



Routing billions of events a day: How we do routing in Schibsted

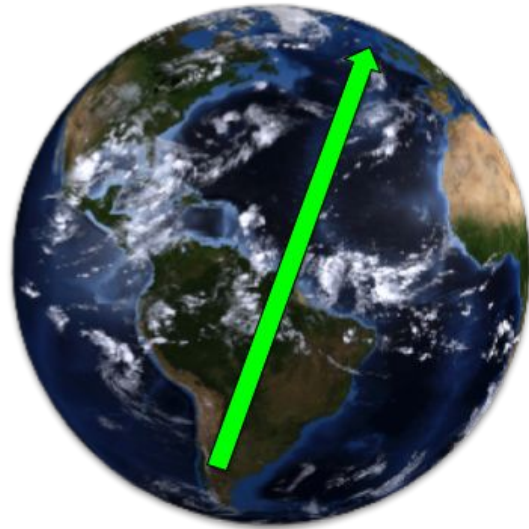


Carlos Manuel Duclos-Vergara, Staff Engineer



About me

Schibsted



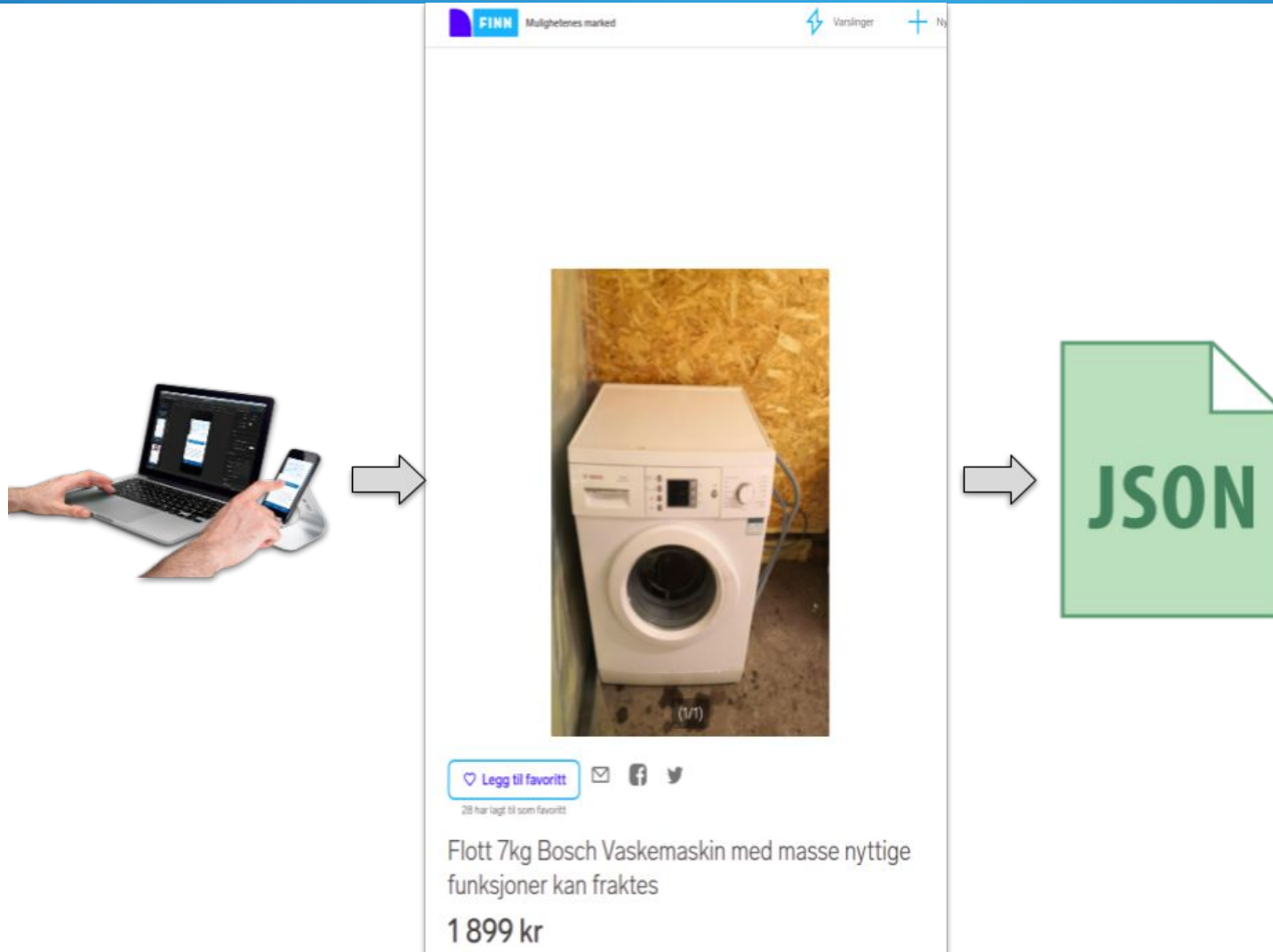
Agenda

- Schibsted
- A short story
- GDPR
- Pulse (our tracking solution)
 - Overview
 - Internals

