Stream-Static Join

Learning Objectives

- ▶ Join Incremental and Static Data
- ► Potential limitations

Stream vs. Static tables

- Streaming tables are ever-appending data sources
- ▶ Static tables contain data that may be changed or overwritten.
 - ▶ breaks the requirement of an ever-appending source for Structured Streaming

Stream-Static join

Students



id	name	course_id
1	Adam	C01

Streaming

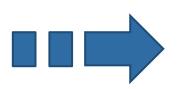
Courses



course_id	course_name
C01	Math

Stream-Static join

Students



id	name	course_id
1	Adam	C01

Streaming

Join result

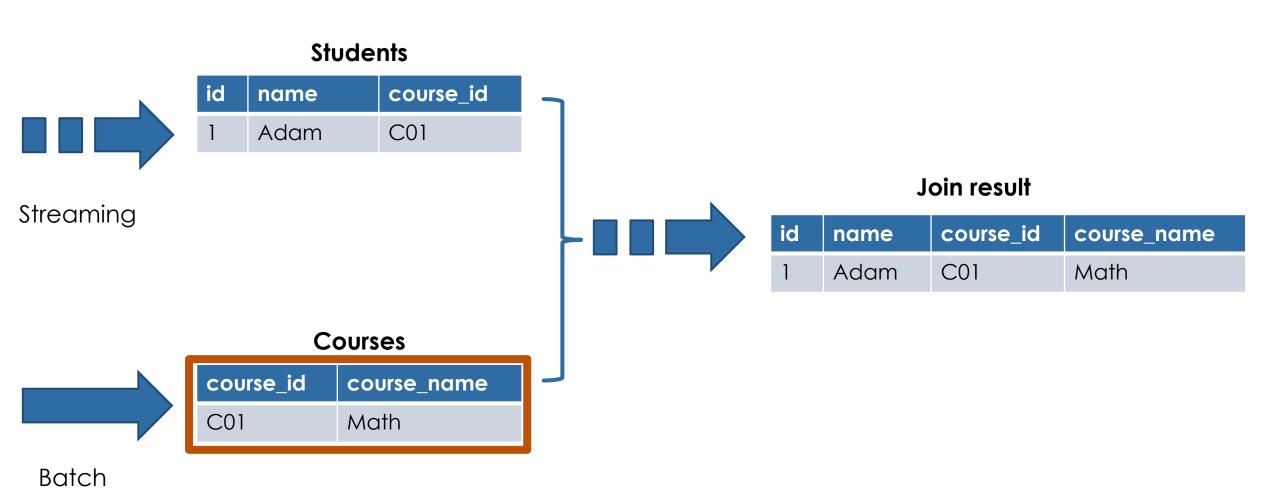
id	name	course_id	course_name
1	Adam	C01	Math

Courses

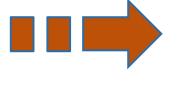


course_id	course_name
C01	Math

Guarantee: latest version of static table



Students



Streaming

id	name	course_id
1	Adam	C01
2	John	C01
3	Sarah	C02

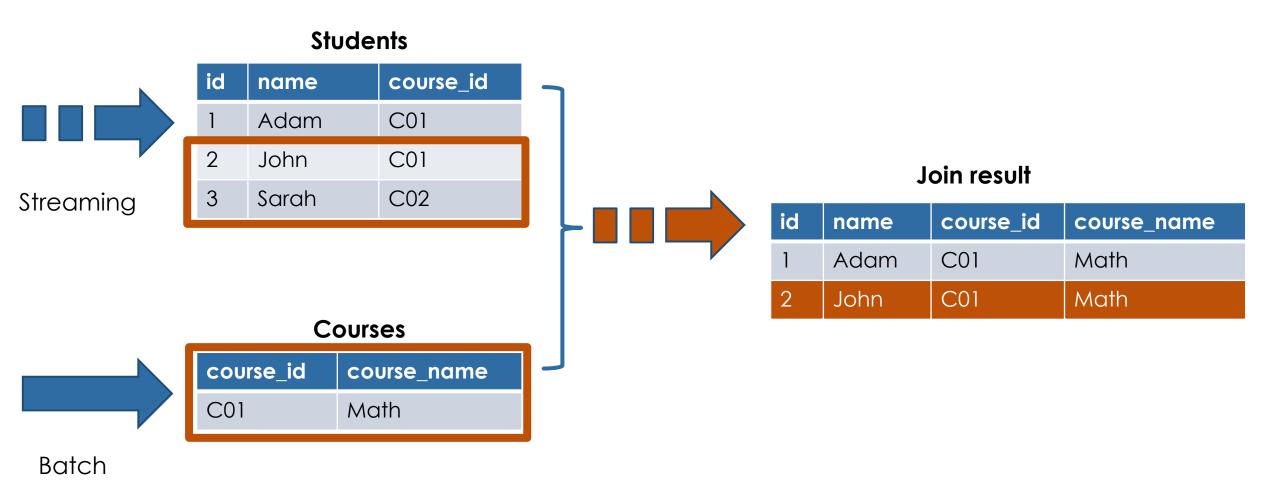
Join result

id	name	course_id	course_name
1	Adam	C01	Math

Courses



course_id	course_name
C01	Math



Students



Streaming

id	name	course_id
1	Adam	C01
2	John	C01
3	Sarah	C02

Join result

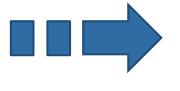
id	name	course_id	course_name
1	Adam	C01	Math
2	John	C01	Math

