

My Awful Ex Data Systems

Data Systems and Relationships

- **Lots of effort**
- **Makes you unhappy**
- **Gives nothing back**

The One Night Stand

- **It's fast and exciting!**
- **It's just this once, no need to take precautions**
- **Mostly happens when your brain is... less rational...**

All Code is Forever
(Until it gets deleted)

All Code is Practise

The Bore

- **Nothing to be excited about**
- **A bit rundown and broken**
- **Probably quite old**
- **“It doesn’t matter”**

Data Warehouses

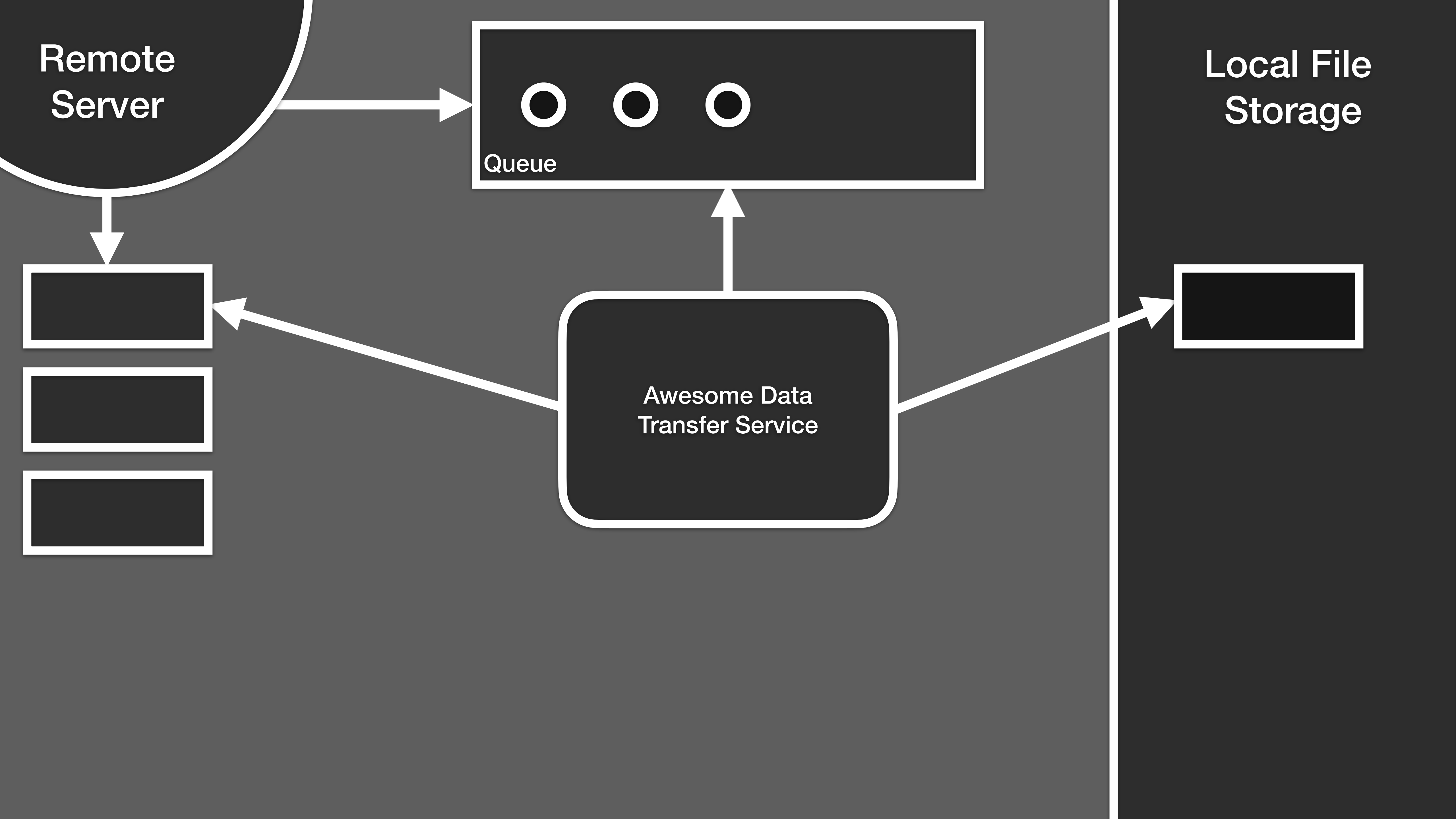
Boring Software Needs Love Too

- **Sometimes boring partners flower with love**
- **Have the fewest moving parts possible**
- **Less code is good code**
- **Focus on making interactions easy and reliable**

Remote
Server



Local File
Storage

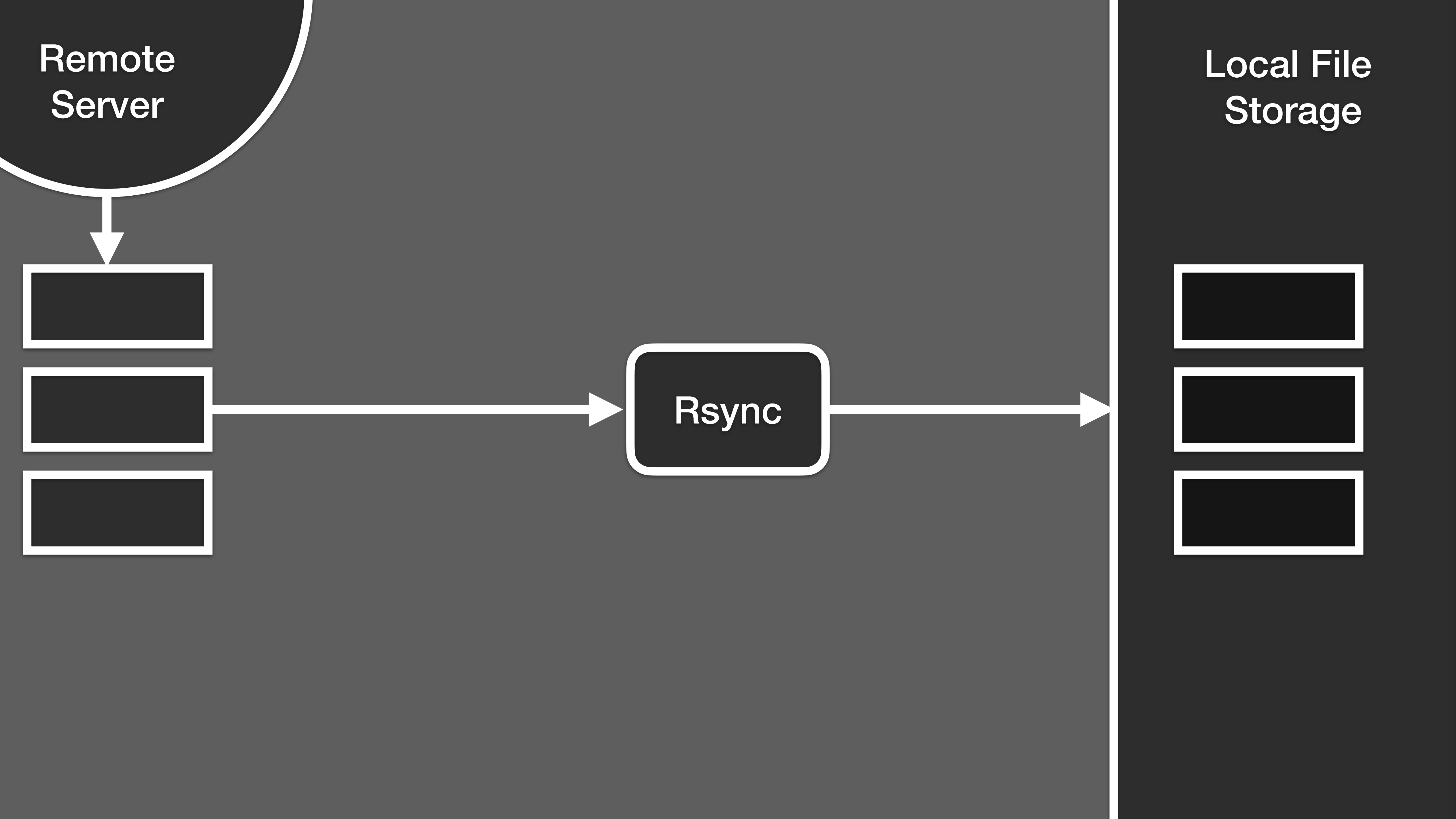


Remote
Server

Queue

Local File
Storage

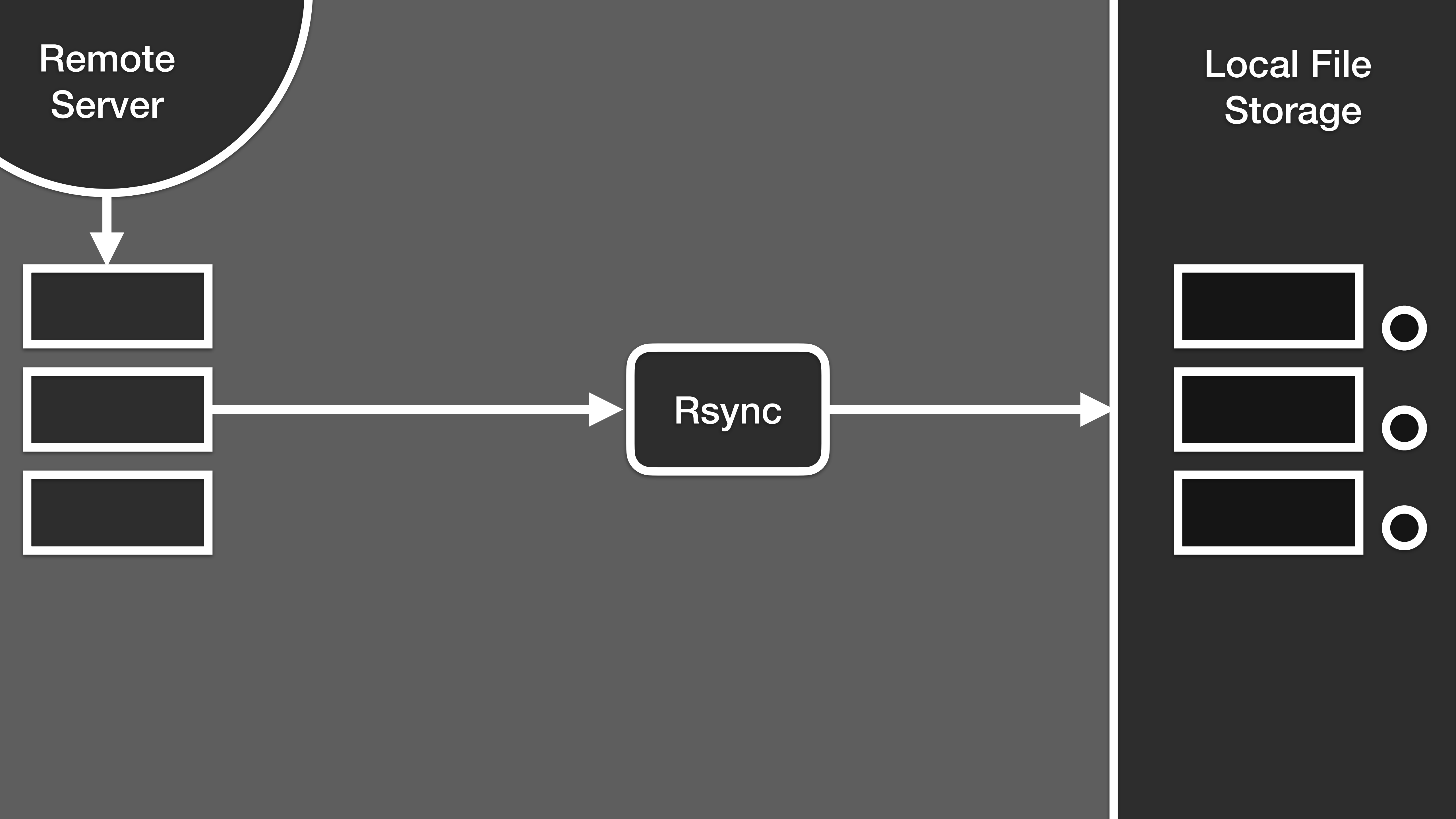
Awesome Data
Transfer Service



Remote
Server

Local File
Storage

Rsync



Remote
Server

Local File
Storage

Rsync

The Backseat Driver

- **Keeps forcing you to do things a certain way**
- **Gets really whiny if you don't give in**
- **Never considers what you want**

