Someone renamed a column. Job still green.

D8A.ACADEMY/RESOURCES/SCHEMA-DRIFT

WHAT HAPPENED

`amount → amount_cents`	renamed upstream, no mail
`the loader`	← job green. amount NULL for 6 days

CATCH IT AT INGEST

`expected schema`	columns and types, in git
`diff on every run`	← missing or retyped: fail the run
`dbt not_null`	in schema.yml: fails the first night
`Great Expectations`	column_to_exist, values_of_type

REMEMBER

`data contract`	the producer runs the check too
`null rate alert`	0% to 100% in one day: page someone

WORTH MEMORIZING

01 The check, before the load

Ten lines of Python in front of the loader. The run fails at 09:41 with the column name in the message.

```
expected = {
    'order_id': 'int64',
    'amount': 'float64',
    'placed_at': 'datetime64[ns]',
}
actual = df.dtypes.astype(str).to_dict()

missing = expected.keys() - actual.keys()
retyped = {c for c in expected
           if c in actual and actual[c] != expected[c]}
if missing or retyped:
    raise ValueError(f'schema drift: {missing} {retyped}')
```

→ A rename shows up as missing. A string where a float was shows up as retyped. New extra columns are allowed here on purpose: additions rarely break a load, they just need a ticket.

02 Where each tool catches it

Same check, four places. The earlier it runs, the fewer rows you have to fix.

TOOL	YOU WRITE	FIRES
ingest check	the expected dict, diffed per run	before the load
dbt	not_null, accepted_values in schema.yml	at dbt test, after the load
Great Expectations	expect_column_to_exist, _to_be_of_type	at the checkpoint
data contract	the schema, versioned, owned upstream	before they deploy

→ A NULL rate alert is the last net: amount went from 0% NULL to 100% NULL in one day. That is not drift, that is a break, and it should page someone.