Courses



course_id	course_name
C01	Math

Students



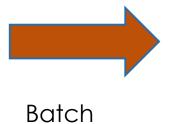
Streaming

id	name	course_id
1	Adam	C01
2	John	C01
3	Sarah	C02

Join result

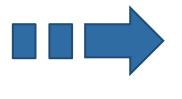
id	name	course_id	course_name
1	Adam	C01	Math
2	John	C01	Math

Courses



course_idcourse_nameC01MathC02Biology

Students

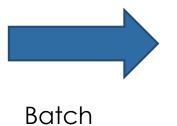


Streaming

id	name	course_id
1	Adam	C01
2	John	C01
3	Sarah	C02

Courses

course_name



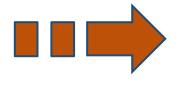
C01	Math
C02	Biology

course_id

id	name	course_id	course_name
1	Adam	C01	Math
2	John	C01	Math

Limitations: Unmatched records

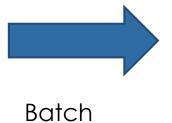
Students



Streaming

id	name	course_id
1	Adam	C01
2	John	C01
3	Sarah	C02
4	Emma	C02
5	Sophia	C03

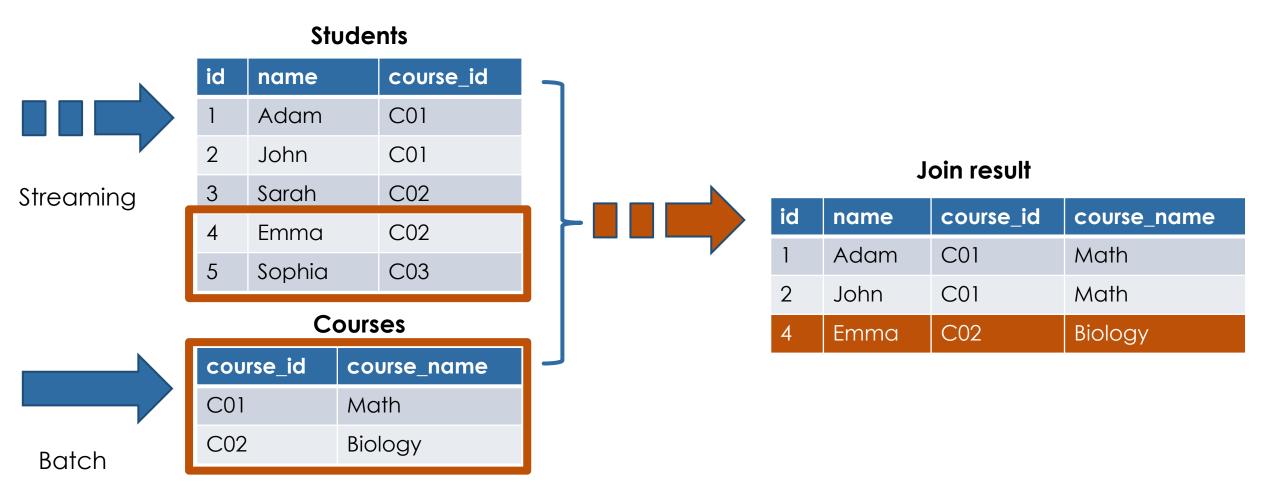
Courses



course_id	course_name
C01	Math
C02	Biology

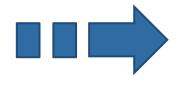
id	name	course_id	course_name
1	Adam	C01	Math
2	John	C01	Math

Limitations: Unmatched records



Stream-static joins are not stateful

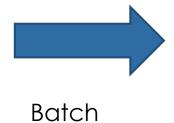
Students



Streaming

id	name	course_id
1	Adam	C01
2	John	C01
3	Sarah	C02
4	Emma	C02
5	Sophia	C03

Courses

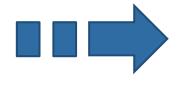


course_id	course_name
C01	Math
C02	Biology

id	name	course_id	course_name
1	Adam	C01	Math
2	John	C01	Math
4	Emma	C02	Biology

Stream-static joins are not stateful

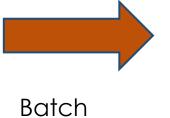
Students



Streaming

id	name	course_id
1	Adam	C01
2	John	C01
3	Sarah	C02
4	Emma	C02
5	Sophia	C03

Courses

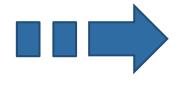


course_id	course_name
C01	Math
C02	Biology
C03	English

id	name	course_id	course_name
1	Adam	C01	Math
2	John	C01	Math
4	Emma	C02	Biology

Stream-static joins are not stateful

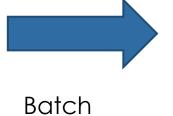
Students



Streaming

id	name	course_id
1	Adam	C01
2	John	C01
3	Sarah	C02
4	Emma	C02
5	Sophia	C03

Courses



course_id	course_name
C01	Math
C02	Biology
C03	English

id	name	course_id	course_name
1	Adam	C01	Math
2	John	C01	Math
4	Emma	C02	Biology