Schibsted



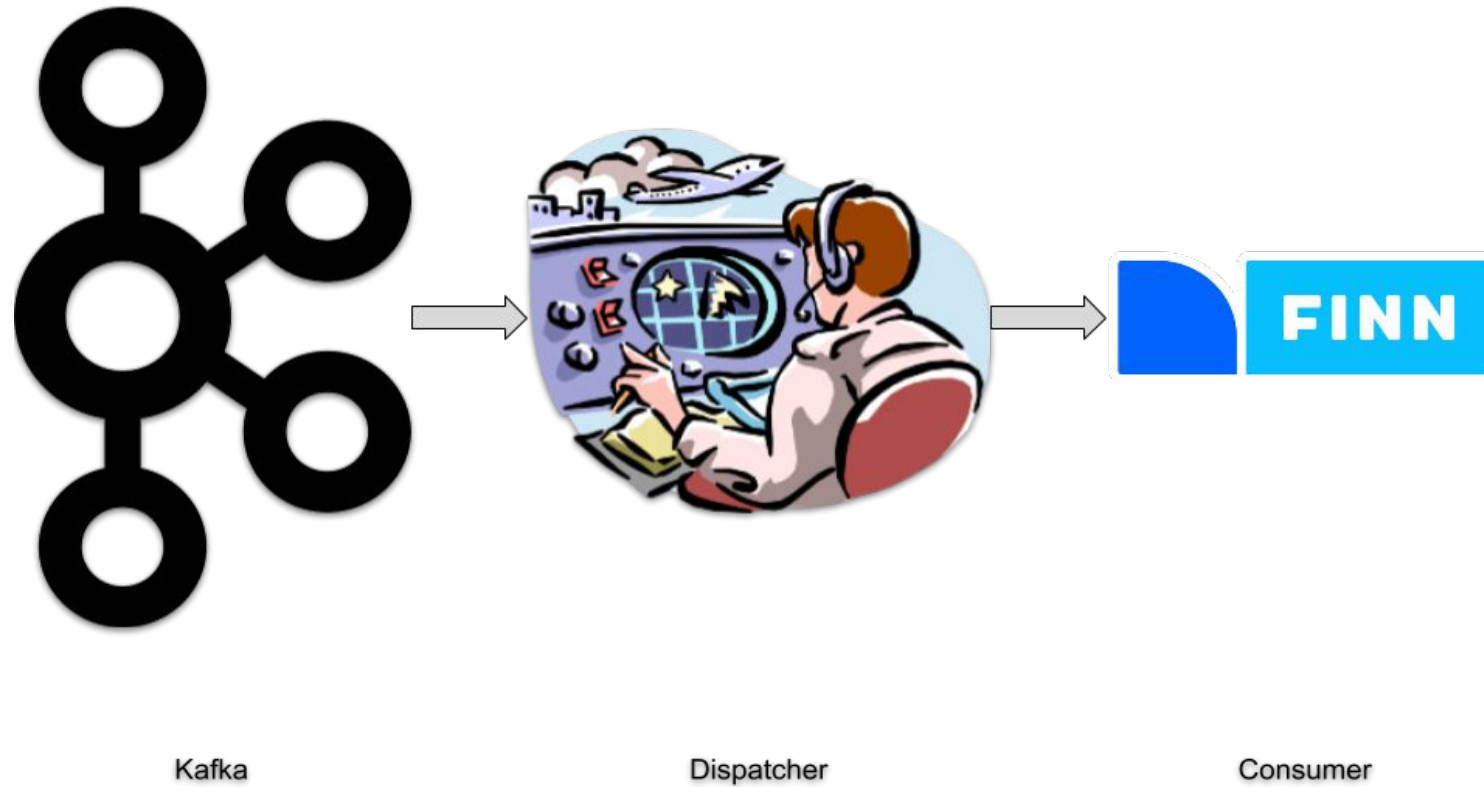
Event generation



Event routing



Event dispatching



Event consumption



GDPR and data collection



User rights

1. Data portability
2. Right to be forgotten

Legal basis for data collection

1. Consent
2. Processing obligation
3. Legal obligation
4. Vital interest
5. Public interest
6. Legitimate interest

End to end event processing solution



Pulse

Pulse ecosystem



Tracking SDK



Data Quality



Data and Statistics

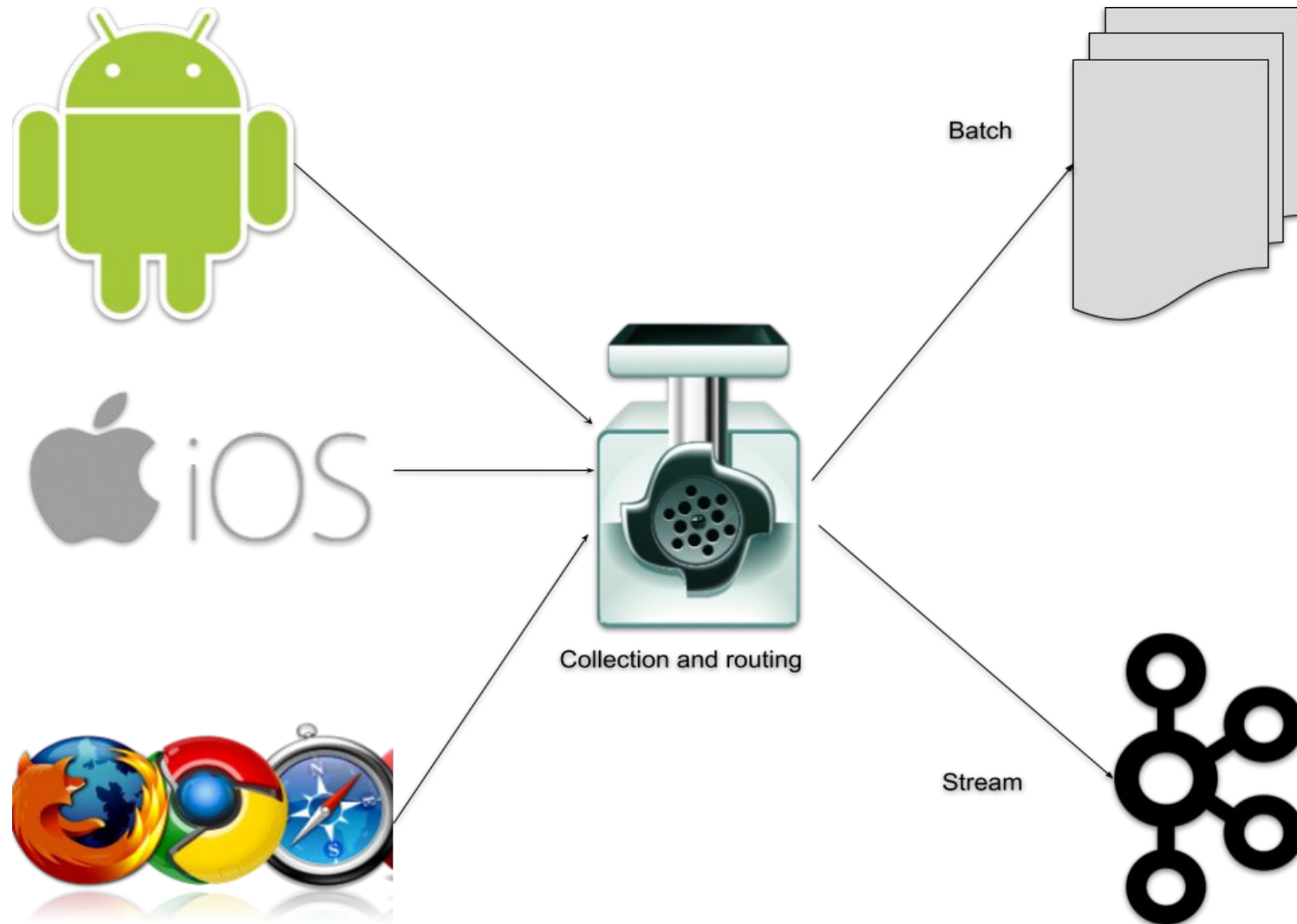
Administration portal

The screenshot shows the Pulse Console Administration portal with the following sections:

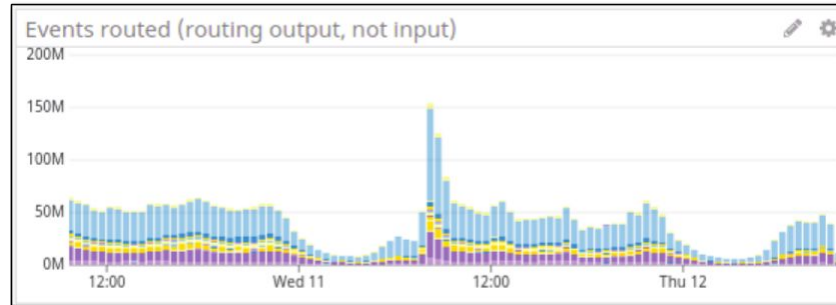
- I want to access data**
 - Datasets**: See available datasets you can request access to.
 - Your permissions**: See your permissions, pending request and grants.
 - AWS keys**: Generate new or delete existing key/secret pairs for getting access to data in our datalake.
- I want to debug my pulse integration**
 - Debug mode**: See a live stream of your debug events with validation results.
 - Get Pulse Unicorn**: See events sent from your website using this Chrome plugin.
 - Validate an event**: Paste an JSON event and check if it validates against the schema.
 - Check an event in routing**: Paste an JSON event and check to which sinks and topics it will be delivered.
- I want to check data quality**
 - DOT Dashboard**: Overview of data quality check results.
 - Validate a check**: Validate a JSLT check.
 - Tagging plans**: See tagging plans for media, marketplaces, common components and other docs.
- I want to start tracking**
 - Register a provider**: If you don't have one, you need a registered provider in order to send Pulse events.
 - All providers**: See all the registered providers.
 - JS SDK**: Start tracking with the Javascript Pulse SDK.
 - iOS SDK**: Start tracking with the iOS Pulse SDK.
 - Android SDK**: Start tracking with the Android Pulse SDK.

