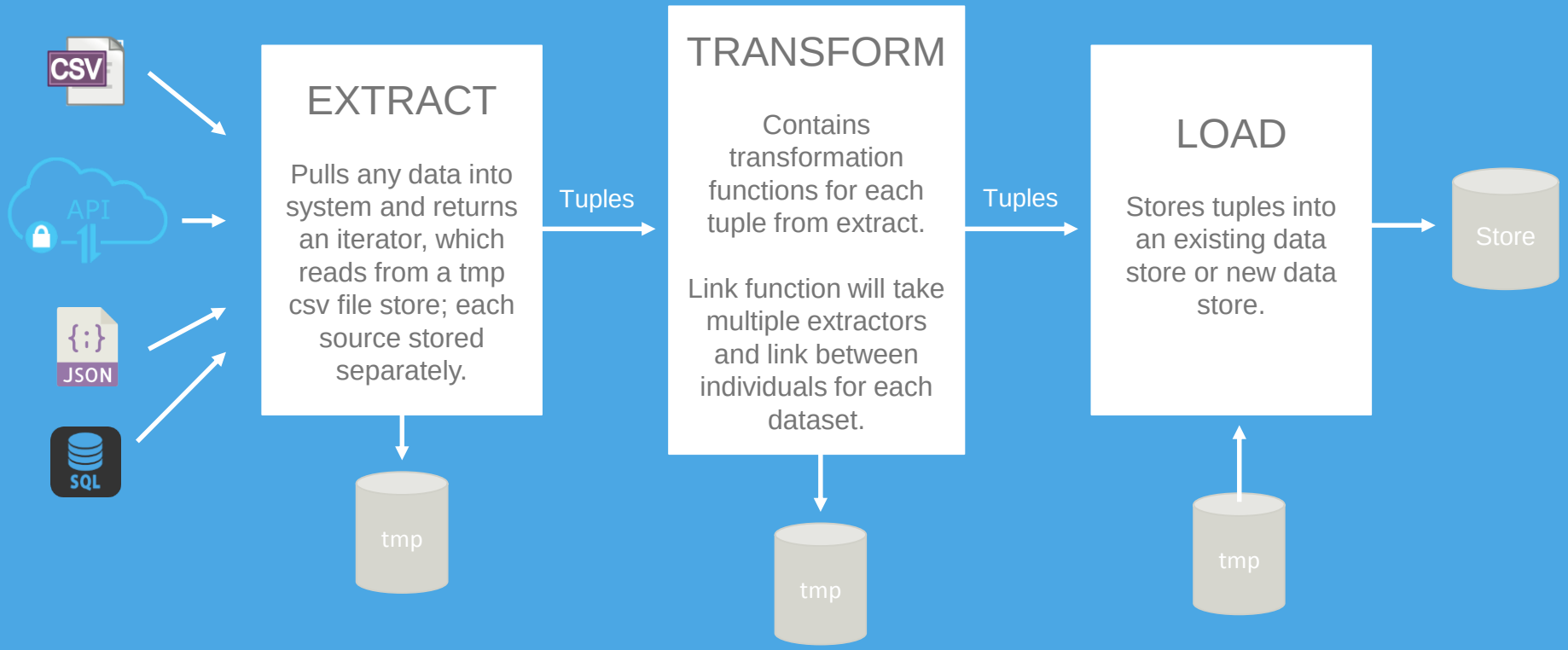
UChicago
Urban Labs

Urban Labs ETL Pipeline
Code Structure

August 2, 2016



Use of temporary files prevents memory errors

Ideal API Structure

```
data = UrbanETL.Extract(source, ‘source_type’)
data2 = UrbanETL.Extract(source2, ‘source_type’)

data.date_time(‘DOB’)
data.normalize_name(‘NAME’)
data2.data_time(‘dob’)
data2.normalize_name(‘name’)

linked_data = data.link([data2], method = ‘eblink’,
columns = [(‘DOB’, ‘dob’), (‘NAME’, ‘name’)])

linked_data.store(destination)
```