- Visualize the data using a Data Visualization tool.



Preliminary Learnings...

...so far

- Centralized master data catalog supporting a medallion (Bronze, Silver, Gold) architecture
- Development of scalable data pipelines leveraging Apache Spark
- Workflow orchestration combined with interactive, AI-powered notebooks (OCI Data Science / Data Flow integration)
- Seamless integration with Oracle Analytics Cloud for advanced reporting and analytics



DEMONSTRATION

Just stream the data into the AI Data Platform.

What we'll demo?

General workflow

- **Stream producer** continuously pulls live bus arrival data from the TfL API and streams new records into an OCI Streaming (Kafka-compatible) topic.
- **Spark Structured Streaming “Bronze layer”** for TfL bus arrivals continuously reads TfL arrival events from OCI Streaming (via Kafka API), parses the JSON payload, adds ingestion metadata and event-time partitions, and appends everything into a Delta “bronze” table.
- **Spark Structured Streaming “Silver layer”** for TfL bus arrivals reads the “bronze” table, cleans and type-casts the data, keeps only the latest prediction per key (vehicleId, naptanId, lineId, direction), upserts it into the Silver table using MERGE, and sends lightweight log records to OCI Logging with each batch.
- **Spark Structured Streaming “Gold layer”** for TfL bus arrivals and their positions
- Visualize stream data in Oracle Analytics



Q&A