Footer: Made by the glorious Data and Privacy Team - [Click here to give us some feedback on this project](#)

Lifetime of an event



Side track: How much is 1 billion events



86400 seconds per day

864000 milliseconds per day

Low traffic hours

10 M / 30 minutes

~ 5000 / s

High traffic hours

50 M / 30 minutes

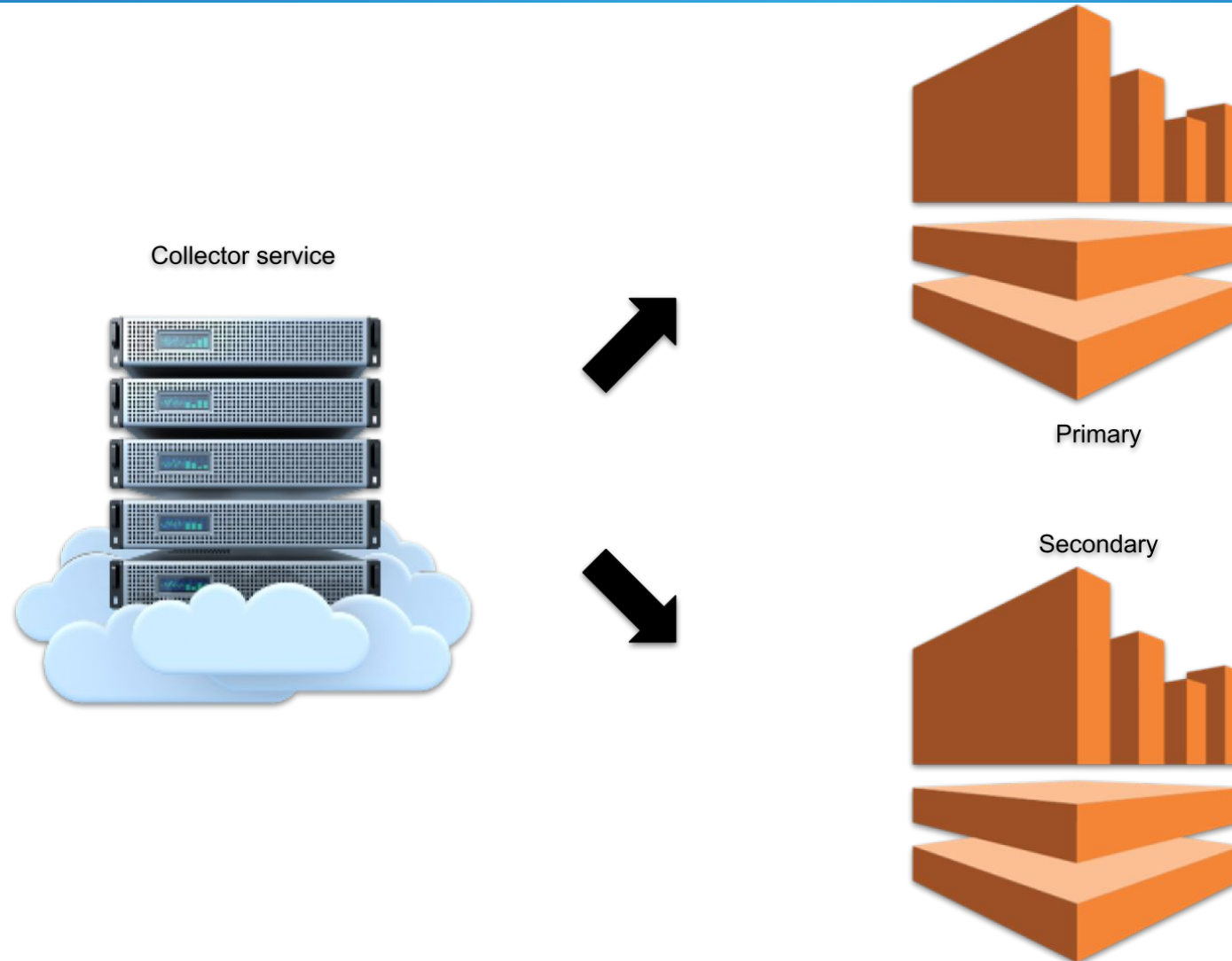
~ 25000 / s

Peaks

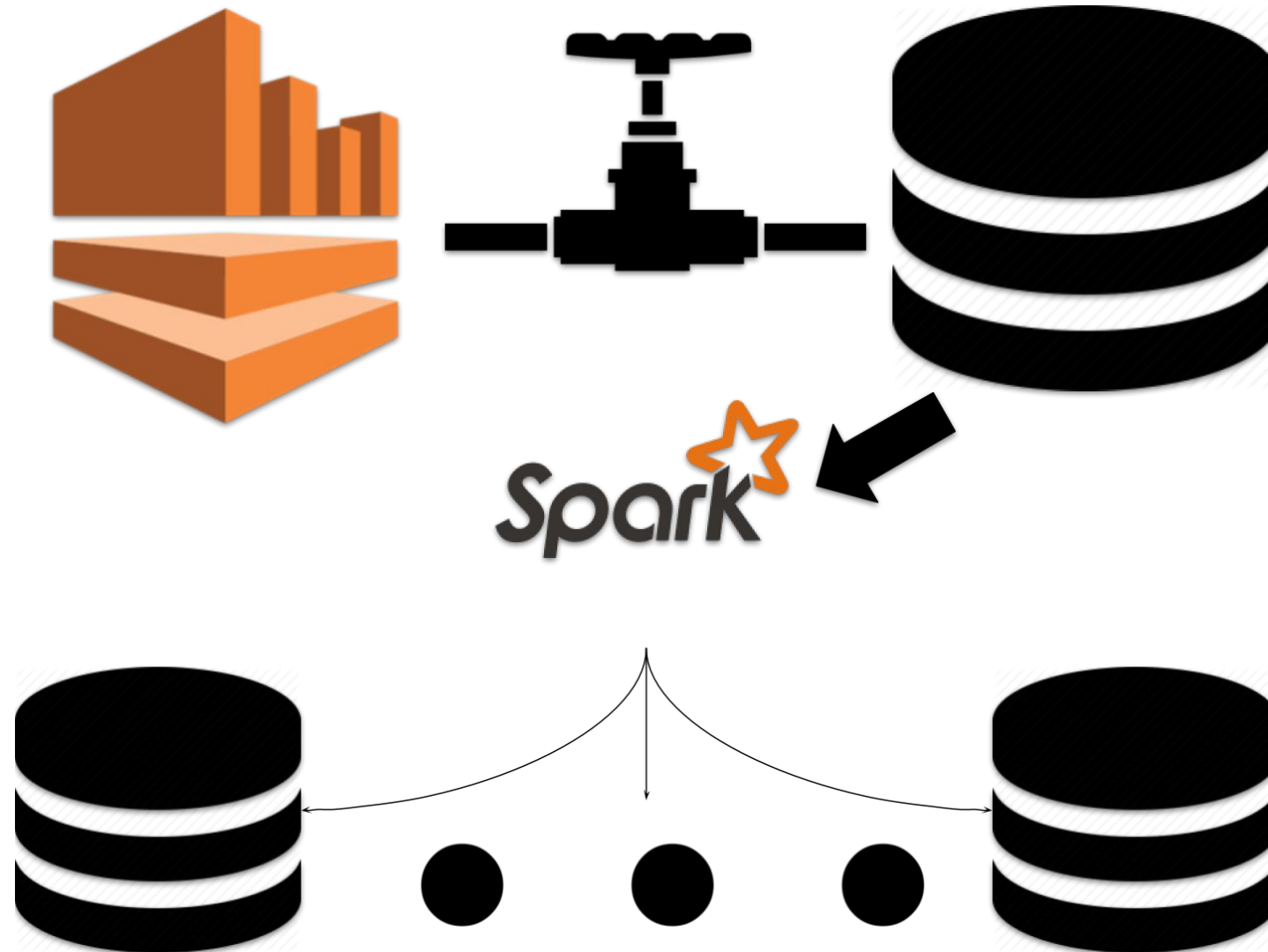
> 100 M / 30 minutes

> 50000 / s