Frameworks Bad
Libraries Good

Everything has a cost

The Shiny New Toy

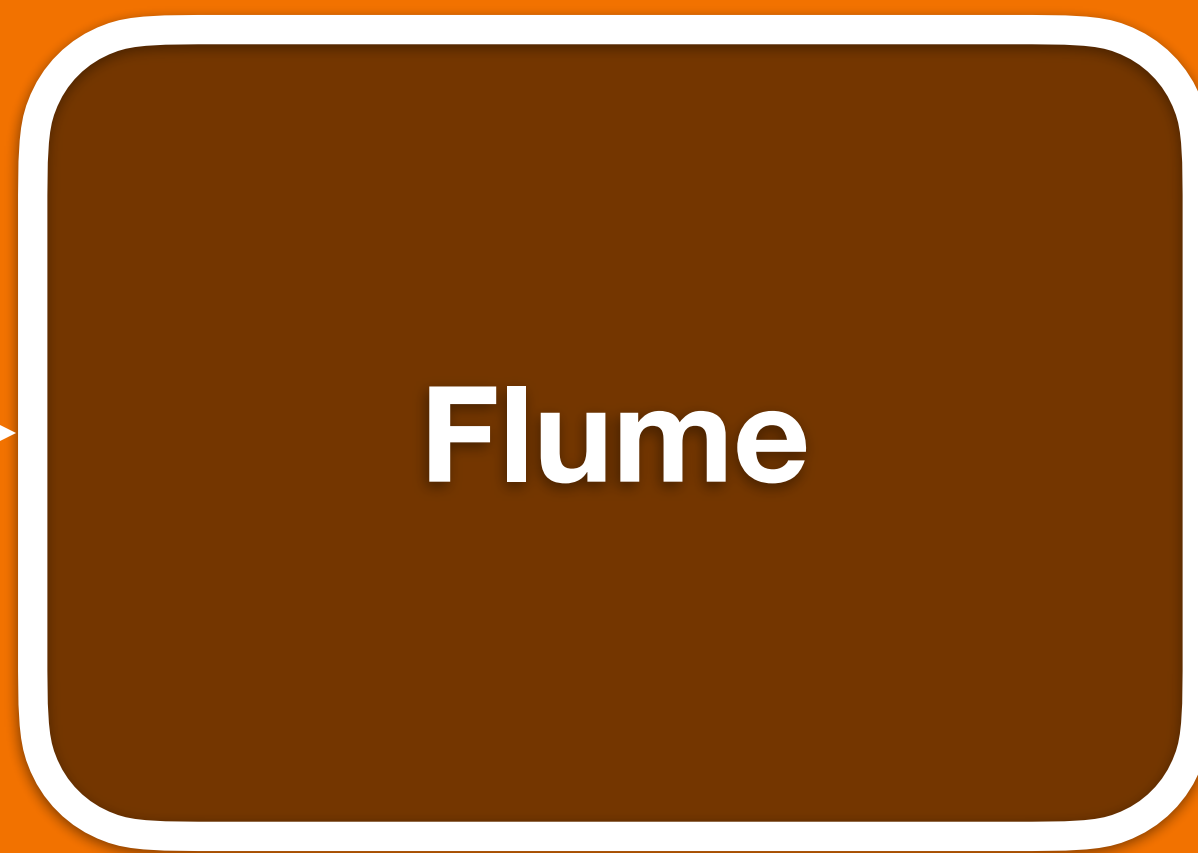
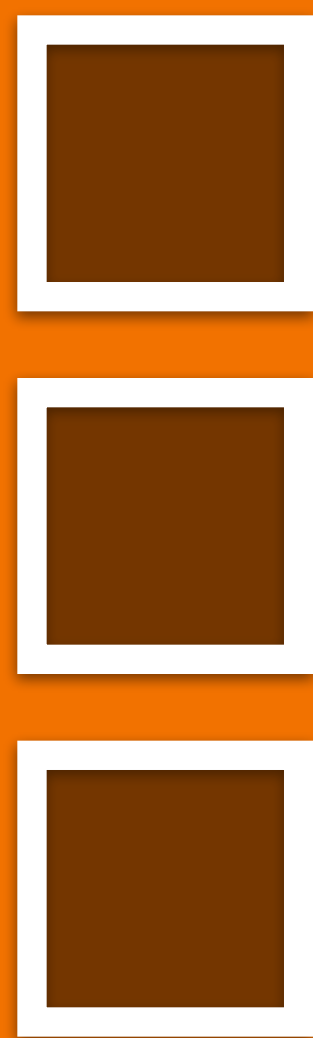
- **Lots of Hype**
- **Promises to solve all problems**
- **No negatives in any possible way**

The Loser

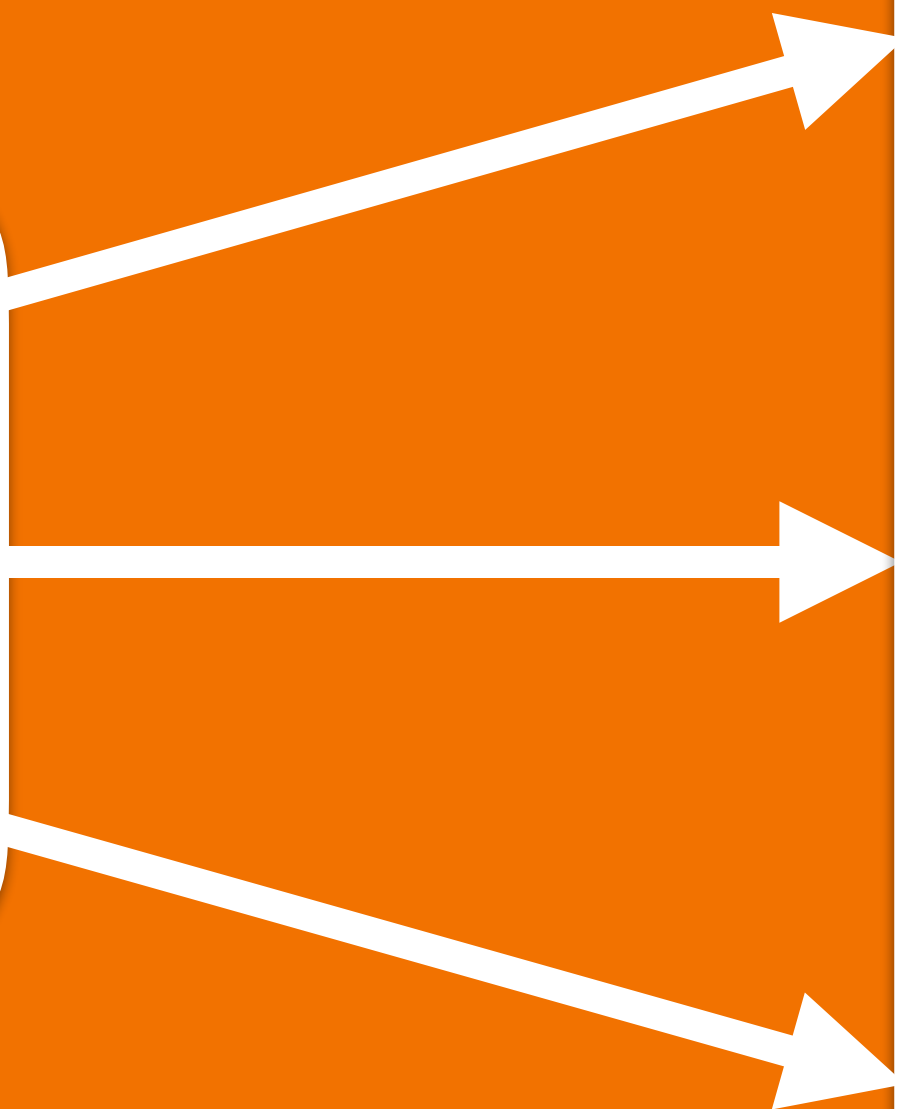
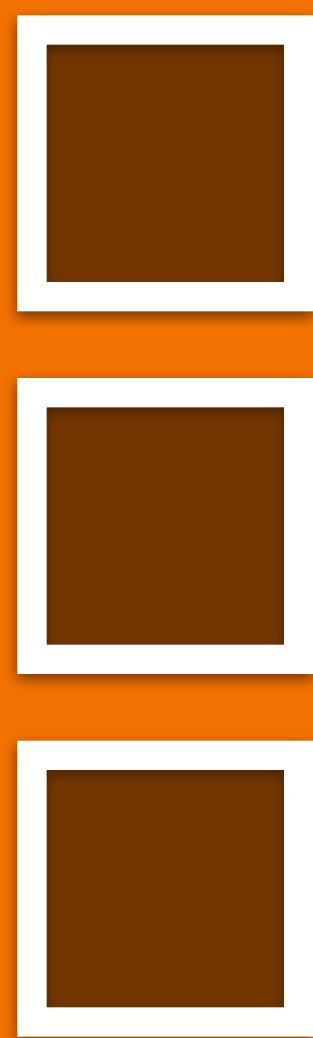
- **Never helps you in any way**
- **Not good at anything**
- **“Why haven’t I left?”**

Flume

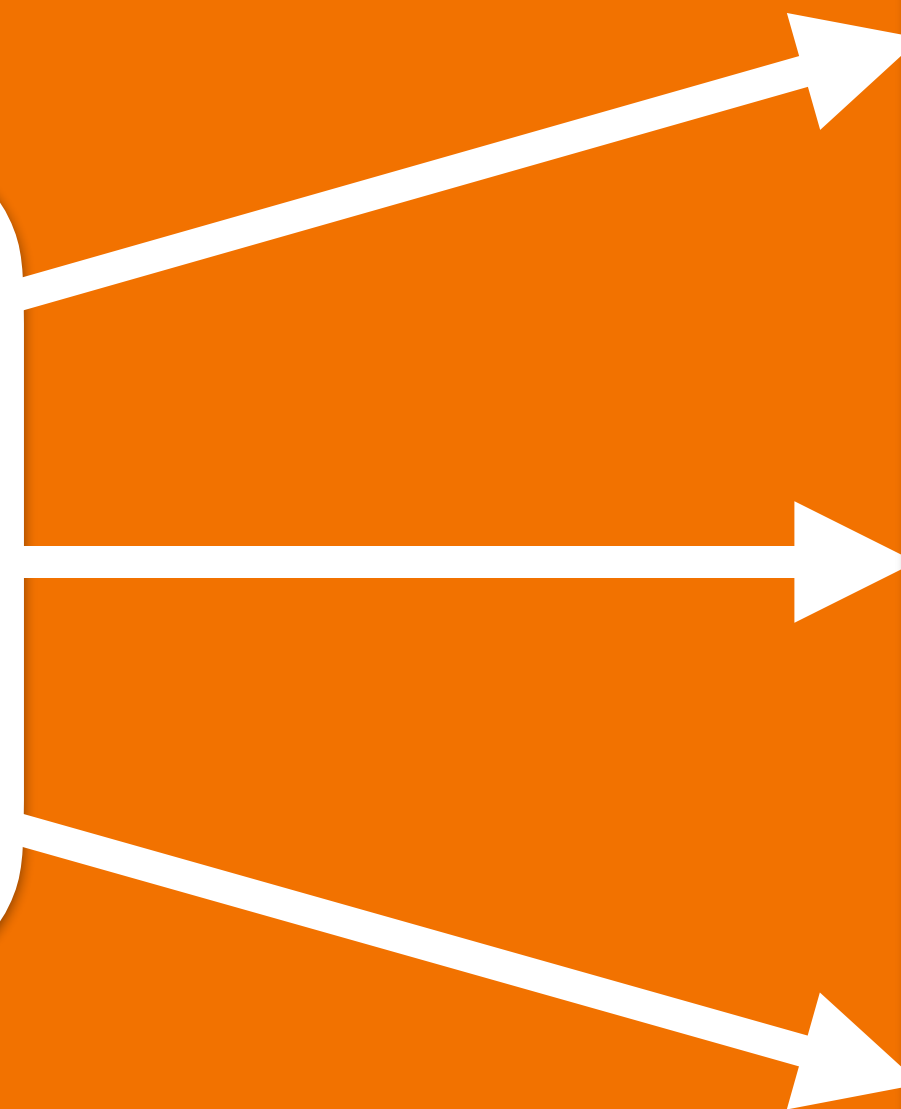
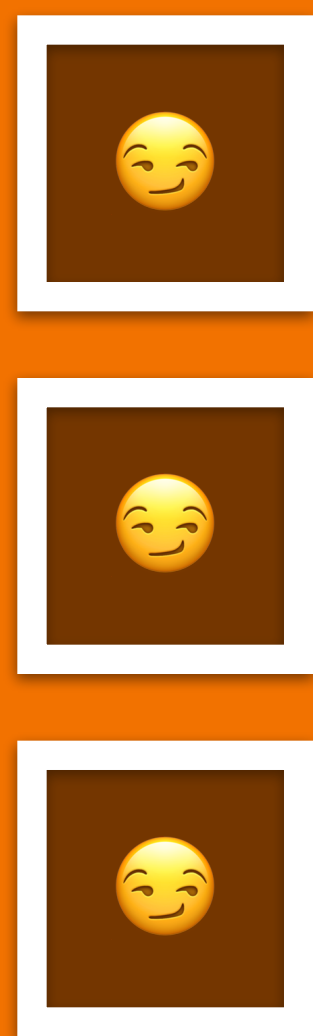
Flume



