DataForge — Data Engineering Platform

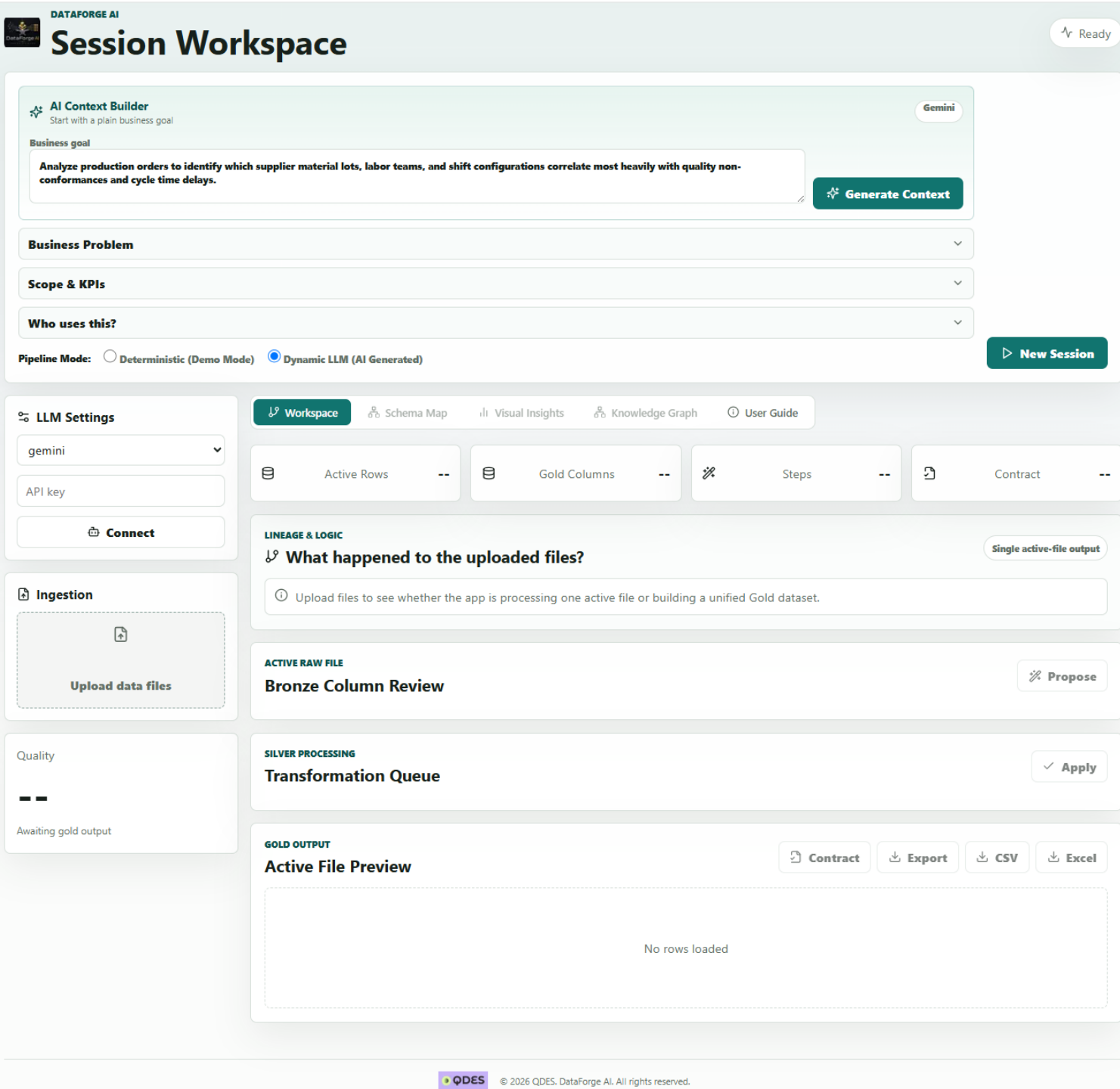
AI-assisted data engineering that turns raw operational data into analytics-ready, governed datasets.