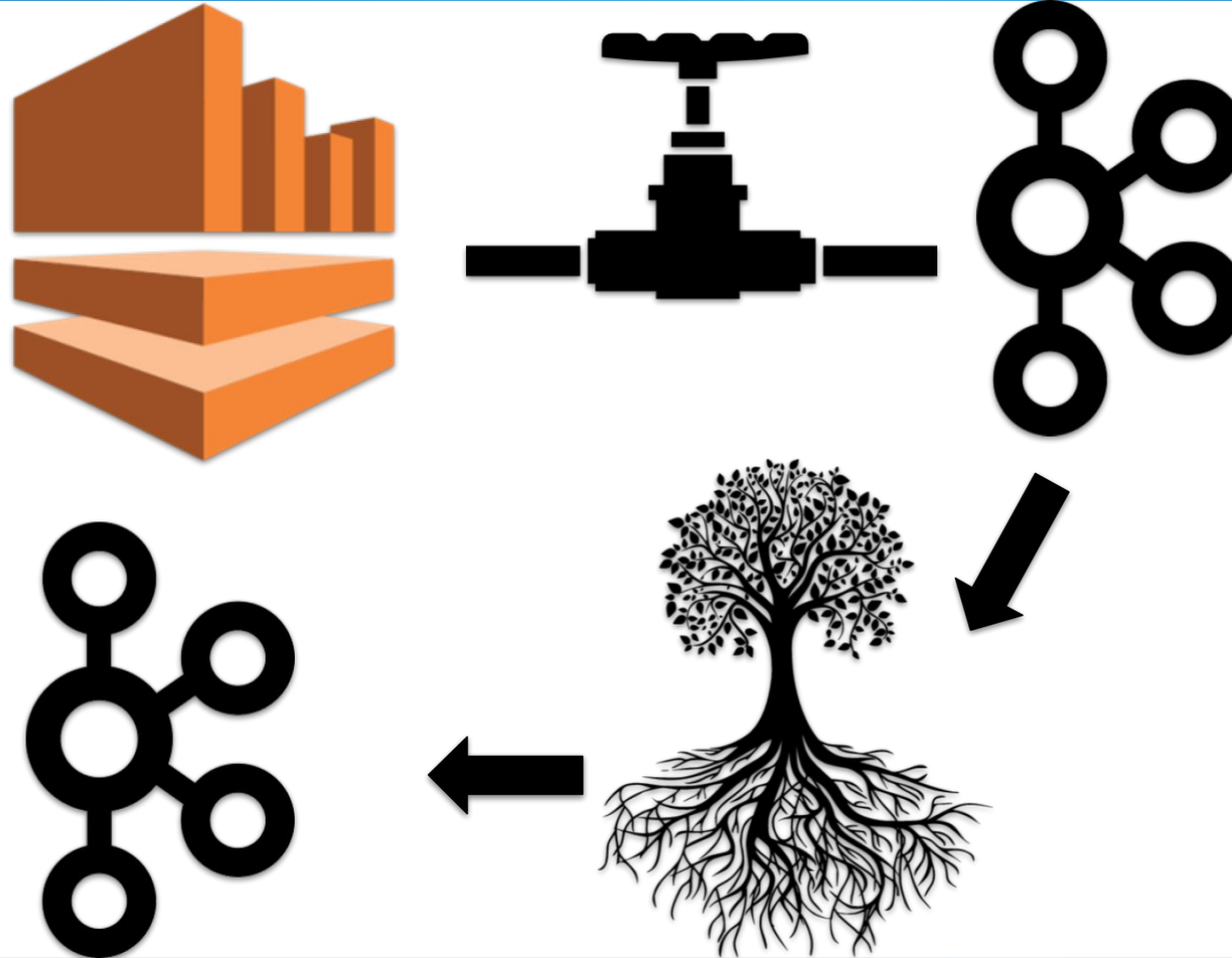
Common pipeline



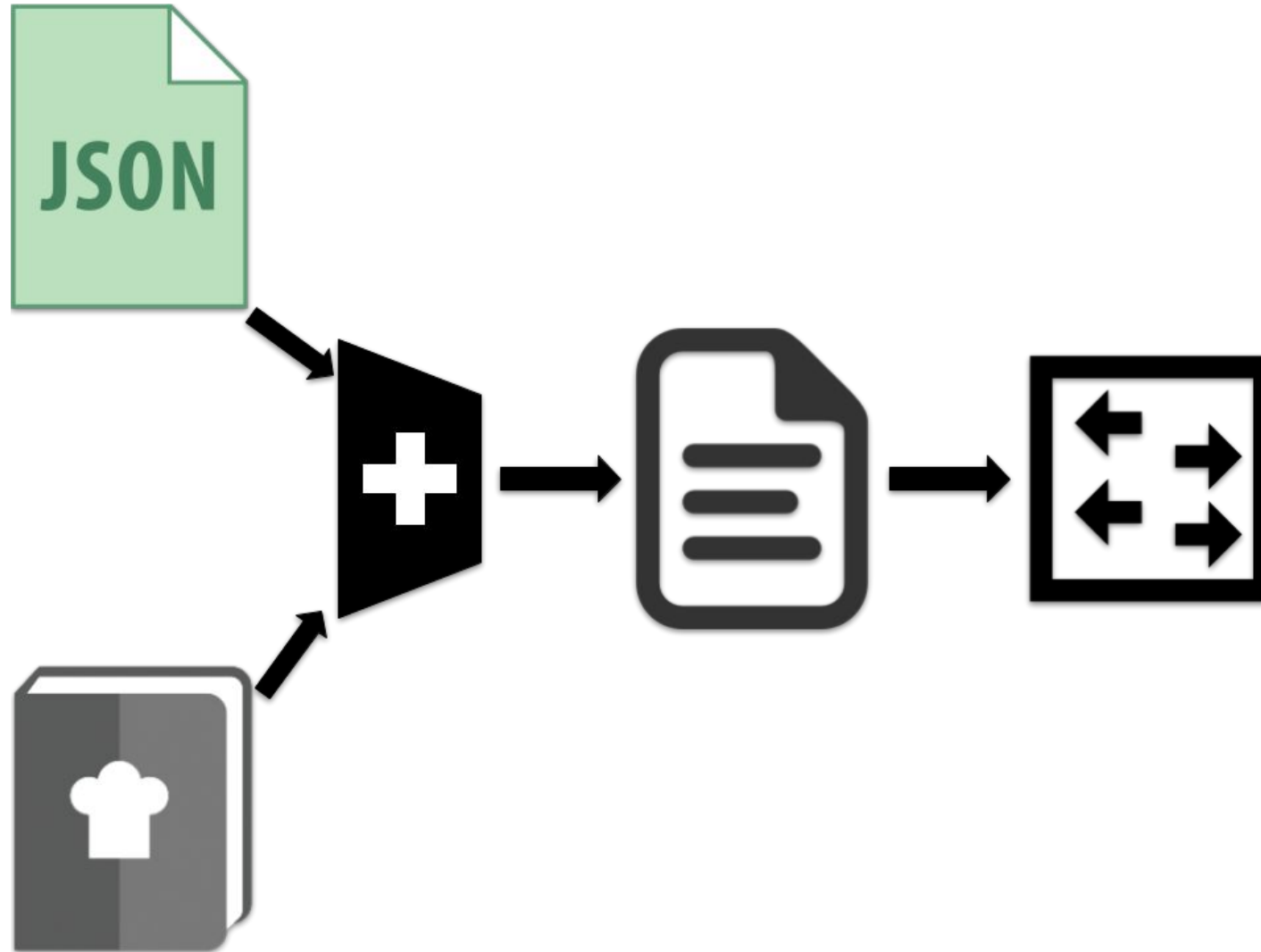
Batch pipeline



Streaming pipeline



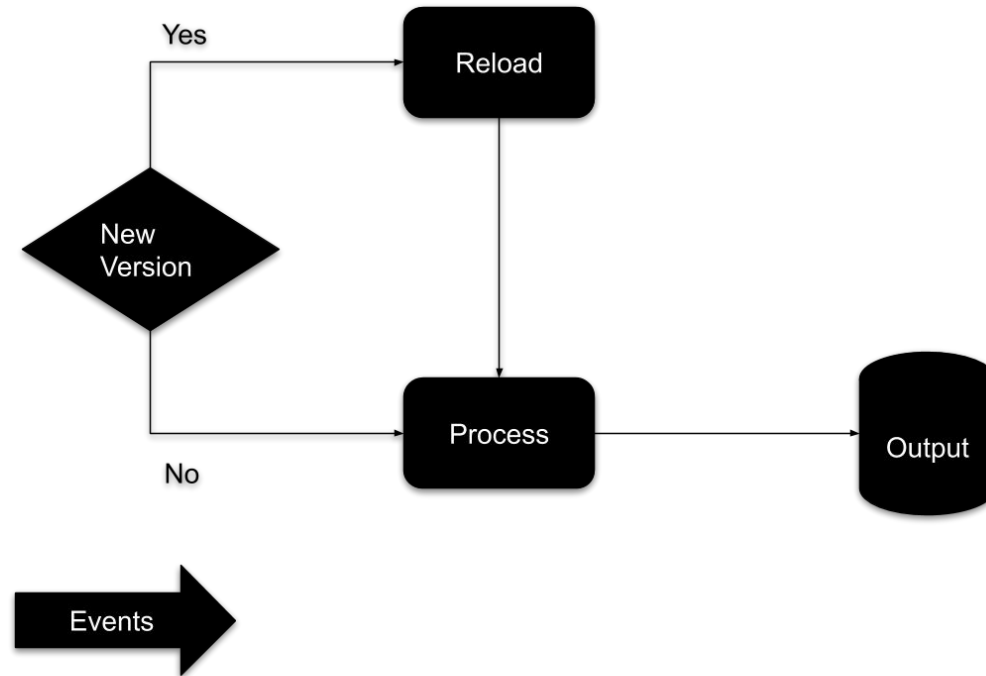
Processing and routing internals



Routing lib



VersionWatcher



Processing: routing language

SinkName:

eventType: event schema

filter: inline || stored || null

transform: stored || null

SinkType:

SinkDetails:

ProbeEvent-1:

eventType: ProbeEvent

kafka:

