

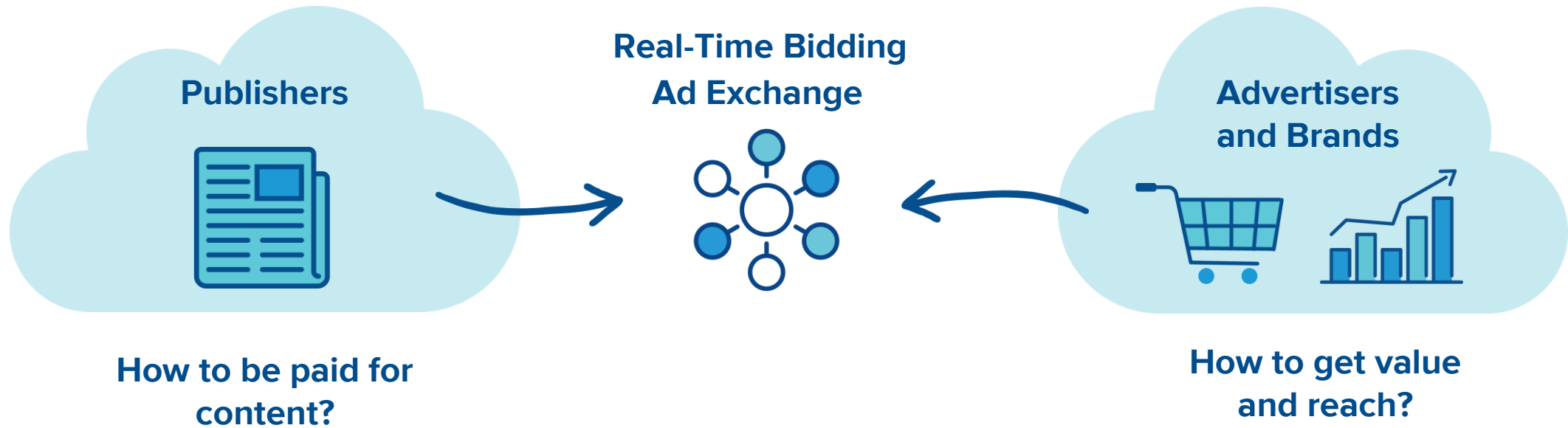
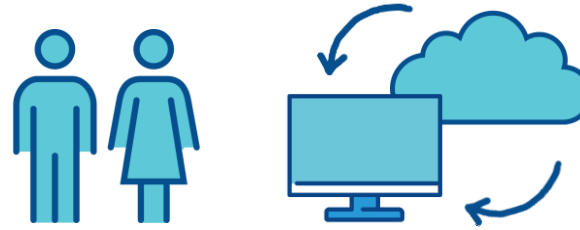


# The Trade Desk's Year with Flink

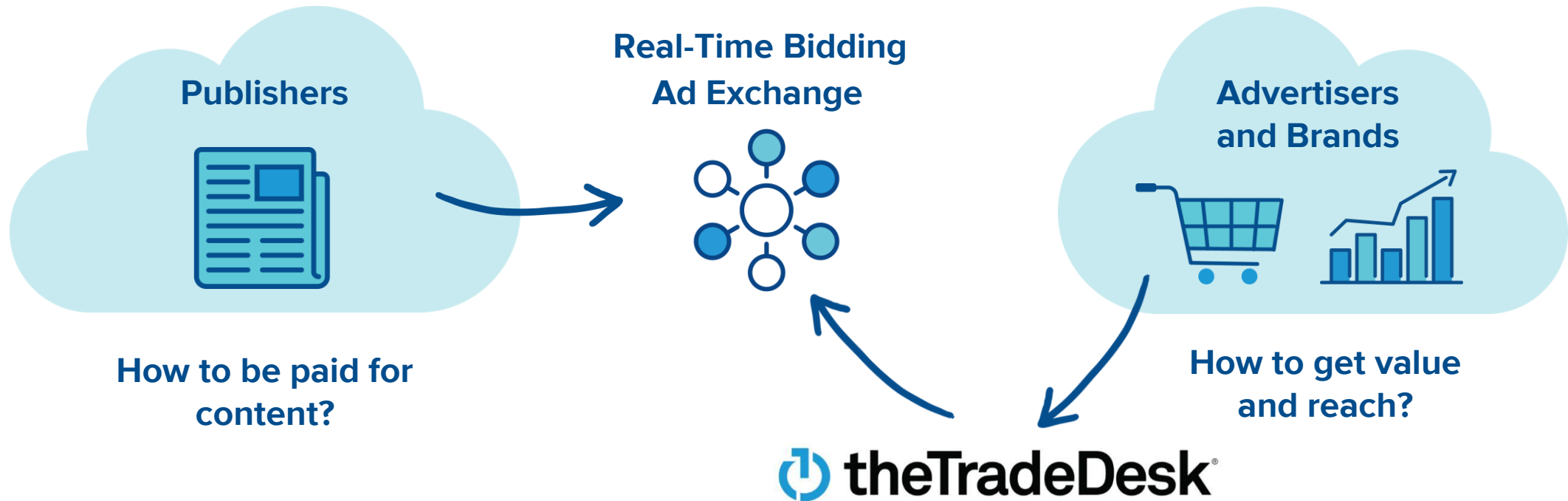
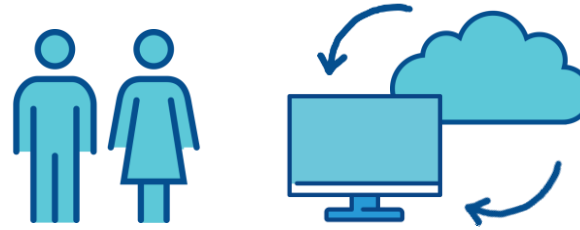
Jonathan Miles