DataForge is a data engineering platform designed to make pipelines repeatable, transparent, and trustworthy. It ingests raw source files and operational data in multiple formats, refines them through a structured bronze → silver → gold (medallion) pipeline, and produces clean, modeled, analytics-ready datasets — with full lineage and documentation for every transformation.

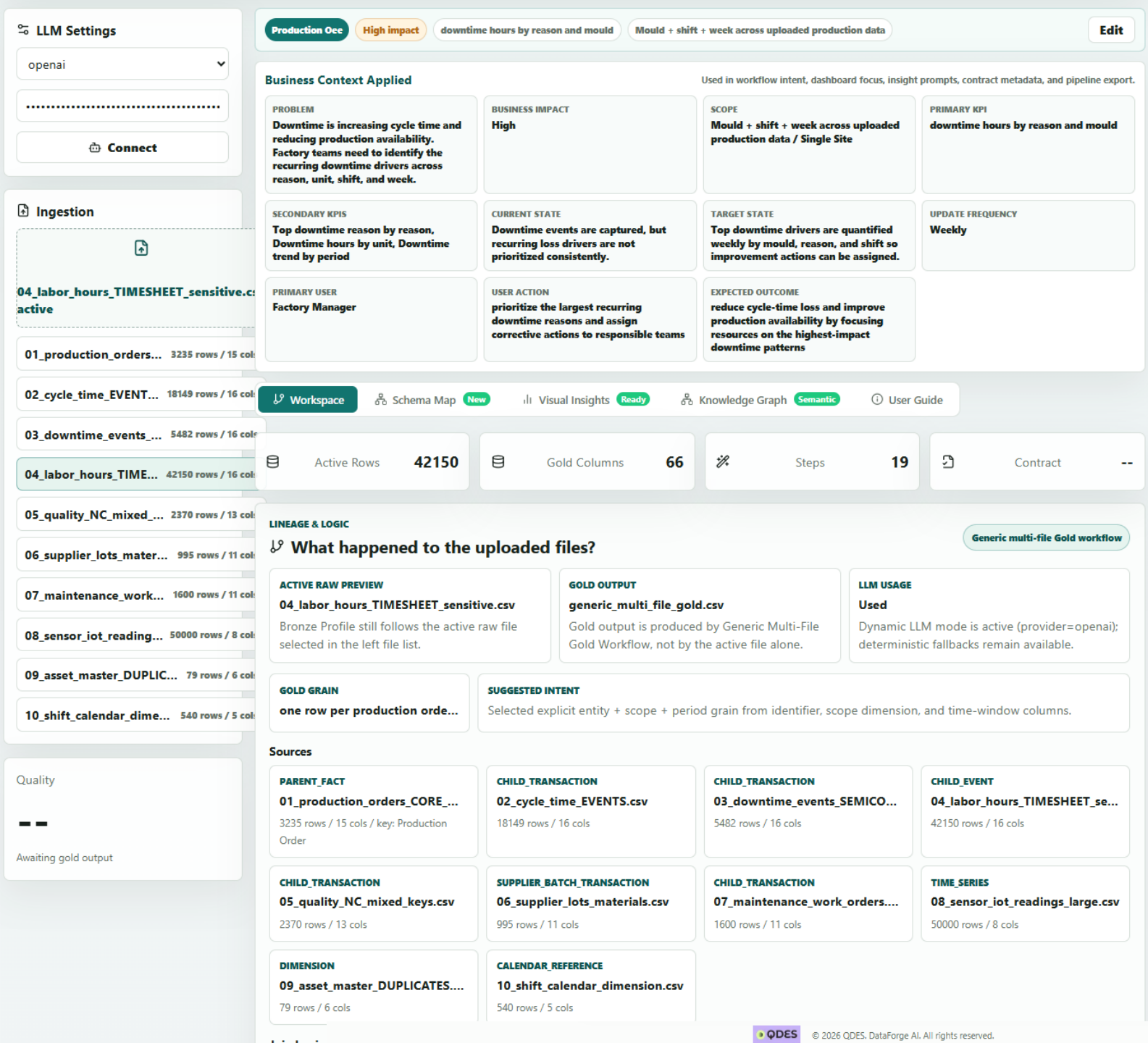
The platform combines automated data cleaning, schema and contract enforcement, semantic modeling, and optional AI LLM assistance. It is built for teams that need more than ad-hoc scripts and spreadsheets: they need traceable lineage, version history, and governed data they can defend and reuse.