HDFS



HDFS



HDFS

O'REILLY®



Using
Flume

FLEXIBLE, SCALABLE, AND RELIABLE DATA STREAMING

Hari Shreedharan

O'REILLY®



Using Flume

FLEXIBLE, ~~SCALABLE~~, AND ~~RELIABLE~~ DATA STREAMING

Hari Shreedharan

DON'T

UIN KINOWA

UIN KINOWA

UIN KINOWA

UIN KINOWA

UIN KINOWA

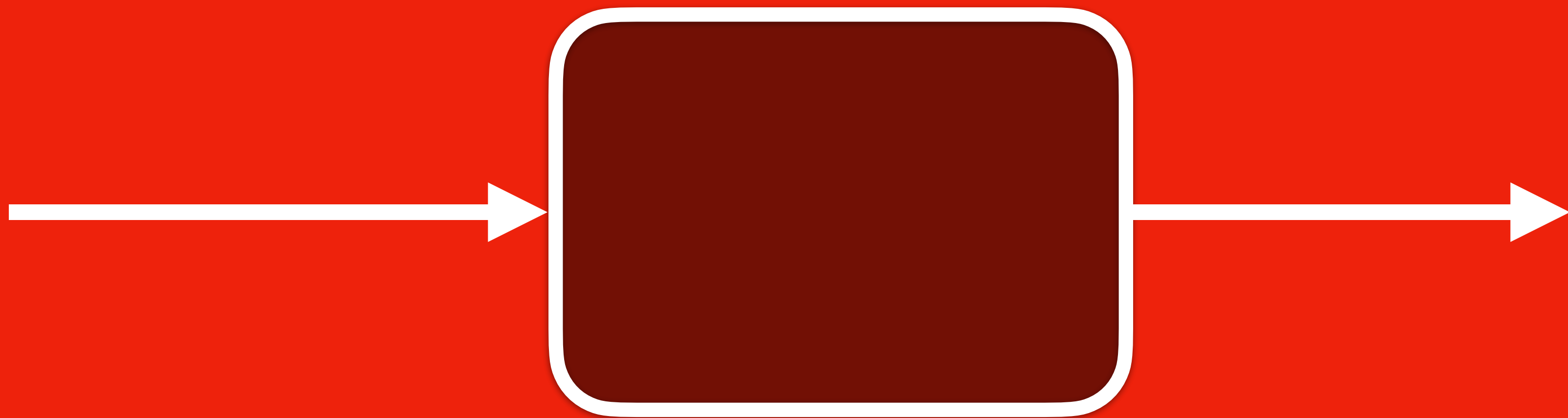
UIN KINOWA

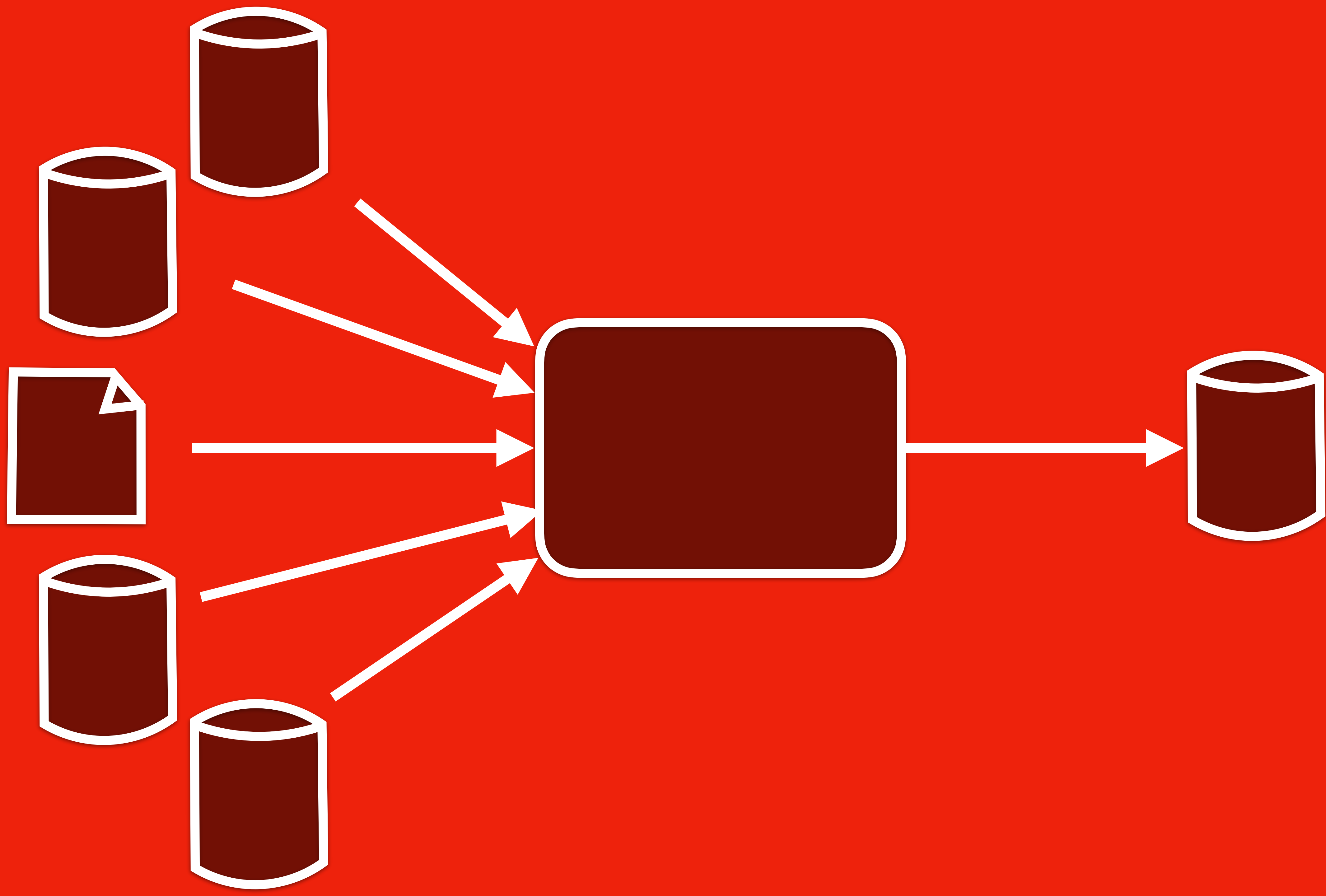
You can do better than Him

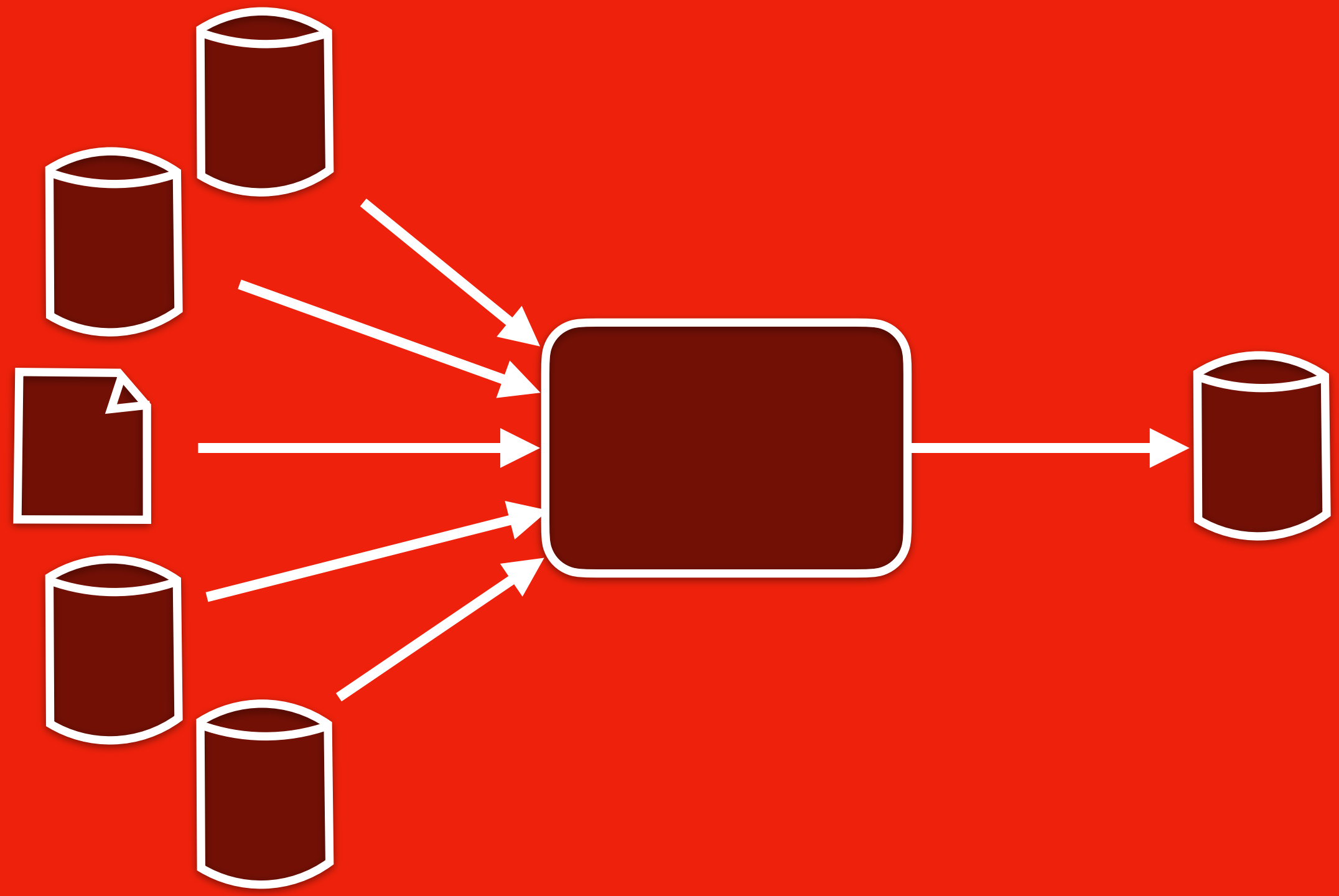
- **Key Value: Redis, Dynamo, BigTable**
- **Document Store: SQL, Elastic Search**
- **Map Reduce: Spark**

