Sweet Streams are Made of These

Data Driven Development for Stream Processing

@CAITO_200_OK | VERVERICA

Introduction

- Caito Scherr

@CAITO_200_OK

Introduction

- Caito Scherr
- Developer Advocate

@CAITO_200_OK

Introduction

- Caito Scherr
- Developer Advocate
- Ververica, GmbH

@CAITO_200_OK

Introduction

- Caito Scherr
- Developer Advocate
- Ververica, GmbH
- Portland, OR, USA

@CAITO_200_OK

Introduction

@CAITO_200_OK

Agenda

- The Challenge
- “Metrics-Driven Metrics”
- Metrics as a Shared Language

@CAITO_200_OK

Agenda

- The Challenge
- “Metrics-Driven Metrics”
- Metrics as a Shared Language

@CAITO_200_OK

Agenda

- The Challenge
- “Metrics-Driven Metrics”
- Metrics as a Shared Language

@CAITO_200_OK

Before We Start...

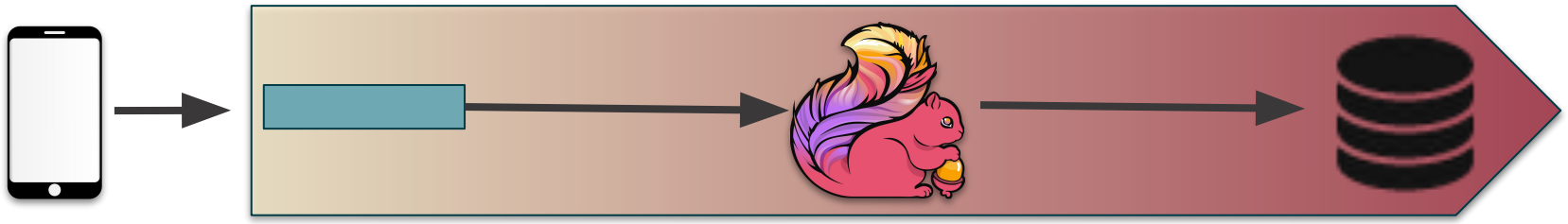
@CAITO_200_OK

Before We Start...