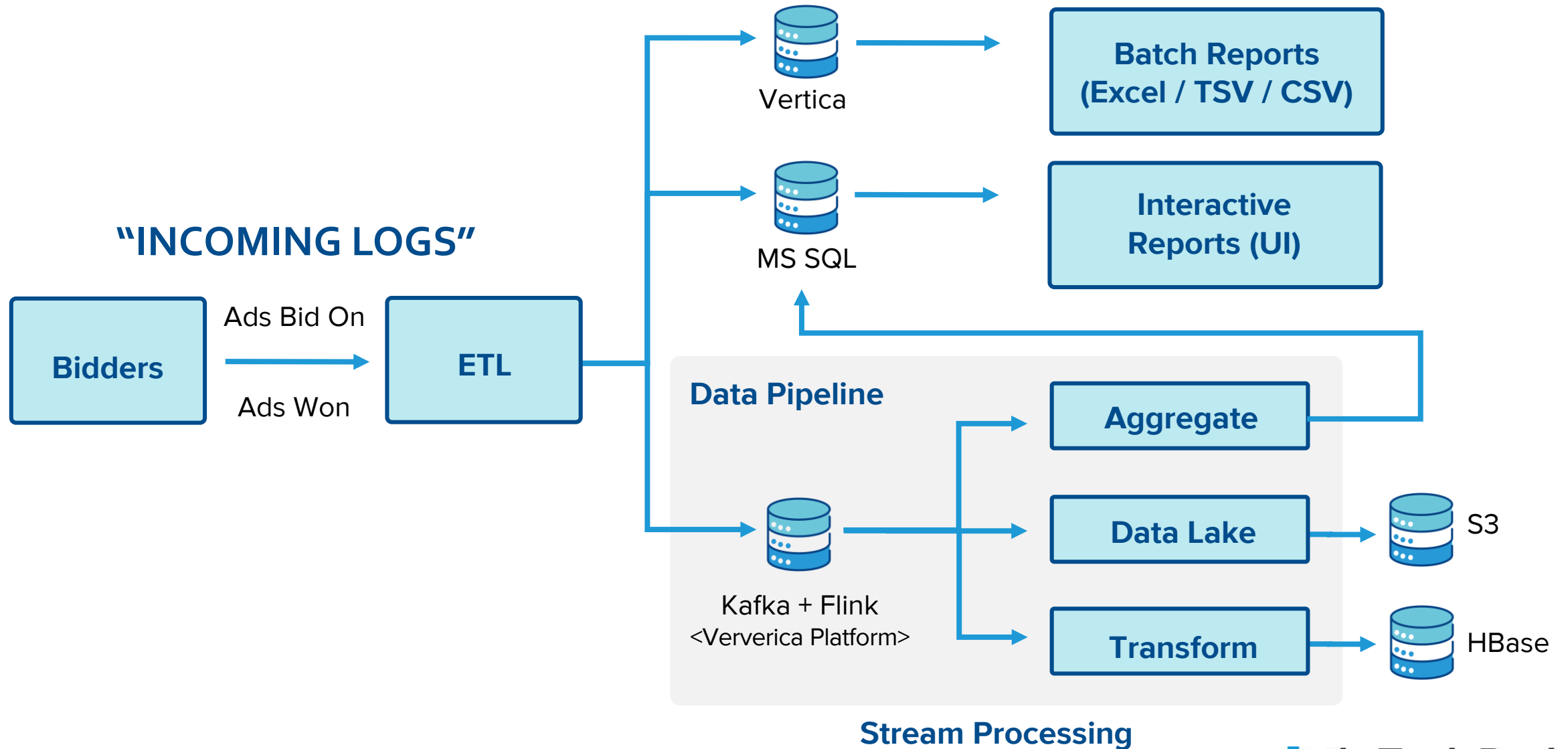
# The Crux of (or Intro to) AdTech



# The Crux of (or Intro to) AdTech



# The Trade Desk Data Systems



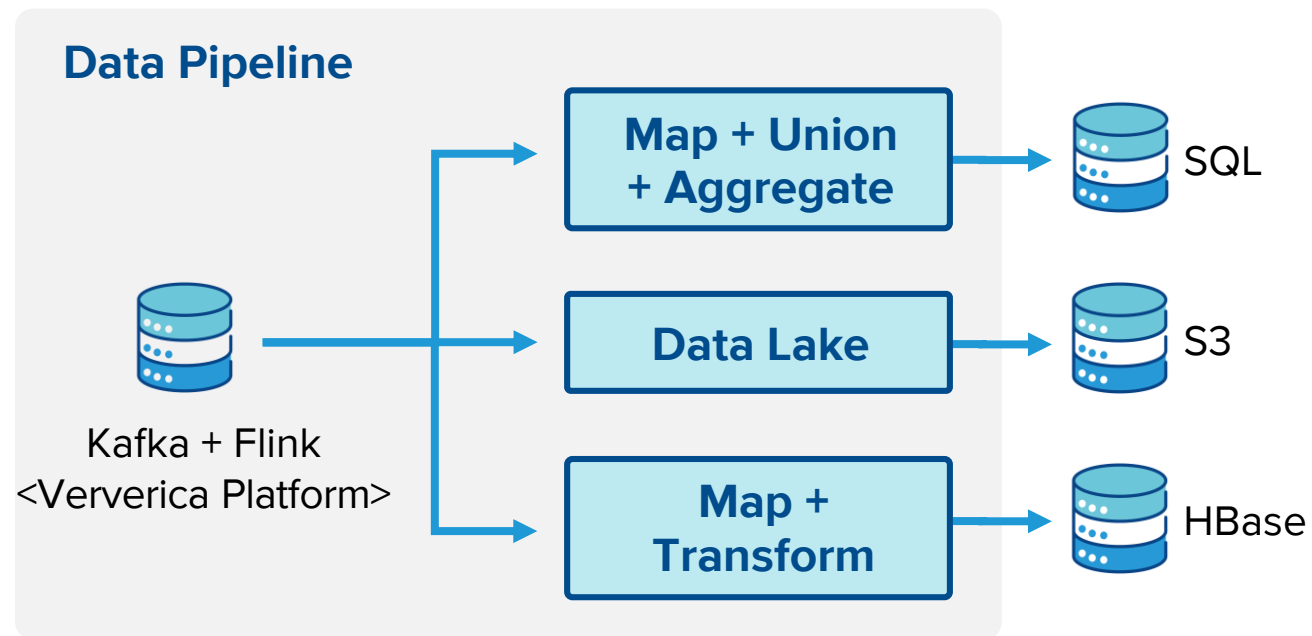




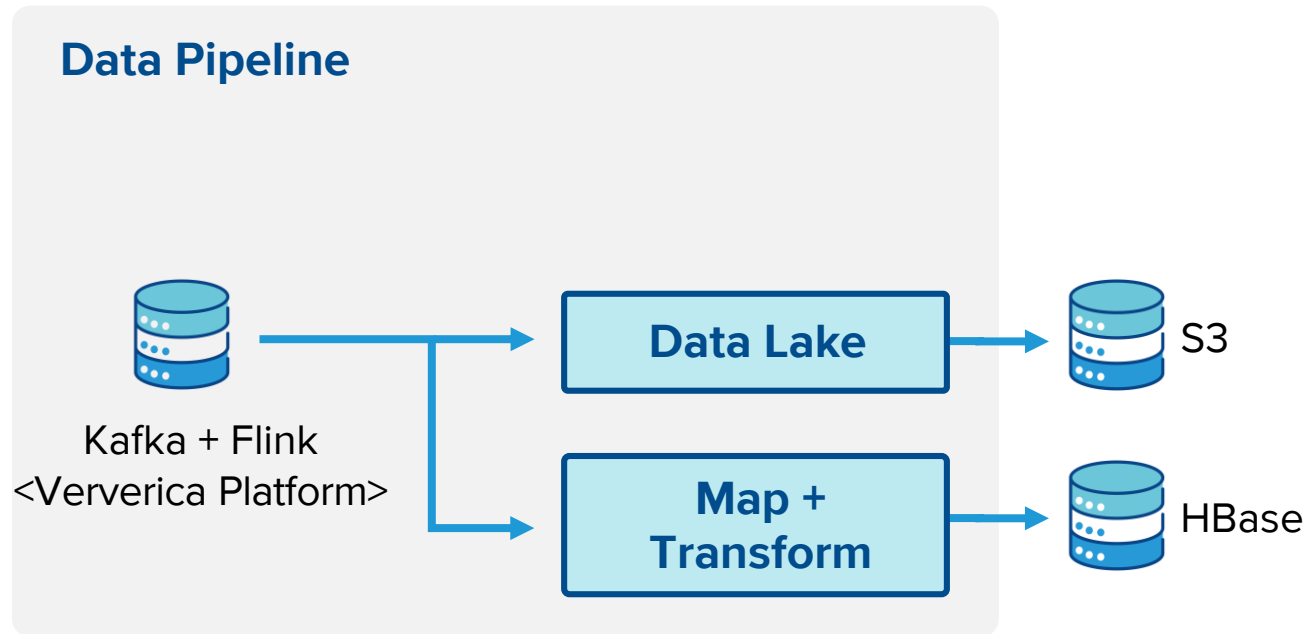
# Flink Highlights

- Apache Flink
  - High-level/abstract, lower testing surface area
  - Most of the distributed systems work is done for you
  - Common sysadmin challenges
- Ververica Platform
  - Orchestration with App Manager on Kubernetes
- Ververica Support
  - When embracing open-source, The Trade Desk needed "enterprise grade" software (revenues)
  - Lean on Ververica, allowing us to work on our business
  - Gateway to Flink community

# Overview of Flink jobs at TTD

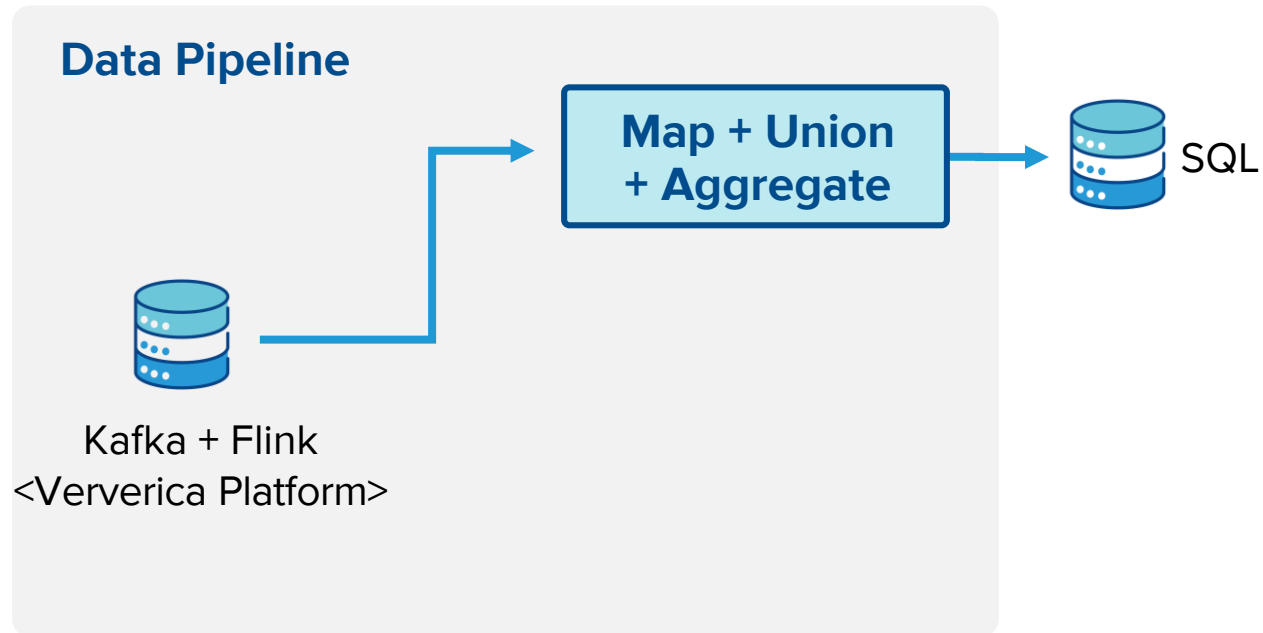


# Anatomy: simpler jobs



- Data Lake
  - Five jobs, one for each input topic
  - Scaled independently
  - Internal Data Science
  - ... also advertising partners
- Map + Transform
  - Two jobs, for different topics
  - Single input and output to Hbase
  - More Data Science use cases