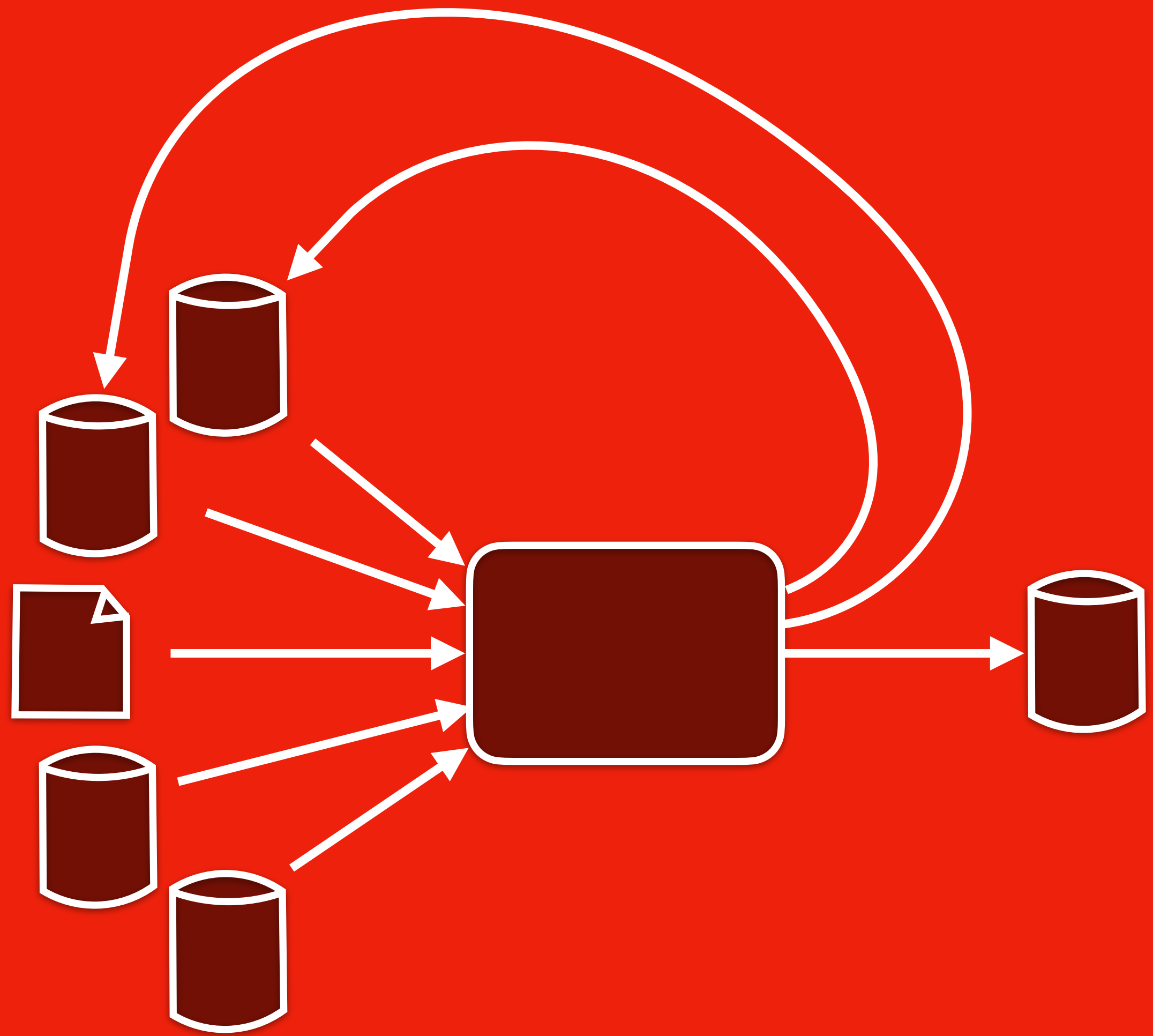
The Monster

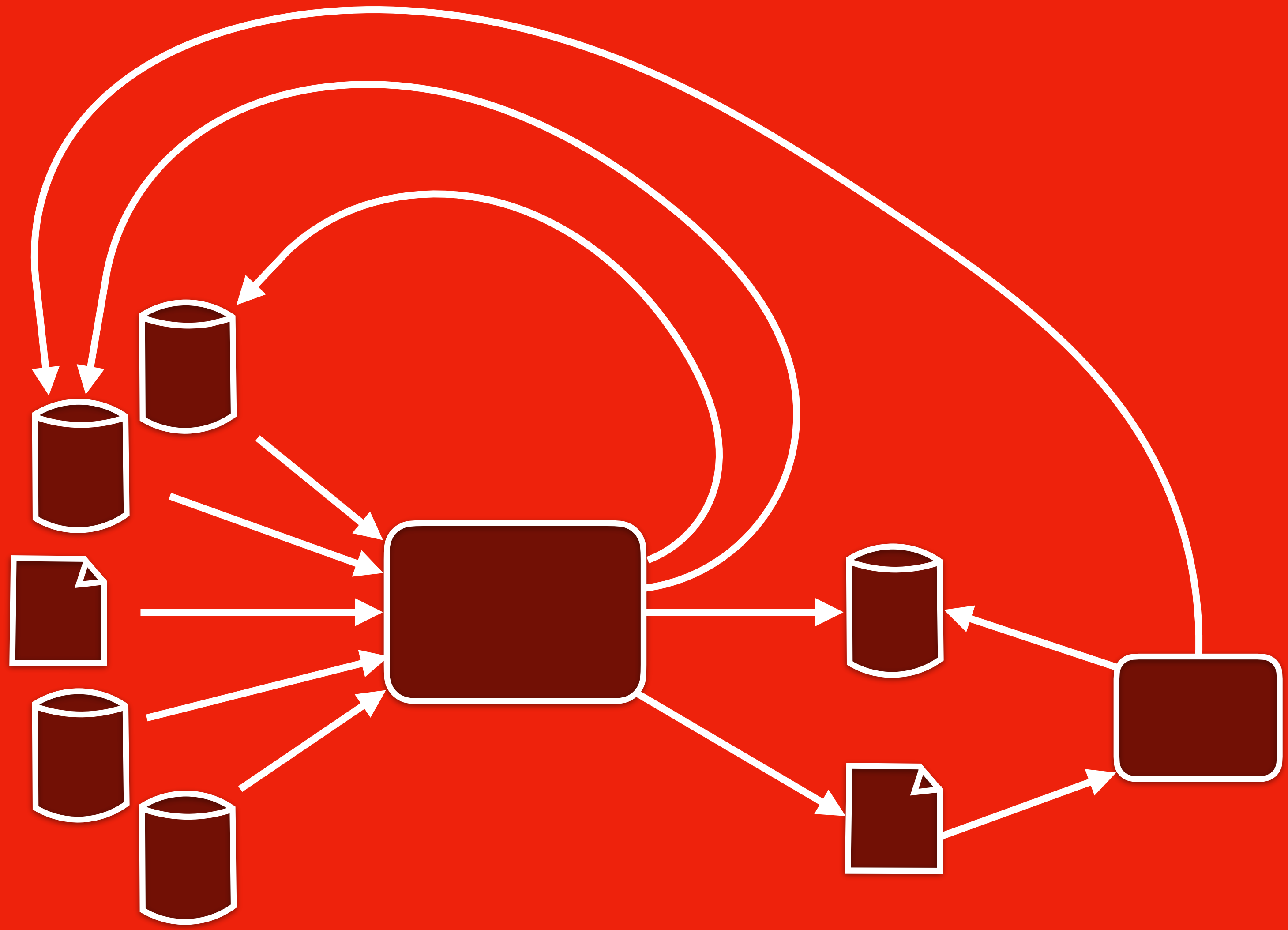
- **Almost nobody understands how it works**
- **Those that do are actually afraid of it**
- **Fixing a problem takes huge amounts of time**
- **You've lost all hope and can't even imagine escape**