Designed for:

Data engineering teams, quality and operations analysts, manufacturing data teams, BI and reporting teams, governance and compliance teams, and digital transformation teams.



Set up Business goal, problem, Scope & KPI



Bring in your uncleaned CSV, Excel, or JSON files.

www.qdes.org

GOLD OUTPUT

Unified Gold Preview

Contract | Export | CSV | Excel

PRODUCTION_ORDER	BLADE_ID	SITE	FACTORY_ID	MOULD	AREA	SHIFT	TEAM
PO-WST-100000	700000	WST	F002	M02	Trim		D
PO-LHR-100001	700001	LHR	F003	M06	Assembly	Weekend	D
PO-WST-100002	700002	WST	F002	m05	Paint	Weekend	2
PO-WST-100003	700003	WST	F002	M04	Assembly	Day	1
PO-LHR-100004	700004	LHR	F003	M05	Infusion	Night	B
PO-NAK-100005	700005	NAK	F001	M11	Assembly	Weekend	2
PO-LHR-100006	700006	LHR	F003	M06	Infusion		3
PO-WST-100007	700007	WST	F002	M03	Trim		2
PO-WST-100008	700008	WST	F002	m06	Infusion	Day	B
PO-NAK-100009	700009	NAK	F001	M04	Infusion	Day	3
PO-NAK-100010	700010	NAK	F001	M09	Trim	Day	B
PO-LHR-100011	700011	LHR	F003	M11	Paint		1

Data Governance

PII: HIGH

Auto-scanned on upload. Click to re-scan or run deeper analysis.

Re-scan PII | Check Schema Drift | Run Quality Rules

PII Scan

RISK: HIGH 3 finding(s)

