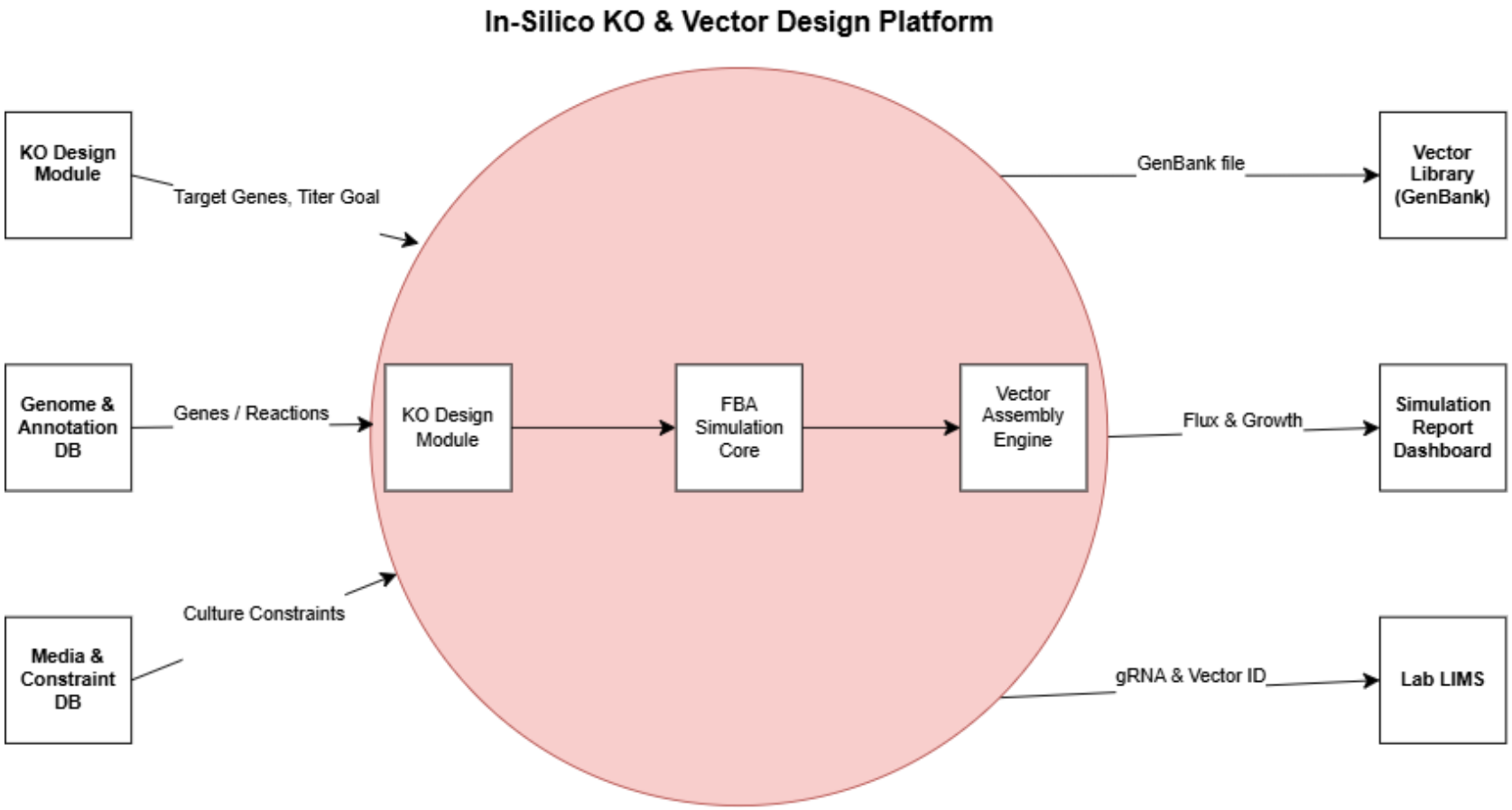
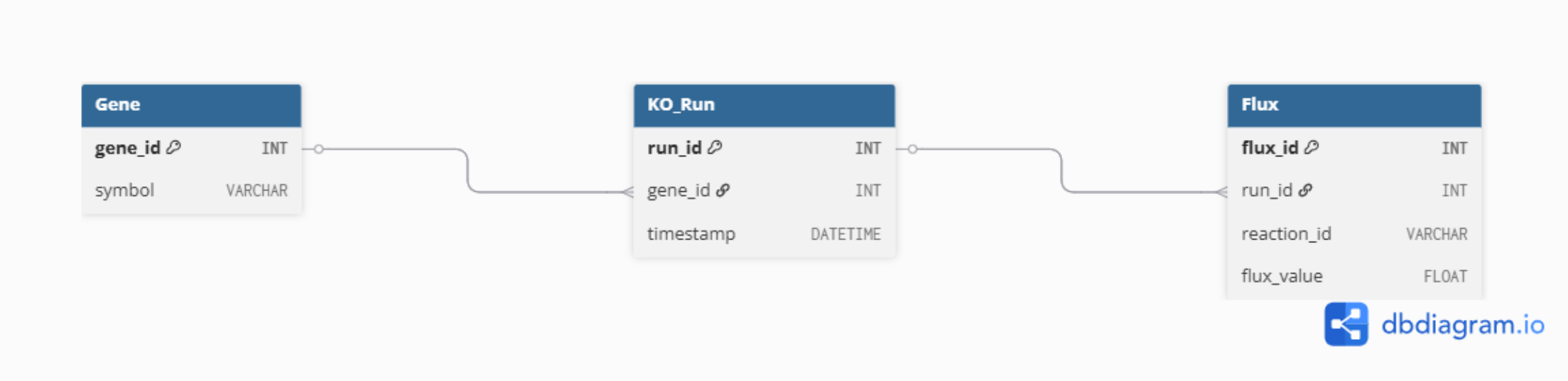
Project Context & Data Flow