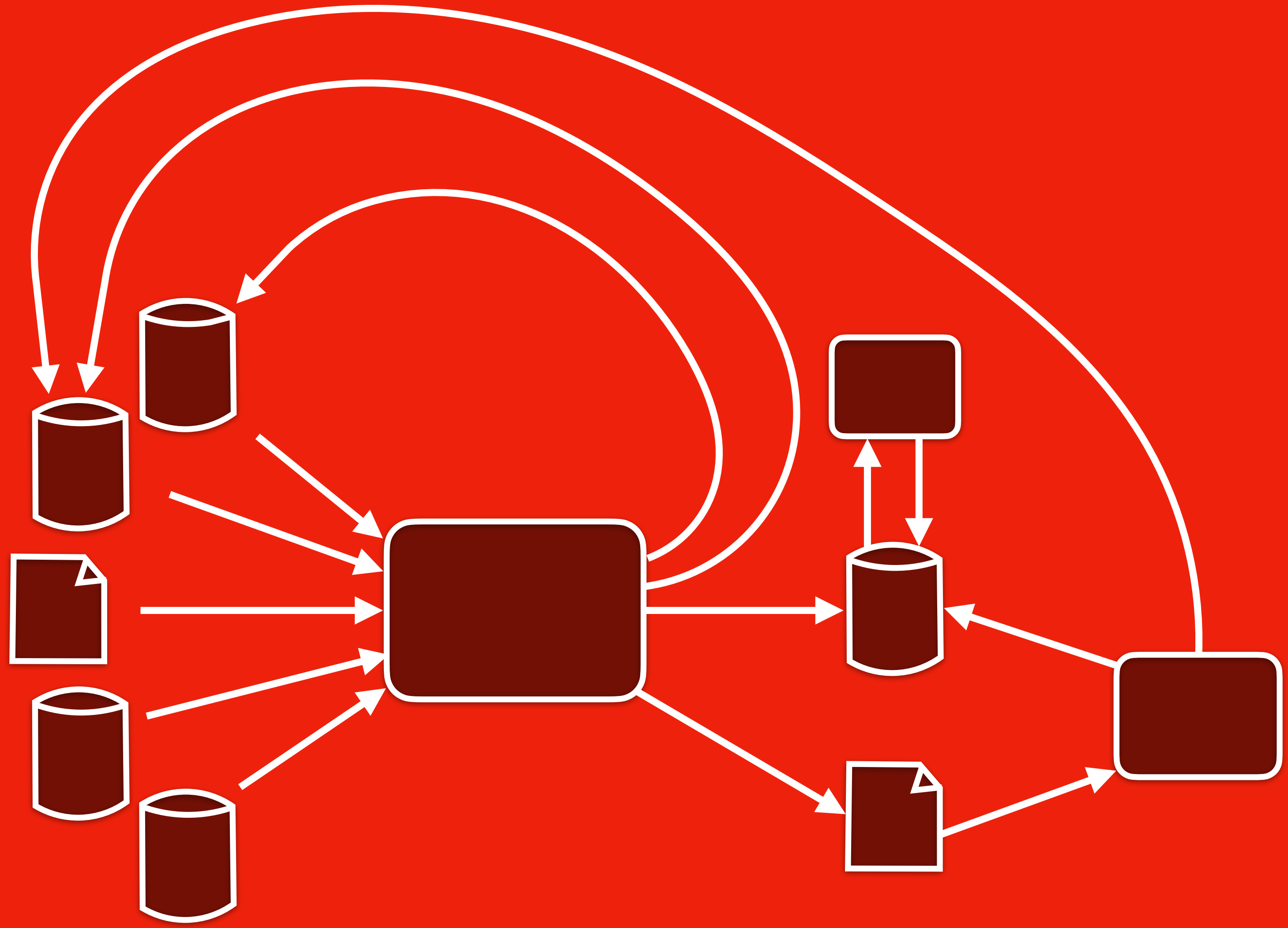


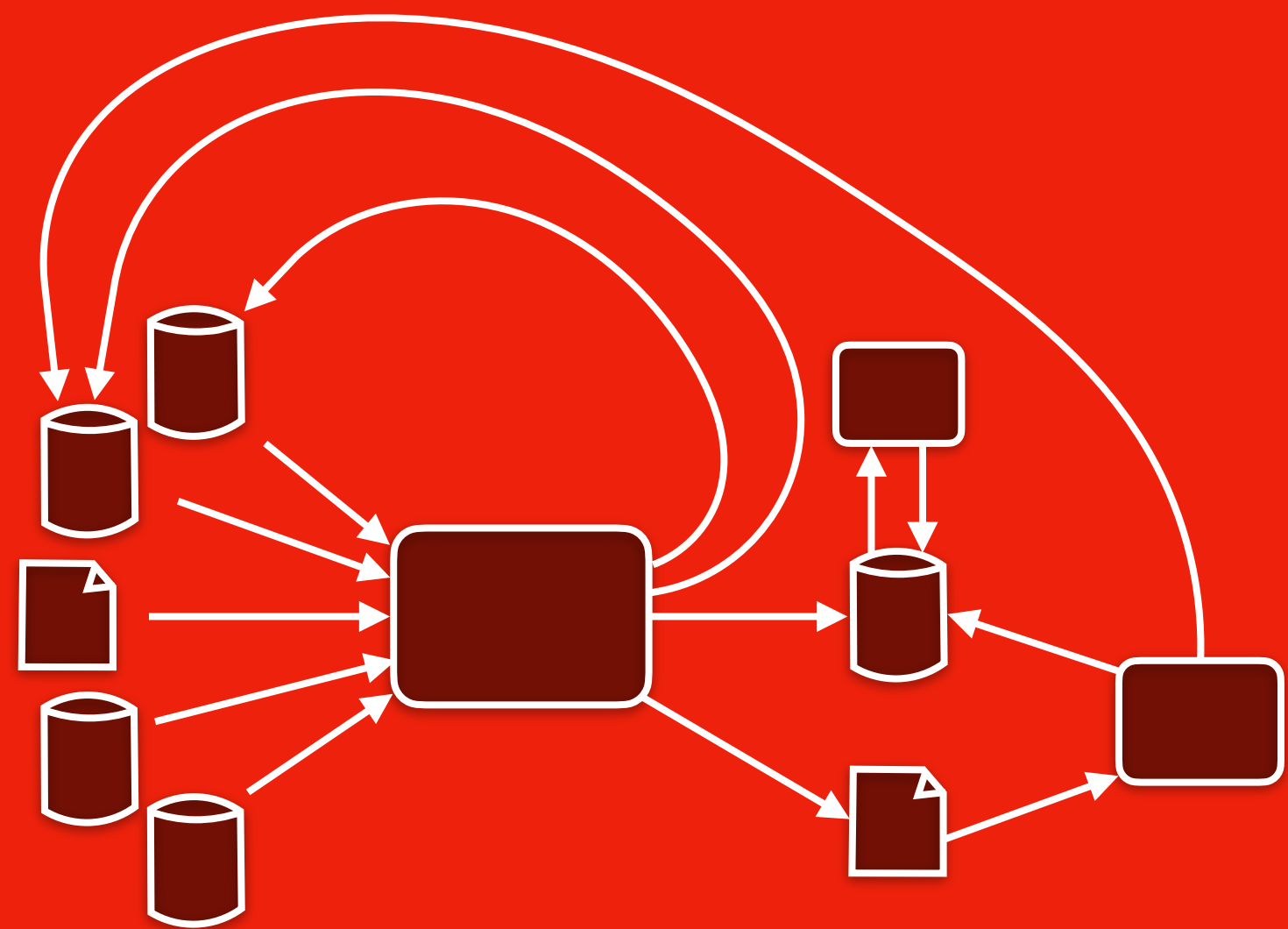


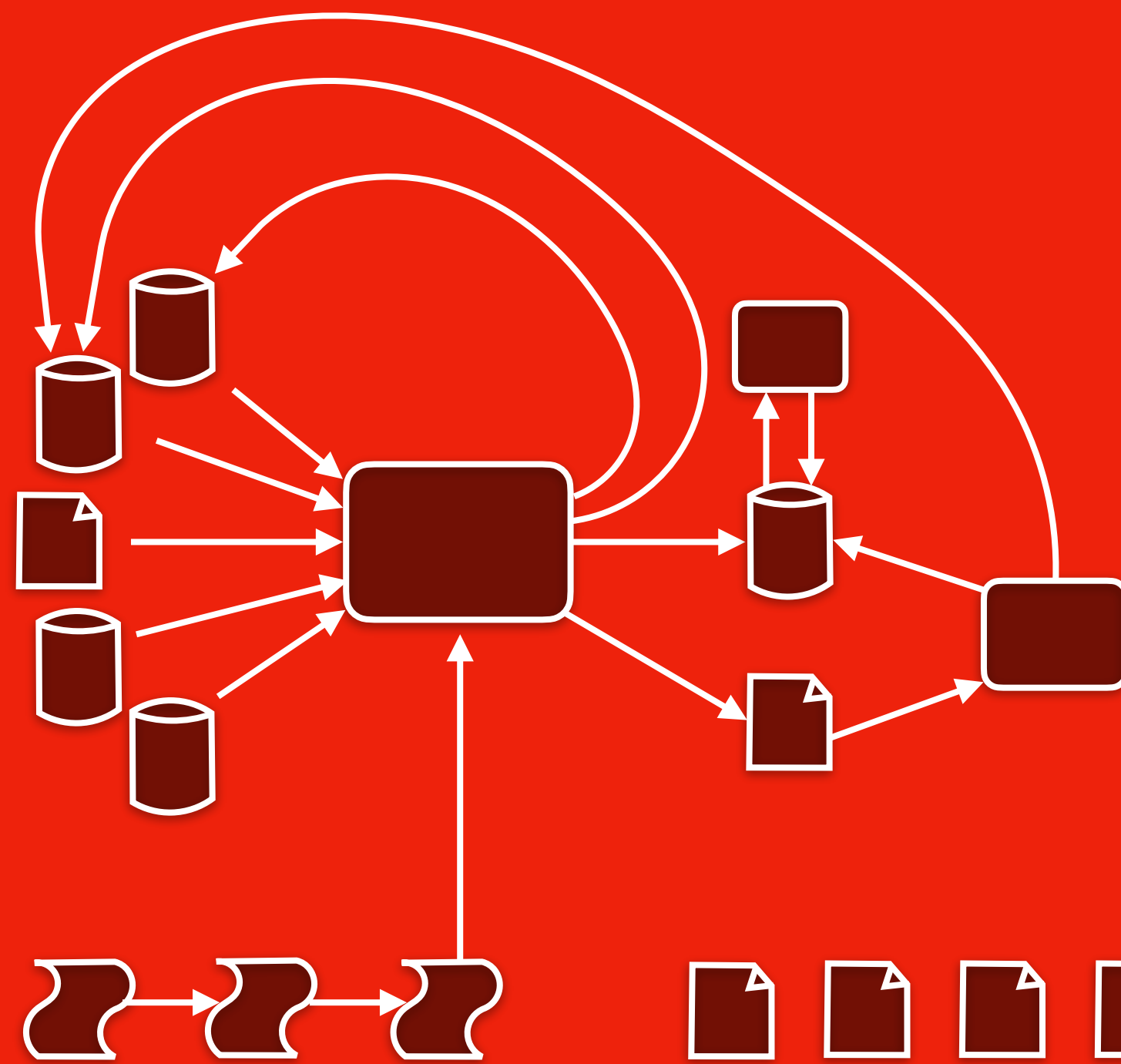












Sometimes killing
is the only option

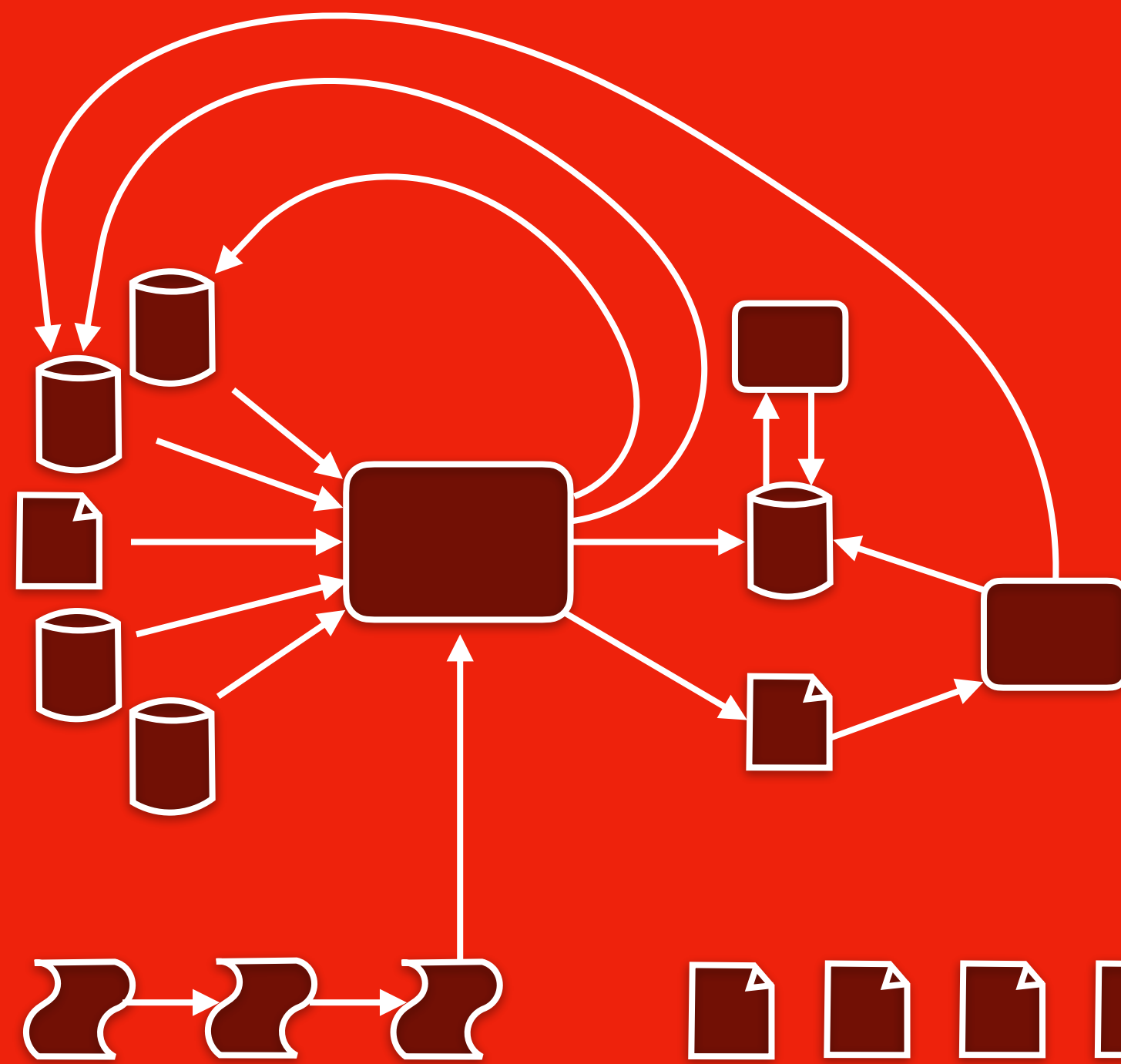
Determinism

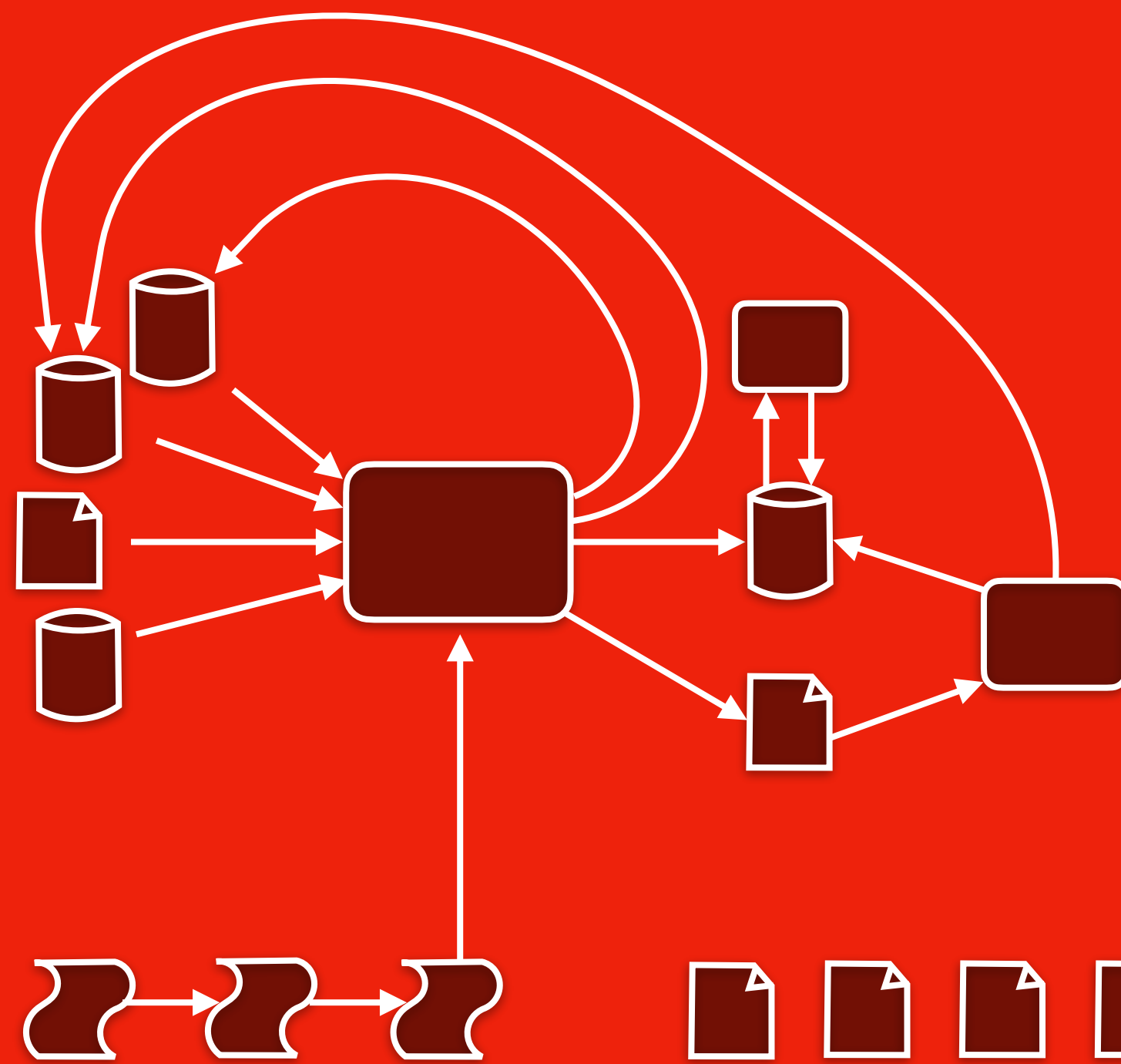