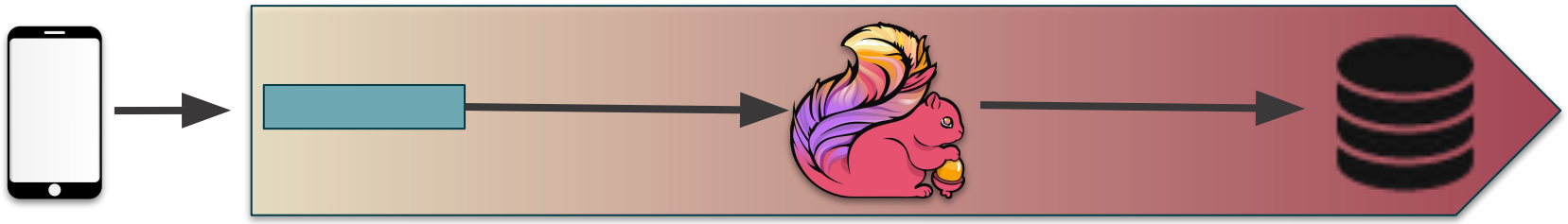
The Challenge

@CAITO_200_OK

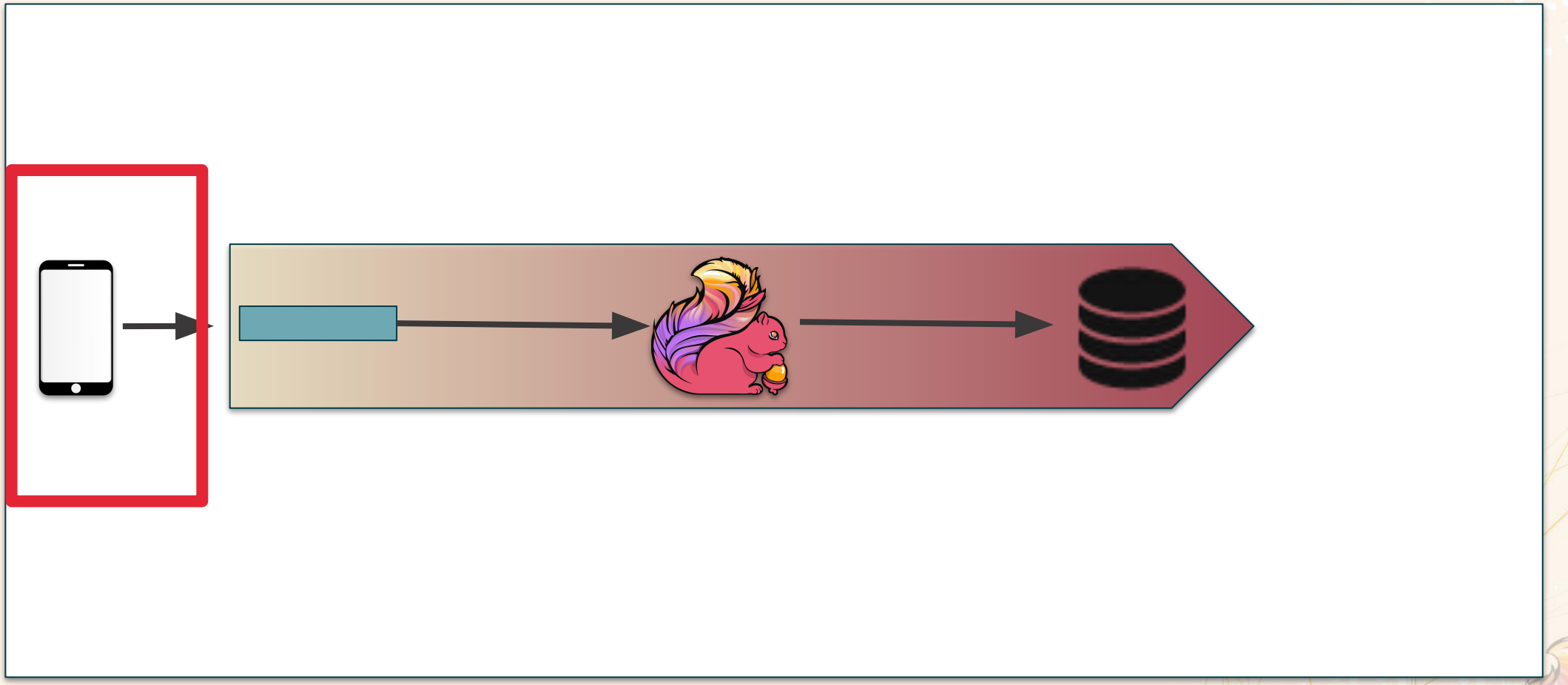
The Challenge



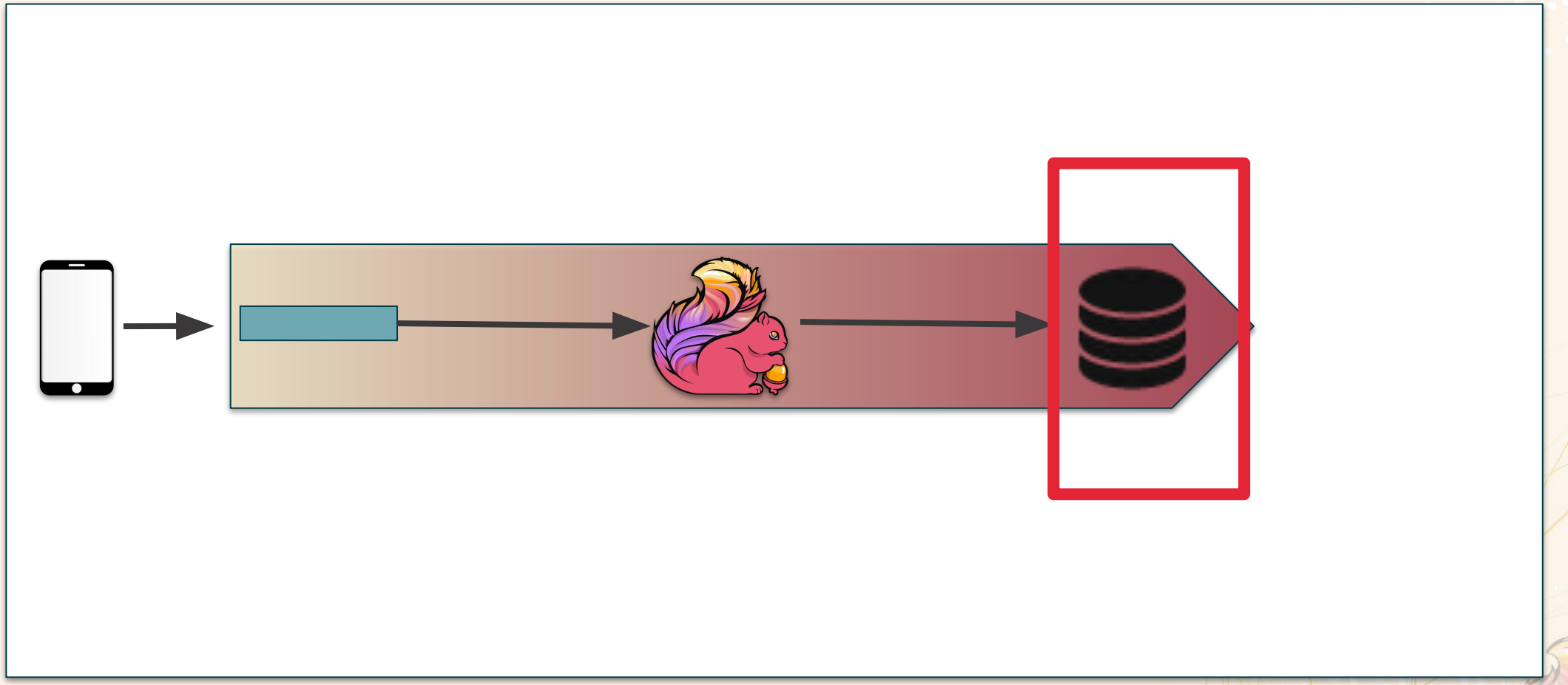
The Challenge



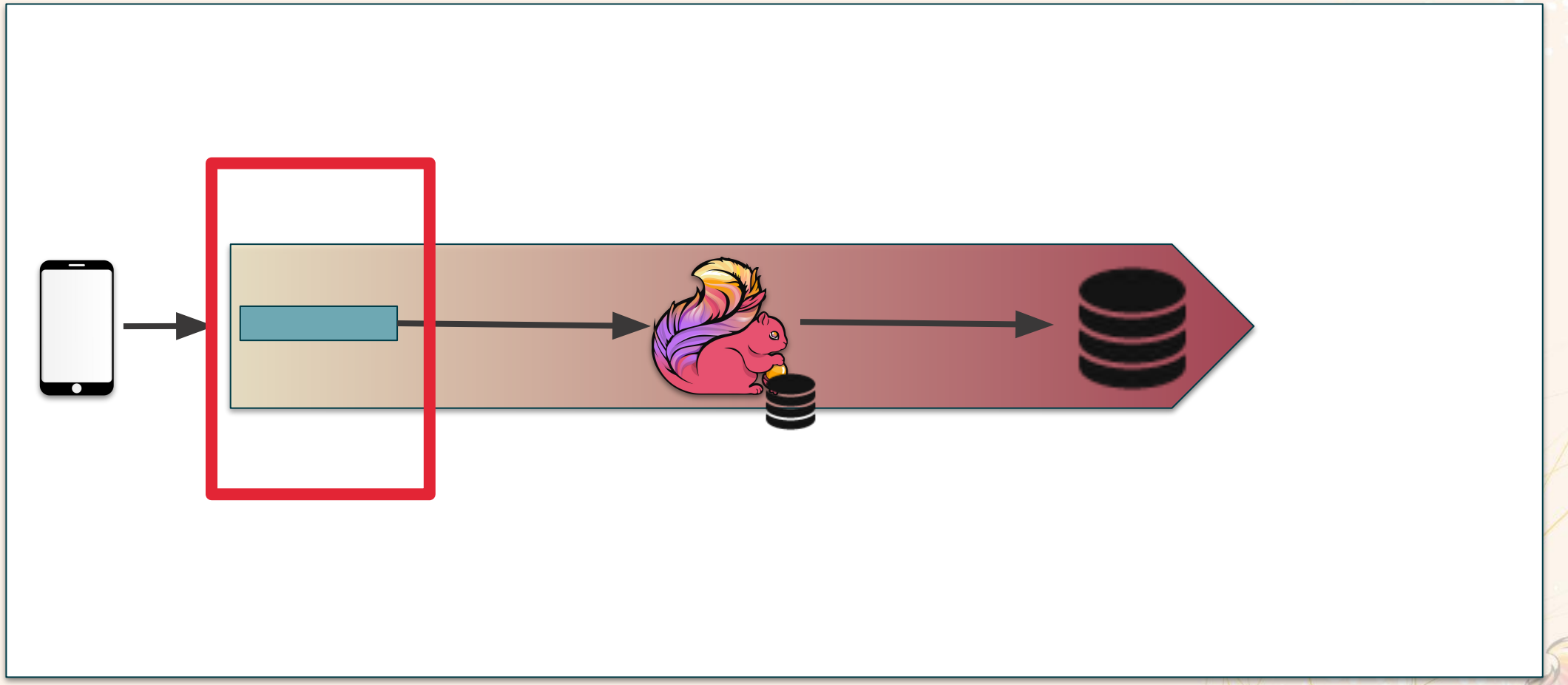
The Challenge



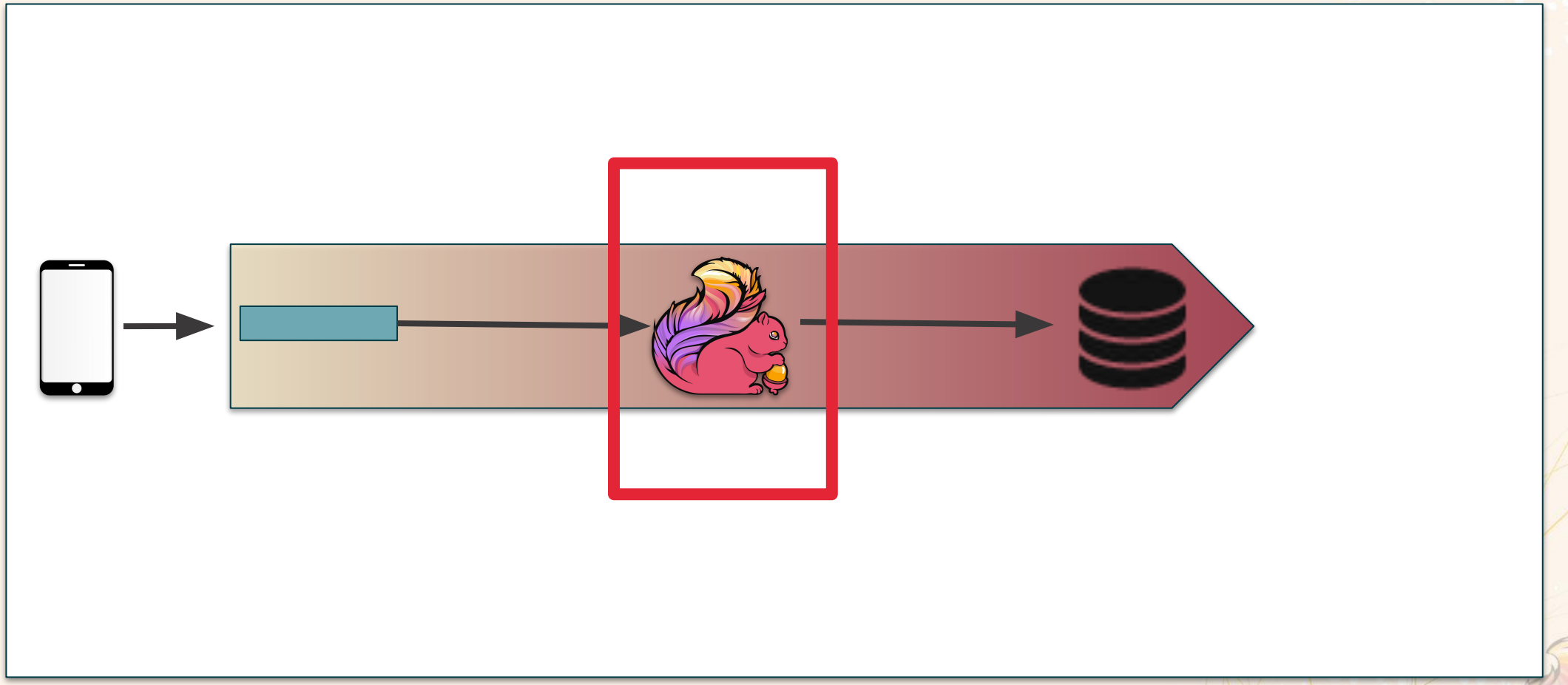
The Challenge



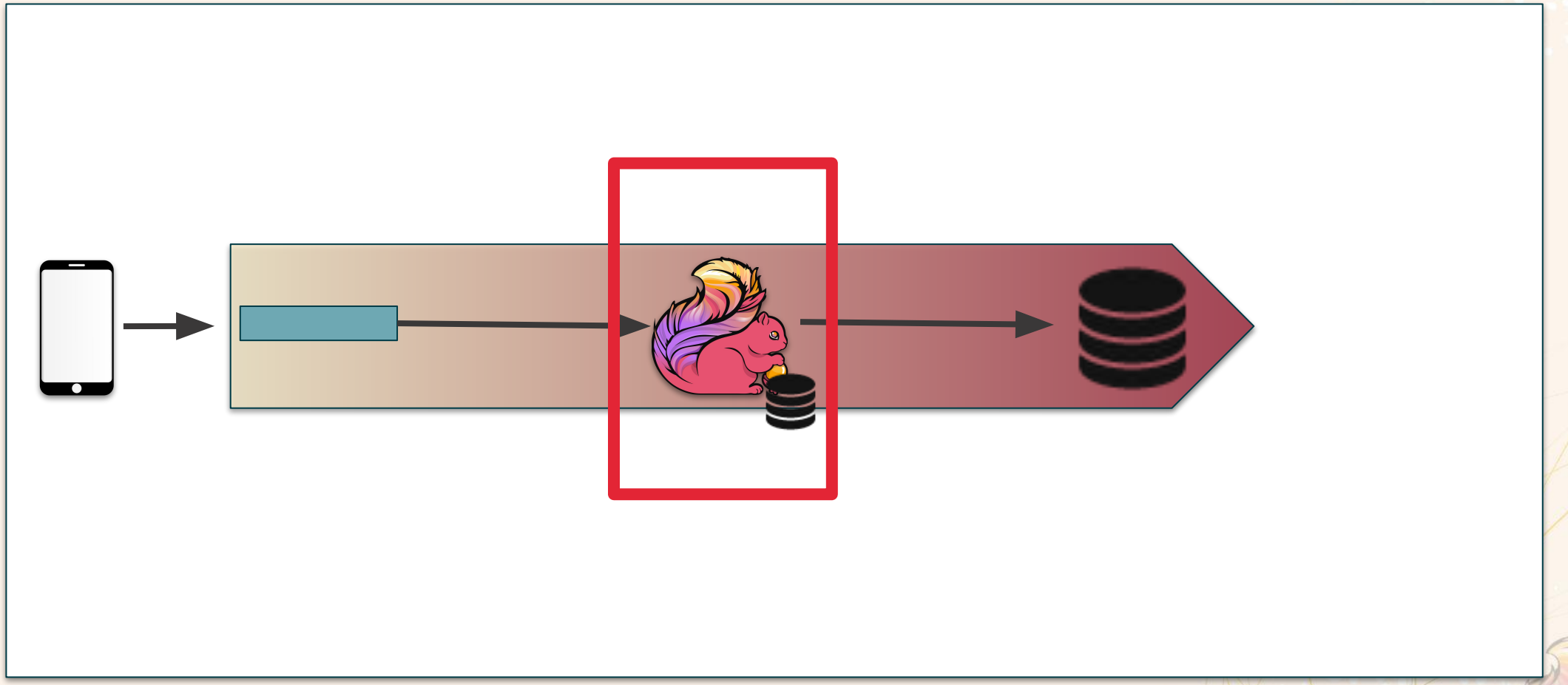
The Challenge



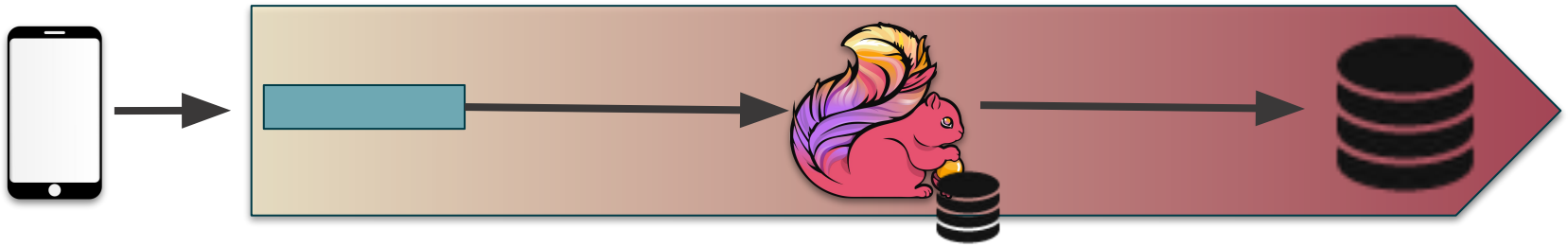
The Challenge



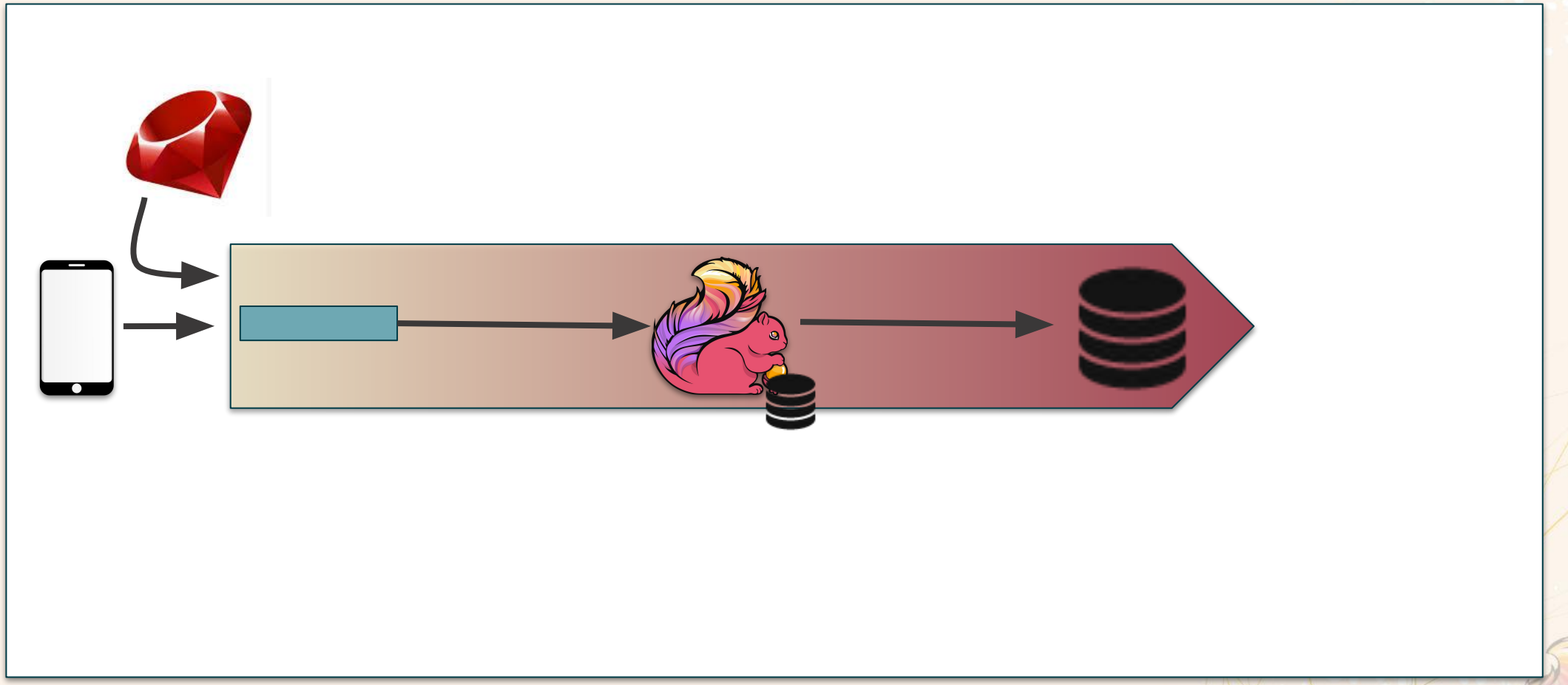
The Challenge



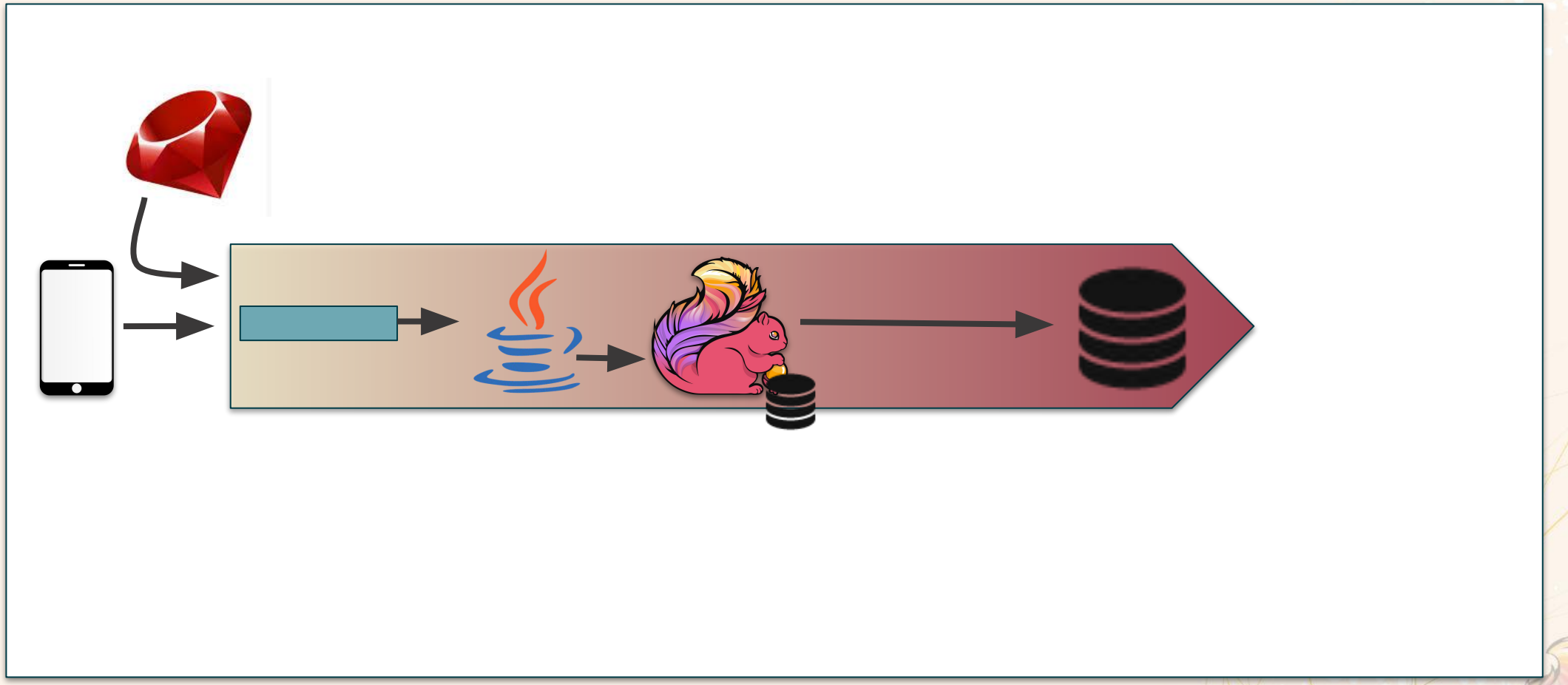
The Challenge