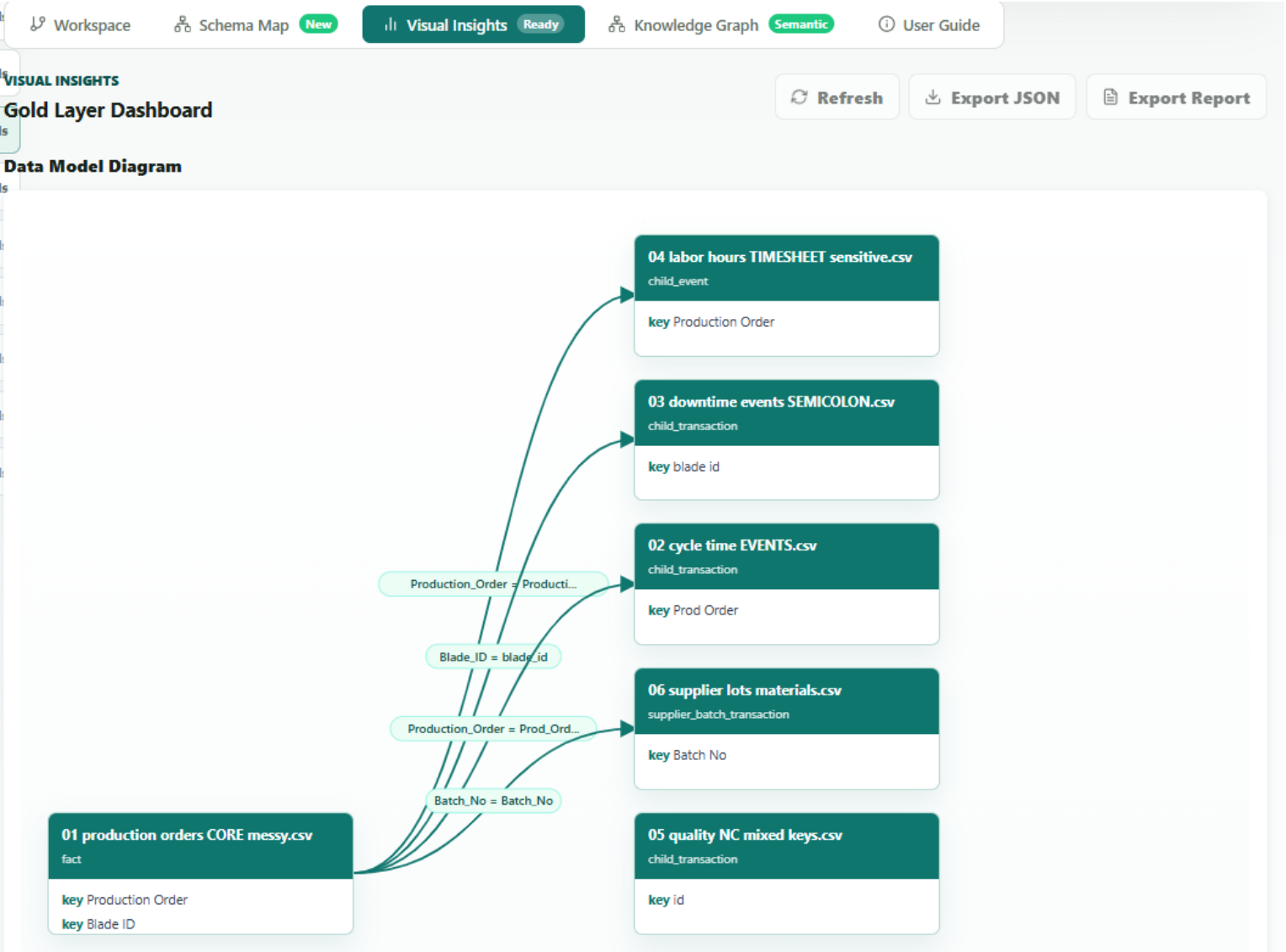
PII Columns: Employee Email, ct_event_id

COLUMN	TYPE	SEVERITY	METHOD
ct_event_id	passport	medium	value_pattern
Employee Email	email	high	column_name
Employee Email	email	high	value_pattern

Schema Drift

No schema drift detected

The AI automatically detects column types, identifies errors, and applies deterministic or LLM cleaning rules.



The system discovers relationships across multiple uploaded datasets and recommends safe joins.

AI INSIGHTS

Dashboard Intelligence

Executive

Use LLM if available (openai)

Regenerate

Generated by: llm | Provider: openai | Model: gpt-4o-mini

Executive Summary

Downtime is significantly impacting cycle time and production availability, necessitating a focused analysis on recurring downtime drivers. Prioritizing these drivers will enable targeted corrective actions to enhance operational efficiency.

Make shorter | Make more executive | Copy summary | Export insight JSON

Key Findings

FINDING
Top Downtime Reason Identified
The leading cause of downtime is 'Quality hold', accounting for 643,725 minutes of downtime.

FINDING
Mould-Specific Downtime
Mould 'M12' has the highest downtime at 308,797 minutes, indicating a need for targeted analysis.

FINDING
Shift Impact on Downtime
The first shift accounts for 1,618,005 minutes of downtime, representing 41.89% of total downtime.