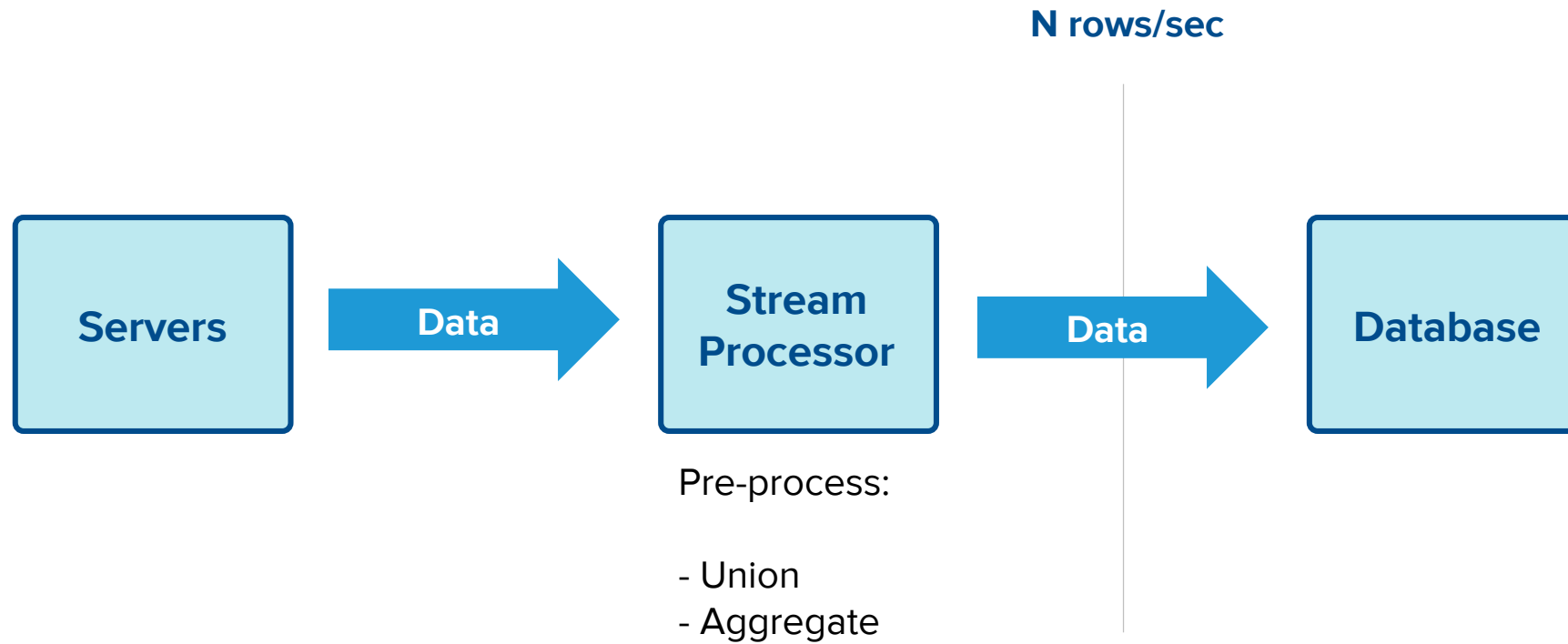
# Anatomy: challenging jobs



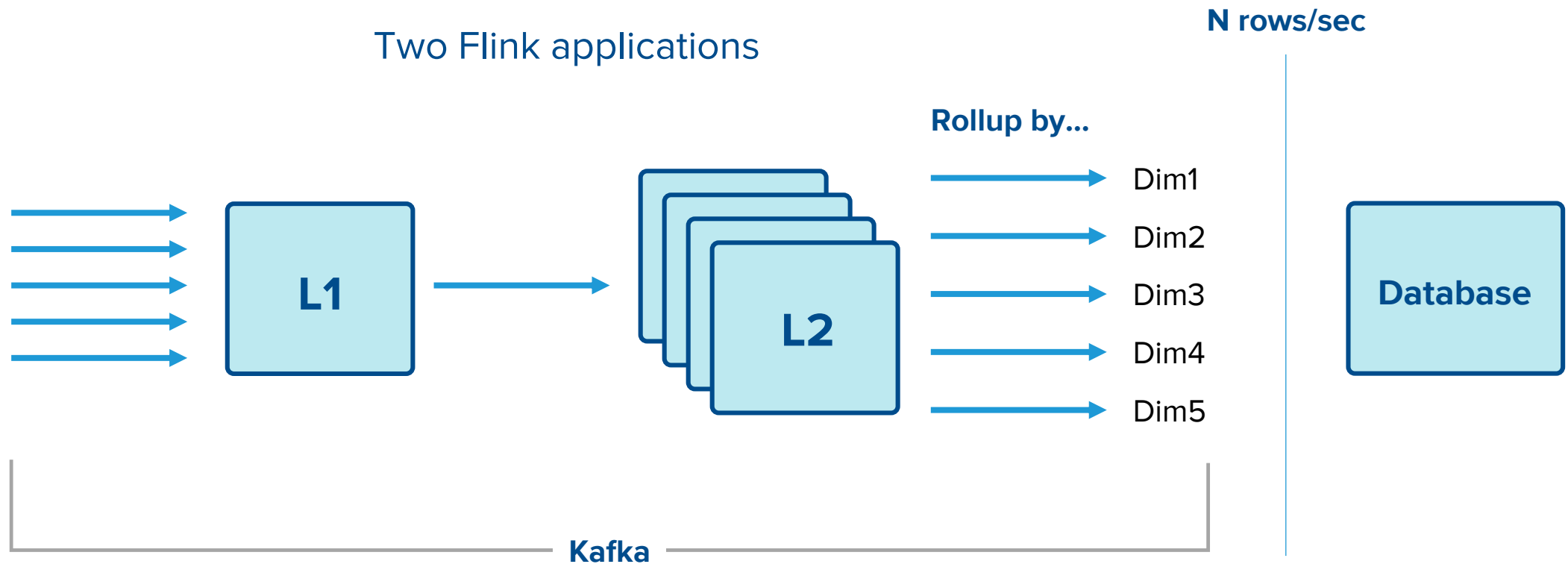
- Map + Aggregate (Real-Time Insights)
  - Aggregate disparate input topics
  - Brands can query DB interactively
- Stream processing
  - Reasonable latency
  - Reduce complexity of management
- Synchronise five input topics with
  - ... different data rates
  - ... different time characteristics