Shows actors, data sources, and core pipeline modules

Key Data Entities & Relationships



Defines Gene, KO_Run, Flux tables for traceable data storage

User Story: Lab Data Ingestion & QC

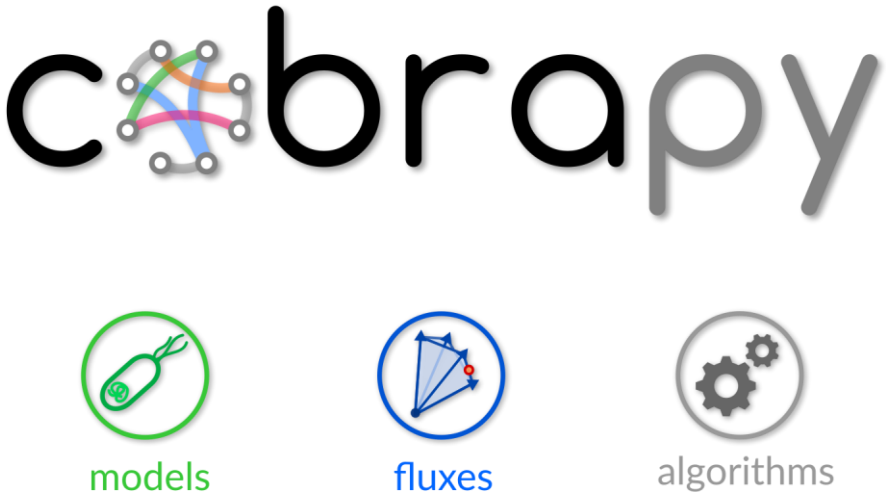
- As a bioinformatician, I want to run barcode QC on raw assay files so that only high-quality cells are analyzed.

- Acceptance Criteria:
    - Flags cells with low read-count or high mitochondrial content.
    - Generates downloadable QC report.

User Story: KO Design & Simulation

- As a data scientist, I want to train the perturbation model on annotated gene datasets so I can predict cell behavior.

- Acceptance Criteria:
    - Model accepts AnnData inputs.
    - Logs and metrics exported per run.



Data Quality Pipeline (Cloud & QC Demo)

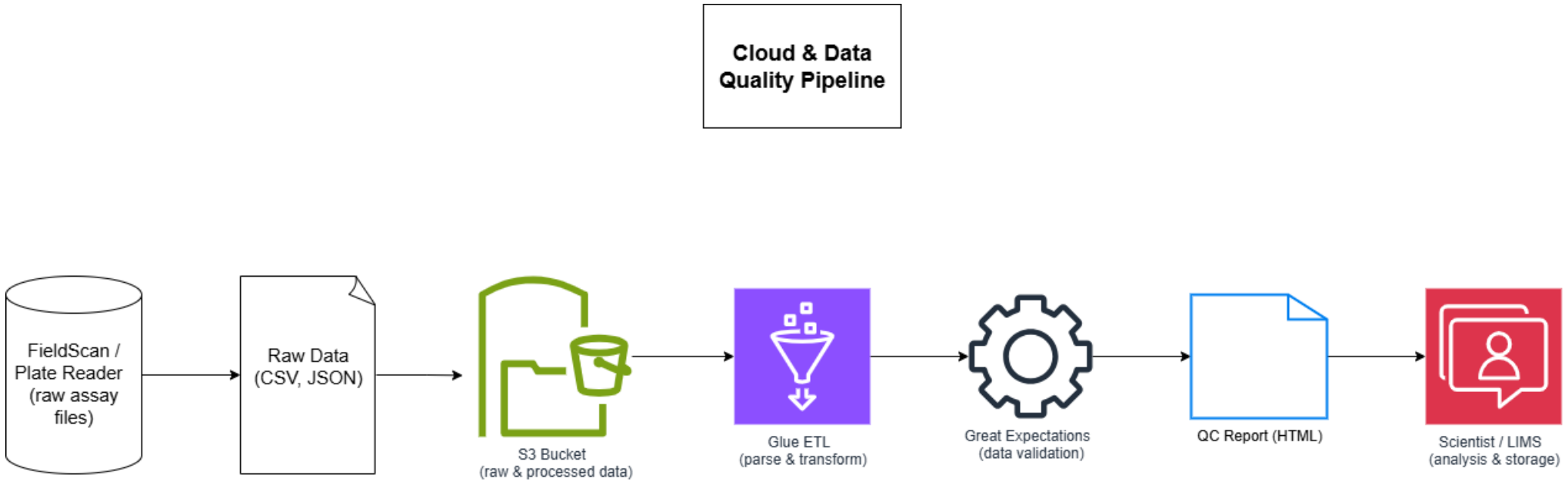


Diagram: FieldScan → S3 → Glue → GE → QC Report → Scientist