Risks

RISK
Inconsistent Prioritization of Downtime Drivers
Failure to consistently prioritize downtime drivers may lead to unresolved issues and continued production inefficiencies.

RISK
Potential for Increased Cycle Time
If downtime drivers are not addressed, cycle time may continue to increase, further reducing production availability.

RISK
Resource Misallocation
Without clear identification of downtime drivers, resources may be misallocated, hindering effective corrective actions.

Recommended Actions

RECOMMENDATION
Prioritize Downtime Reasons
Establish a systematic approach to prioritize and address the top downtime reasons weekly.

RECOMMENDATION
Assign Corrective Actions
Assign specific corrective actions to responsible teams based on the prioritized downtime drivers.

RECOMMENDATION
Monitor Downtime Trends
Regularly monitor downtime trends to identify emerging issues and adjust strategies accordingly.

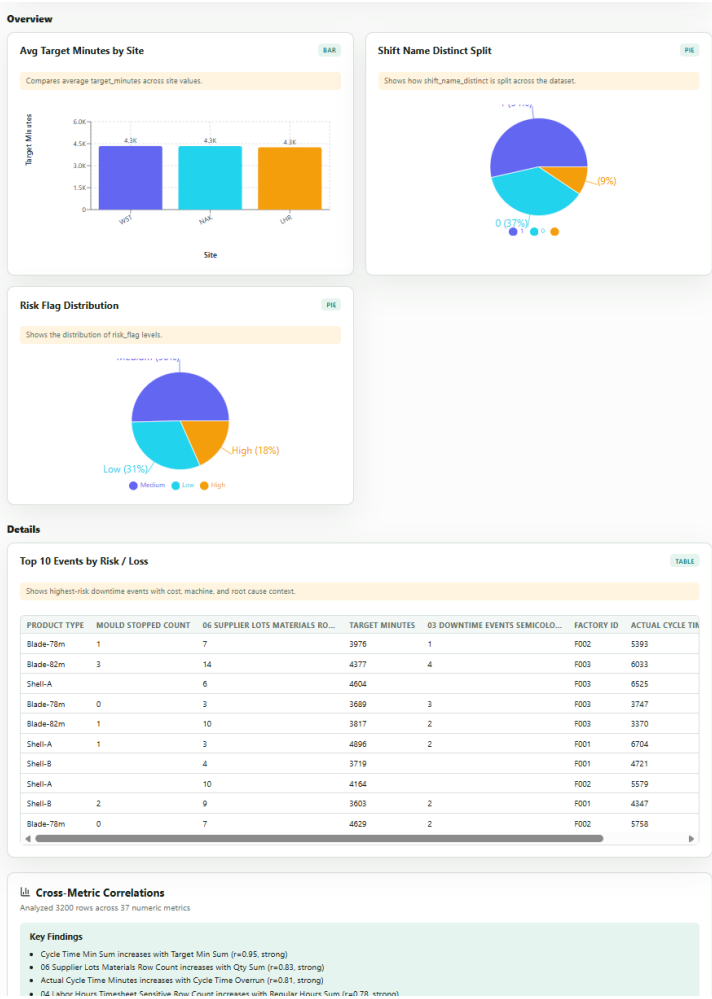
Data Quality Notes

DATA_QUALITY
Data Completeness
The dataset contains 3,200 rows, providing a comprehensive view of production data.