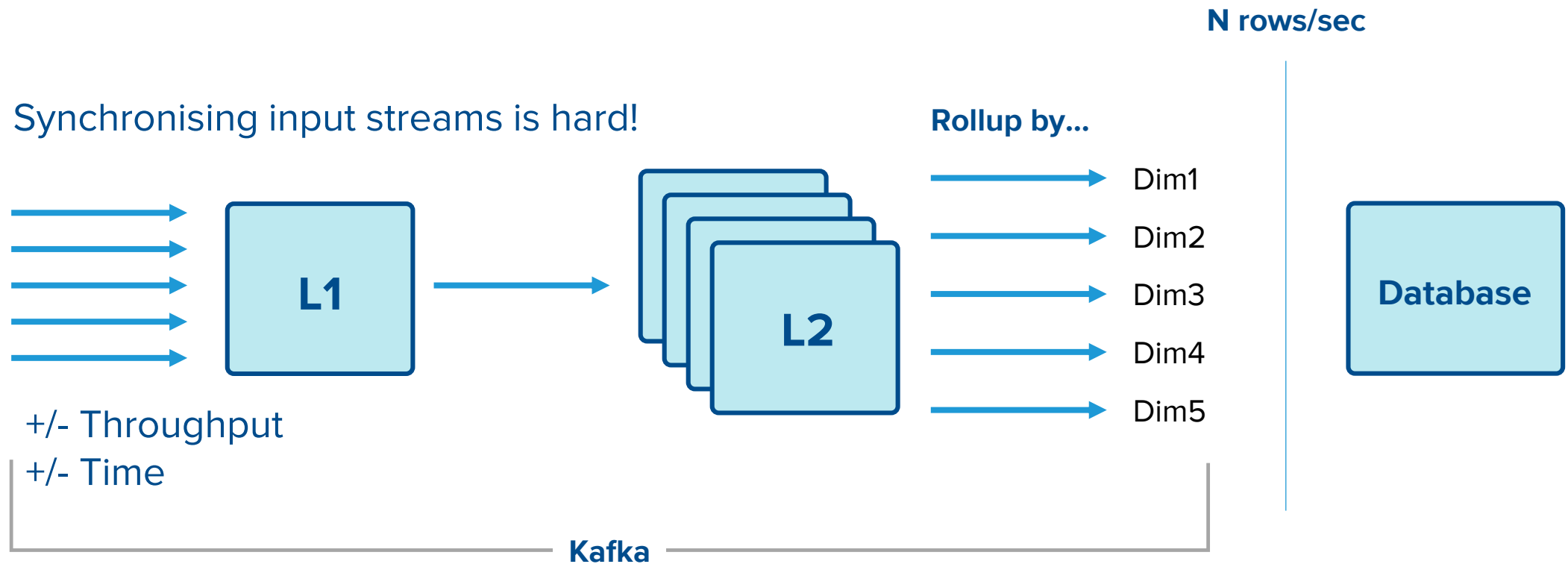
# Anatomy: challenging jobs



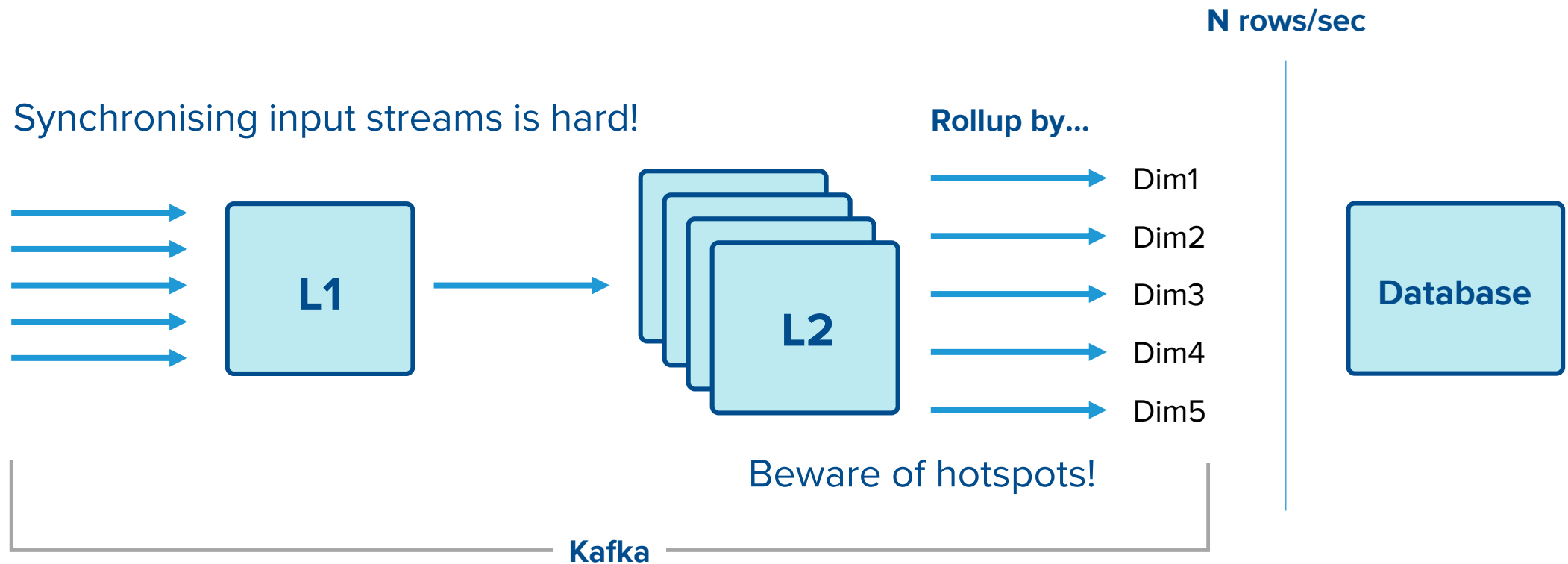
# Anatomy: challenging jobs



# Anatomy: challenging jobs

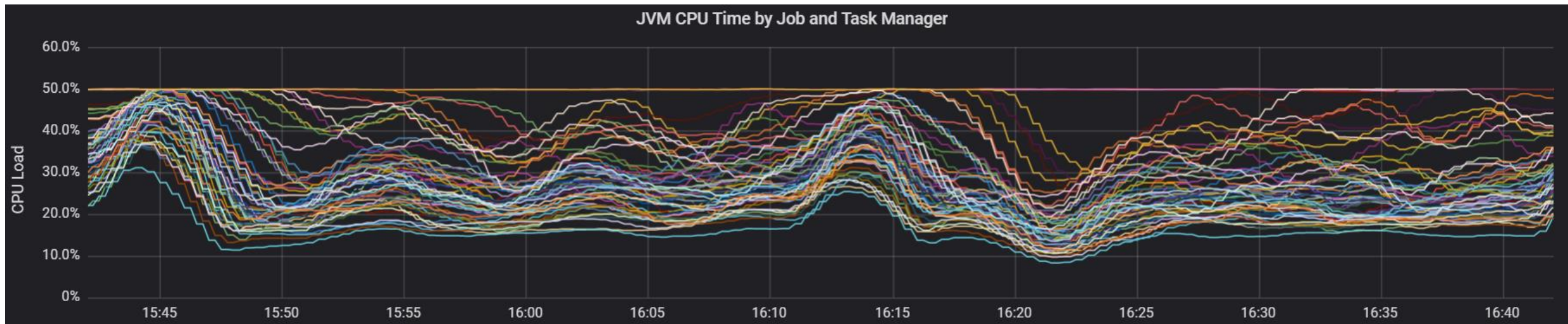


# Anatomy: challenging jobs



# Anatomy: challenging jobs

Hotspots...

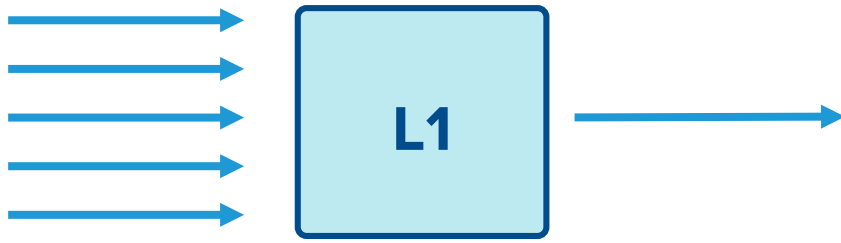


... caused global slow-down during catchup.



# Tips: synchronizing input streams is hard

Synchronising input streams is hard!



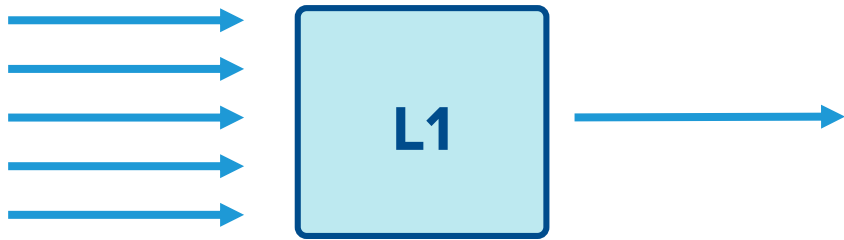
**SQL Table**

Key	T	I1C1..n	I2C1..n	I3C1..n	I4C1..n	I5C1..n
K1	T1	Value	Value	Value	Value	Value
K2	T2	Value	Value	Value	Value	Value

```
SELECT SUM(I1.C1), SUM(I1.C2)...  
FROM Input1 I1  
  JOIN Input2 I2 ON (I2.key=I1.key...)  
  JOIN Input3 I3 ON (I3.key=I1.key...)  
  JOIN Input4 I4 ON (I4.key=I1.key...)  
  JOIN Input5 I5 ON (I5.key=I1.key...)  
WHERE Input1.key = @key  
GROUP BY Key, Window
```

# Tips: synchronizing input streams is hard

Synchronising input streams is hard!