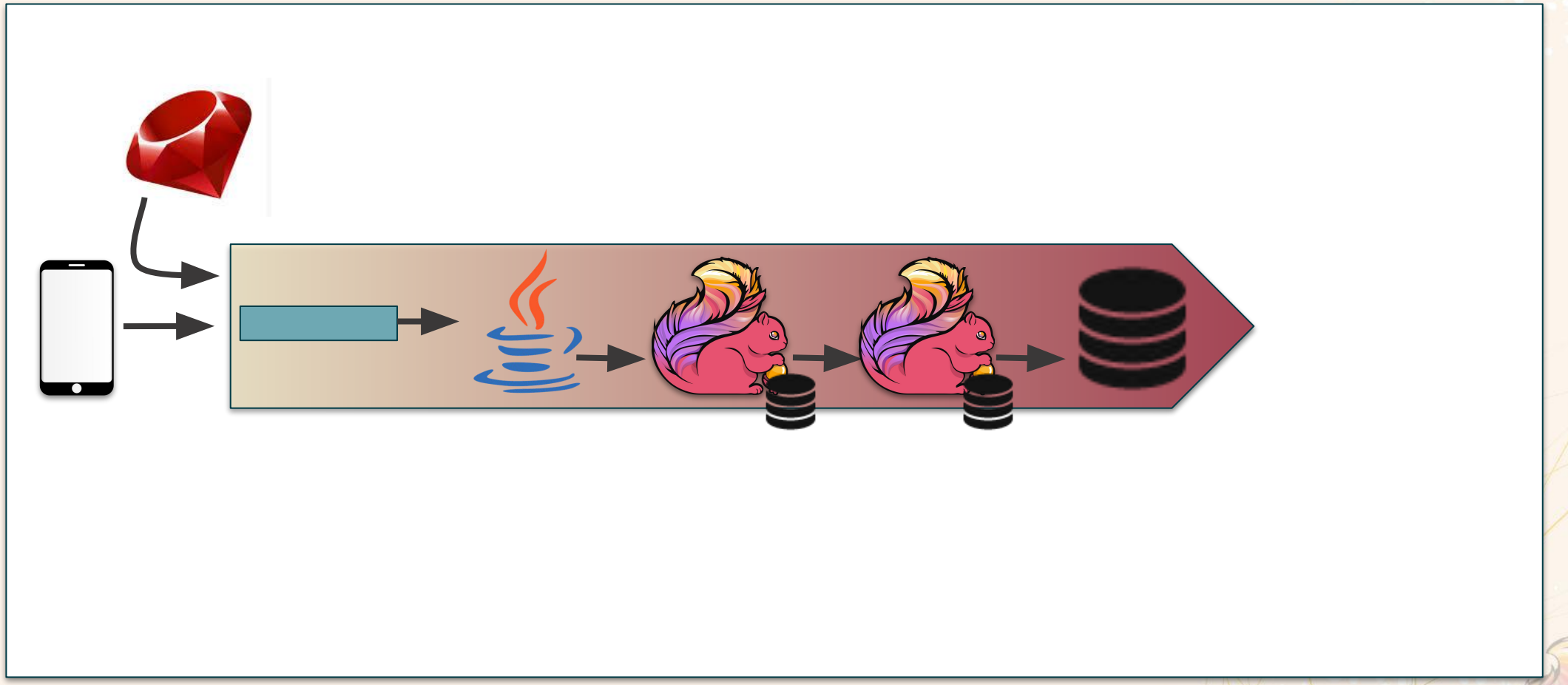
The Challenge



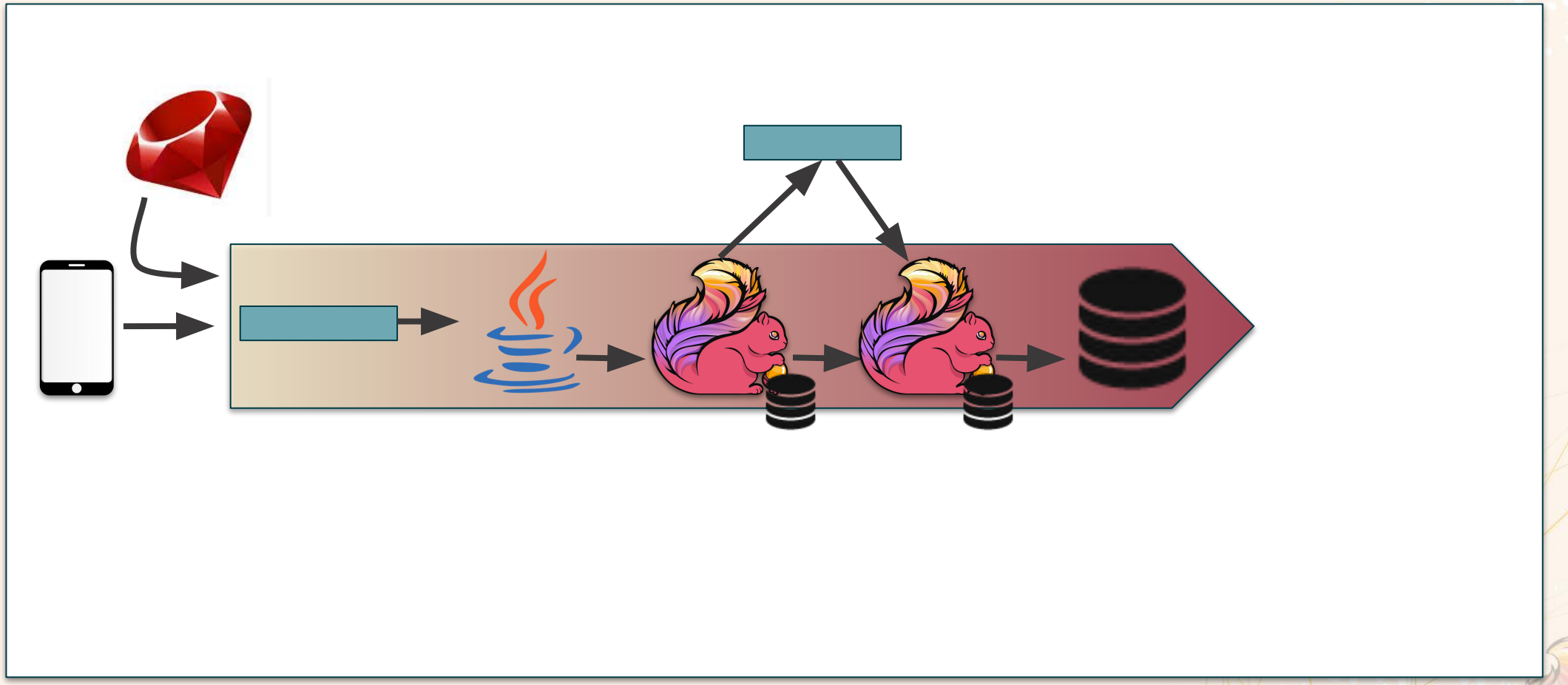
The Challenge



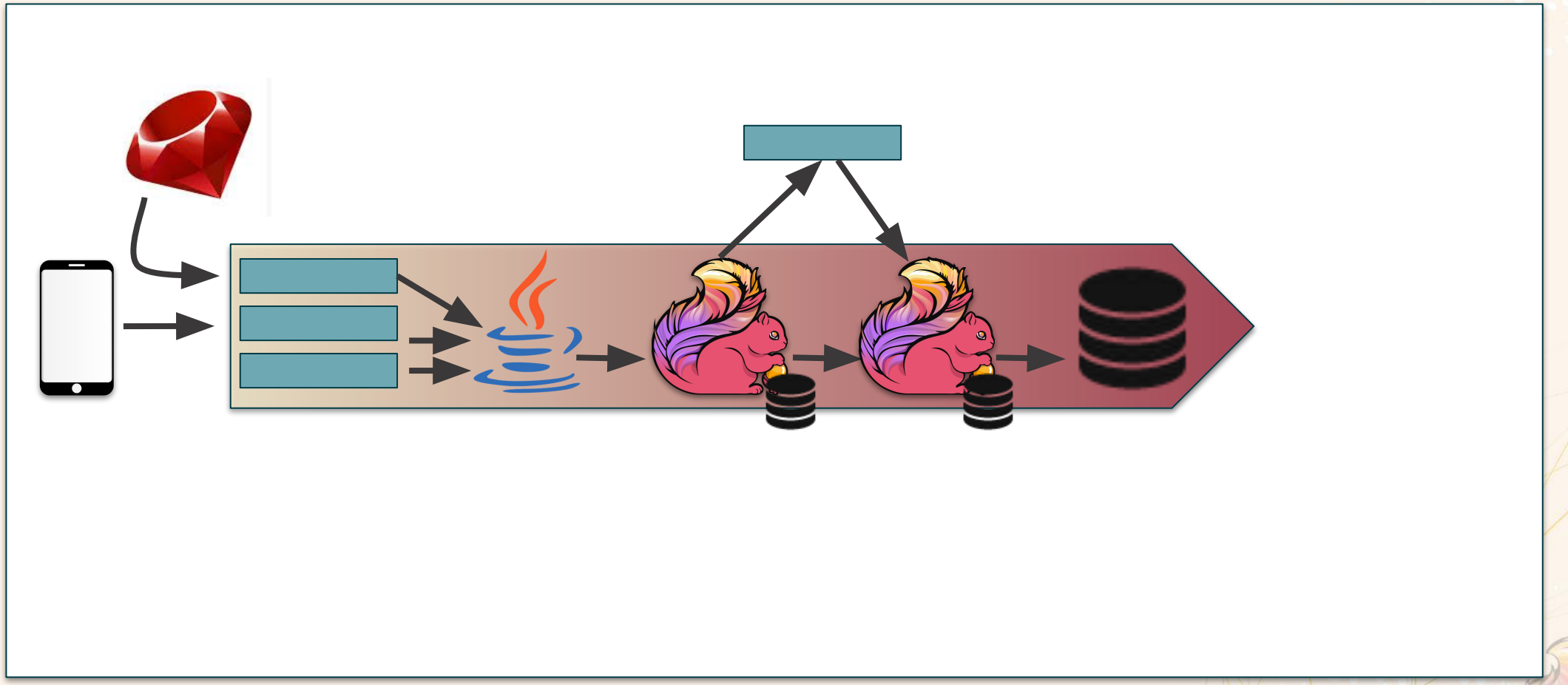
The Challenge



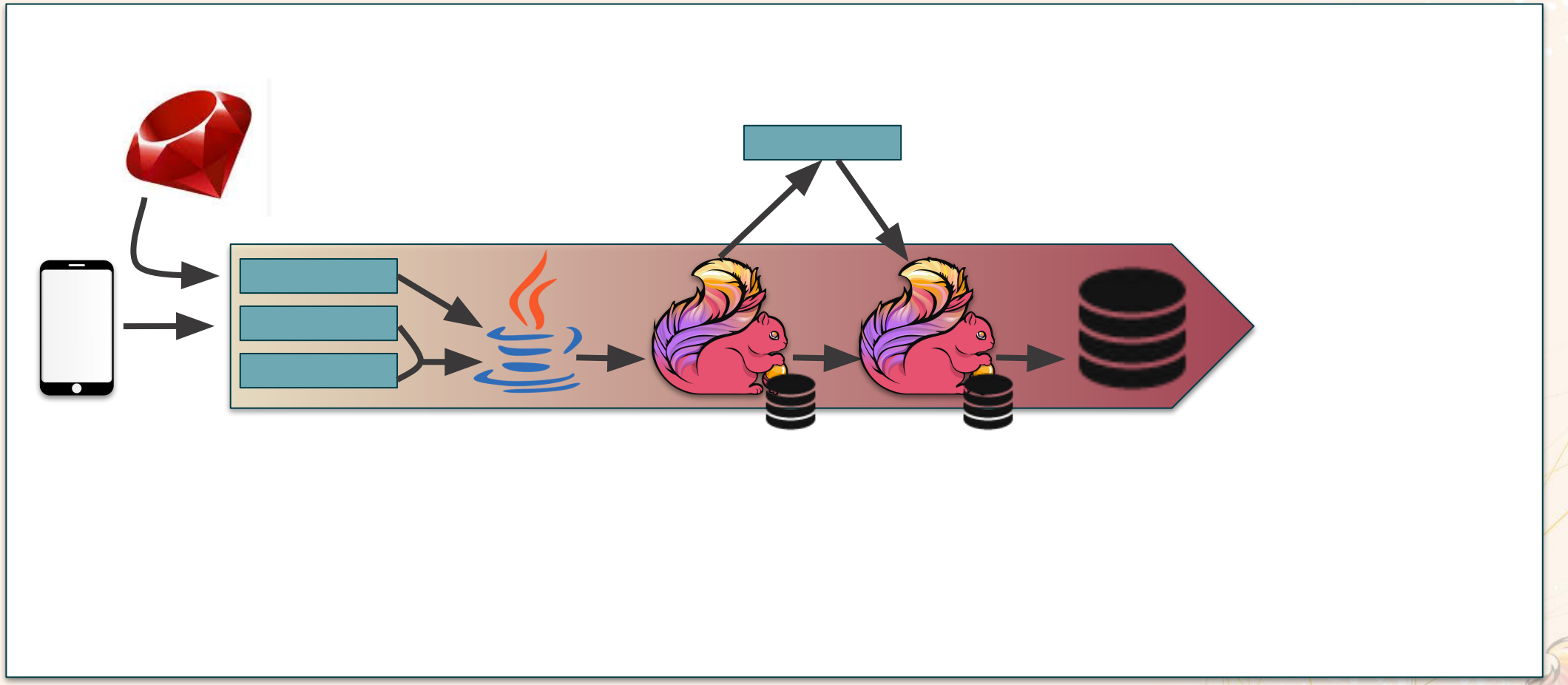
The Challenge



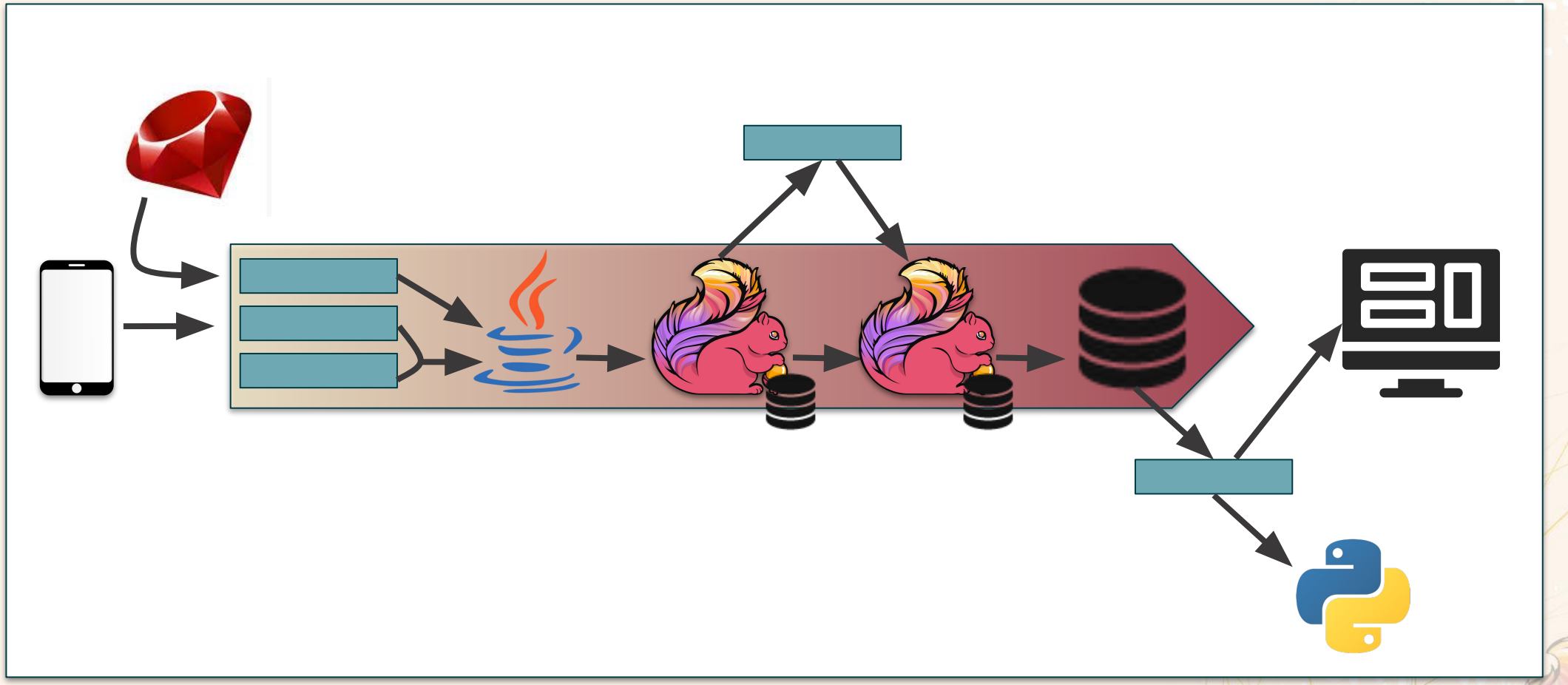
The Challenge



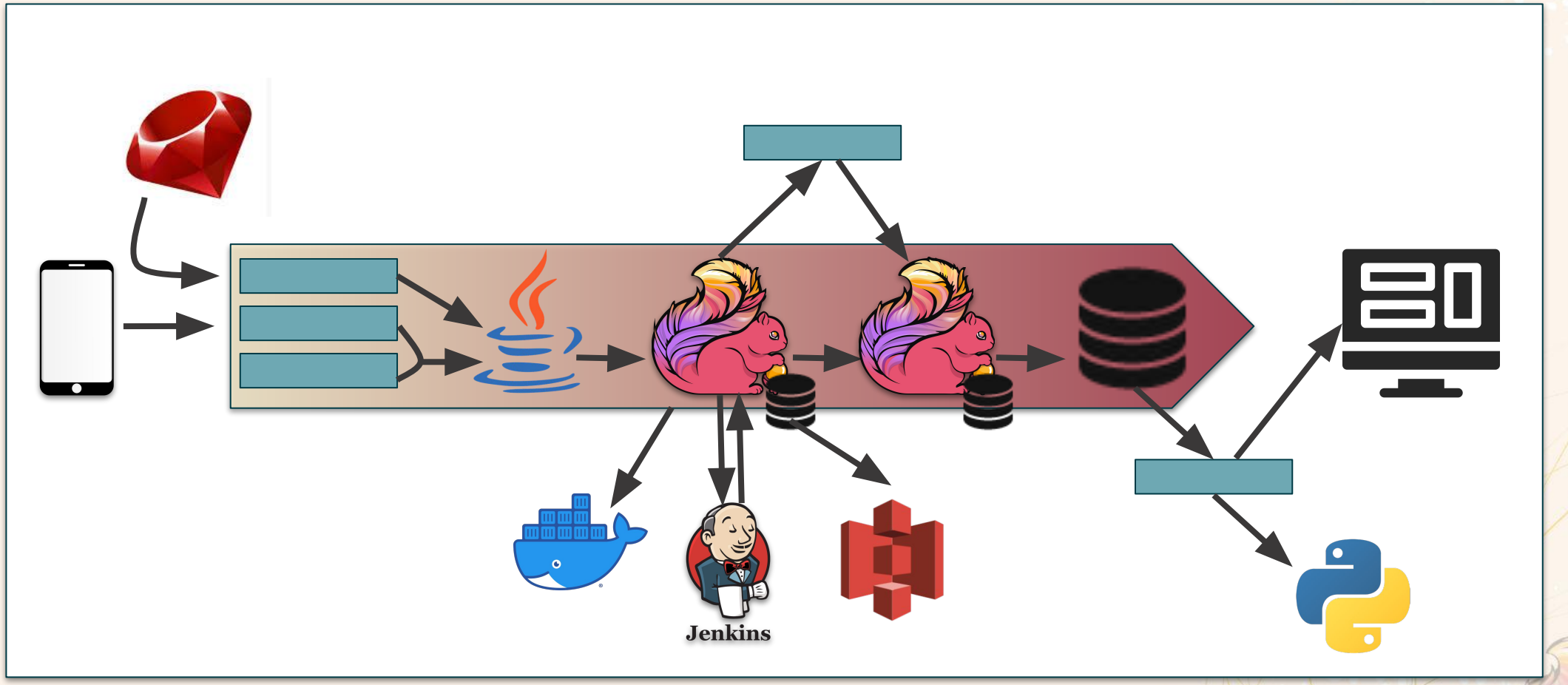
The Challenge



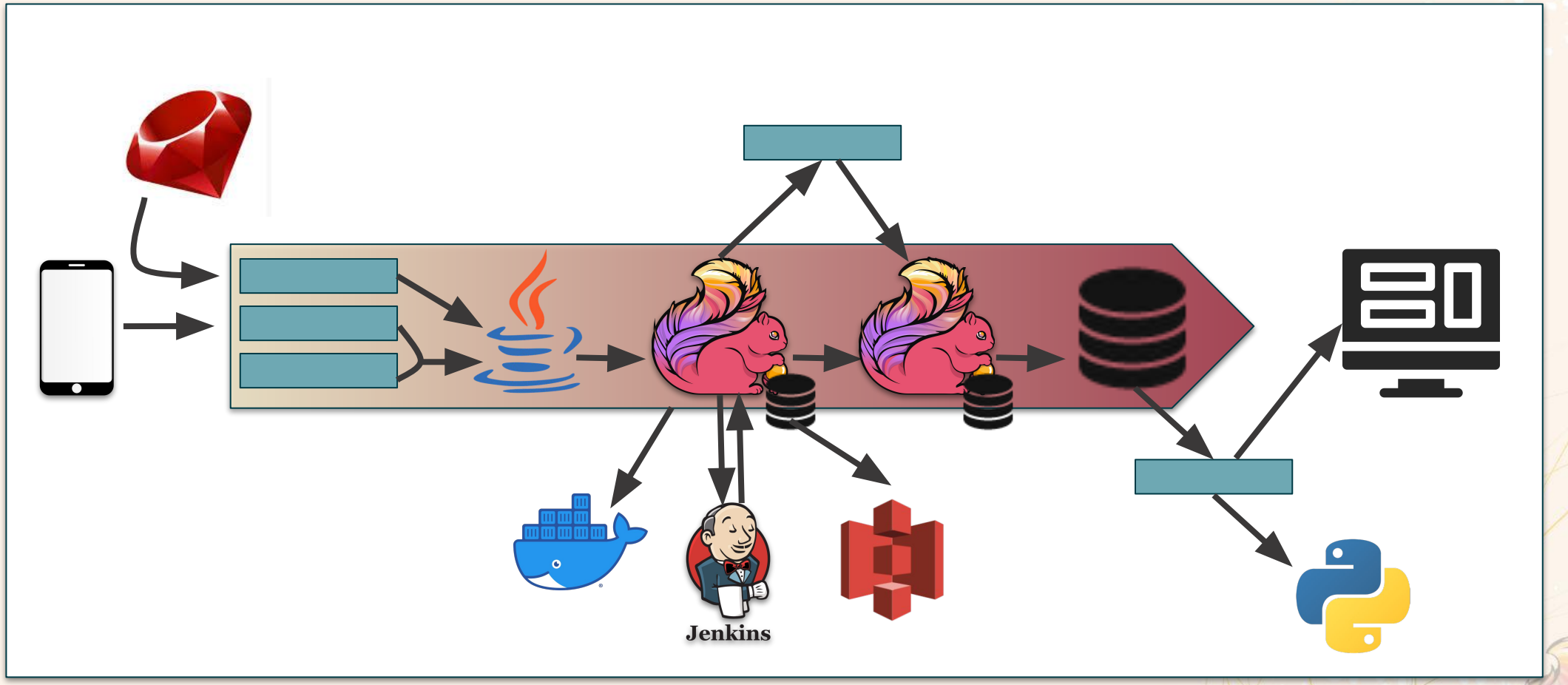
The Challenge



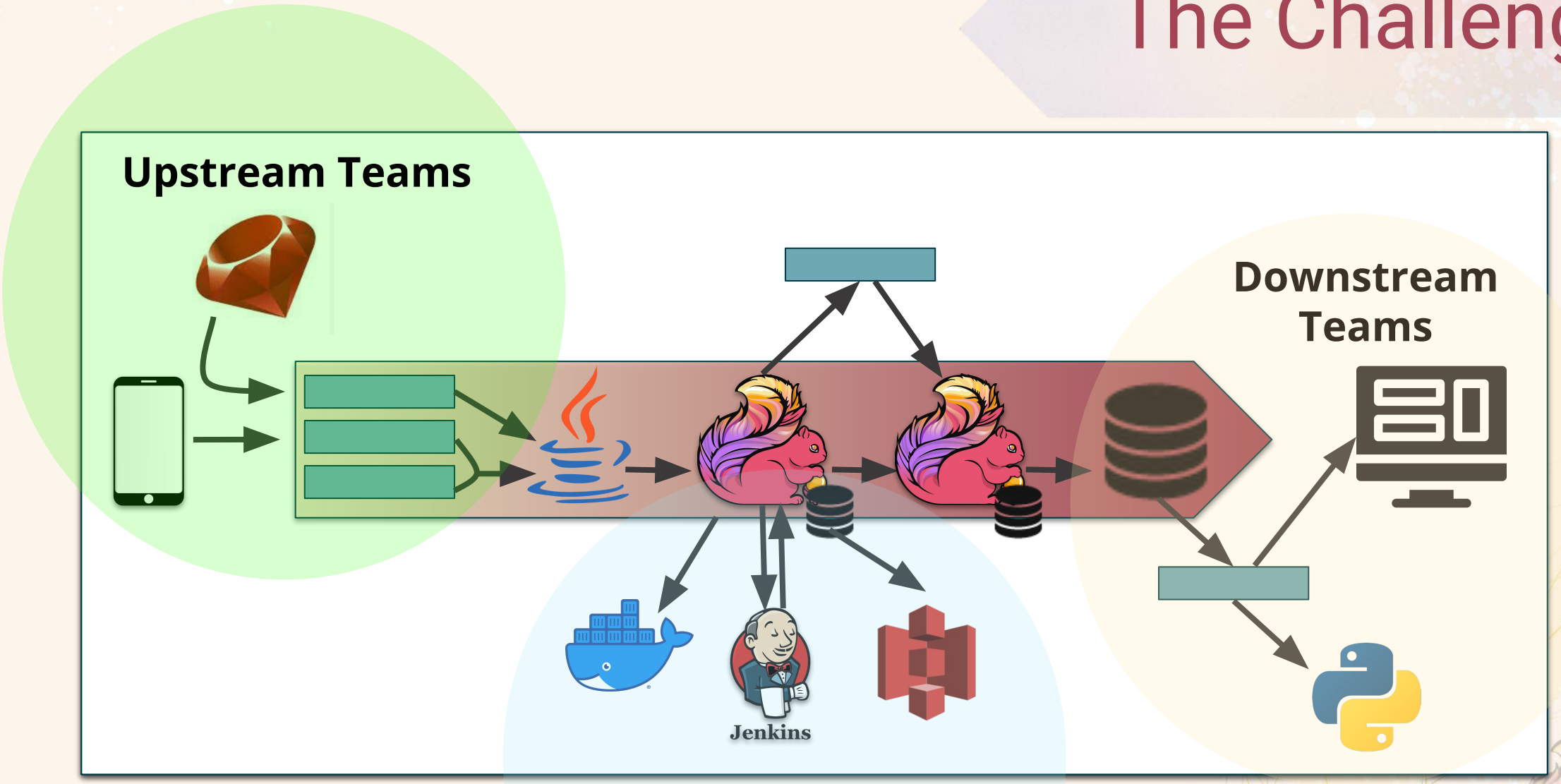
The Challenge



The Challenge



The Challenge



@CAITO_200_OK

"Metrics-Driven Metrics"

