1. **A screenshot of a computer

   Description automatically generatedDraw an ER model for the 5 five tables created in Athena**

All numeric values are stored as bigint in Athena, except for days\_since\_prior\_order, which is stored as double.

Text features are stored as string.

1. **Design a query to join orders table and order\_products table together, filter on eval\_set = ‘prior’**

SELECT

op.\*,

o.\*

FROM "raw"."order\_products" as op

LEFT JOIN "raw"."orders" as o

ON op."order\_id" = o."order\_id"

WHERE o."eval\_set" = 'prior';

**Put all the glue table definitions into cloudformation, use github action to deploy the table**

*GitHub:* [*https://github.com/langou3/project\_ibma\_difan*](https://github.com/langou3/project_ibma_difan)

The data will be loaded into the Glue Catalog through a Glue Crawler from S3. I am using the Athena-generated DDL to define the table.

The Glue Crawler is pre-defined using the web service.

The action job is stored in .github/workflow/deploy-table.yml, and the template for defining the table is stored in .github/cfn/glue-table.yml