STREAMING SETUP

Event Streams

Hadoop

Hadoop (pre-processing and storage) → Druid

Kafka → Samza → Druid

Insight

2015
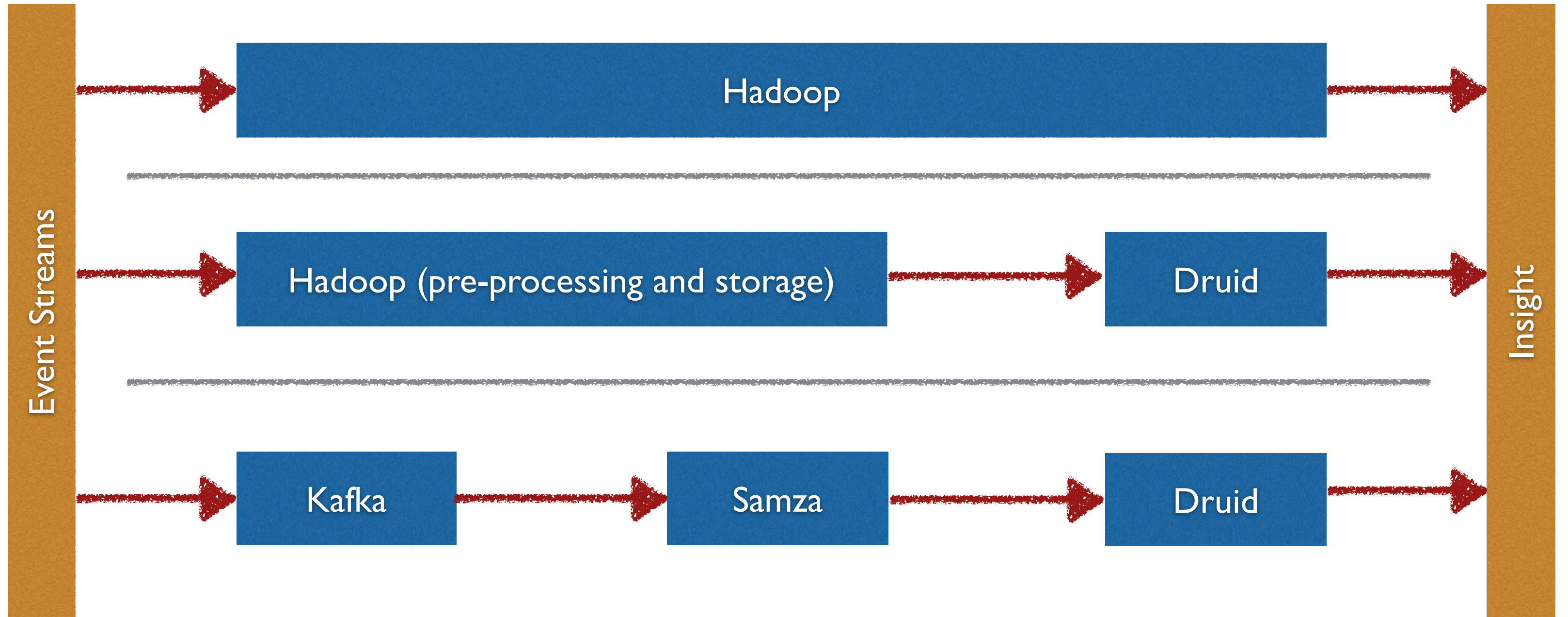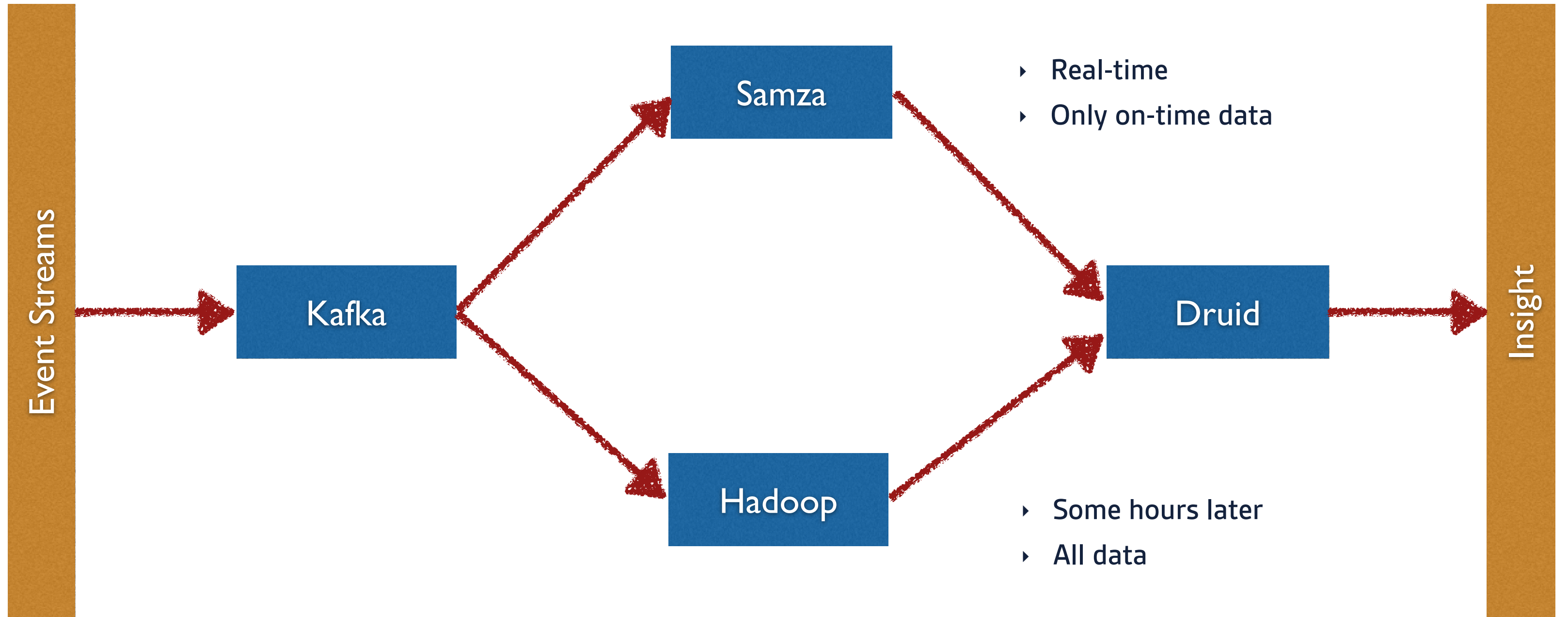
STREAMING ONLY INGESTION

‣ Stream processing isn't perfect

‣ Difficult to handle corrections of existing data

‣ Windows may be too small for fully accurate operations

‣ Hadoop was actually good at these things

OPEN SOURCE LAMBDA ARCHITECTURE

Event Streams → Kafka

Kafka → Samza → Druid
- Real-time
- Only on-time data

Kafka → Hadoop → Druid
- Some hours later
- All data

Druid → Insight

2015

TAKE-AWAYS

‣ When Druid?

- You want to power user-facing data applications
- You want to do your analysis on data as it's happening (realtime)
- Arbitrary data exploration with sub-second ad-hoc queries
- OLAP, BI, Pivot (anything involved aggregates)
- You need availability, extensibility and flexibility

2015

DRUID IS OPEN SOURCE

WWW.DRUID.IO

twitter **@druidio**

irc.freenode.net **#druid-dev**

MY INFORMATION

FJ@IMPLY.IO

twitter @fangjin

LinkedIn fangjin

THANK YOU