Consistency over Correctness

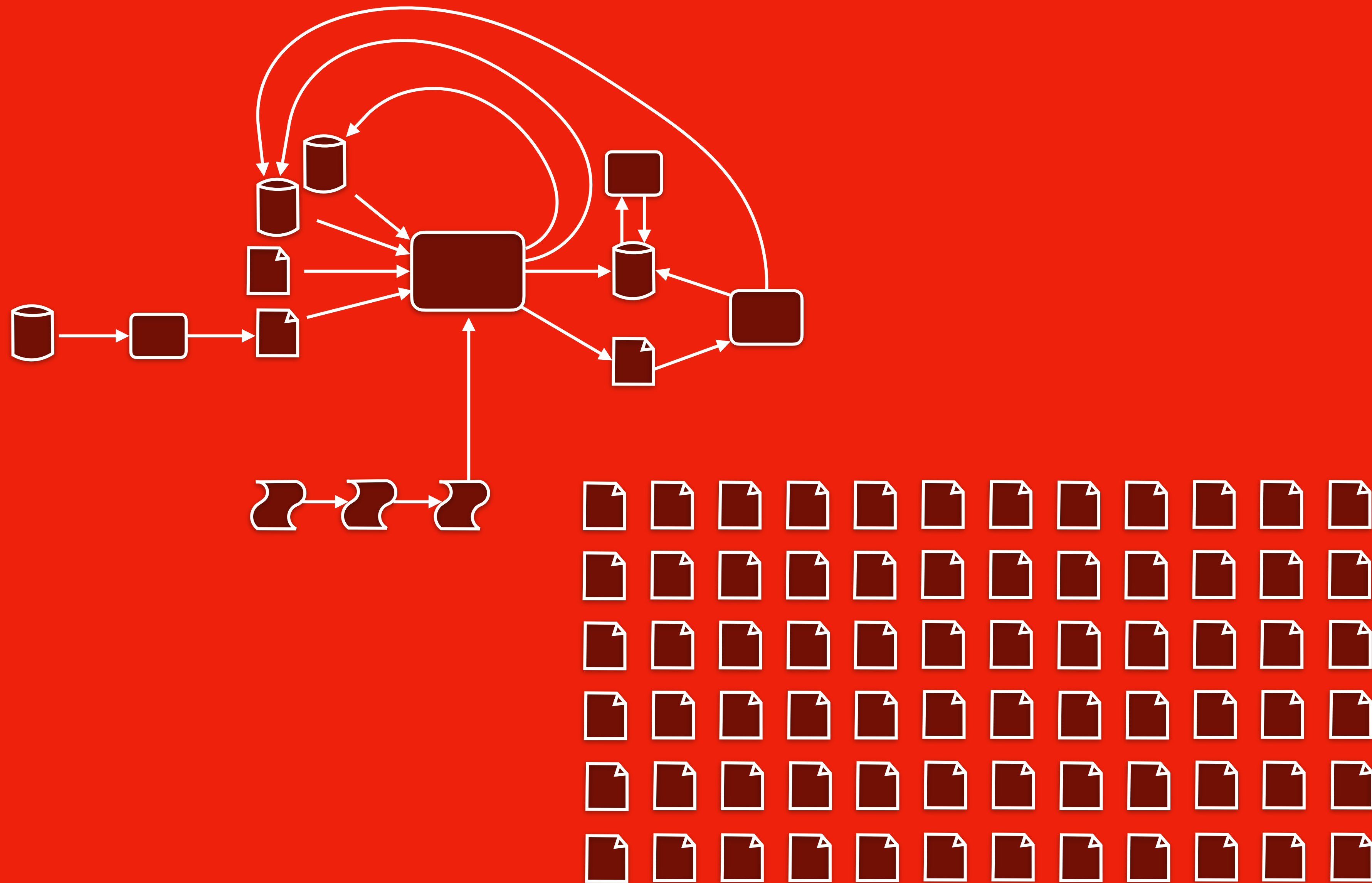
One thing well

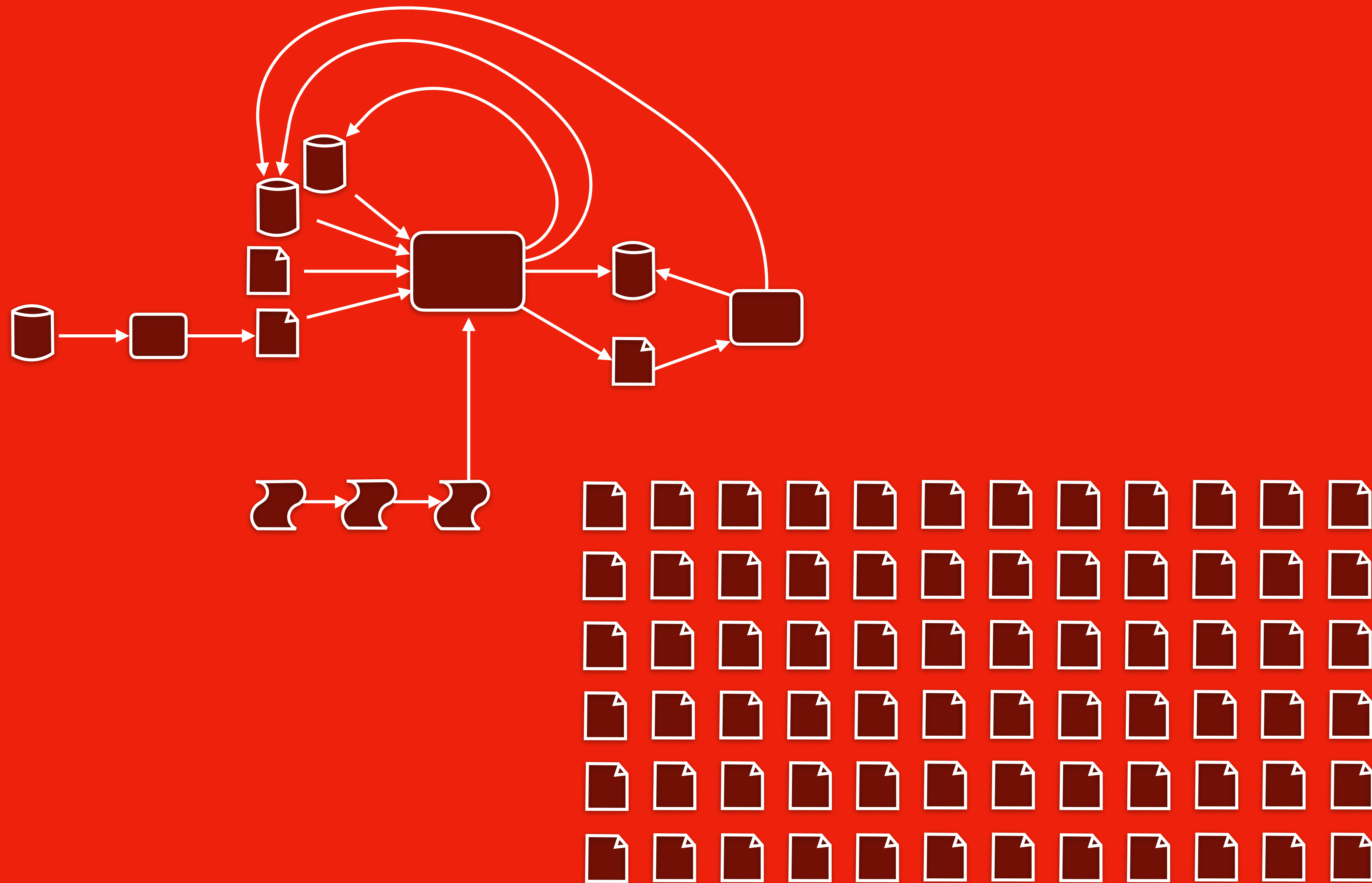
In batch processing
your data IS your logs

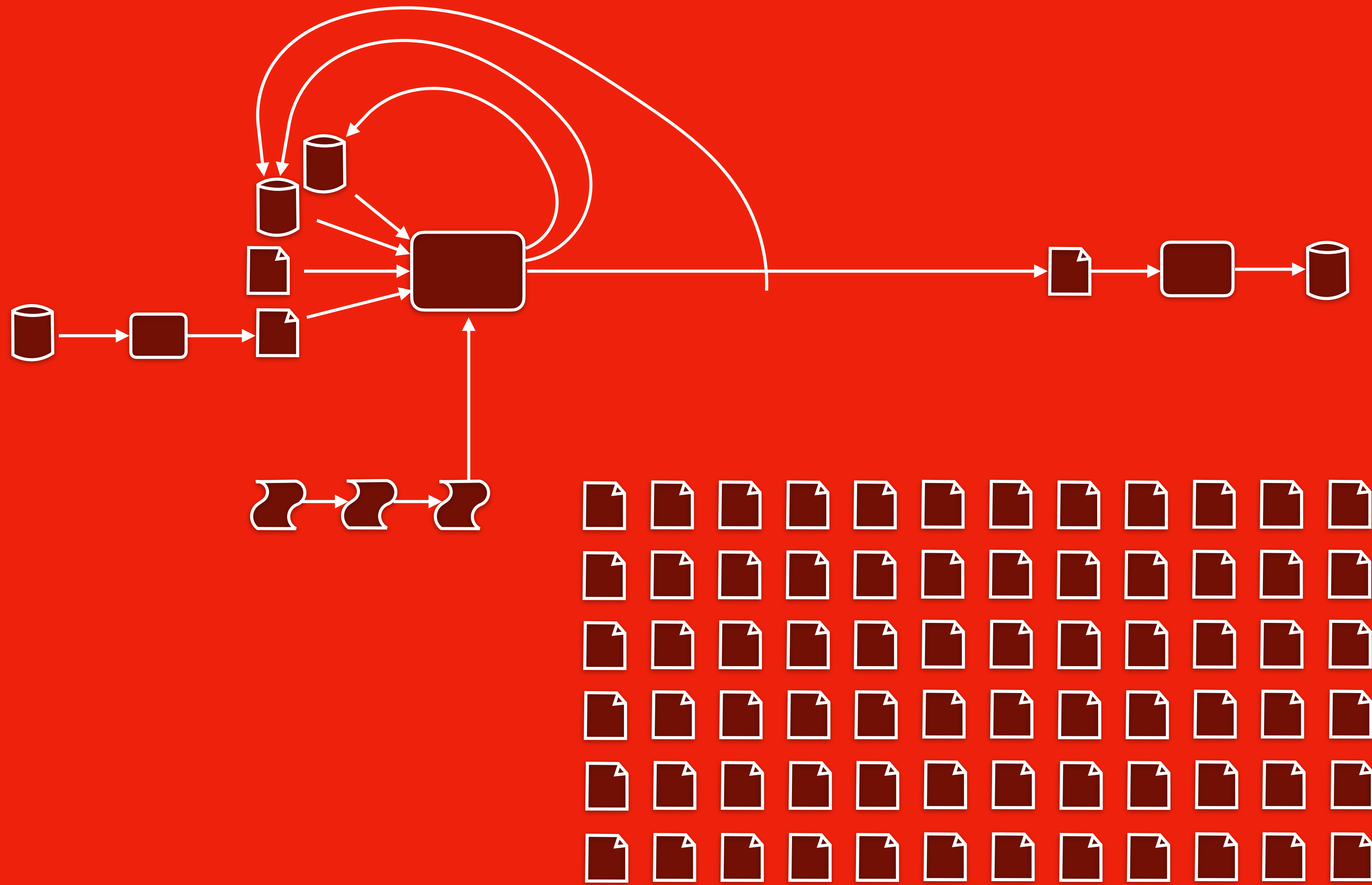
How easily can you
see what is happening?