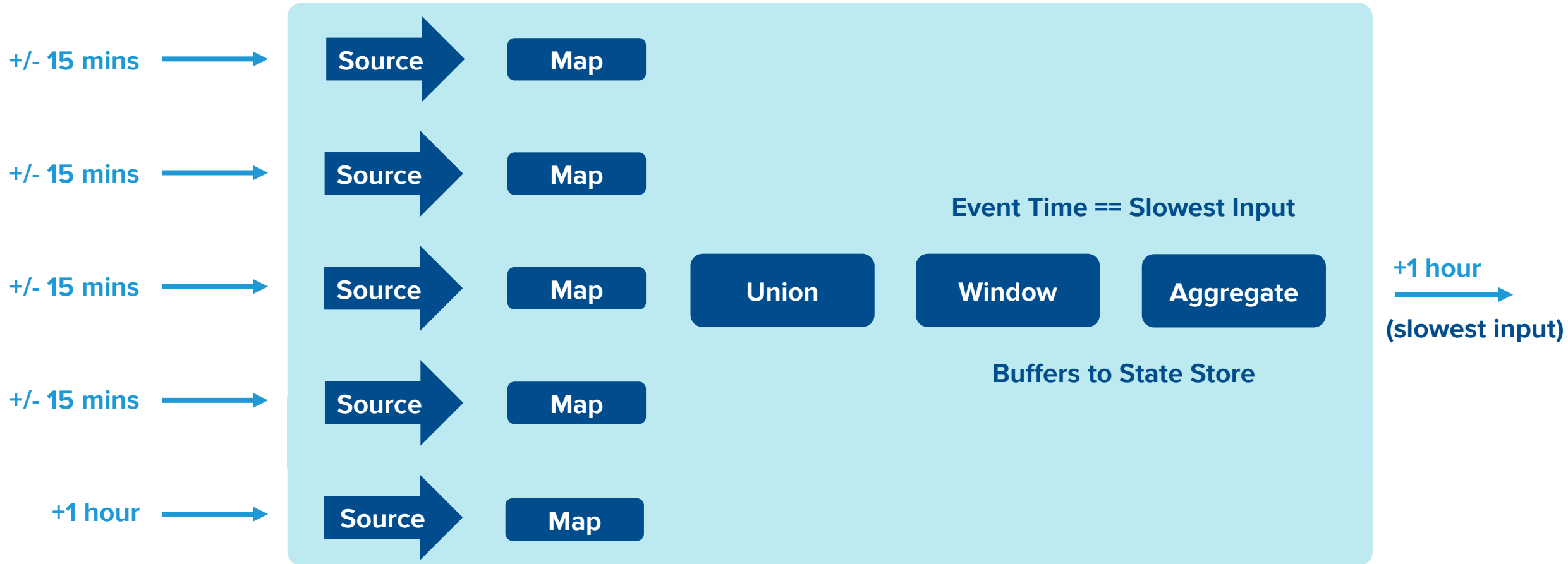


THREE CHALLENGES:

- **Time synchronisation**
- Throughput synchronisation
- Late data

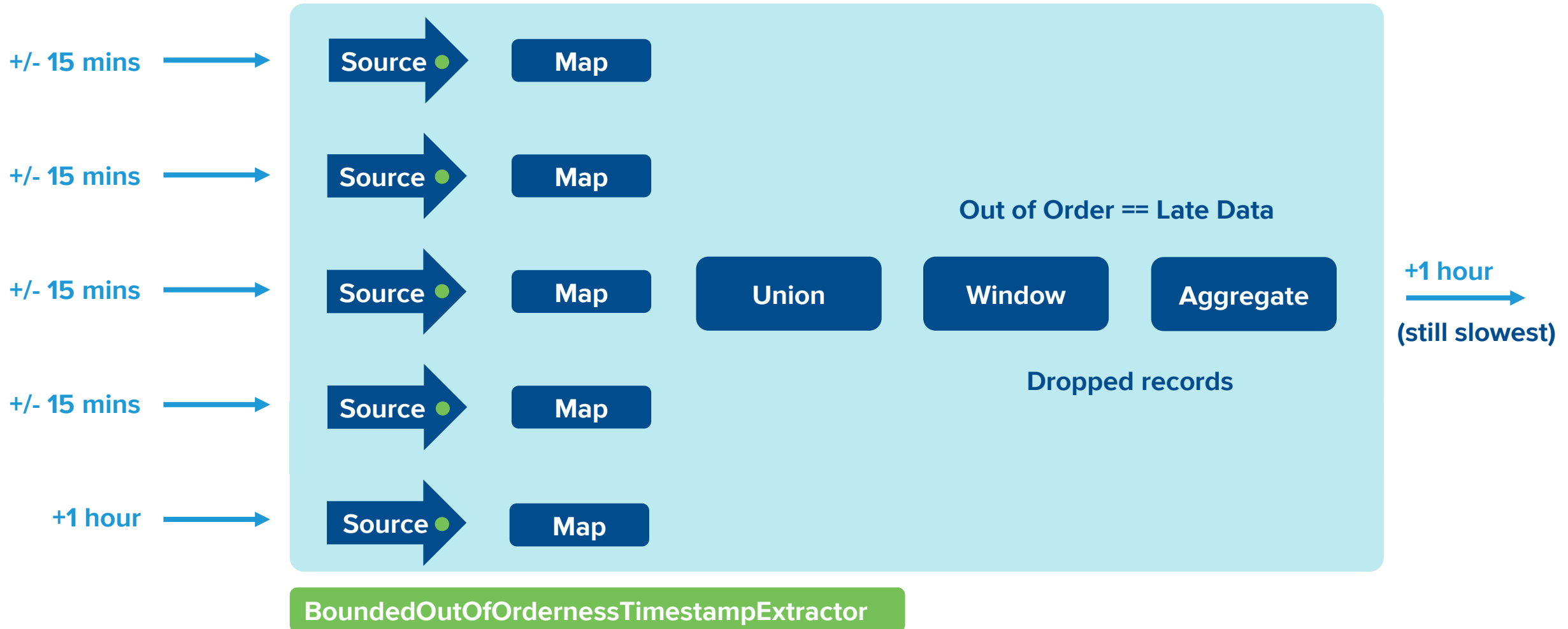
# Tips: synchronizing input streams is hard

## Time synchronisation



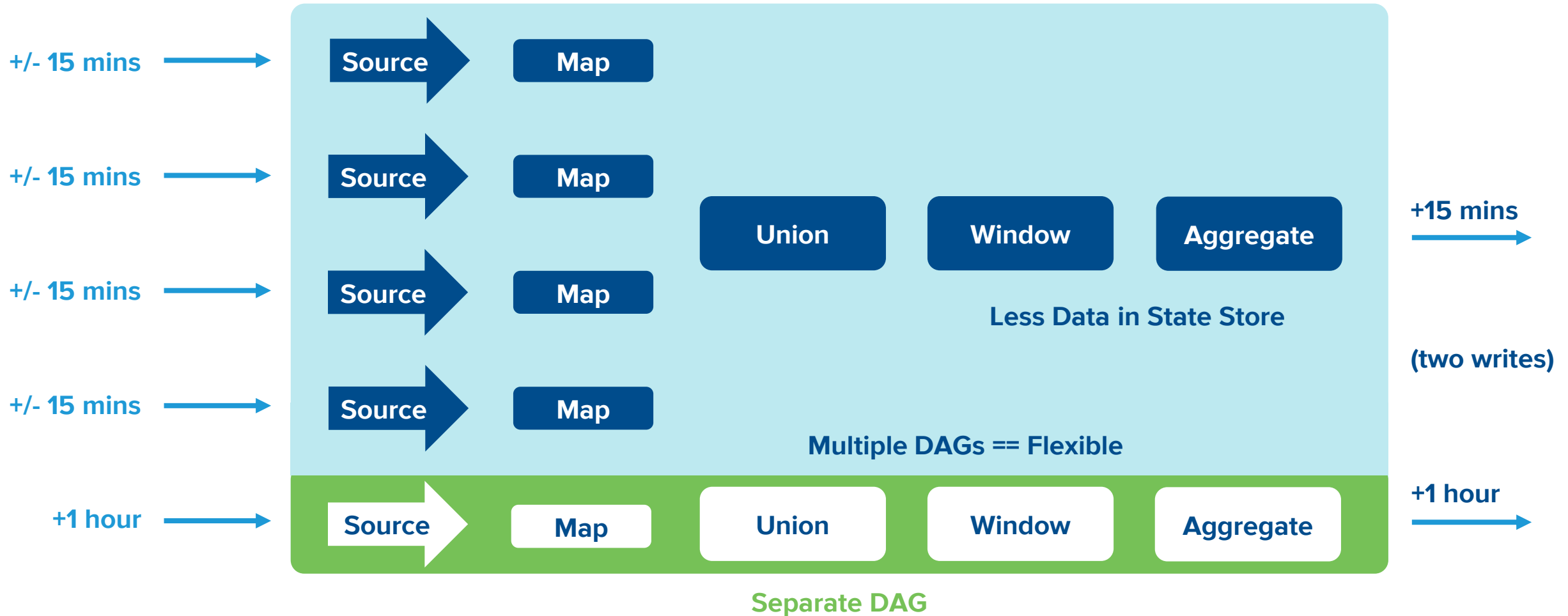
# Tips: synchronizing input streams is hard

## Time synchronisation (out of order)



# Tips: synchronizing input streams is hard

## Time synchronisation (different latency)





# Tips: synchronizing input streams is hard

Synchronising input streams is hard!