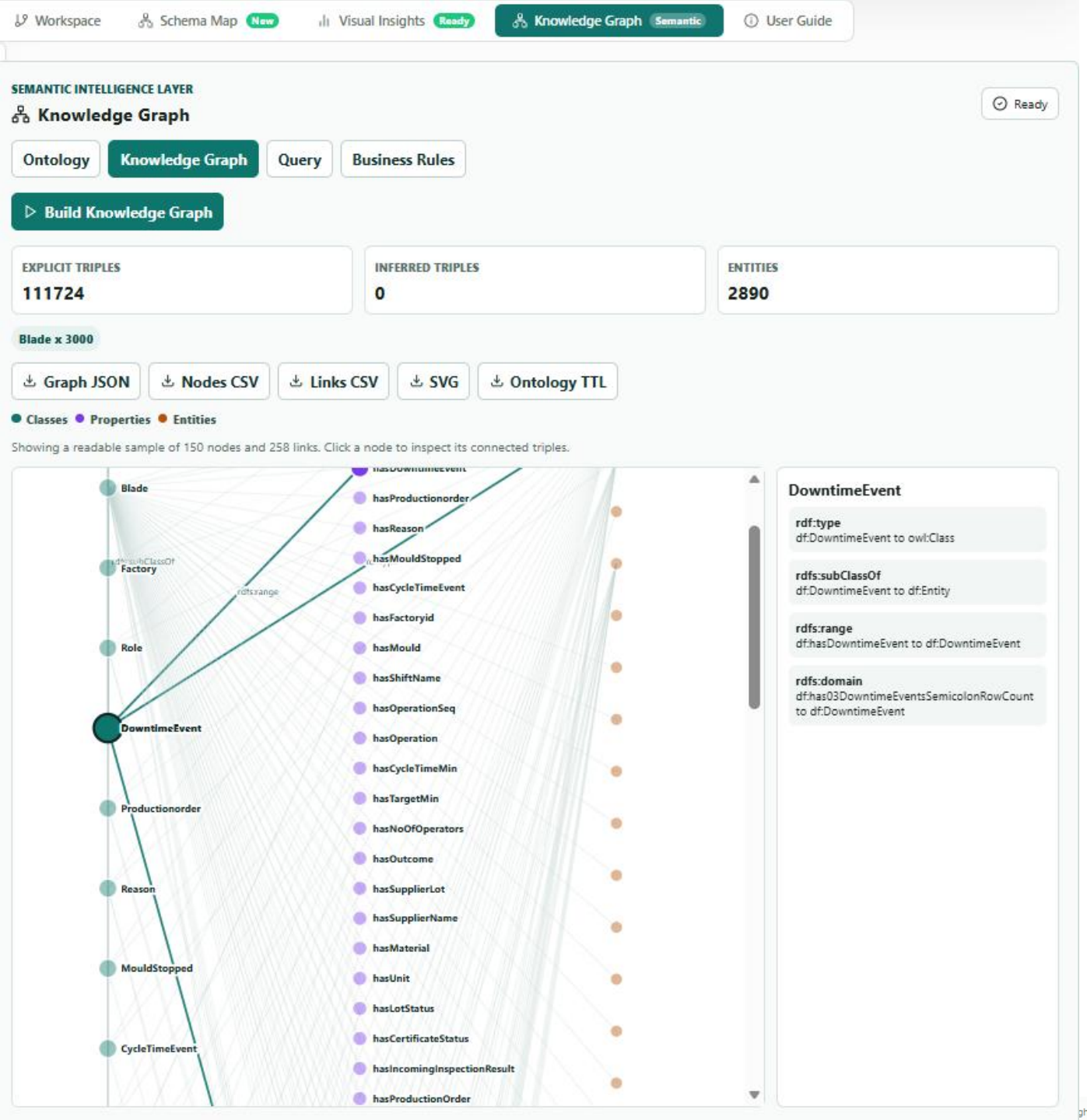
DATA_QUALITY
High Confidence in Metrics
Key metrics have a confidence level of 0.98, indicating reliable data for analysis.

DATA_QUALITY
Potential Data Gaps
Ensure ongoing data collection processes are robust to prevent gaps in future analyses.

Build a unified dataset, auto-generate visual dashboards, and create AI Insights.