@CAITO_200_OK

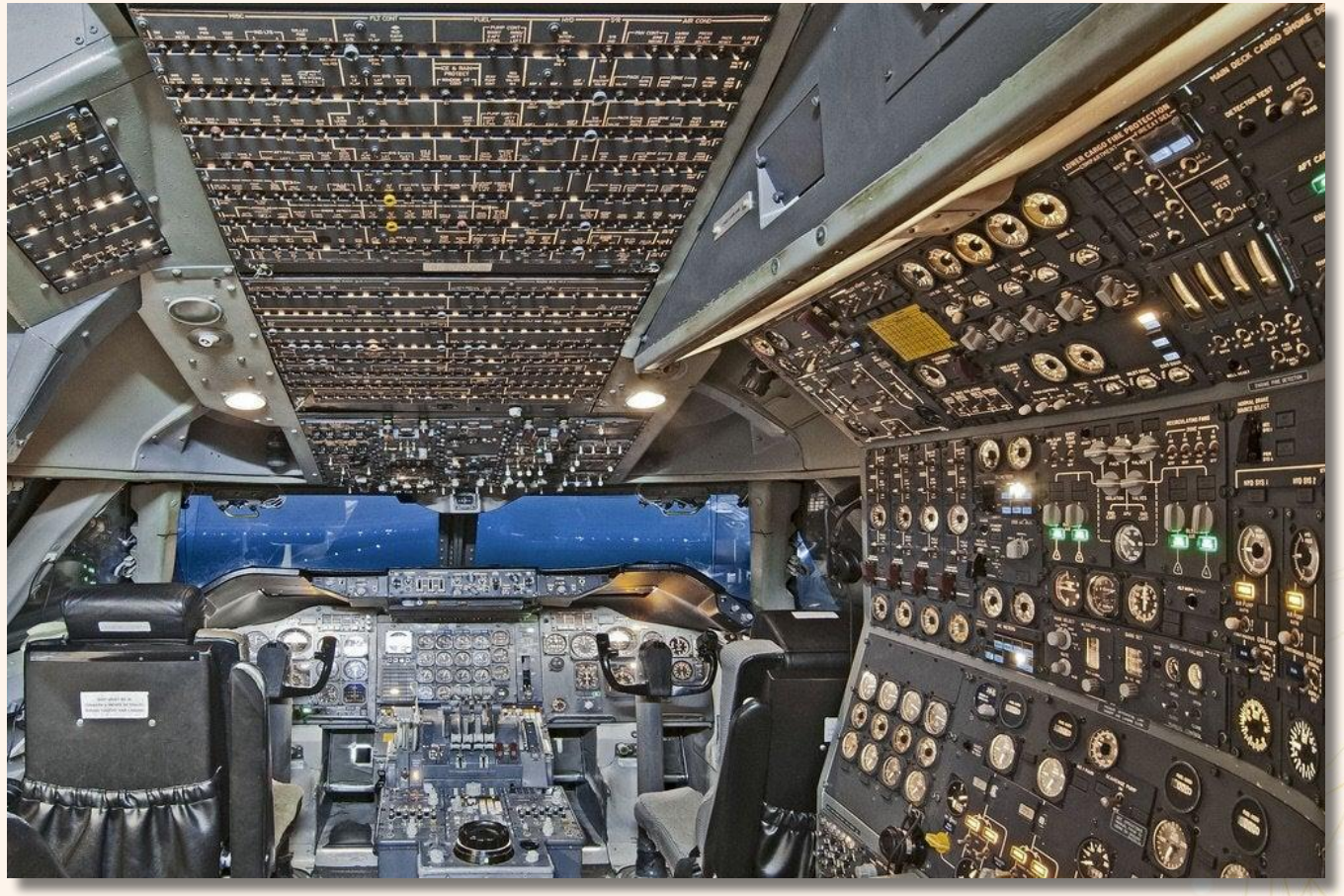
Metrics-Driven Metrics

"You can measure almost anything, but you cannot pay attention to everything"

Metrics-Driven Metrics

"Any situation where people create their own dashboards without structure, quickly starts to look like the cockpit of a 747"

Metrics-Driven Metrics



@CAITO_200_OK

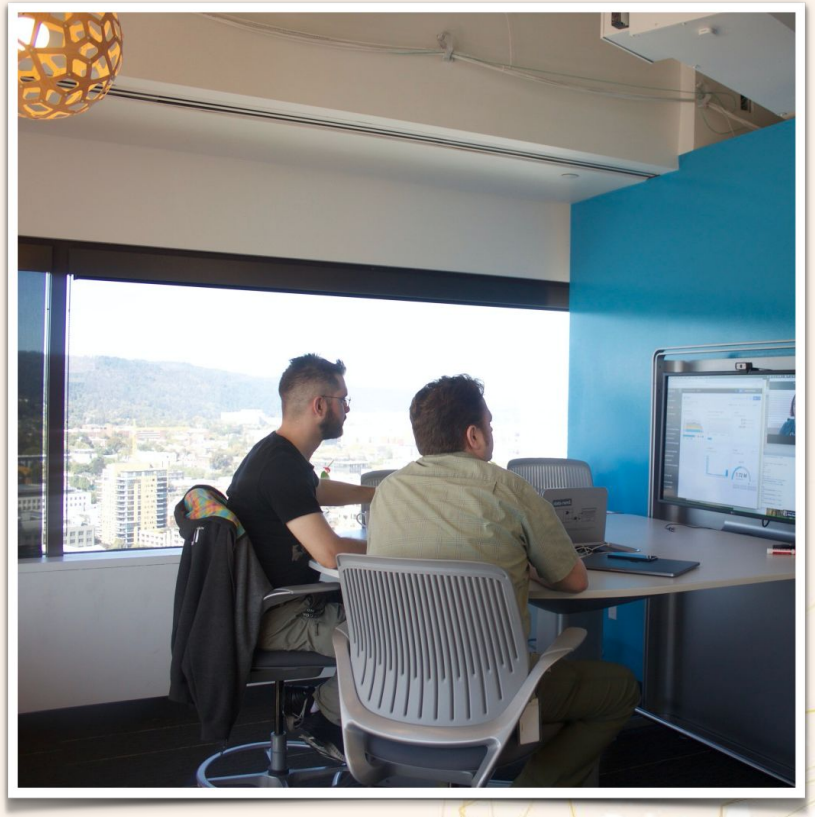
Metrics-Driven Metrics

@CAITO_200_OK

Metrics-Driven Metrics

The dashboard should **be**:

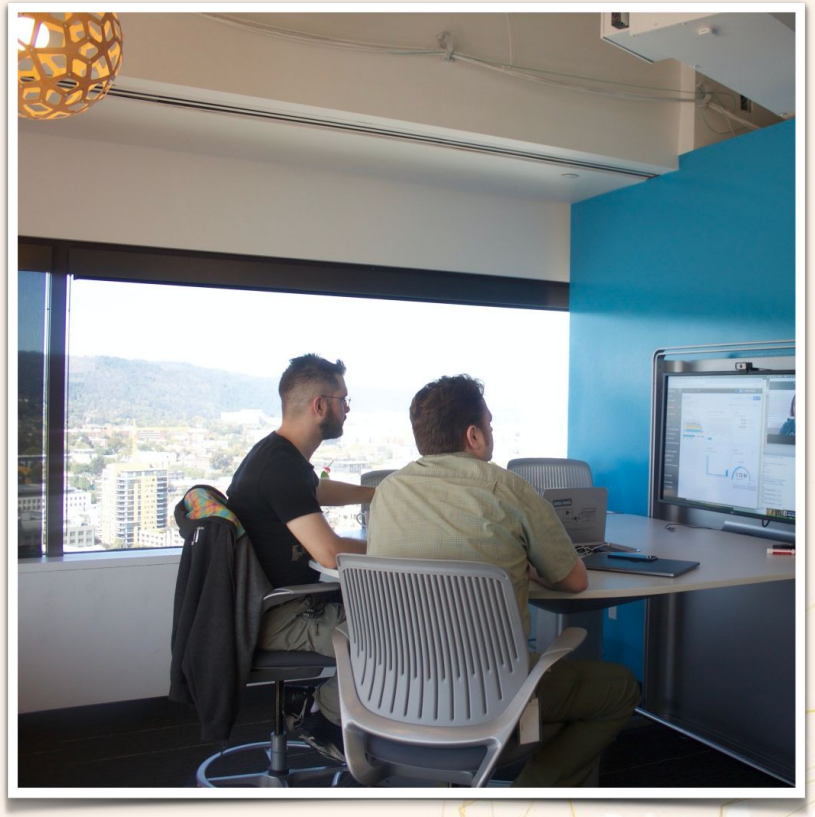
- Meaningful
- Iterative
- Accessible



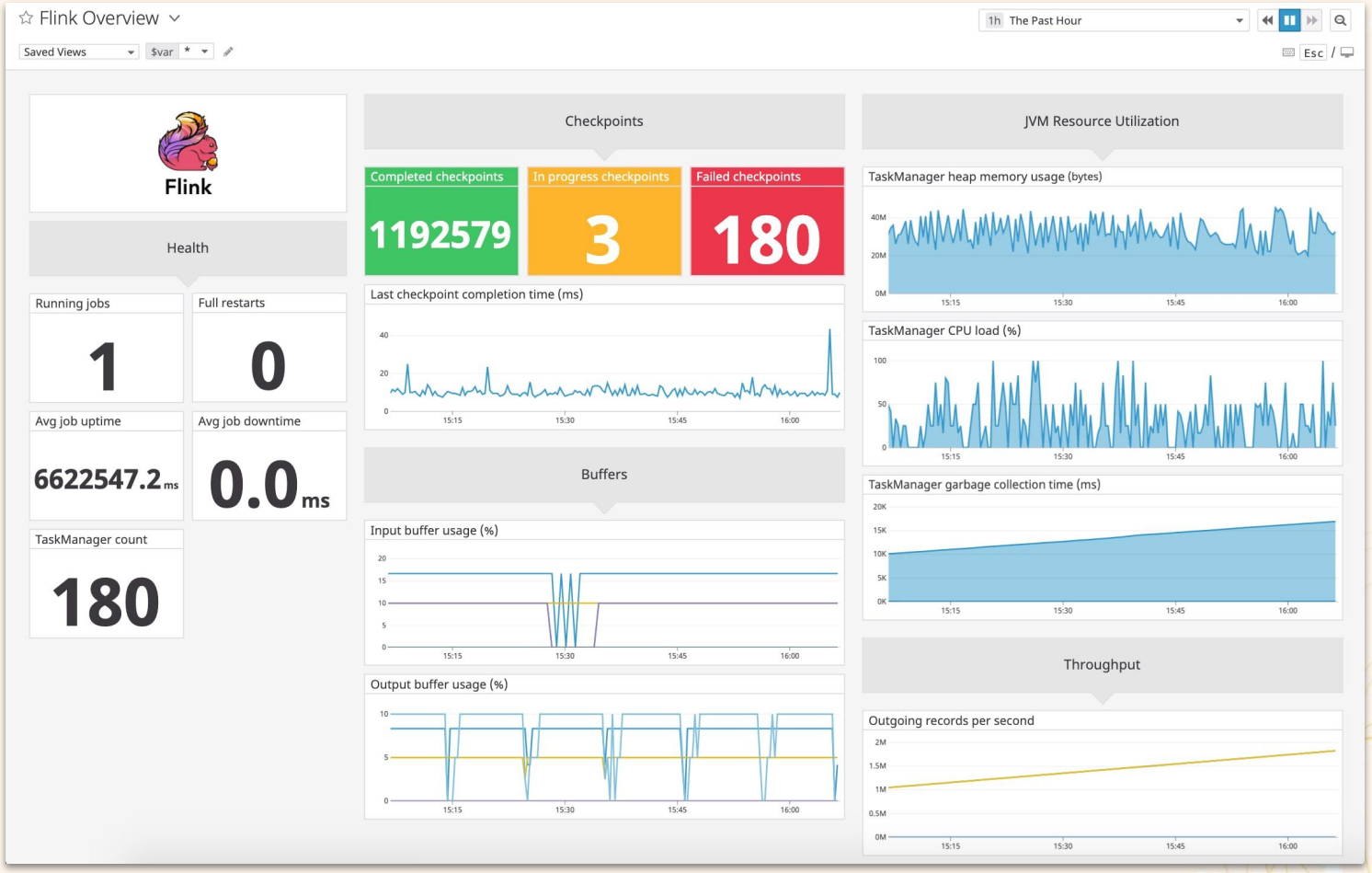
Metrics-Driven Metrics

The dashboard should **reflect** your:

- Roadmap
- Highest Risk
- Most uncertain metrics



Metrics-Driven Metrics



@CAITO_200_OK

Metrics-Driven Metrics

Flink Overview

Saved Views | $var *

1h The Past Hour

Esc /

Flink

Health

Running jobs	Full restarts
1	0

Avg job uptime	Avg job downtime
6622547.2 ms	0.0 ms

TaskManager count

180

Checkpoints

Completed checkpoints	In progress checkpoints	Failed checkpoints
1192579	3	180

Last checkpoint completion time (ms)

Buffers

Input buffer usage (%)

Output buffer usage (%)

JVM Resource Utilization

TaskManager heap memory usage (bytes)

TaskManager CPU load (%)

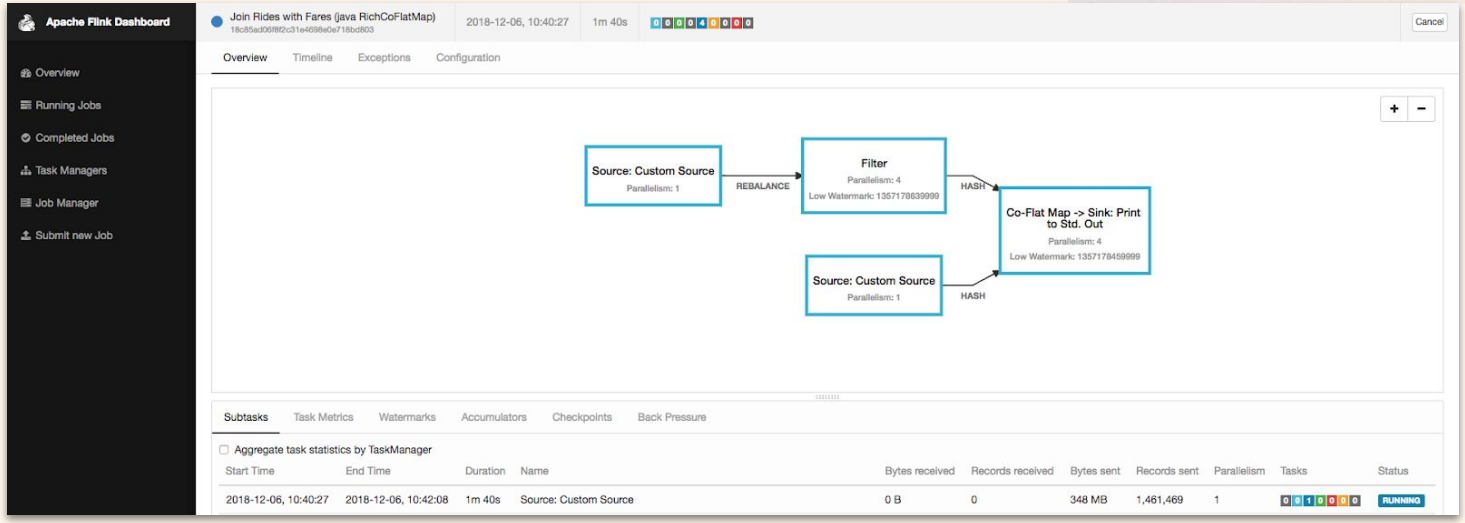
TaskManager garbage collection time (ms)