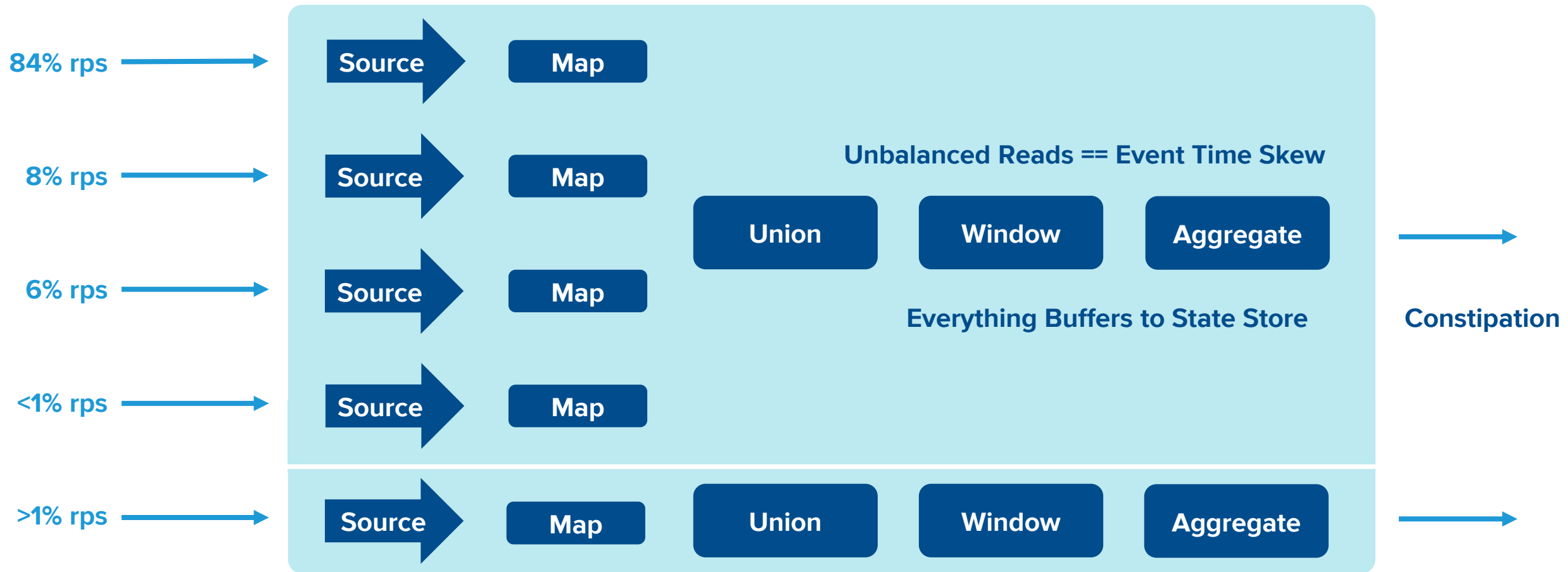


THREE CHALLENGES:

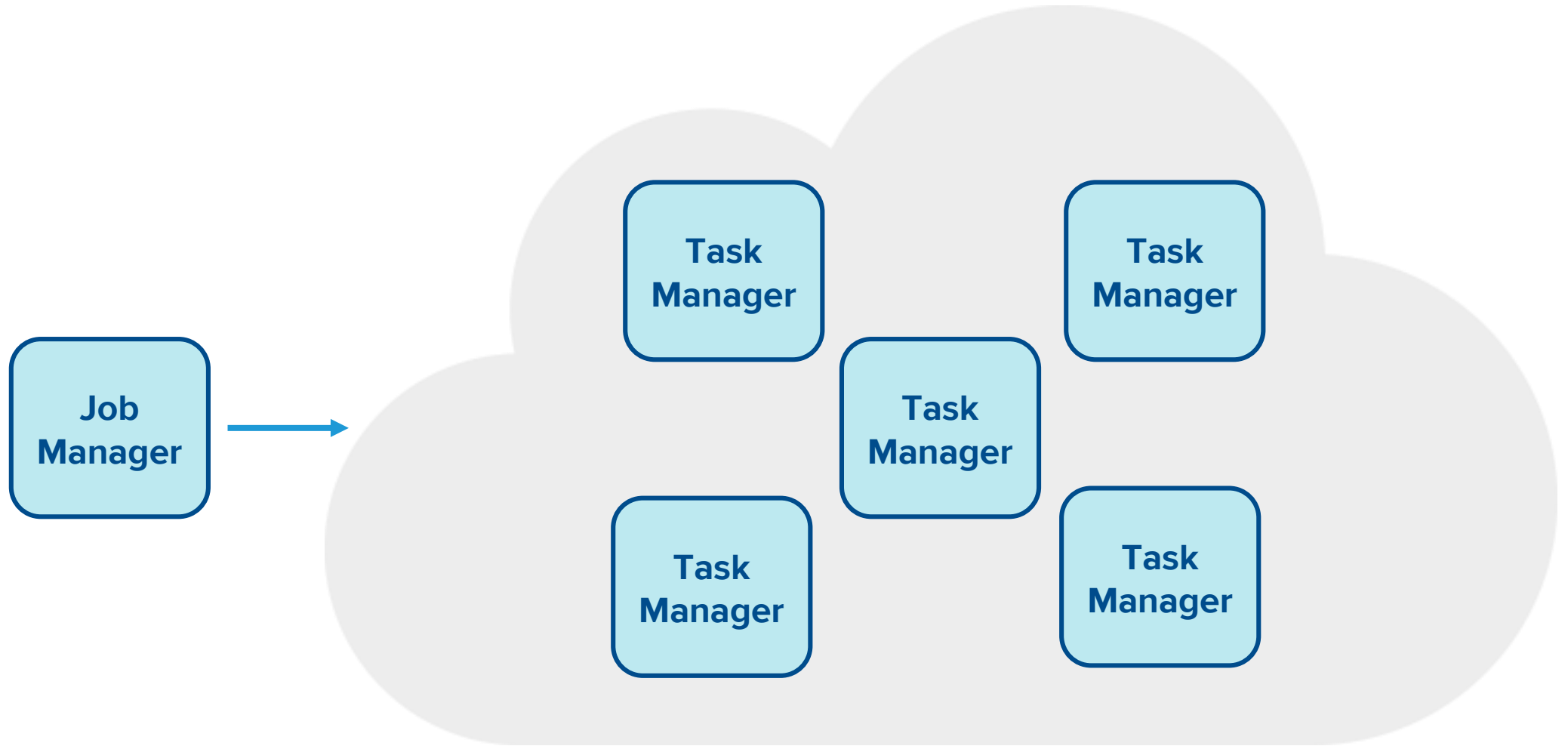
- ~~Time synchronisation~~
- **Throughput synchronisation**
- Late data