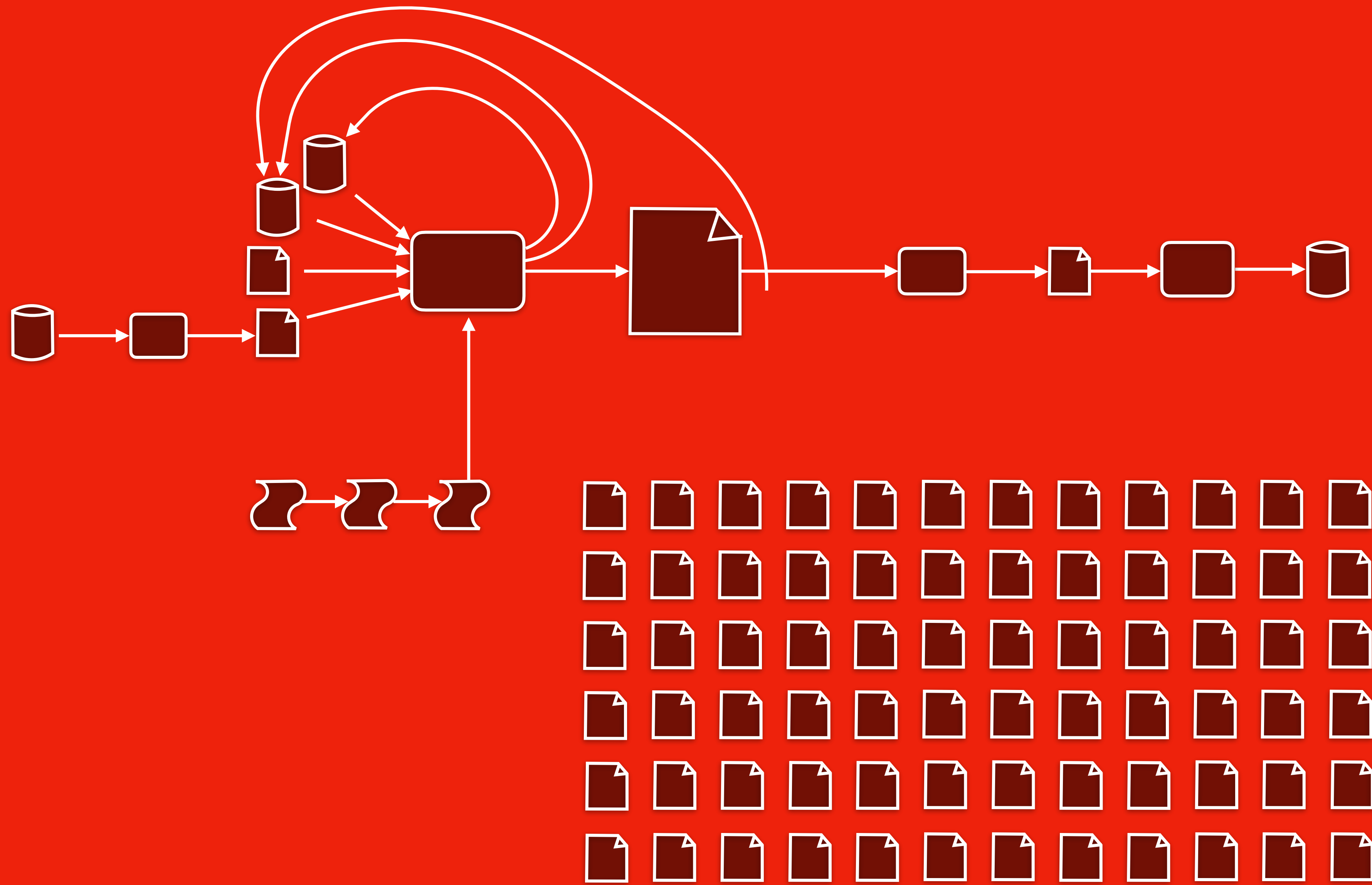


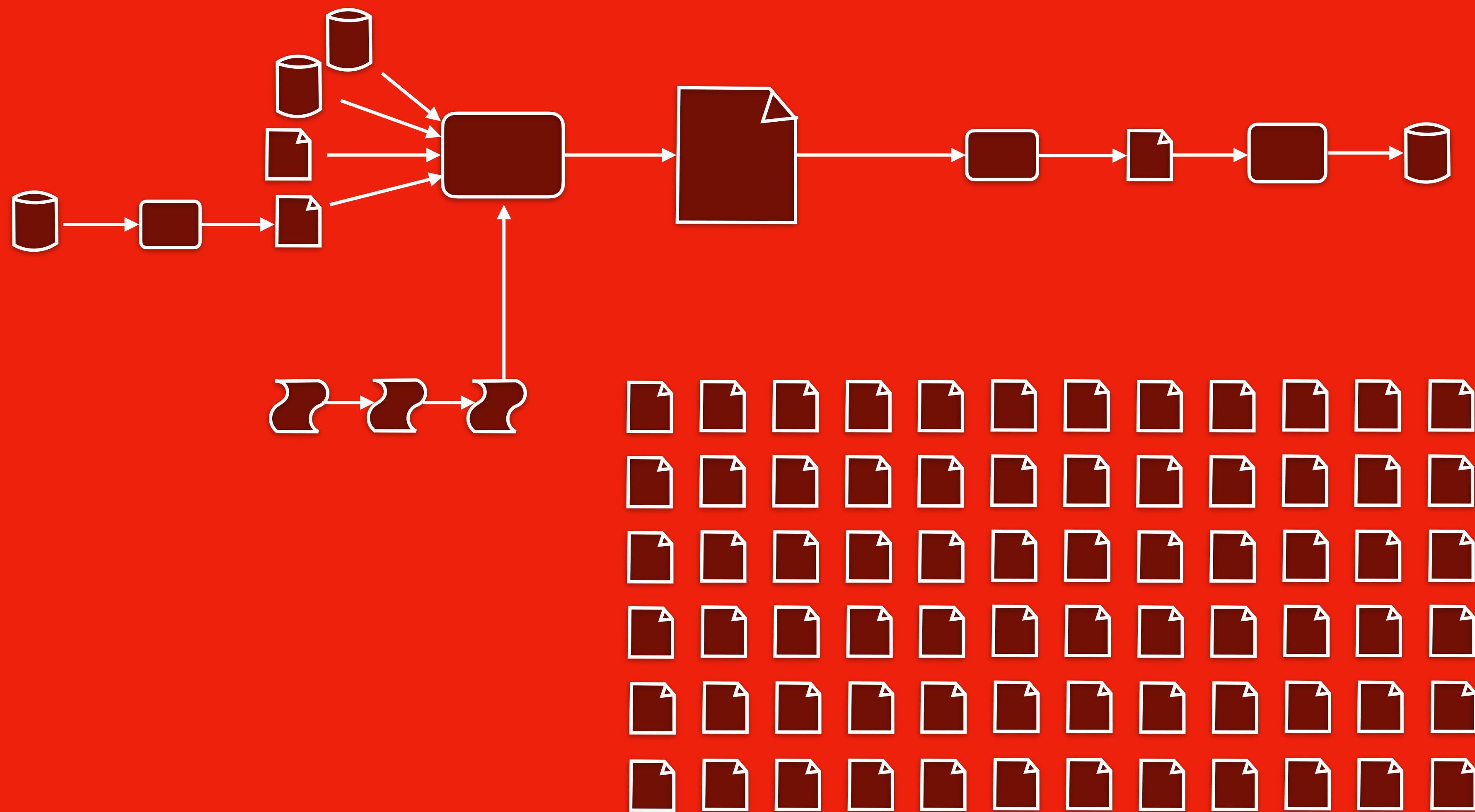


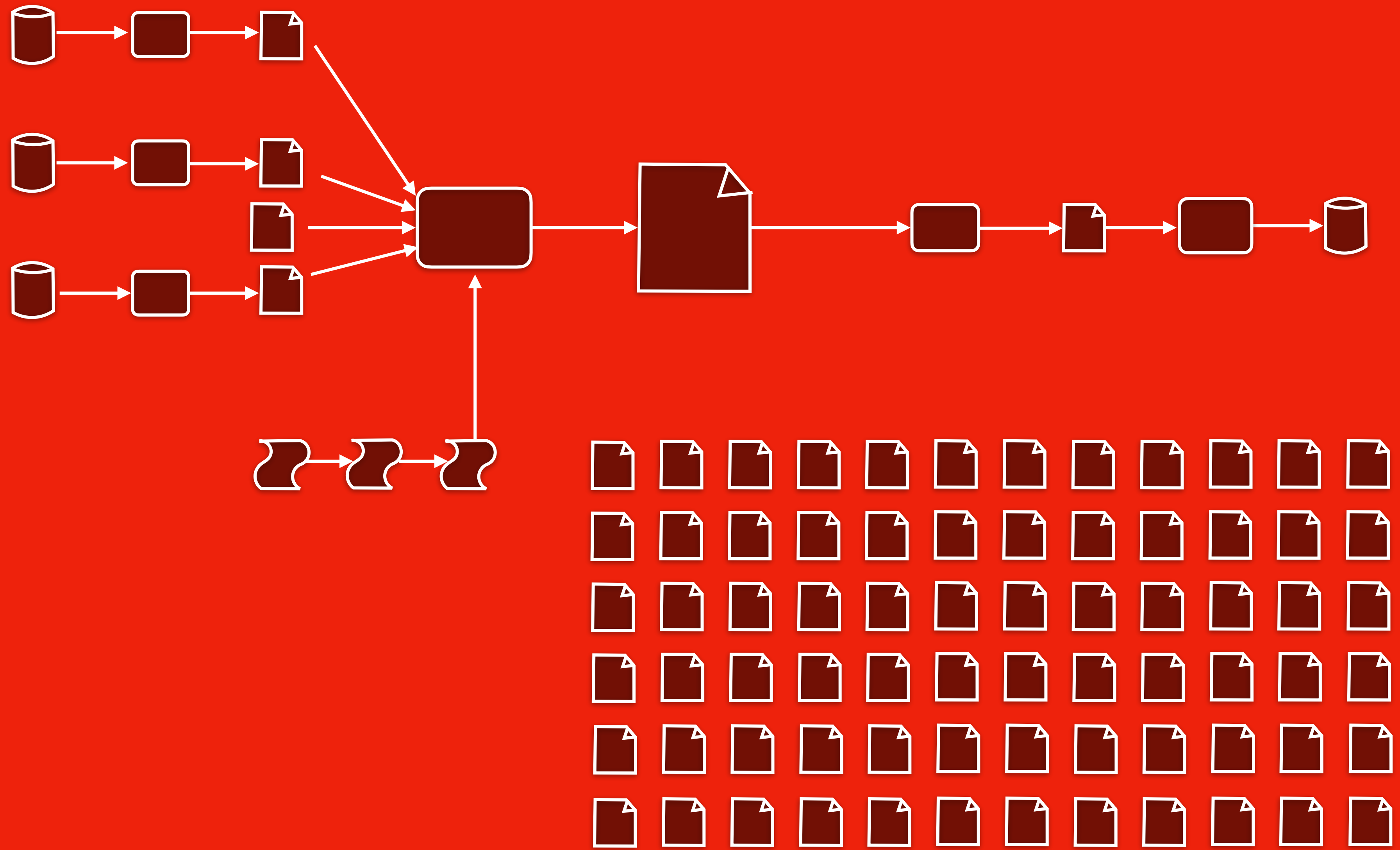


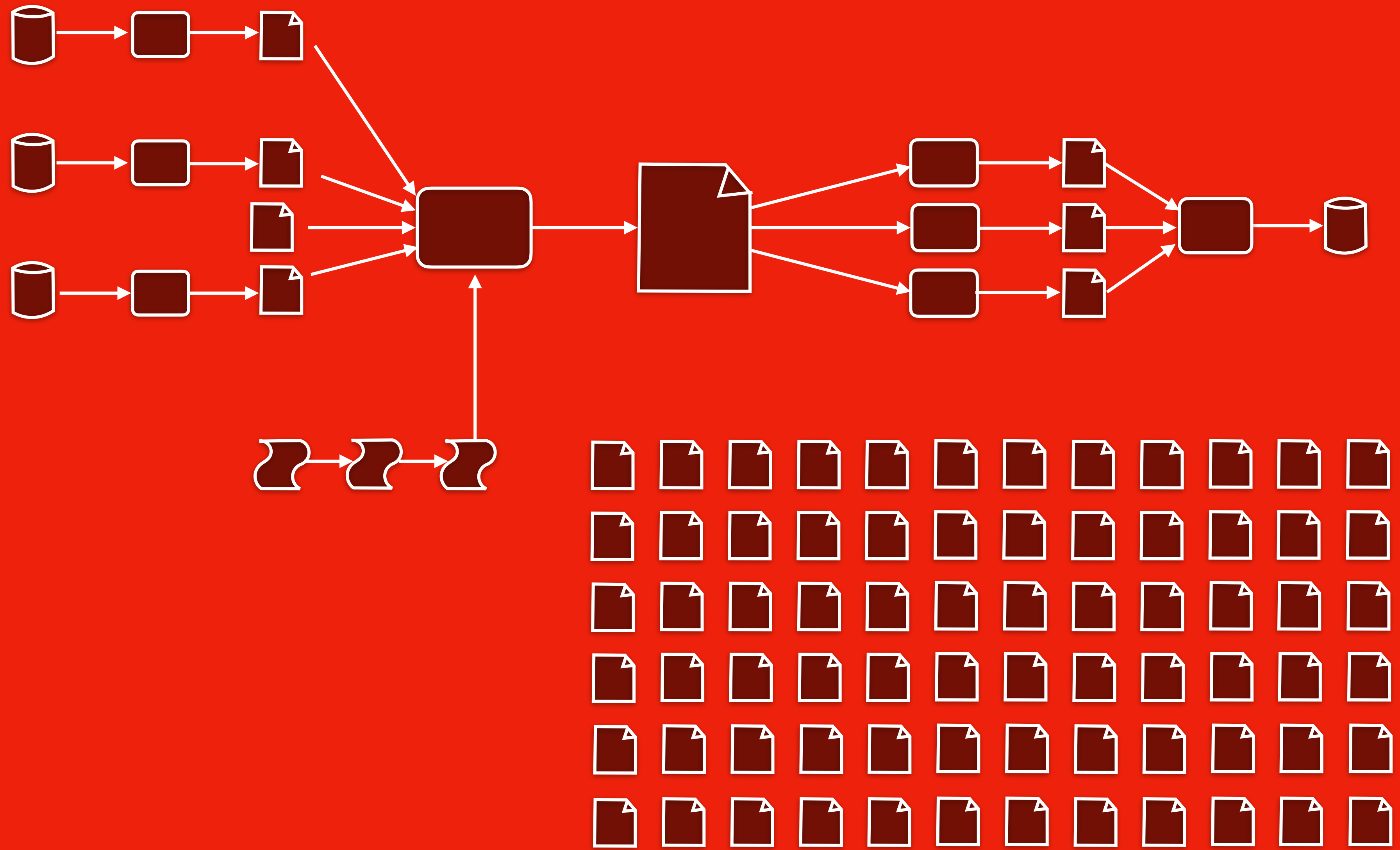


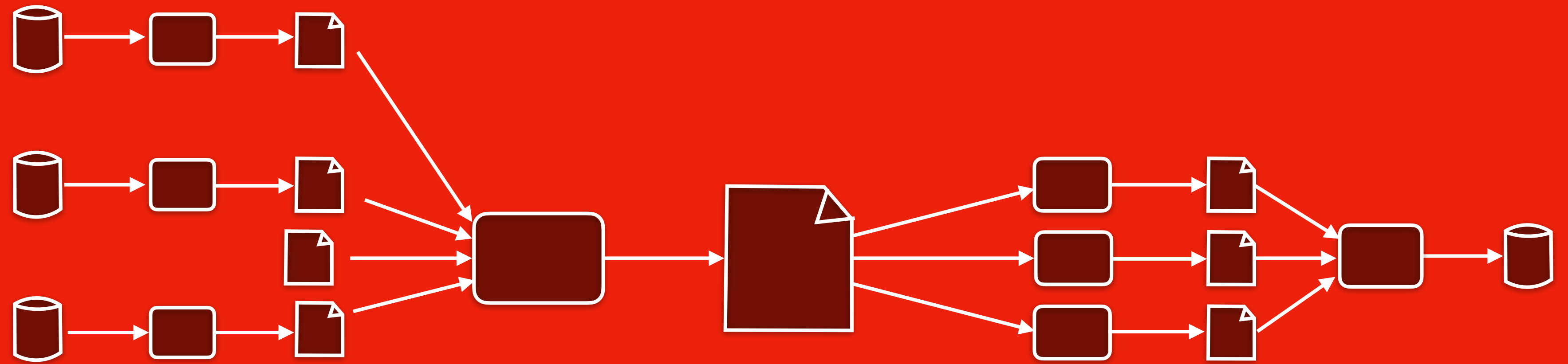










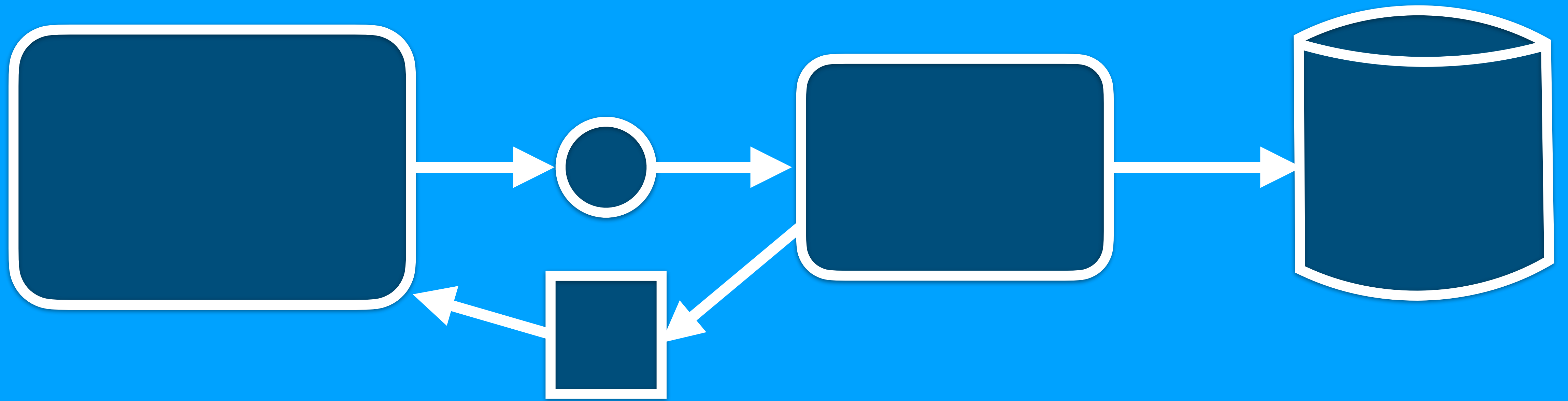


Database

The Mirror

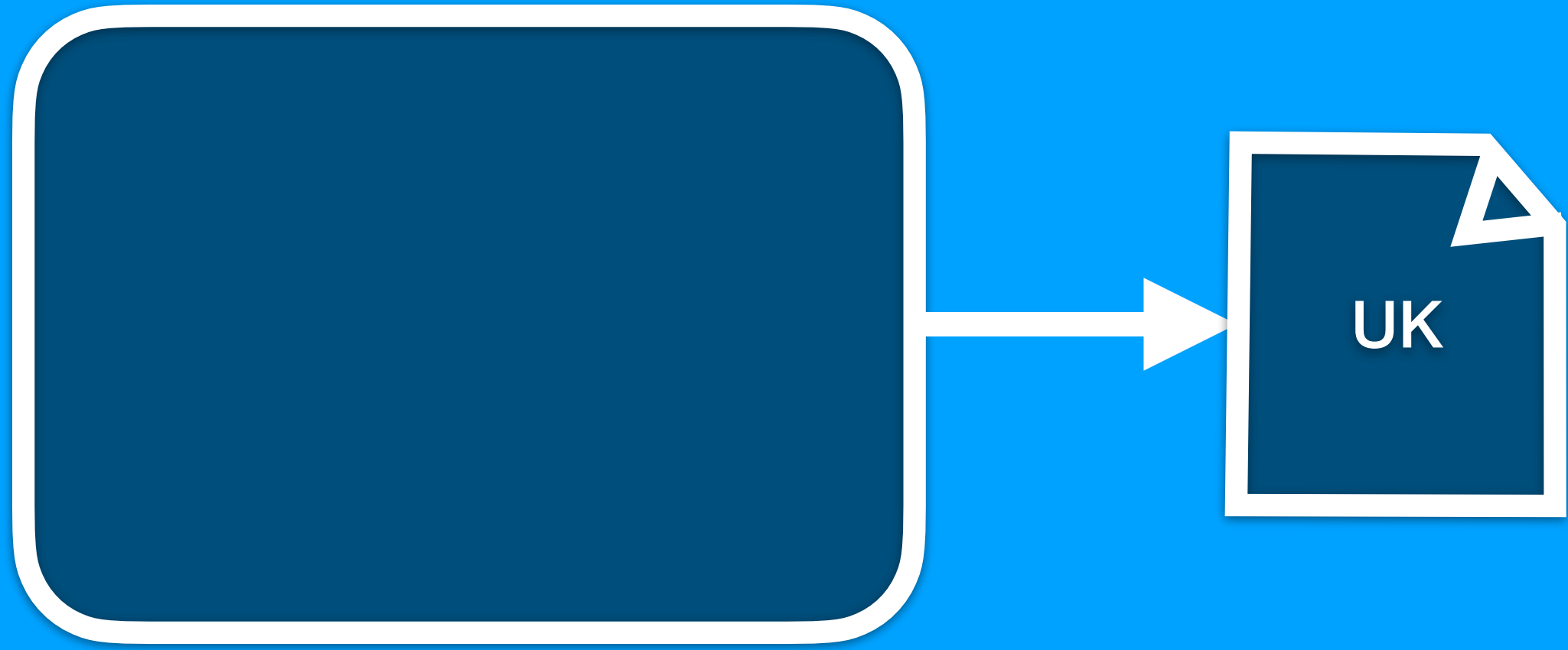
