Week 2

Elements of Data Integration (CS 333E)

Jan 23, 2026

Data Pattern: Full Loader (Ch 2)

What is it?

Why does it exist?

What are its pros and cons?

How is it implemented?

Implementation Option 1

```
begin transaction;

delete from t where 1 = 1;

insert into t
    values (123, 'foo'),
           (456, 'bar'),
              (789, 'baz'),
                …

commit transaction;
```

Implementation Option 2

Day 1:

```
create or replace view emp as
    select * from emp_2026_01_22;

select * from emp;
  ->  Seq Scan on table emp_2026_01_22
```

Day 2:

```
create or replace view emp as
    select * from emp_2026_01_23;

select * from emp;
  ->  Seq Scan on table emp_2026_01_23
```

Open Table Formats

- Provides a table abstraction for a dataset's underlying files
- Started emerging in 2009 with Hive
- Modern examples: Iceberg, Delta, Hudi

- Key features:
  - ACID transactions, table statistics, performance optimization, time travel, schema evolution

Source: O'Reilly, The Architecture of Apache Iceberg