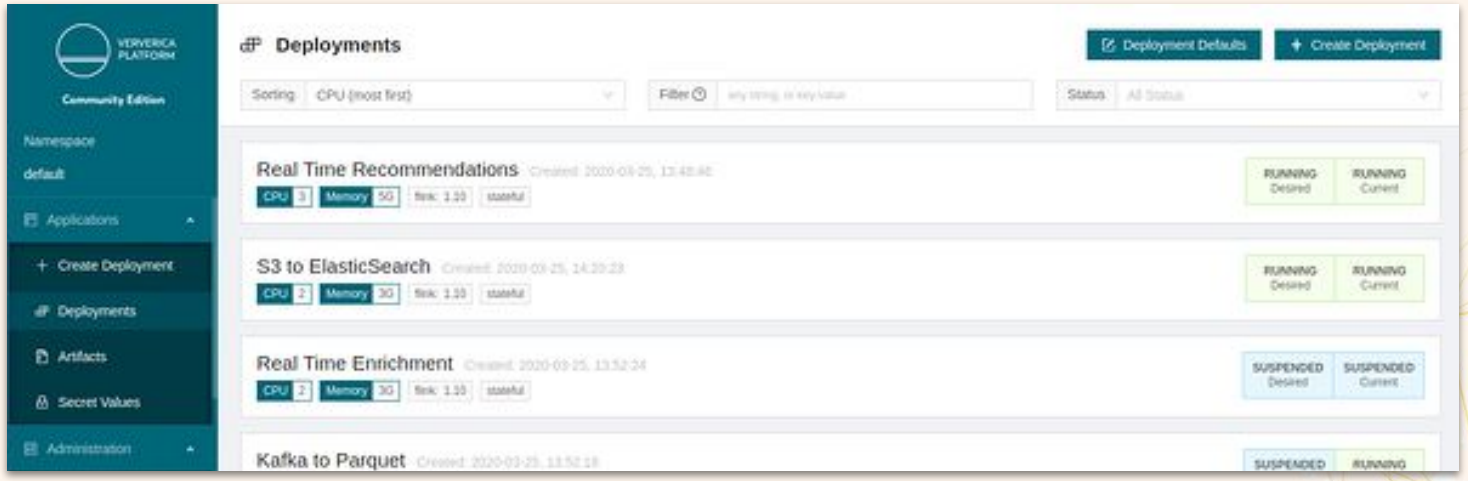
Throughput

Outgoing records per second

@CAITO_200_OK

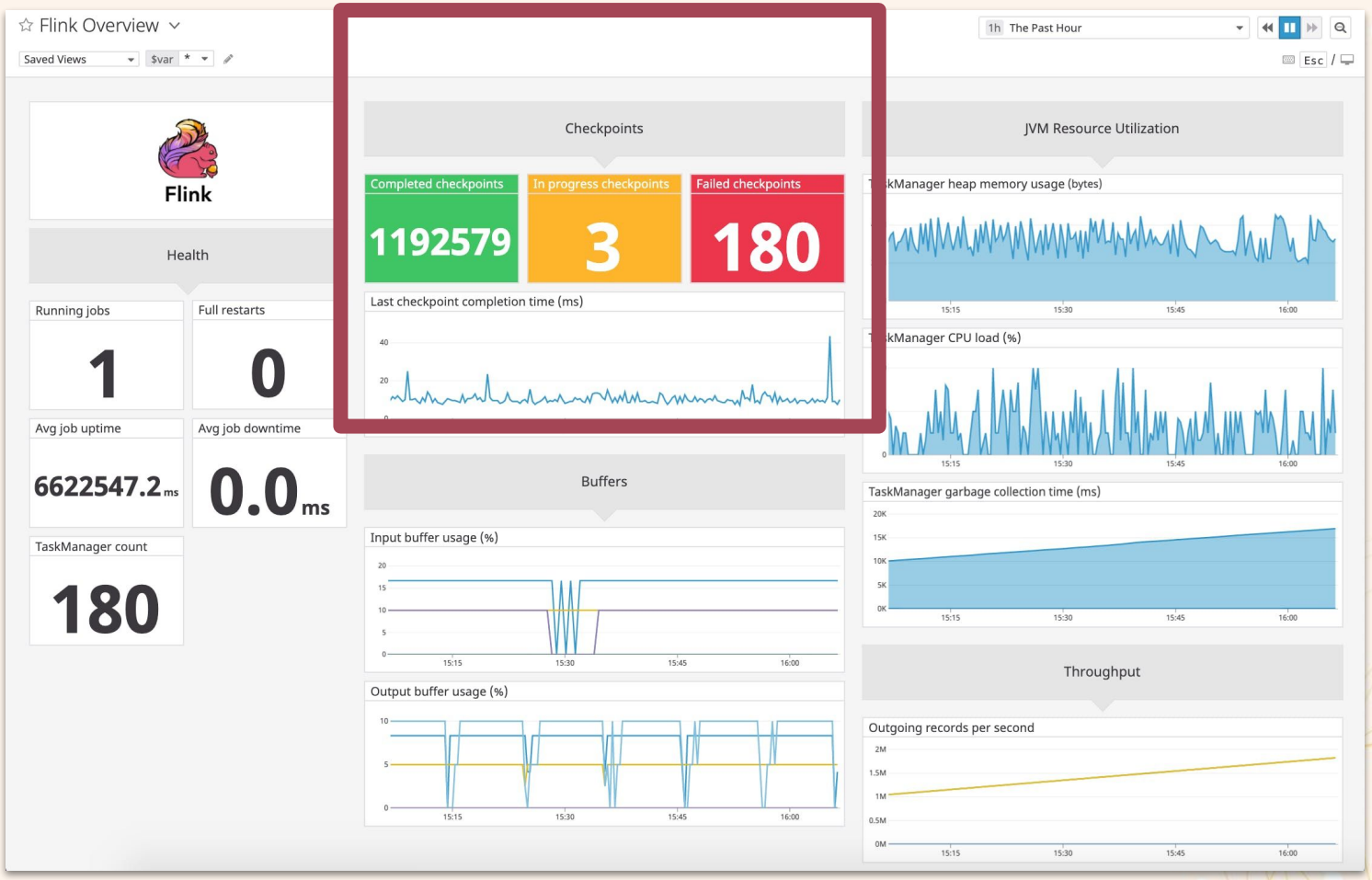
Metrics-Driven Metrics





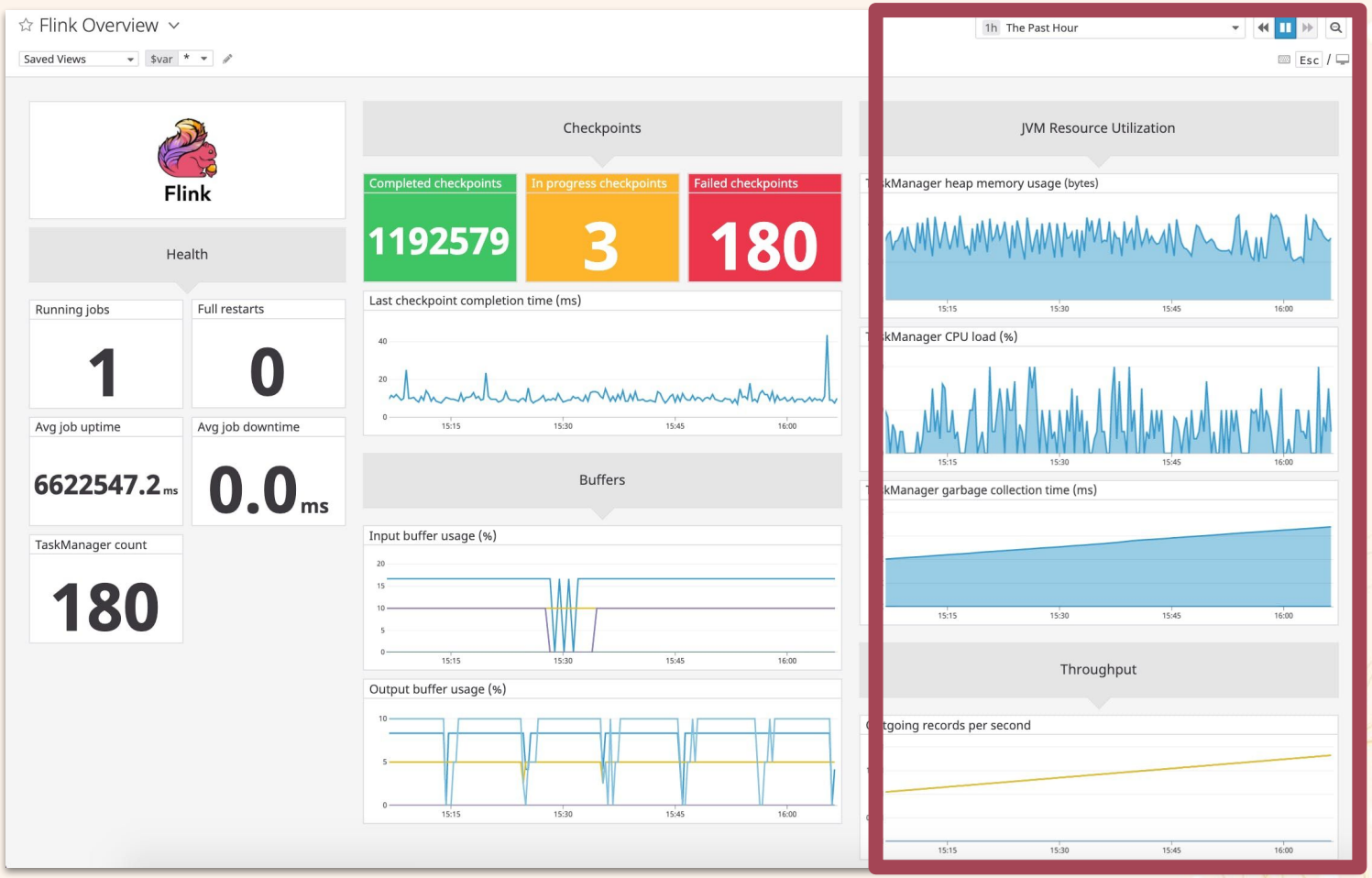
@CAITO_200_OK

Metrics-Driven Metrics



@CAITO_200_OK

Metrics-Driven Metrics



@CAITO_200_OK

Metrics-Driven Metrics

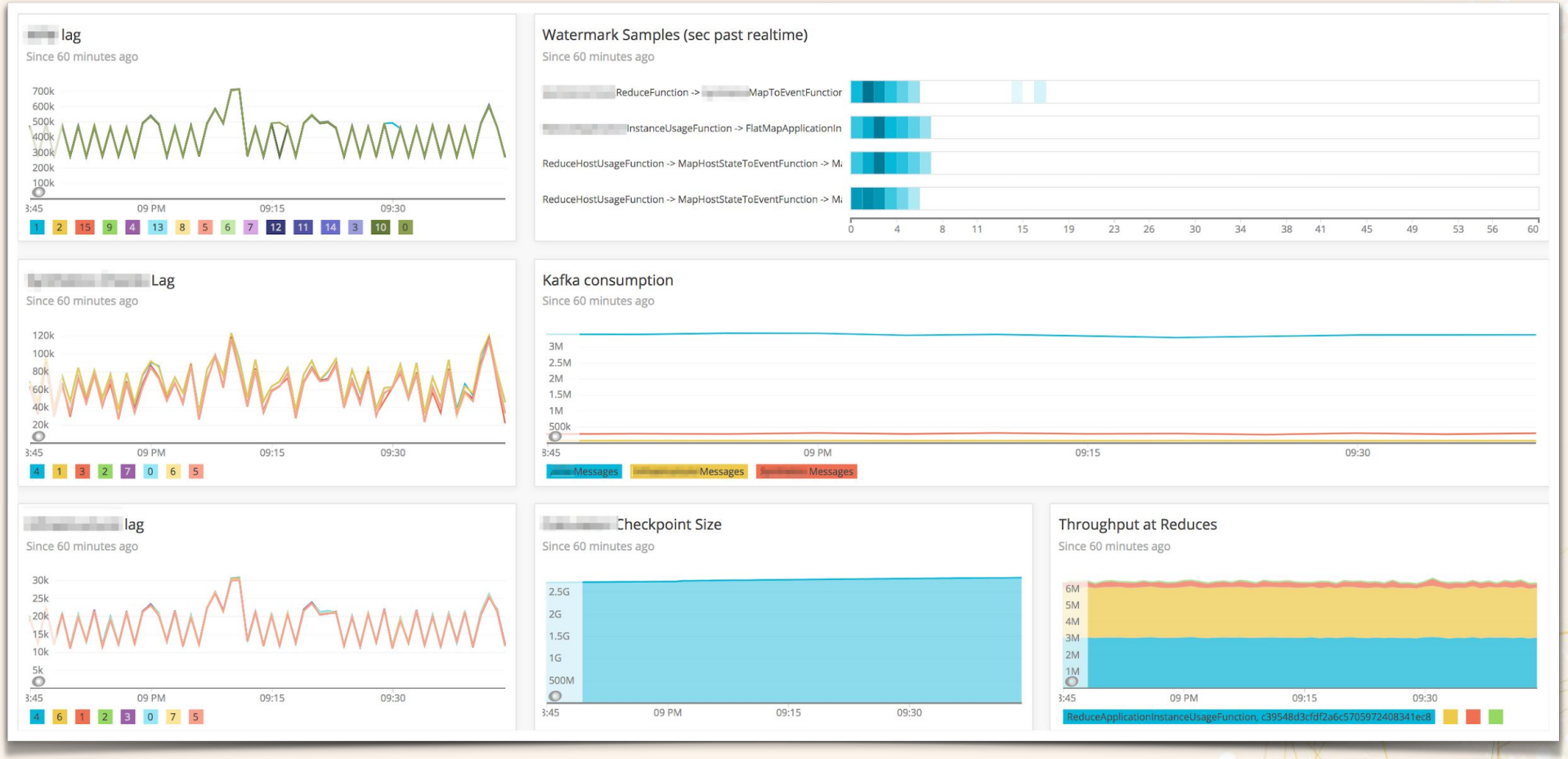
Prometheus

```
metrics.reporter.prom.class: org.apache.flink.metrics.prometheus.PrometheusReporter
```

DataDog

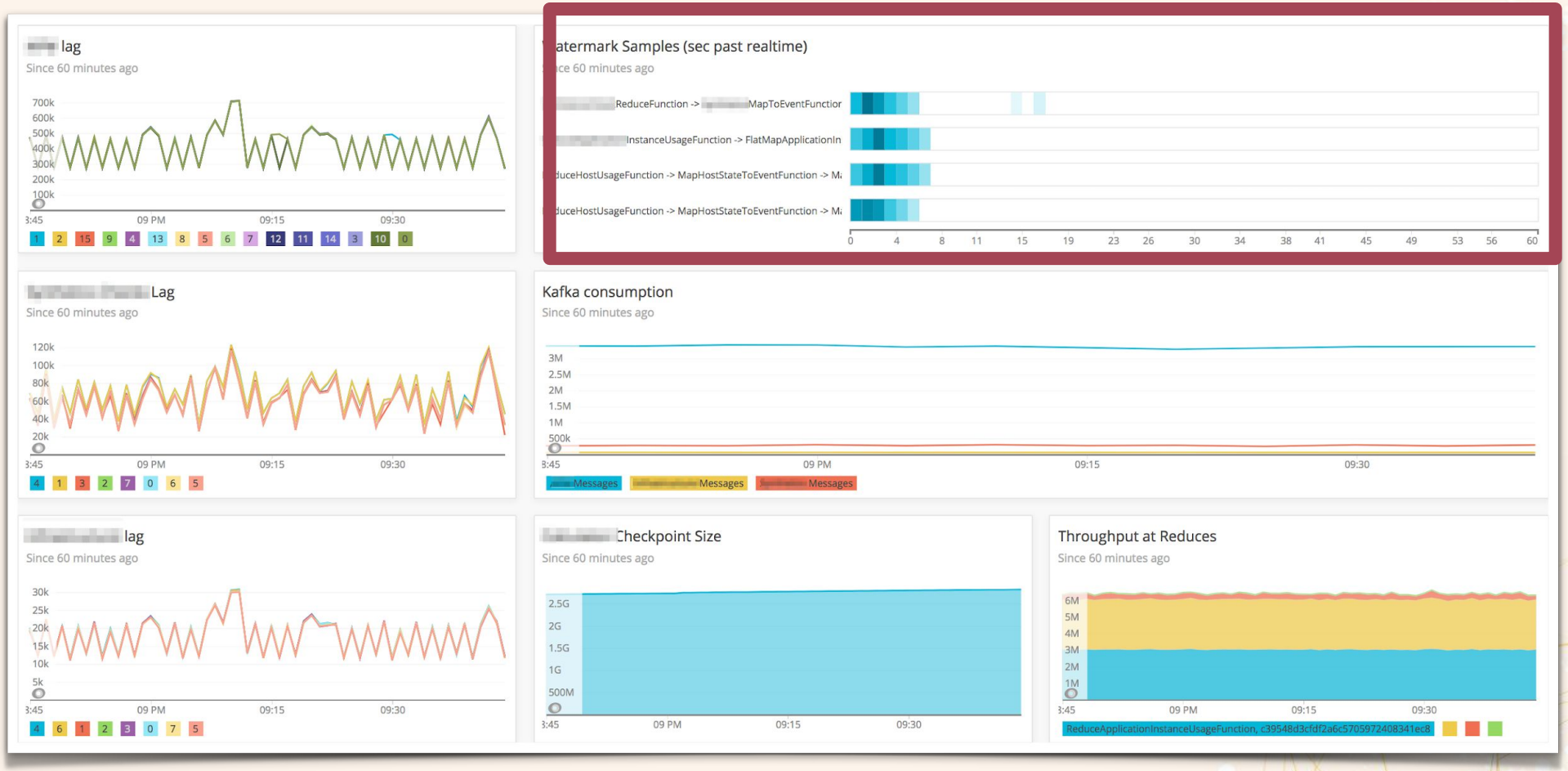
```
metrics.reporter.dghttp.factory.class: org.apache.flink.metrics.datadog.DatadogHttpRepo
rterFactory
metrics.reporter.dghttp.apikey: xxx
metrics.reporter.dghttp.tags: myflinkapp,prod
metrics.reporter.dghttp.proxyHost: my.web.proxy.com
metrics.reporter.dghttp.proxyPort: 8080
metrics.reporter.dghttp.dataCenter: US
metrics.reporter.dghttp.maxMetricsPerRequest: 2000
metrics.reporter.dghttp.interval: 60 SECONDS
```