- **Are you repeating the same mistakes over and over?**
- **What is the common factor with these systems?**
- **“It’s not you, it’s me”**

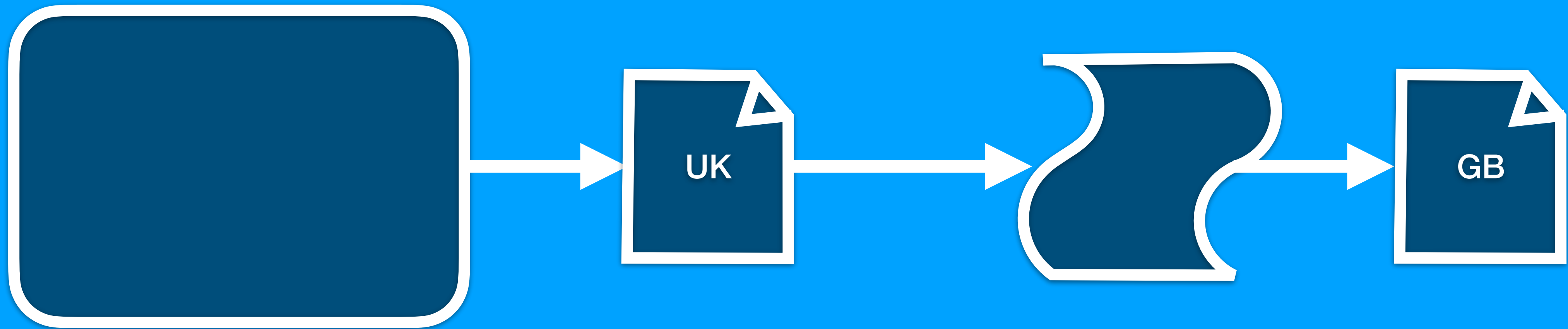
Unnecessary Abstraction

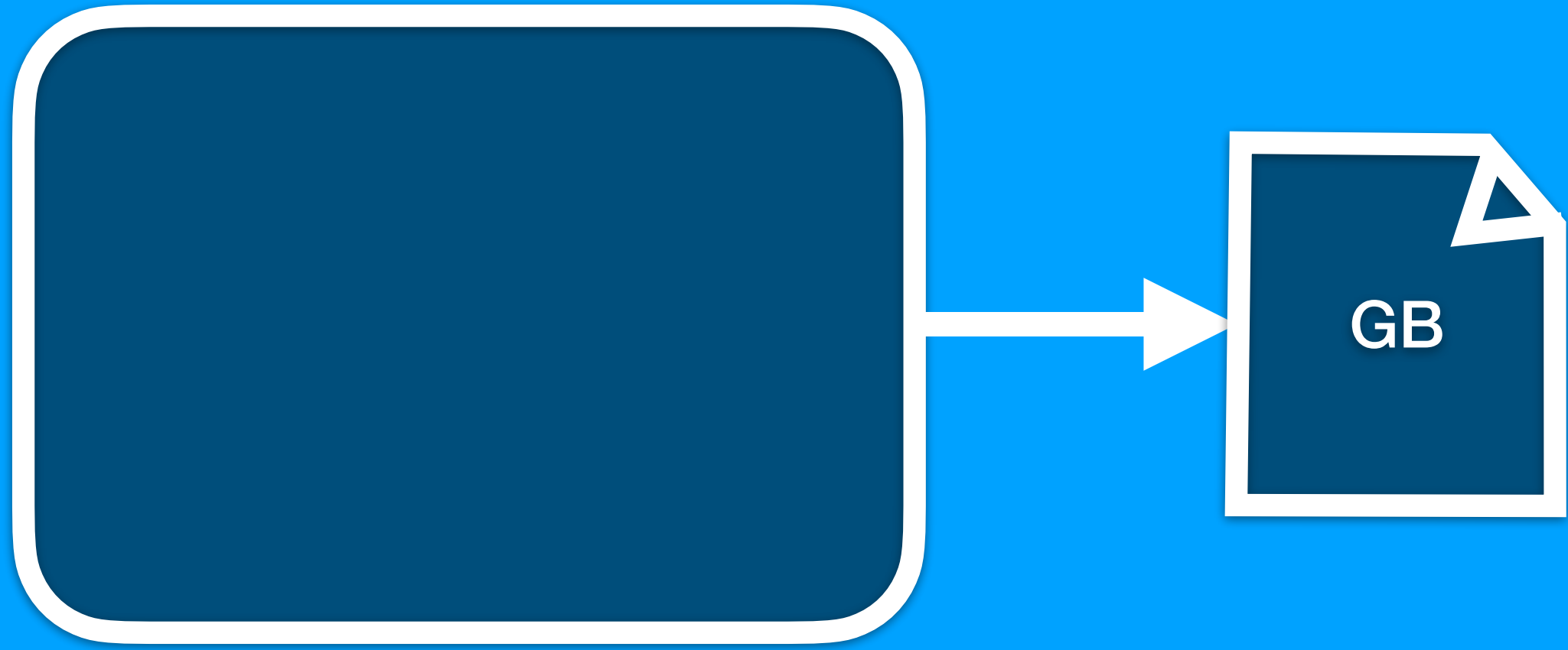




Solve the problem where it occurs







Communication