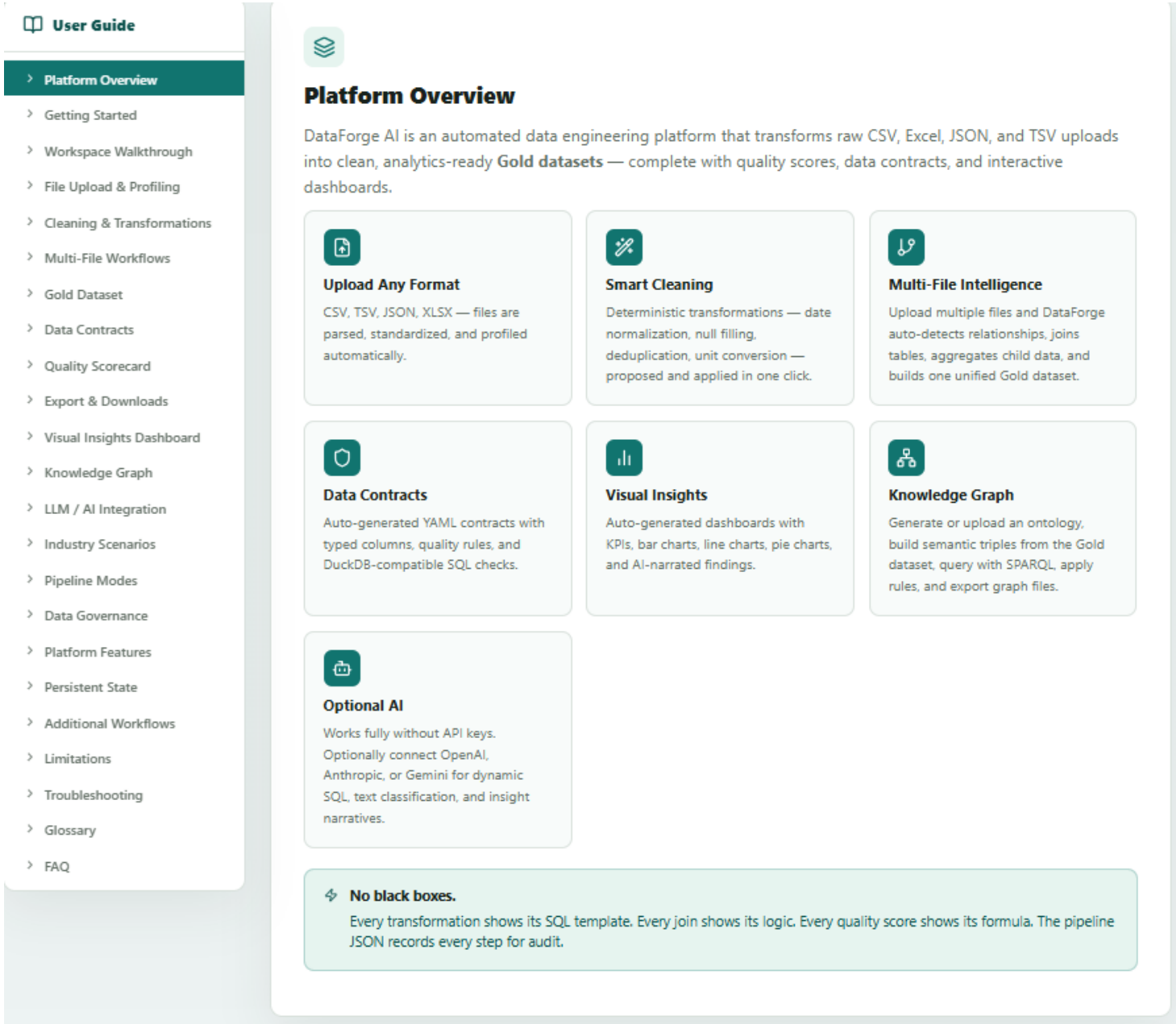


www.qdes.org



Generates ontology, Knowledge graph, Query and Business rules

www.qdes.org



Detailed User Guide