Metrics-Driven Metrics



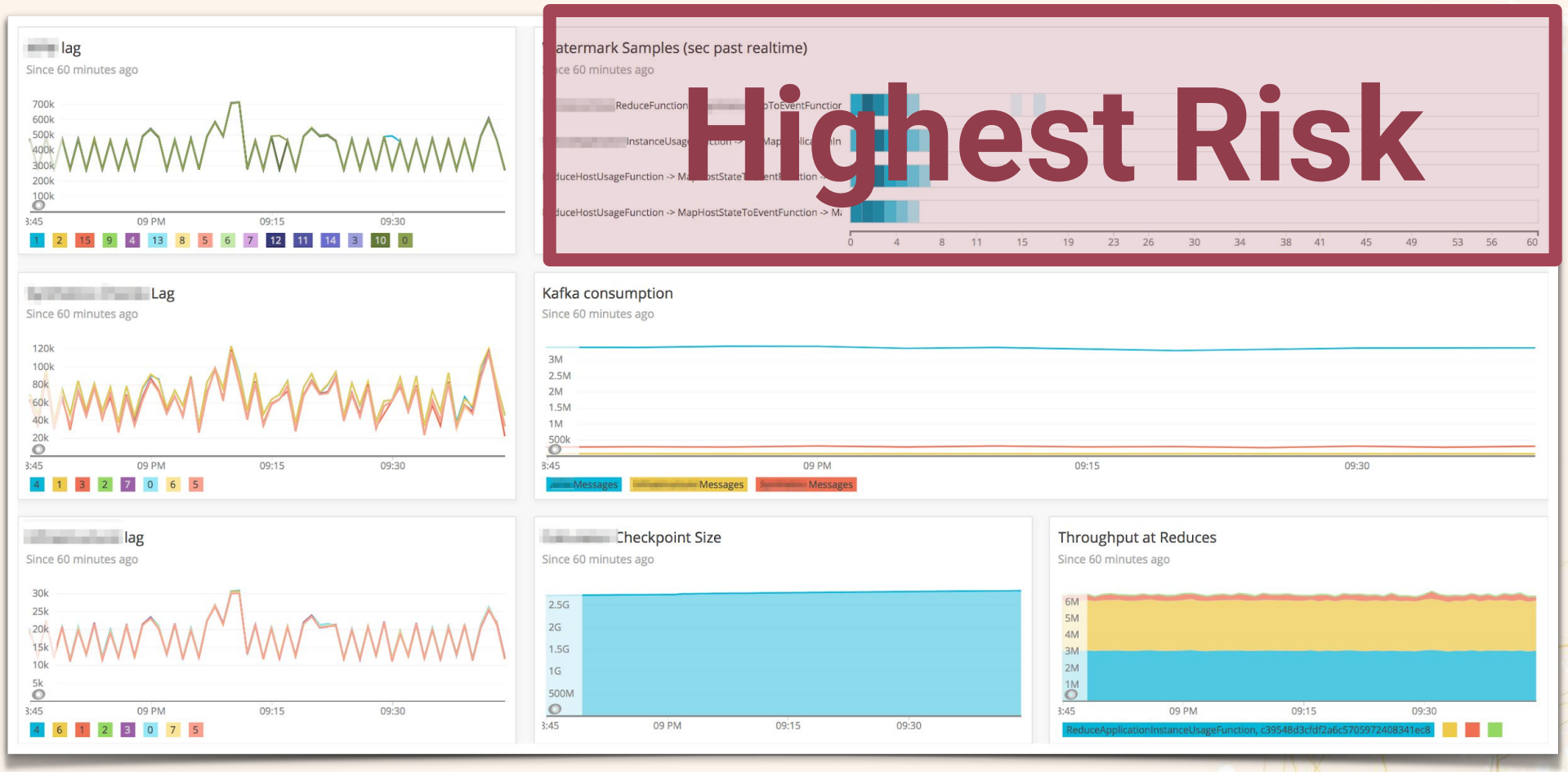
@CAITO_200_OK

Metrics-Driven Metrics



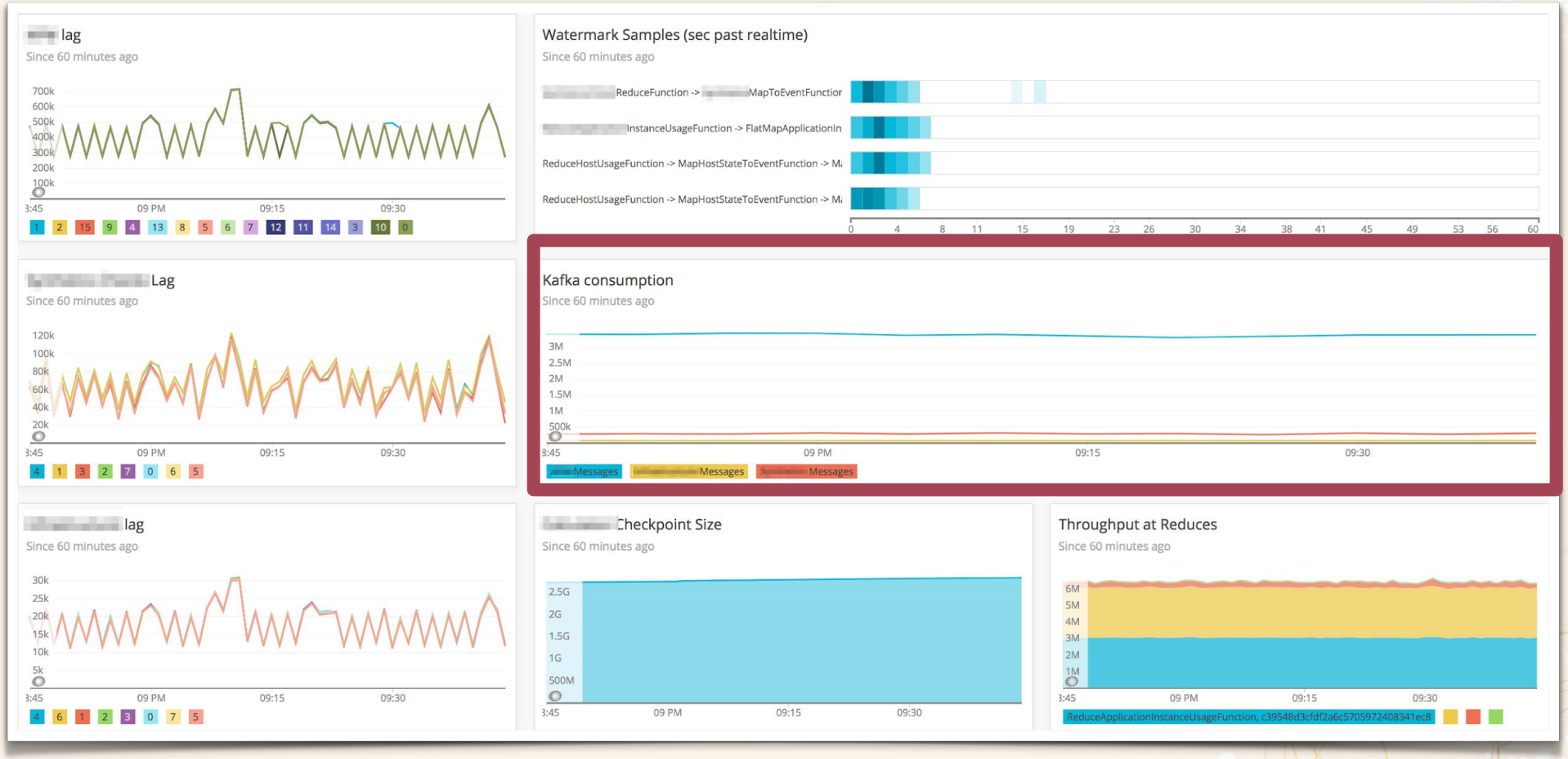
@CAITO_200_OK

Metrics-Driven Metrics



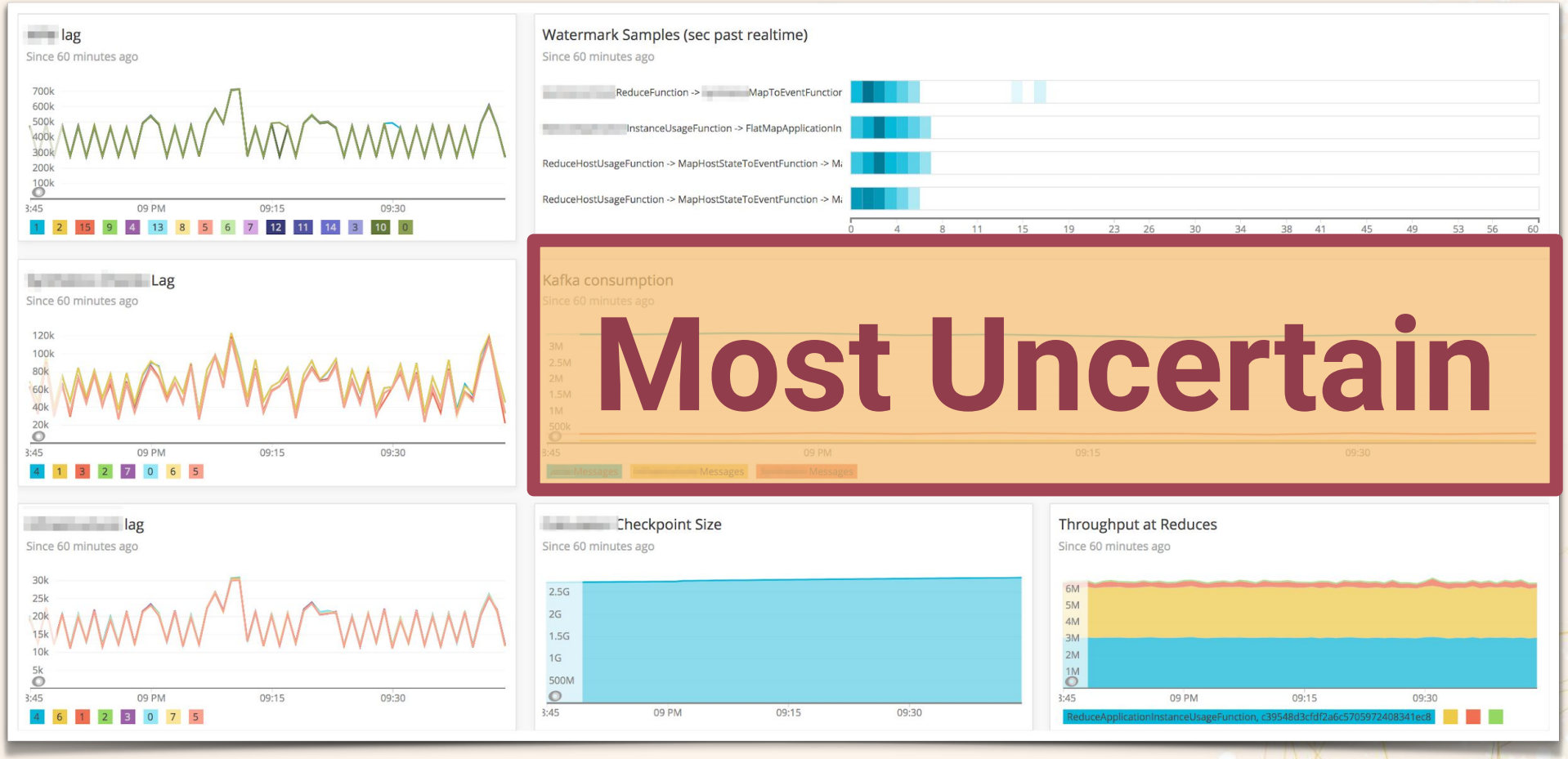
@CAITO_200_OK

Metrics-Driven Metrics



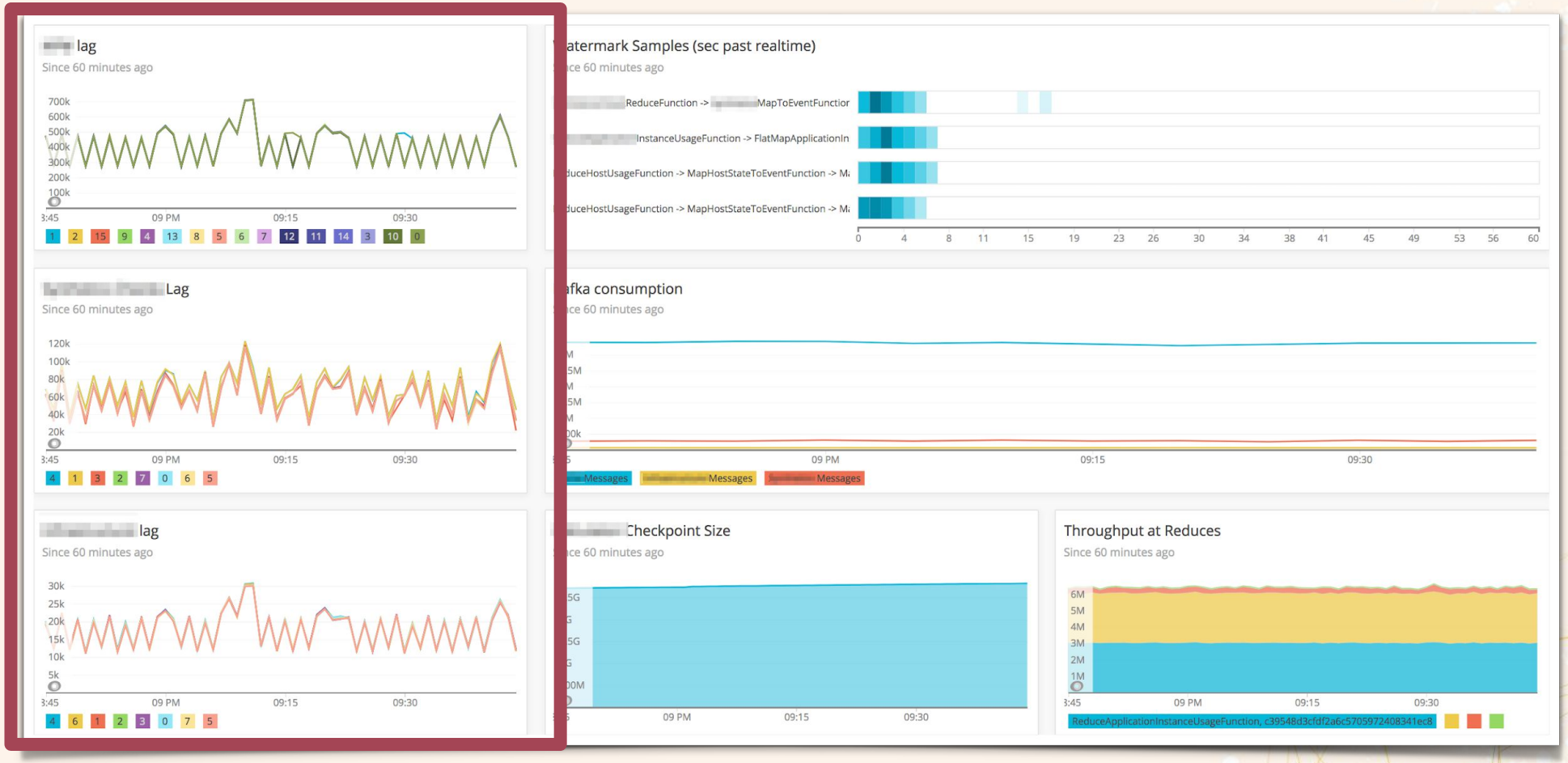
@CAITO_200_OK

Metrics-Driven Metrics



@CAITO_200_OK

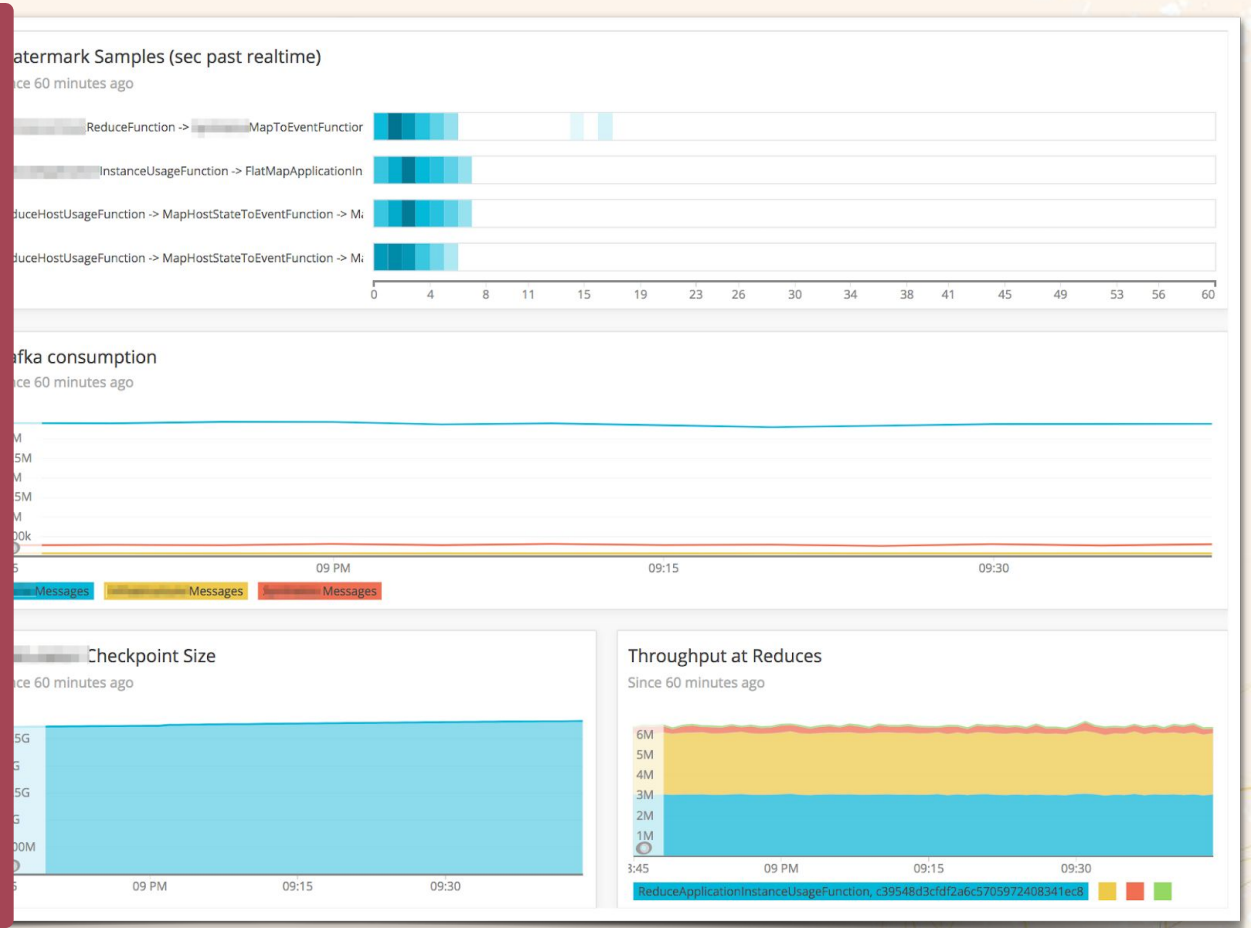
Metrics-Driven Metrics



@CAITO_200_OK

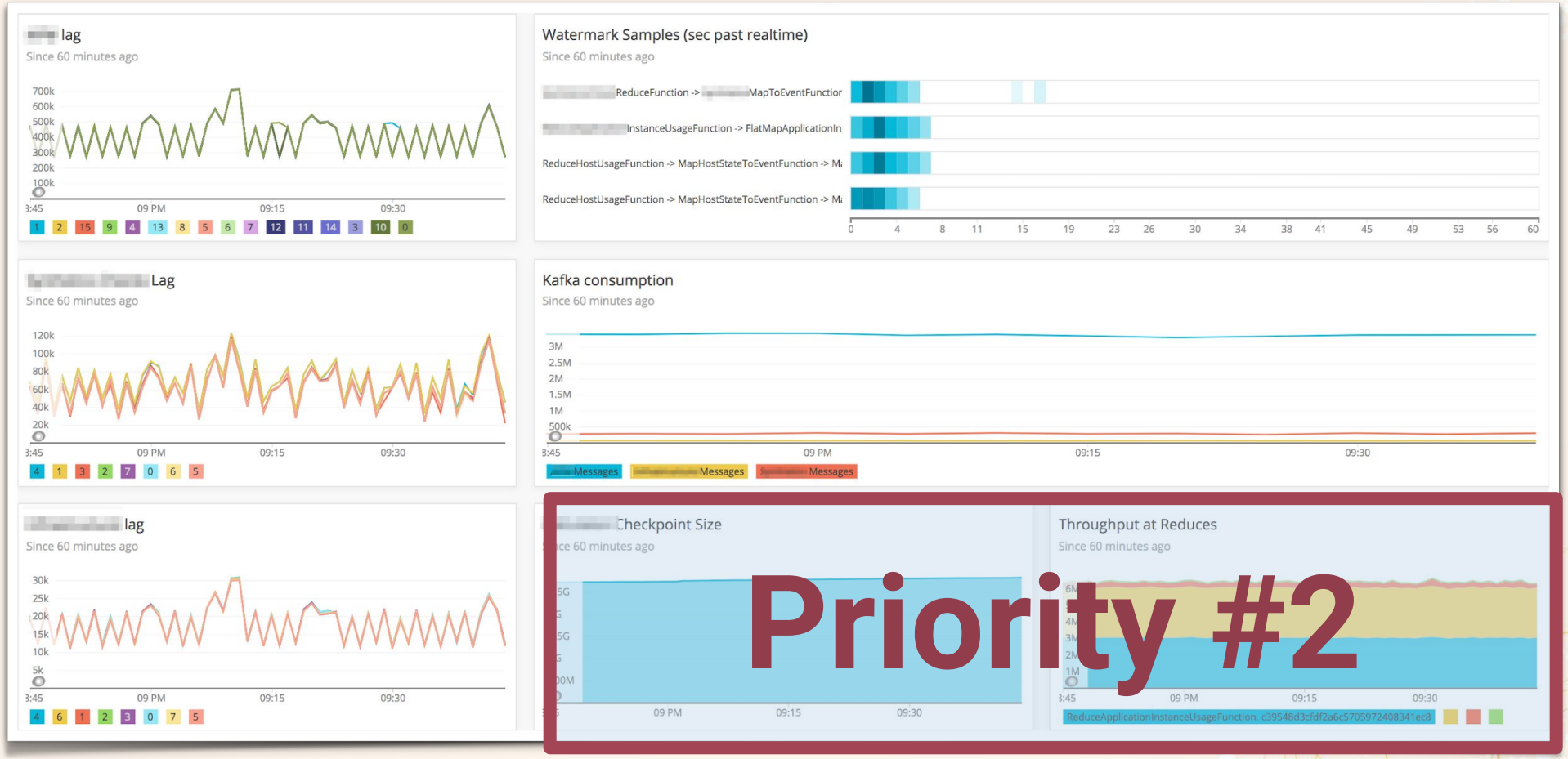
Metrics-Driven Metrics

Current Roadmap Priority



@CAITO_200_OK

Metrics-Driven Metrics



@CAITO_200_OK

Metrics-Driven Metrics

Current Roadmap Priority

Highest Risk

Most Uncertain

Priority #2

@CAITO_200_OK

Metrics as a Shared Language

@CAITO_200_OK

Shared Language

"Even the best metrics driven development fails easily, when it's only built for yourself"

