UK OUG | CONFERENCE DISCOVER 2025

Just Streams: Real-Time Data Pipelines in OCI

With a Live Demo Twist

Powered by Oracle AI Data Platform

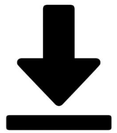
Sandi Holub
Principal Consultant

Žiga Vaupot
Director & Lead Analytics Consultant

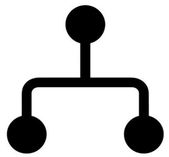
smartQ

We master Data. Every kind of Data.

source data



bulk load



incremental refresh

real-time & streaming

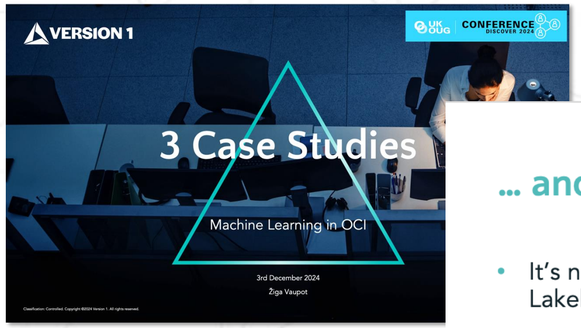
data platform

analytics

machine learning & artificial intelligence

smartQ

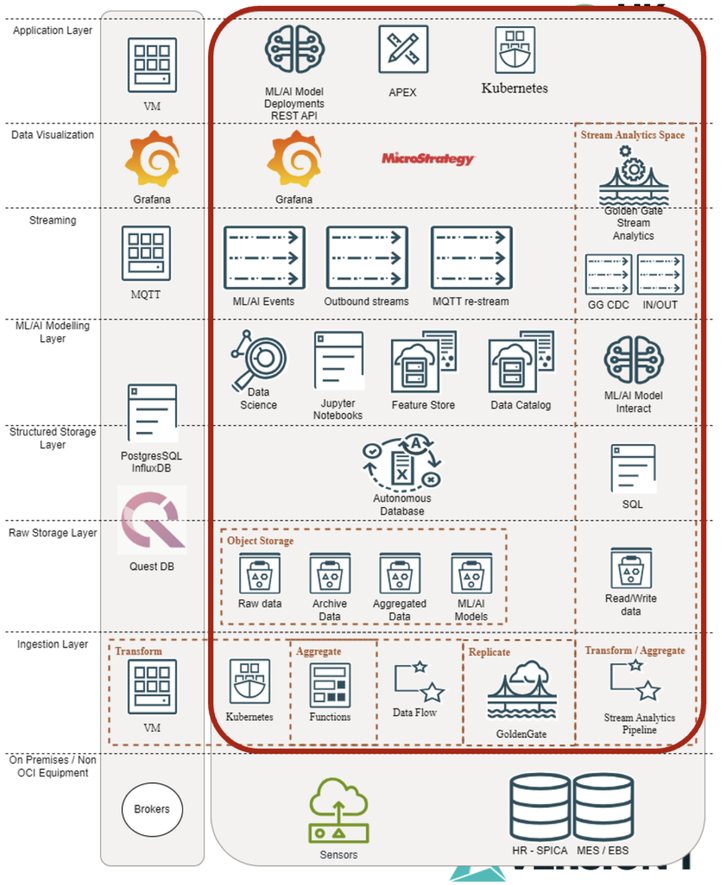
In my presentation at last year's UKOUG event...



... 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...

ORACLE Products Industries Resources Customers Partners Developers Company View Accounts Contact Sales

Oracle News

Press Release

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

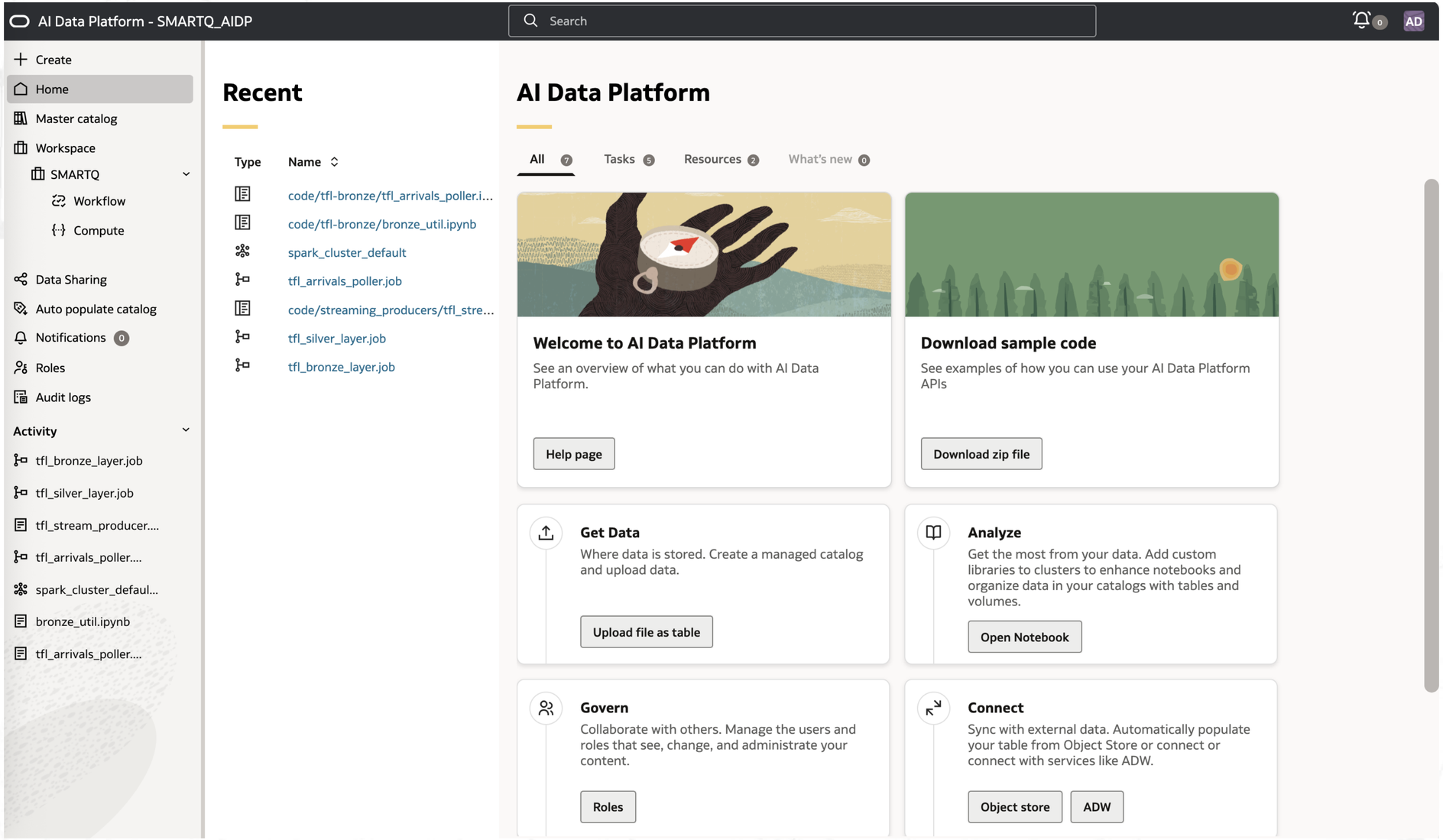
Availability of Oracle AI Data Platform helps customers accelerate AI initiatives with secure, unified data and agentic automation

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

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



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.

smartQ

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

smartQ