

DATA ENGINEERING

INTERMEDIATE

2 DAYS

Cursor for Data and ML Engineering

Cursor applied to data work rather than application development. Covers notebook and exploratory workflows, pipeline and SQL generation with proper review, data quality testing, documentation and lineage, and the guardrails that matter when a wrong query is expensive rather than just broken.

Why Cursor for Data and ML Engineering?

Built for data work

Notebooks, pipelines and SQL rather than generic application development examples.

Documentation that finally happens

Pipeline and model documentation produced as a by-product rather than a backlog item.

Generated SQL needs scrutiny

A wrong query is expensive and often silent. Learn what to check every time.

Quality built in

Data tests and validation generated alongside the transformations they protect.

What you'll be able to do

- ✓ Configure Cursor for data projects, notebooks and analytics repositories
- ✓ Accelerate exploratory analysis without losing analytical rigour
- ✓ Generate and refactor ETL and ELT pipeline code
- ✓ Produce and critically review SQL and transformation logic
- ✓ Build data quality tests, schema validation and fixtures
- ✓ Generate pipeline, model and dataset documentation
- ✓ Apply guardrails around PII, sample data and expensive queries
- ✓ Establish review standards for AI-assisted data work

Who this is for

- Data engineers and analytics engineers
- Machine learning engineers and data scientists
- BI developers working in code-based transformation tools
- Analytics teams adopting AI-assisted development

Curriculum

01 Cursor in a Data Workflow

- Setting up Cursor for notebooks, scripts and analytics repositories
- Environment, dependency and kernel considerations
- Indexing a data project so context is actually useful
- Where AI assistance helps in data work and where it misleads
- **Hands-on:** configure Cursor against a real data repository

02 Exploratory Analysis and Notebooks

- Accelerating initial exploration of an unfamiliar dataset
- Generating transformations and iterating on them quickly
- Plotting, profiling and summarising with AI assistance
- Keeping analytical reasoning yours rather than delegating judgement
- **Hands-on:** profile and explore a dataset end to end

03 Building and Refactoring Pipelines

- Generating ETL and ELT pipeline code from a specification
- Refactoring legacy transformation logic safely
- Working with transformation frameworks and model files
- Feature engineering code and reusable transformation patterns
- **Hands-on:** refactor an existing pipeline with test coverage first

04 SQL Generation and Review

- Generating queries from a natural language description
- Reviewing generated SQL for correctness, not just syntax
- Join logic, granularity and the errors that produce plausible wrong numbers
- Query cost, scan volume and performance considerations
- **Hands-on:** review a set of generated queries and find the broken one

05 Testing and Data Quality

- Generating data tests alongside transformations
- Schema validation, constraints and freshness checks
- Test fixtures and representative sample data
- Catching silent failures that pipelines do not surface
- **Hands-on:** add a quality test suite to an untested pipeline

06 Documentation, Lineage and Guardrails

- Generating pipeline, dataset and model documentation
- Column descriptions, lineage notes and handover artefacts
- PII, sample data and what must never reach a prompt
- Review standards for AI-assisted data work
- **Hands-on:** document a pipeline and agree team guardrails