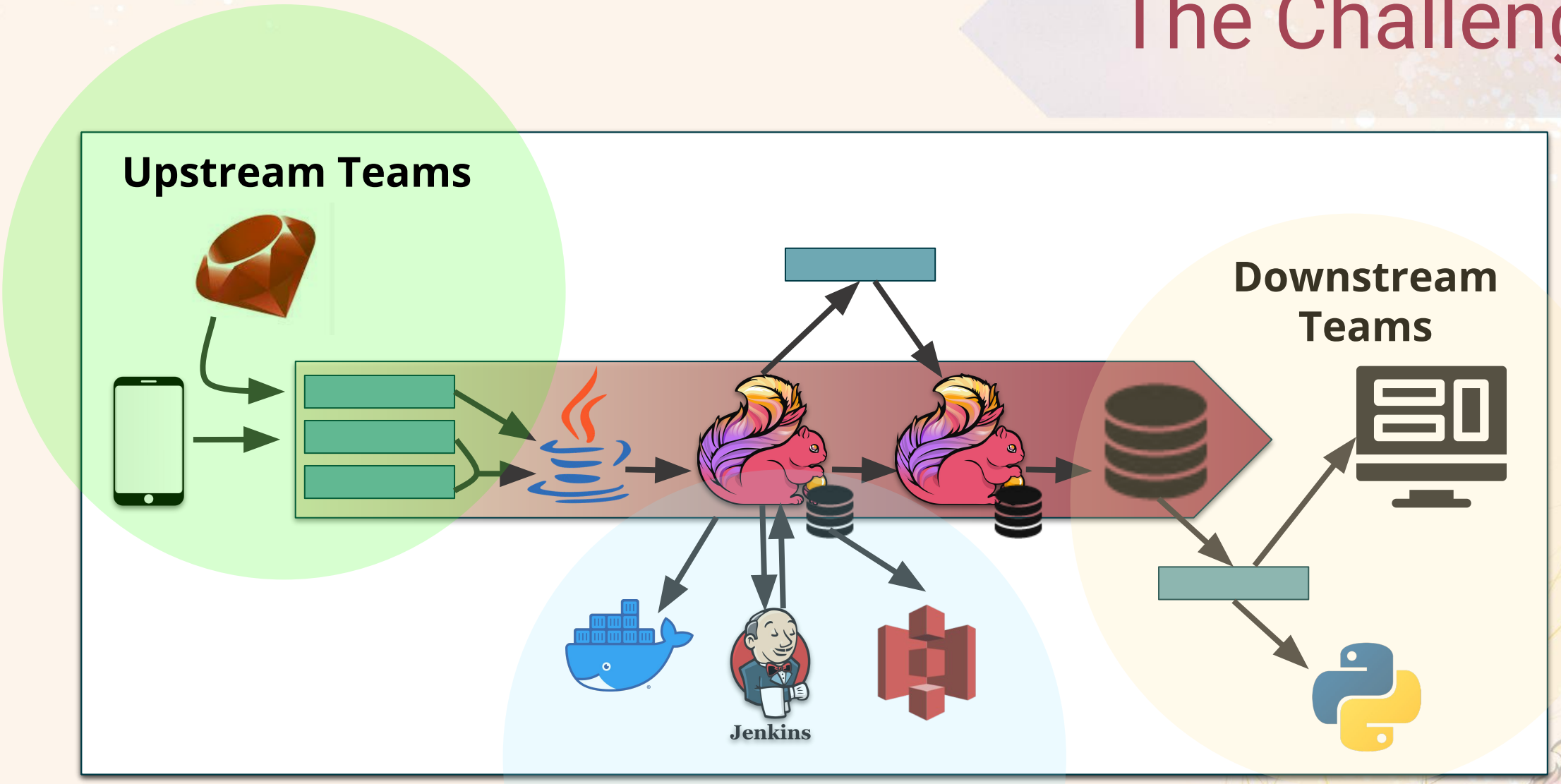
Shared Language

Implementation:

- Identify your impacted groups
- Identify the most effective tools for each group
- Enable automation