# Tips: synchronizing input streams is hard

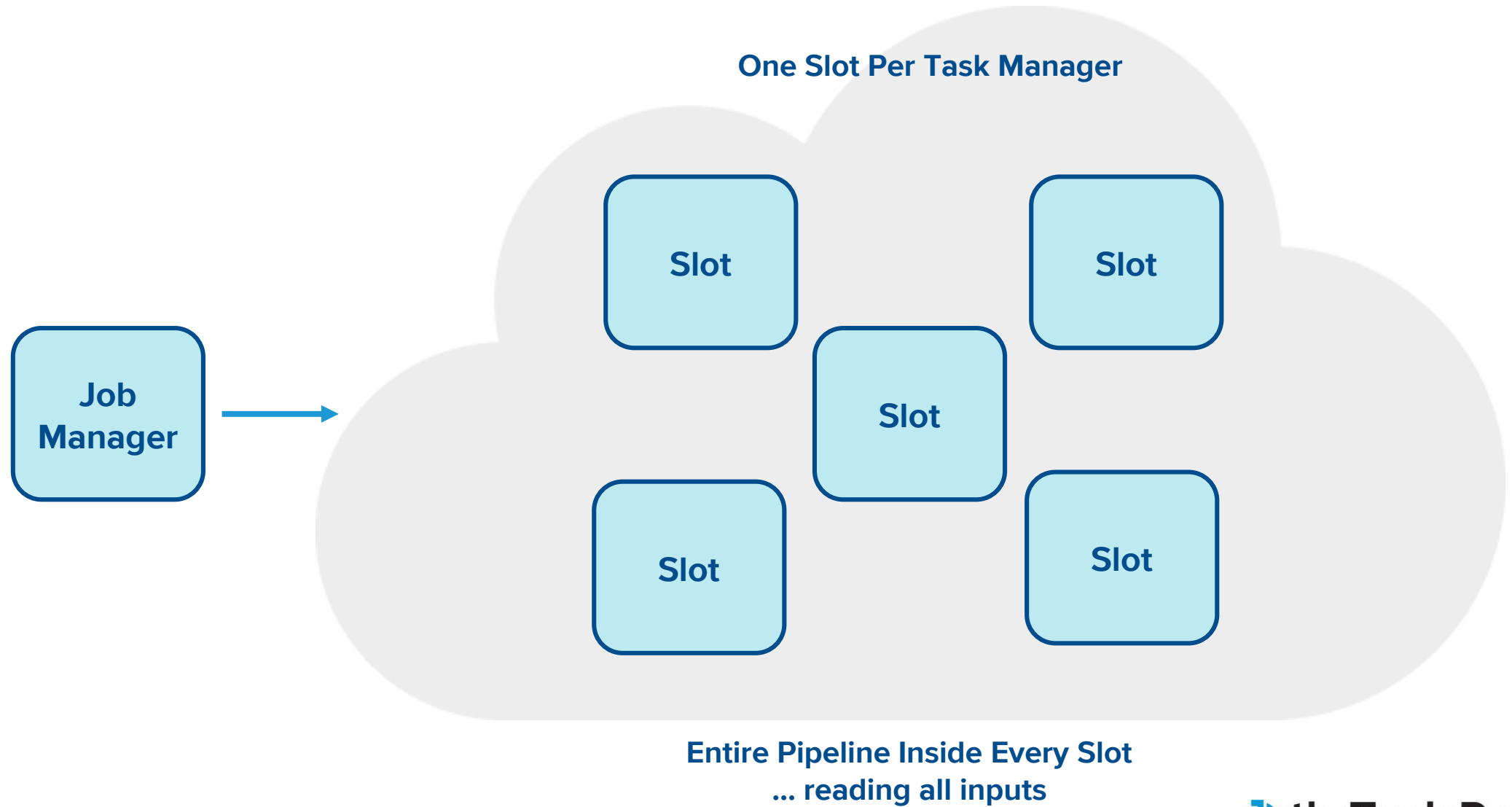
## Throughput balancing



Tips: synchronizing input streams is hard



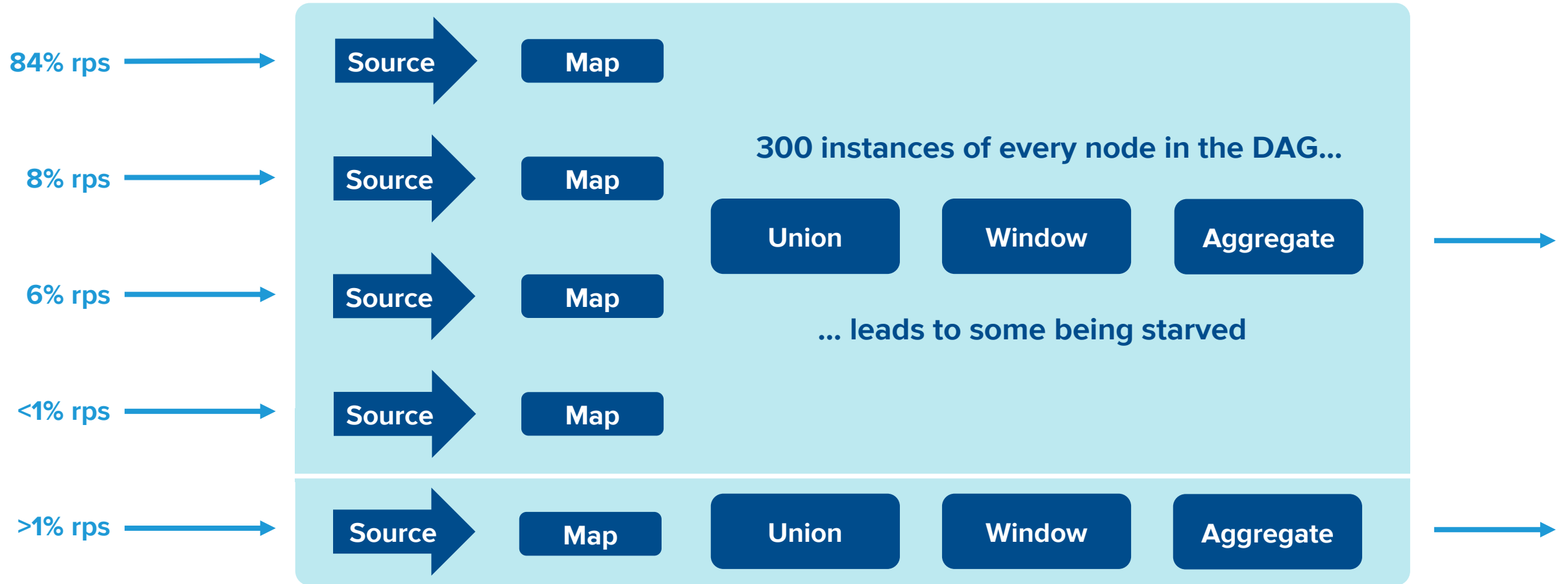
Tips: synchronizing input streams is hard



