

Databricks: Data Warehousing, AI/BI and Genie

Databricks: Data Warehousing, AI/BI and Genie is a hands-on course on turning Gold layer data into an analytical product for the business. The program focuses on the analytical model, shared metric definitions in Unity Catalog (metric views), working with a SQL warehouse and building AI/BI dashboards.

Participants go through the whole process: from designing fact and dimension tables, through defining metrics as a metric view and choosing and costing a SQL warehouse, to publishing a dashboard and preparing context for Genie, which lets users ask questions about data in natural language. The course ends with a sales dashboard built on shared, documented metric definitions that all users work from. The exercises use RetailOps data (a fictional retail company whose data we use for all exercises).



Training recipients

The training is intended for people who prepare data for reporting and analysis in Databricks, and for people who make data available to business users.

The course is particularly suitable for:

- data analysts and BI specialists working in SQL,
- analytics engineers designing the Gold layer,
- report authors who want to use AI/BI dashboards or connect Databricks to Power BI,
- teams that want to give business users access to Genie in a controlled way.

The training will be especially valuable for organizations where the same metrics are calculated differently in different reports.



Benefits

- The analytical model – you will learn to design fact and dimension tables with the right granularity.
- Shared metric definitions – you will learn Unity Catalog metric views and how to define KPIs once and use them in SQL, AI/BI dashboards, Genie and alerts.
- SQL warehouse – you will learn how to choose the warehouse type and size and how to control performance and cost.
- AI/BI dashboards – you will build a dashboard with filters and drill-through to details and learn the rules for publishing it.
- Integration with BI tools – you will learn patterns for connecting to Power BI and how to keep metrics consistent.
- Genie – you will learn how to prepare context, test questions and set permissions so that natural language answers can be trusted.



Training program

Module – From the Gold layer to an analytical model

Module goal: the participant designs tables so that analysts and BI tools can use them easily.

- Data grain – the level of detail at which we store data (for example, an order line or a whole order)
- Fact and dimension tables (star schema) – sales in one table, customers, products and dates in separate ones
- Primary and foreign keys in Unity Catalog – a description of relationships between tables that BI tools and Genie use
- Table and column descriptions, tags and marking trusted tables – documentation for people and for Genie
- Lab: the participant adds descriptions and keys to a prepared star schema

2. Module – Unity Catalog metric views

Module goal: the participant defines business metrics in one place so that every report calculates them the same way.

- The problem – the same metric (for example, revenue) calculated differently in different reports and teams
- Metric view – a definition of metrics and dimensions stored once in Unity Catalog
- Building the definition: data source, table relationships, dimensions (for example, region, month) and measures (for example, revenue, margin, as well as running totals and moving averages)
- Using metric views in SQL queries, dashboards and Genie
- Permissions – who can use the metric definitions
- Lab: the participant defines revenue, margin and average order value as a metric view

3. Module – SQL warehouse: performance and cost

Module goal: the participant can choose and configure a SQL warehouse and find the cause of slow queries.

- SQL warehouse – processing power for SQL queries and dashboards; types: serverless, pro, classic
- Size and scaling – how a warehouse responds to many users at once
- Query history and query profile – how to find a slow query and understand why it is slow
- Cost – automatic shutdown, result caching and warehouse usage in system tables
- Lab: the participant analyzes a slow query and proposes a better warehouse configuration

4. Module – AI/BI dashboards

Module goal: the participant builds and publishes a dashboard for business users.

- Dashboard data sources – SQL queries and metric views
- Visualizations, filters, parameters, cross-filtering and drill-through to details on another dashboard page
- AI assistant (Genie Code) – creating charts, filters and dashboard pages from a description in natural language
- Publishing – who sees the dashboard, and whether they see data under their own permissions or the author's
- Refresh schedule and subscriptions – sending the dashboard by email, to Slack or to Microsoft Teams
- Demo: building a dashboard (participants build it on their own in the workshop, module 7)

5. Module – Integration with Power BI and other BI tools

Module goal: the participant knows how to connect Databricks to Power BI without metrics drifting apart.

- Power BI connection modes: importing data or live queries (DirectQuery) – the benefits and costs of each (overview and demo)
- Publishing data from Databricks to Power BI
- Keeping metrics consistent with Power BI – the Power BI connector does not support metric views natively today; we use an intermediate view with MEASURE() or a native query
- When an AI/BI dashboard is enough and when an external BI tool is needed; moving an existing Power BI or Tableau report into an AI/BI dashboard, with the metric definitions recreated as metric views (import with Genie Code, /importBI)
- Lab: the participant chooses how to share data with three types of consumers

6. Module – Genie: asking questions about data in natural language

Module goal: the participant sets up Genie so that its answers are correct and respect permissions.

- Genie Agent (formerly Genie space) – a set of tables and metric views that users can ask about in natural language; business users work with it in Genie One
- Instructions and sample queries – how to teach Genie the company's vocabulary and correct calculations
- Trusted queries – ready, verified answers to the most common questions
- Quality testing – a set of control questions and comparing answers with the expected ones
- Permissions and limitations – Genie sees only the data available to the user; when not to give users Genie
- Lab: the participant prepares a Genie Agent for sales data and tests it on 5 control questions

(benchmark)

7. Module – End-to-end workshop

Module goal: the participant combines metrics, a dashboard and Genie into one analytical product.

- Workshop: metric view → KPI dashboard → Genie Agent built on the same metric definitions. The outcome is a published sales dashboard and a Genie Agent based on shared, documented metrics.
- Checklist before giving business users access to the dashboard and Genie



Expected preparation of the participant

- Completion of DBX-FND (Databricks Fundamentals) or equivalent knowledge of Unity Catalog and Delta tables
- Practical knowledge of SQL (JOIN, GROUP BY, aggregate functions)
- For people building their own Gold layer, we recommend completing DBX-ETL first

To increase the comfort of work and training's effectiveness we suggest using an additional monitor. The lack of additional monitor does not exclude participation in the training, however, it significantly influences the comfort of work during classes.

You can find information and requirements of participation in Distance Learning trainings at:

<https://www.altkomakademia.pl/en/how-it-works/about-courses/>



Training Includes

Training includes

- access to the Altkom Akademia student portal
- a prepared Databricks environment with a SQL warehouse and RetailOps scenario Gold data for the duration of the training
- permissions to create dashboards and Genie Agents in the training environment
- exercise materials in electronic form

Note: During selected training sessions, the Academy provides Participants with access to paid tools and services necessary to complete the training program (in particular Azure, Fabric, and other cloud resources). The Participant is required to use the provided resources solely within the scope of the training agenda and in accordance with the trainer's instructions. Use of these resources beyond the training scope—particularly for private projects, testing additional services, modifying the environment configuration, or undertaking activities that generate costs unrelated to the exercises—will result in the Participant being charged for costs exceeding the allocated limit, based on expenses incurred by the Academy.



Duration

1 days / 7 hours

Language

- Training: English
- Materials: English