topic: probe-topic

Event formats: probe event

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "allOf": [
    {
      "$ref": "base-routable-event.json#"
    }
  ],
  "description": "Events sent by Data Platform Probe to measure latencies and missing events in the pipeline",
  "id": "http://schema.schibsted.com/events/backend-probe-event.json#",
  "properties": {
    "senderId": {
      "description": "Sender ID, in case several instances of Probe is running",
      "type": "integer"
    },
    "sequenceNumber": {
      "description": "Probe sequence number",
      "type": "integer"
    },
    "timeSent": {
      "$ref": "../common-definitions.json#/definitions/timestamp",
      "description": "UTC timestamp of when the event is generated by Probe"
    }
  },
  "title": "BackendProbeEvent",
  "type": "object"
}
```

JSLT: The magic sauce of processing

JSON query and transformation language

Github repo: <https://github.com/schibsted/jslt>

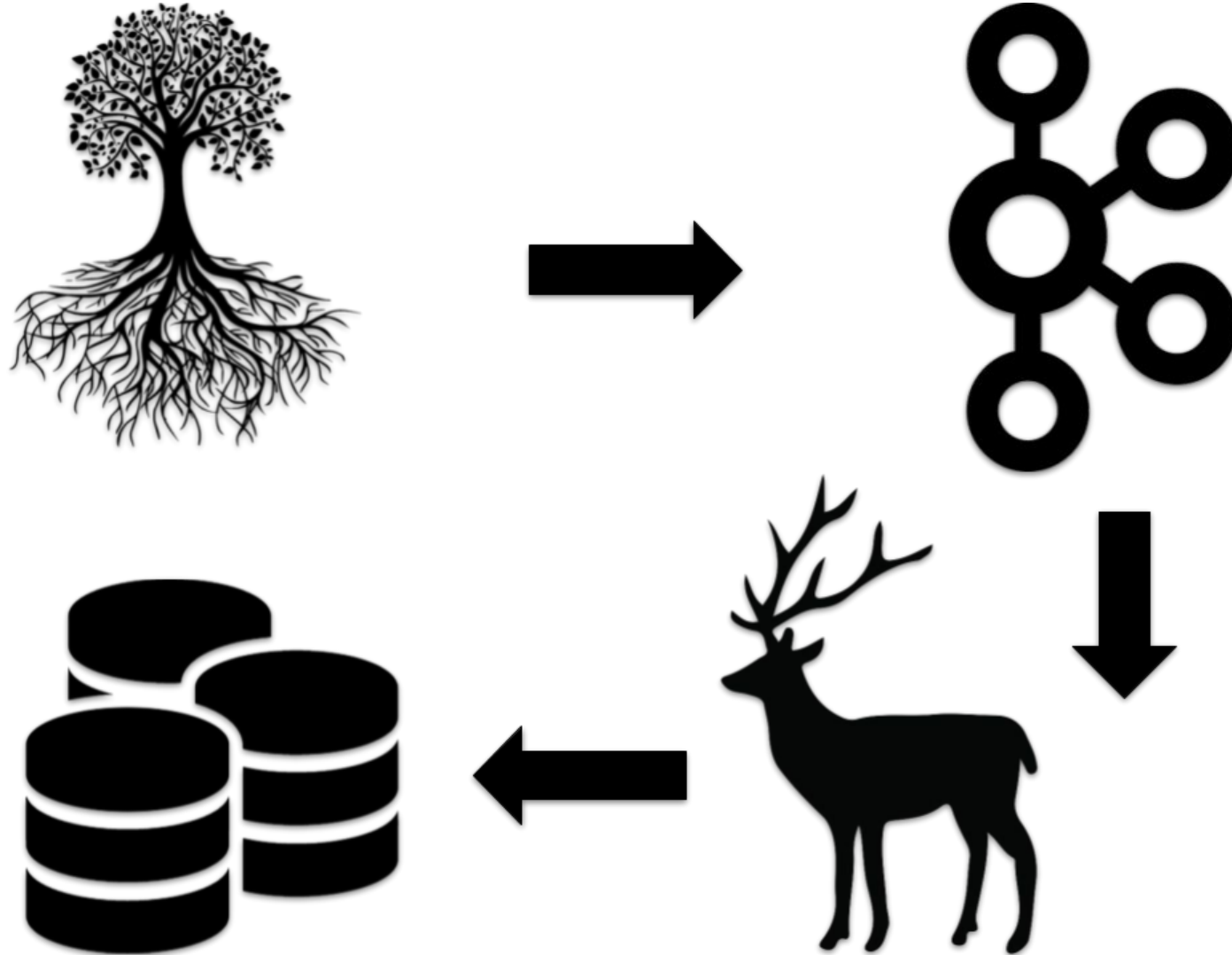
License: Apache 2.0

```
{
  "time": round(parse-time(.published, "yyyy-MM-dd'T'HH:mm:ssX") * 1000),
  "device_manufacturer": .device.manufacturer,
  "device_model": .device.model,
  "language": .device.acceptLanguage,
  "os_name": .device.osType,
  "os_version": .device.osVersion,
  "platform": .device.platformType,
  "user_properties": {
    "is_logged_in" : boolean(.actor."spt:userId")
  }
}
```

Routing: batch



Routing: streaming



Lessons learned (so far...)

- Schemas and versions
- Backfilling and recovery
- Logging and metrics
- Auditing

And finally

WE'RE

HIRING!



Extra

About Schibsted



1839 - Schibsted Forlag established



1860 - Aftenposten is founded



1992 - Listed in Oslo exchange



1966 - VG acquired



2000 - Finn.no launched

Today - Leader in Media and marketplaces in the Nordics

Marketplaces

blocket



shpock

tori



News Media

VG

Aftenposten

AFTONBLADET

SvD

Bergens Tidende

Omni

Fædrelandsvennen

Stavanger Aftenblad

Some of our Next companies