# Tips: synchronizing input streams is hard

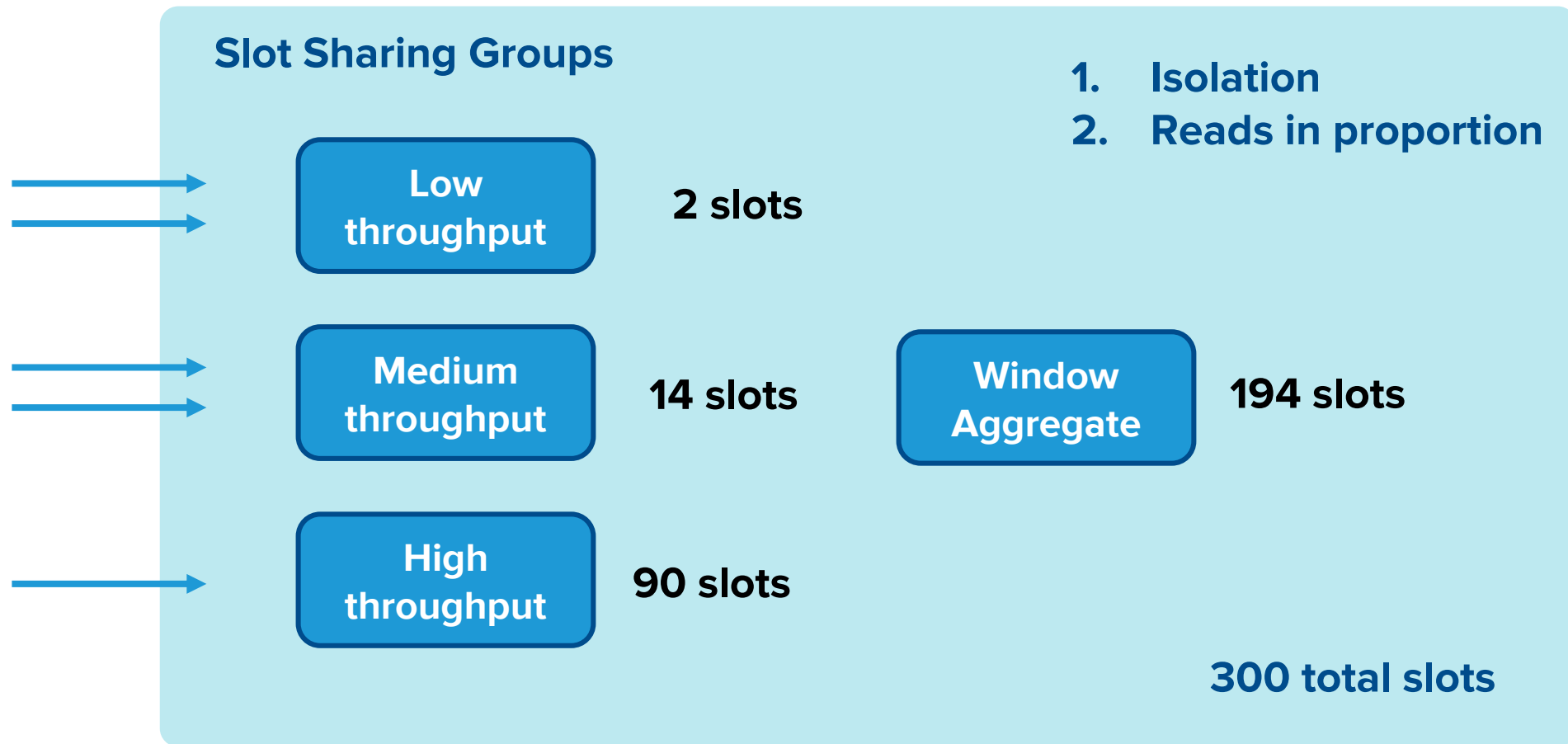
## Throughput balancing

### Every Slot Has These Two DAGs (e.g.)

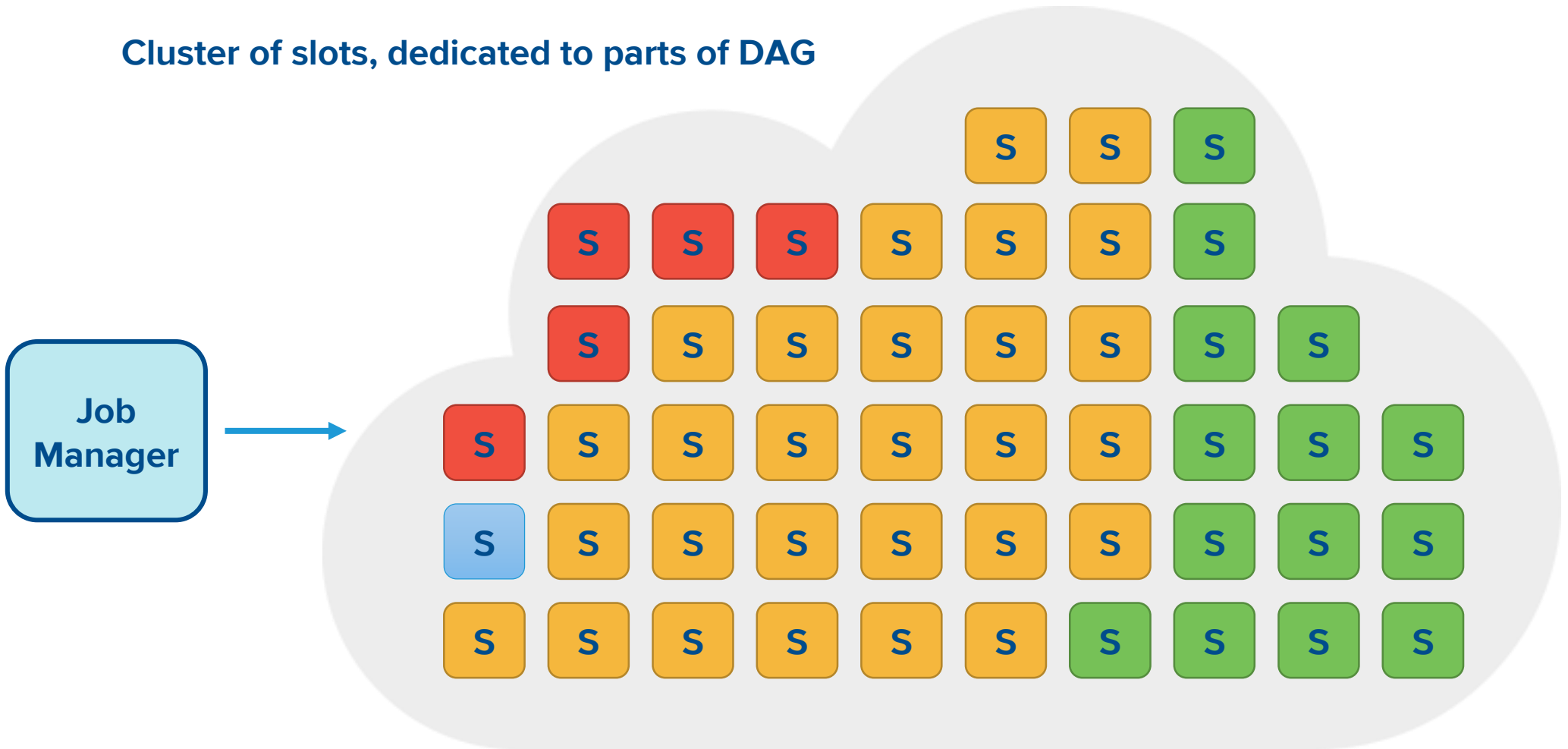




Tips: synchronizing input streams is hard









Tips: synchronizing input streams is hard



... ongoing work to make this easier out of the box

## Tips: synchronising input streams is hard

	Shared Slots (default)	Slot Groups
Control rate of consumption		
Easy attribution of effects		
Even distribution of resource		

# Tips: synchronizing input streams is hard

Synchronising input streams is hard!

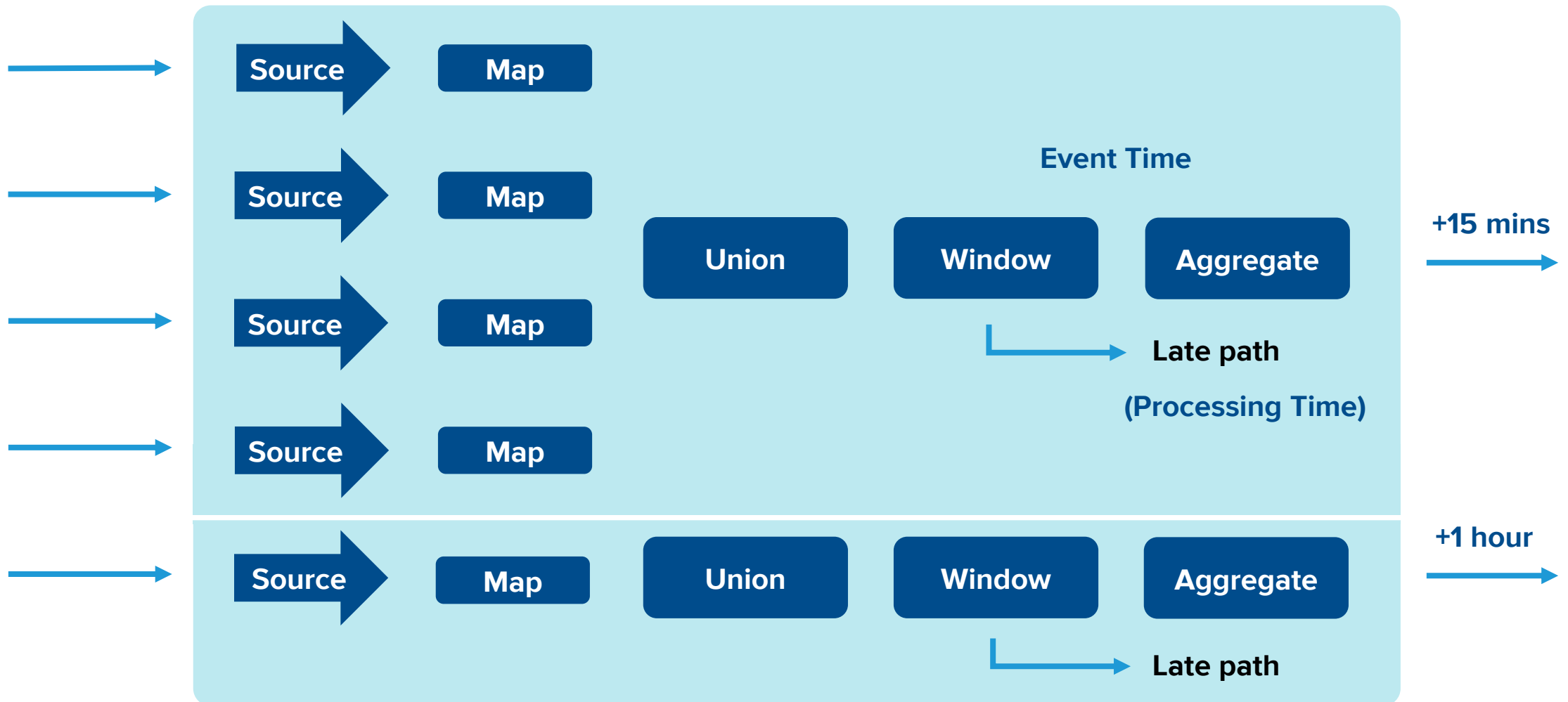


## THREE CHALLENGES:

- ~~Time synchronisation~~
- ~~Throughput synchronisation~~
- **Late data**

Tips: synchronizing input streams is hard

Late Data







## Tips: synchronizing input streams is hard

- Flink's flexibility and demanding use-case
- Unbalanced time
  - BoundedOutOfOrdernessTimestampExtractor (+/- 15 minutes)
  - Separate DAG in same job (+1 hour)
  - Actual late data shunted to separate path
- Unbalanced throughput, Flink won't balance reads for you
  - TTD: Slot Model → Slot Groups
  - Future/community: Event time synchronization across sources ([FLINK-10886](#))
- Constipation and laxative
  - Time-based distribution of individual topic
  - Processing-time eviction
- Real-time data is okay, catchup is main problem



# Future

- Why was Flink a good choice?
  - Generalised framework, well defined semantics
  - Most distributed systems problems under control
  - Flexibility
- CDC project
  - Generalised/democratised access to snapshots of database tables
- What Flink features are we excited about?
  - Event time synchronization across sources ([FLINK-10886](#))
  - Ability to take checkpoints during high backpressure (alignment sentinels stuck in same queue as messages)
- We can dream...
  - Generalised solution to hotspots problem



# Plug

- Diagnostic tool, flink-diag.py
  - GitHub?
  - Come talk to us?
- We're hiring
  - Come change the world of advertising with us!
- Questions?
  - Many people from The Trade Desk attending today
  - Feel free to come talk to us!

Thank you