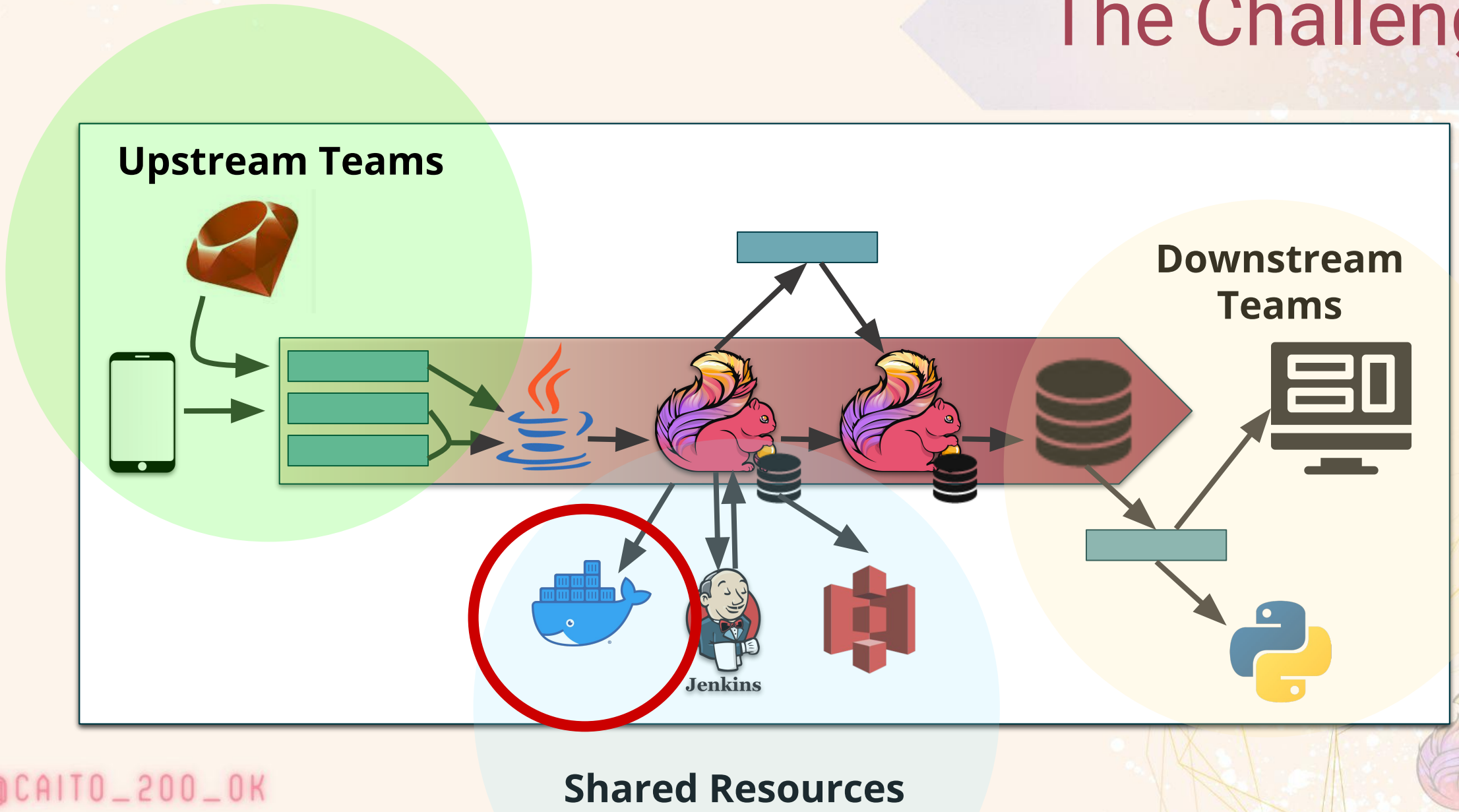
@CAITO_200_OK

The Challenge



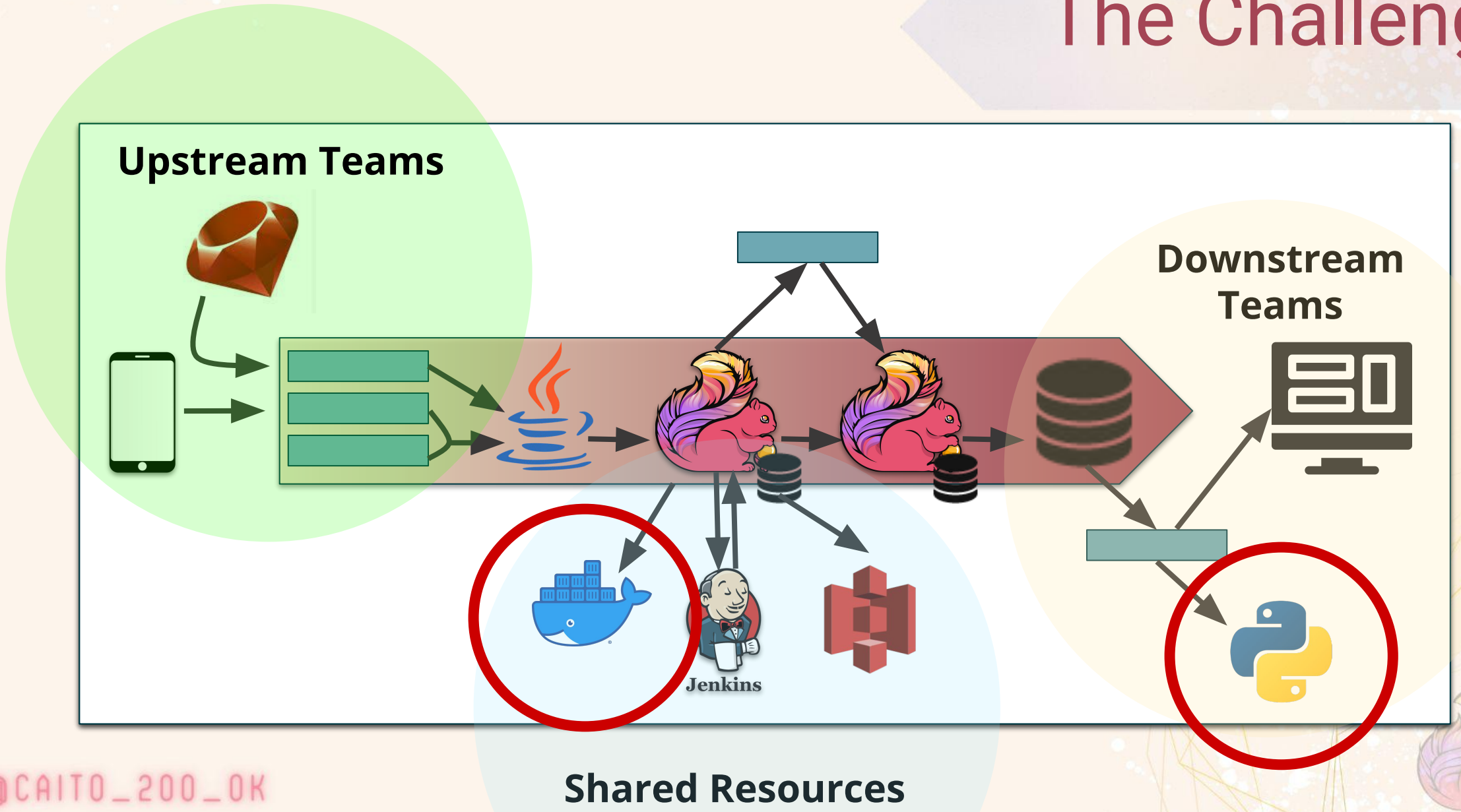
@CAITO_200_OK

The Challenge



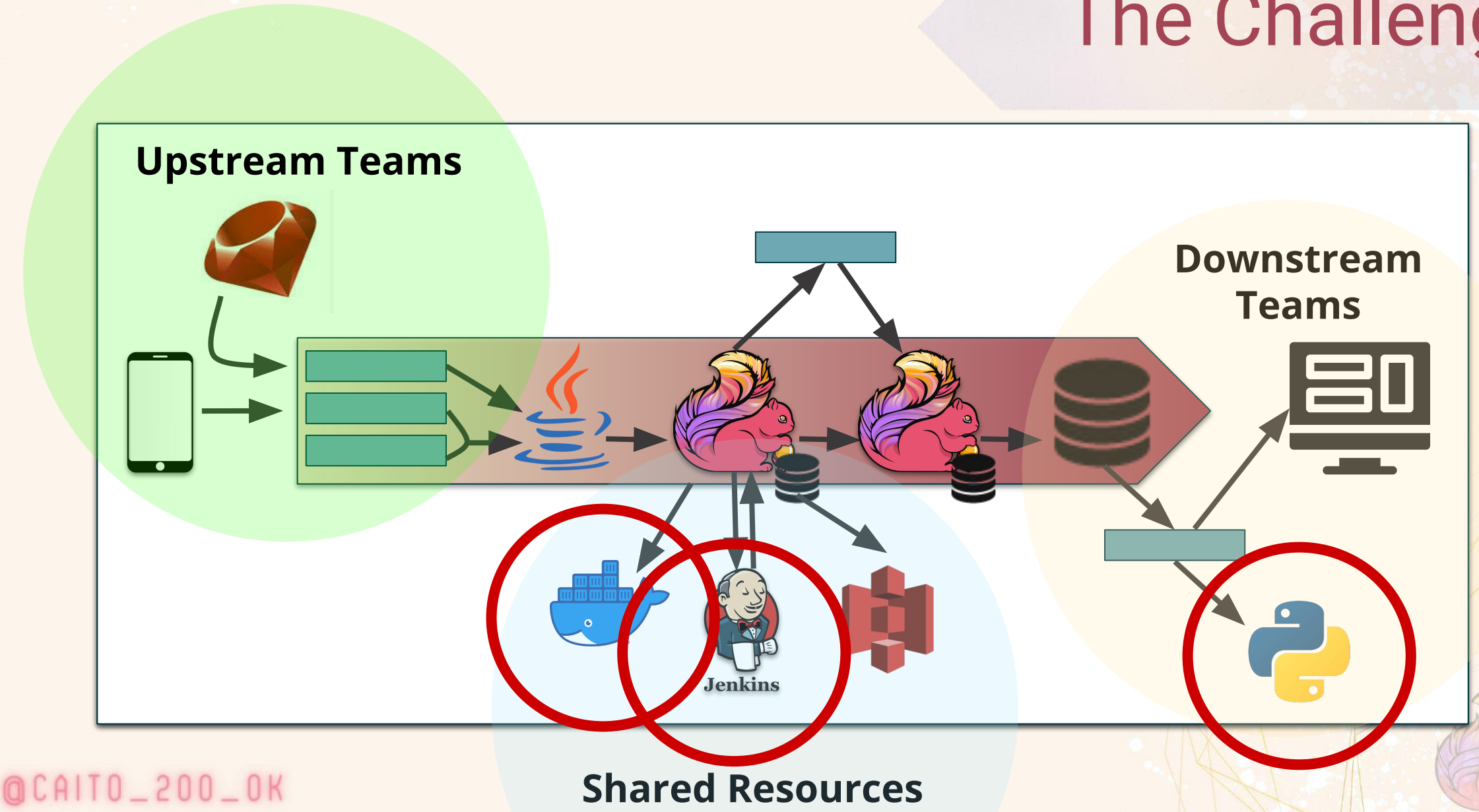
@CAITO_200_OK

The Challenge



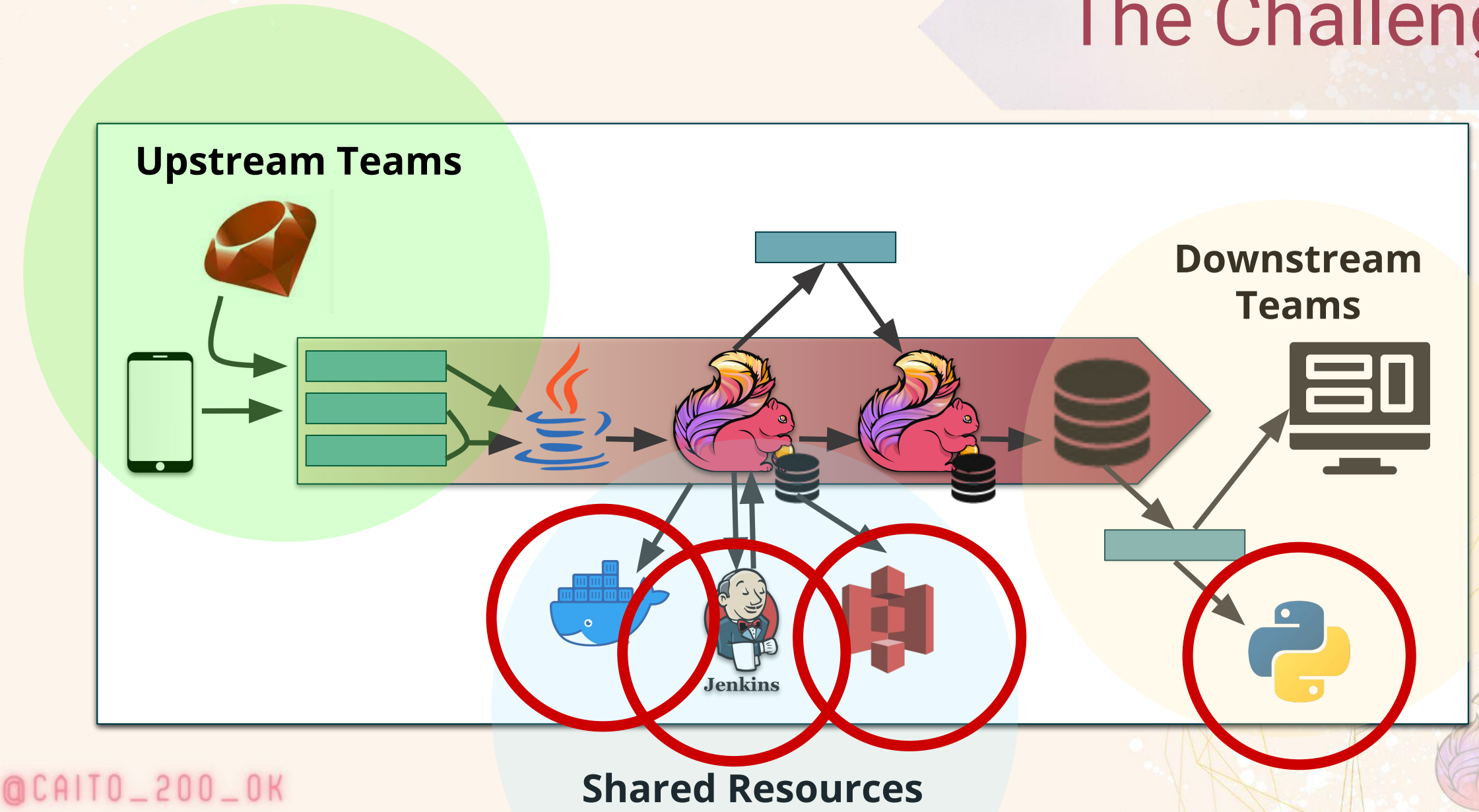
@CAITO_200_OK

The Challenge



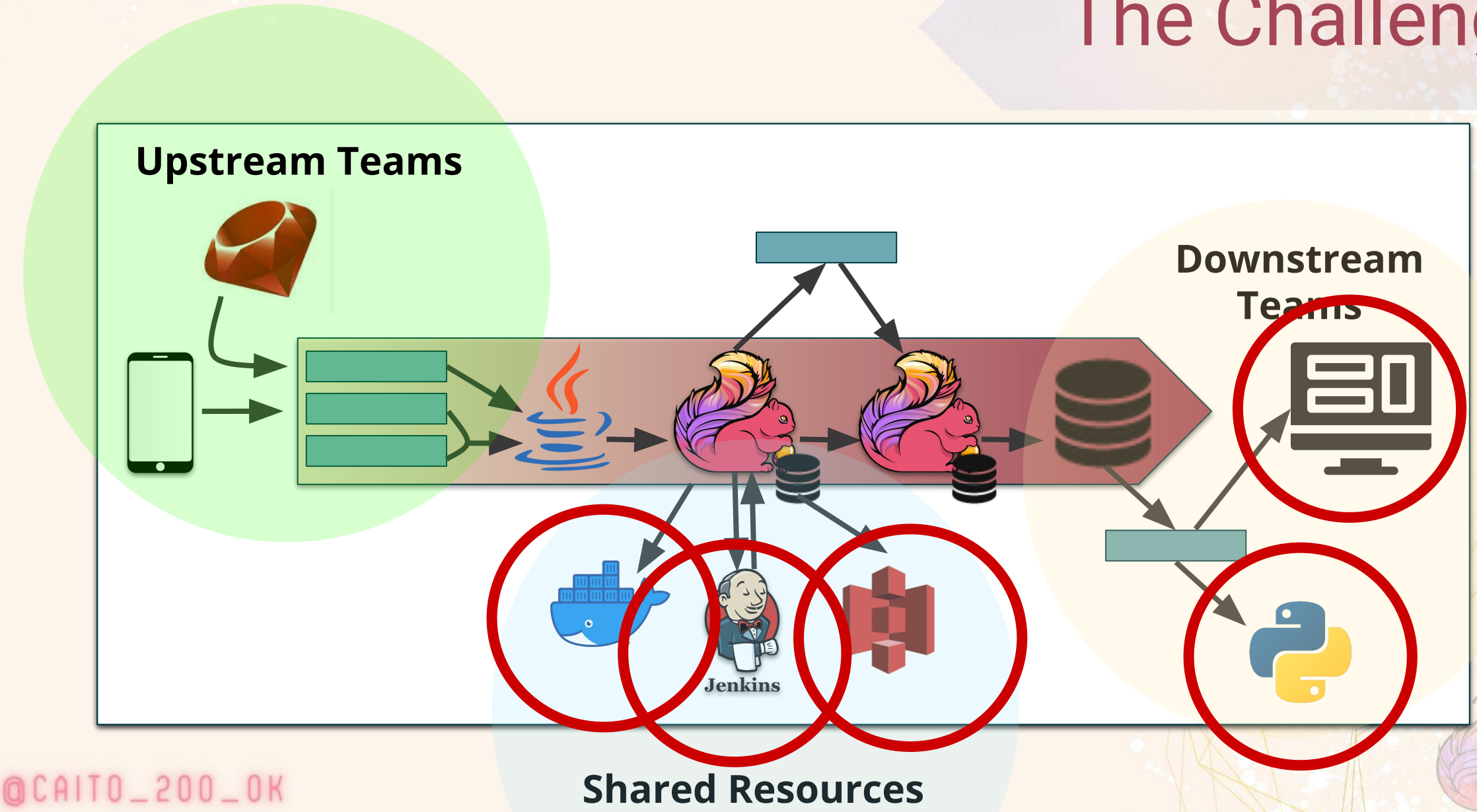
@CAITO_200_OK

The Challenge



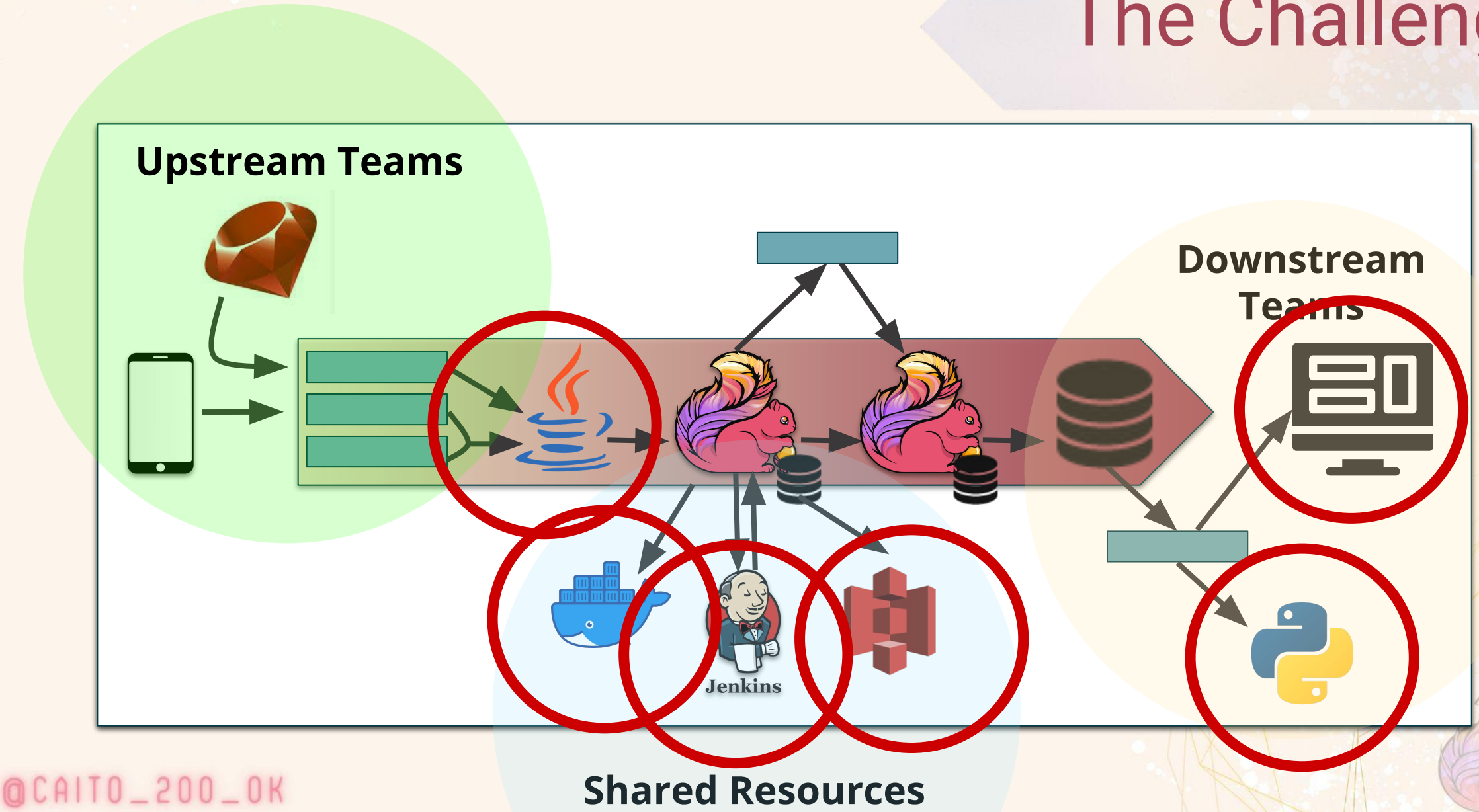
@CAITO_200_OK

The Challenge



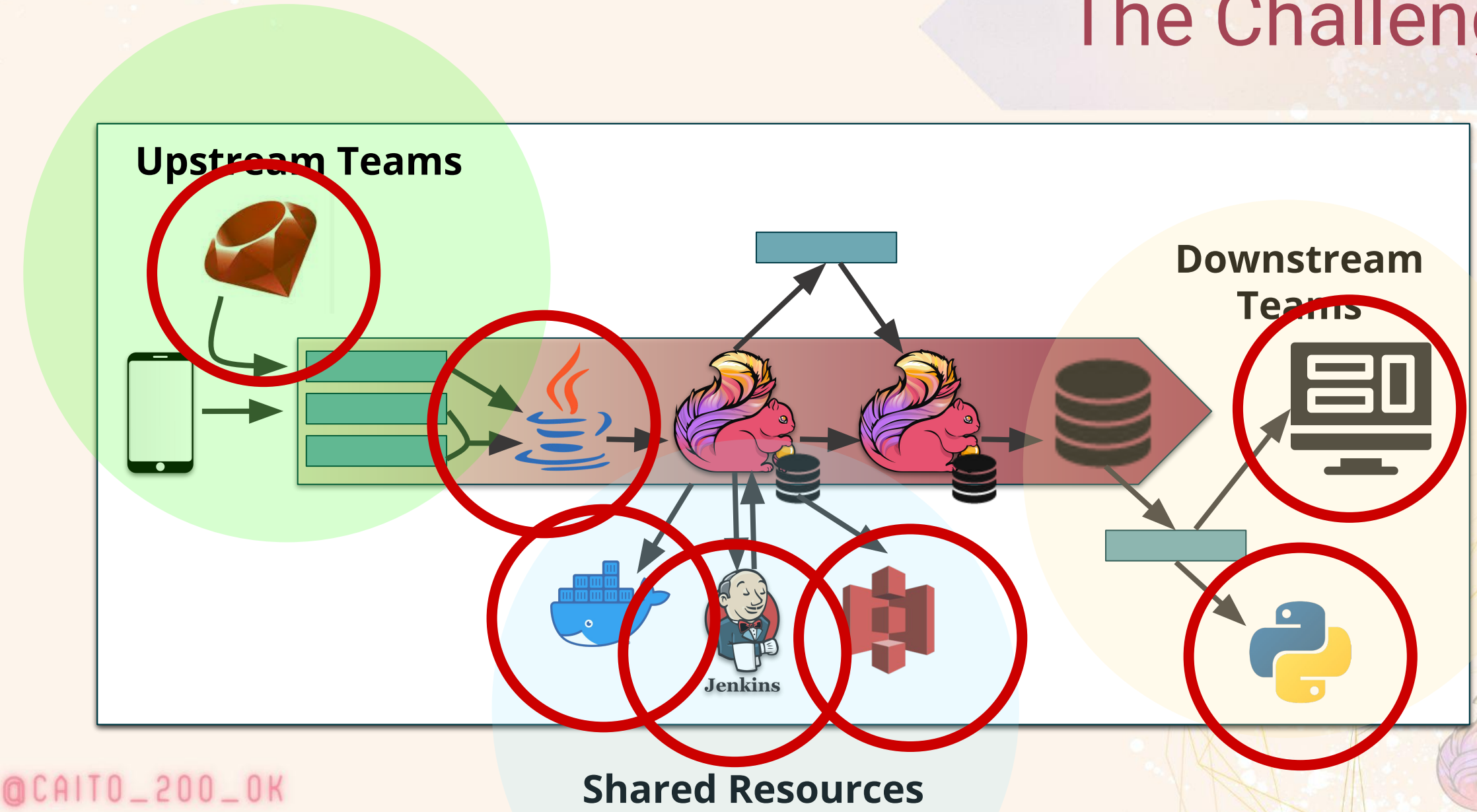
@CAITO_200_OK

The Challenge



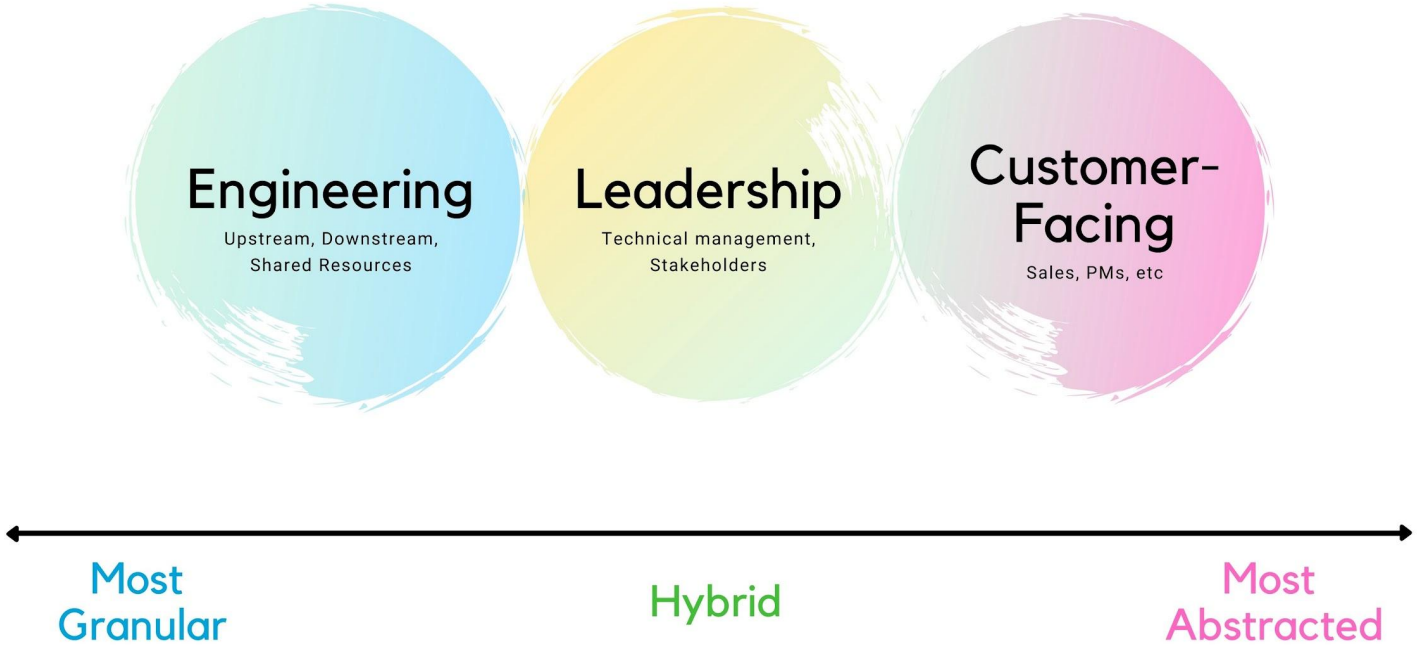
@CAITO_200_OK

The Challenge



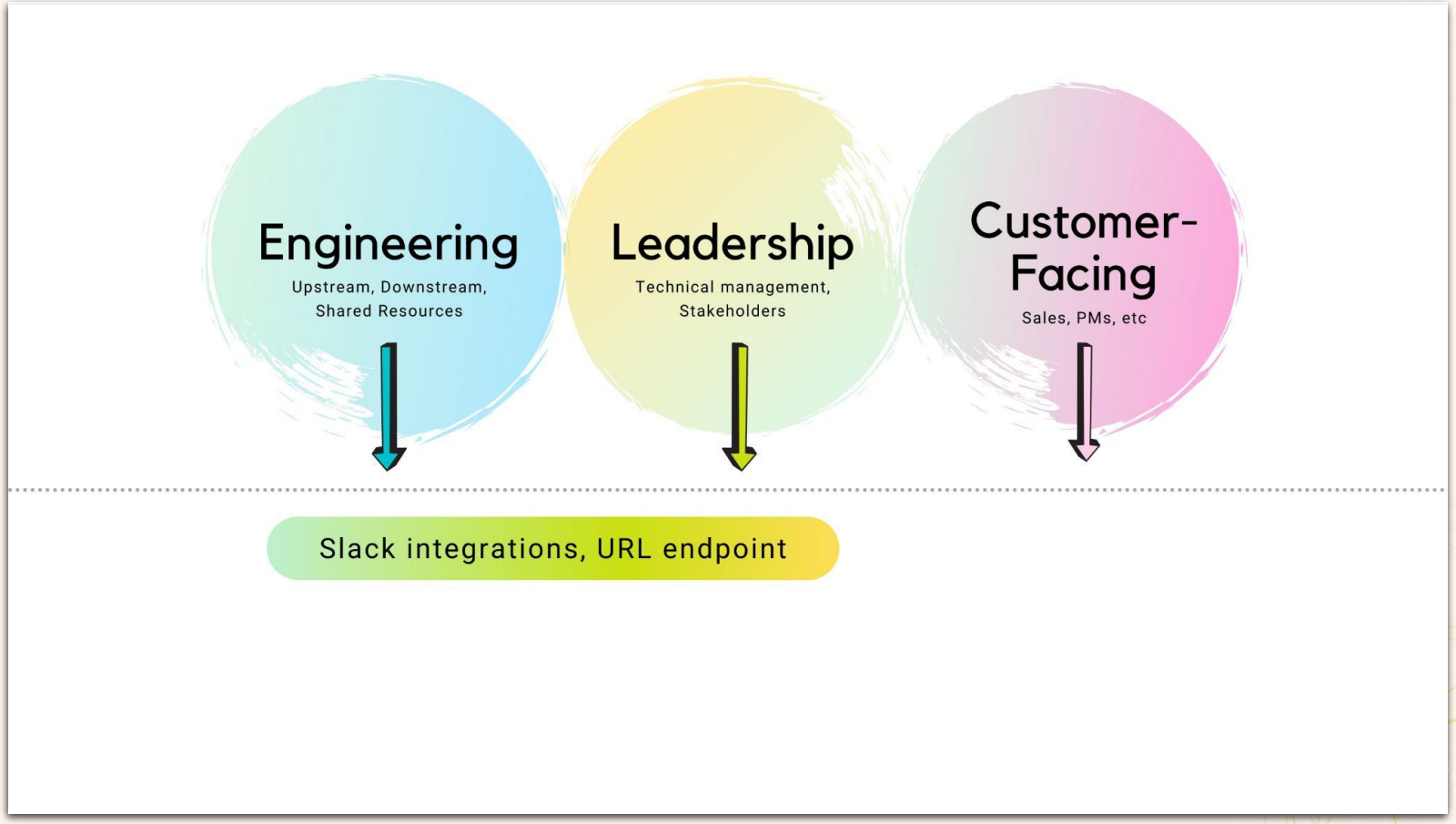
@CAITO_200_OK

Shared Language



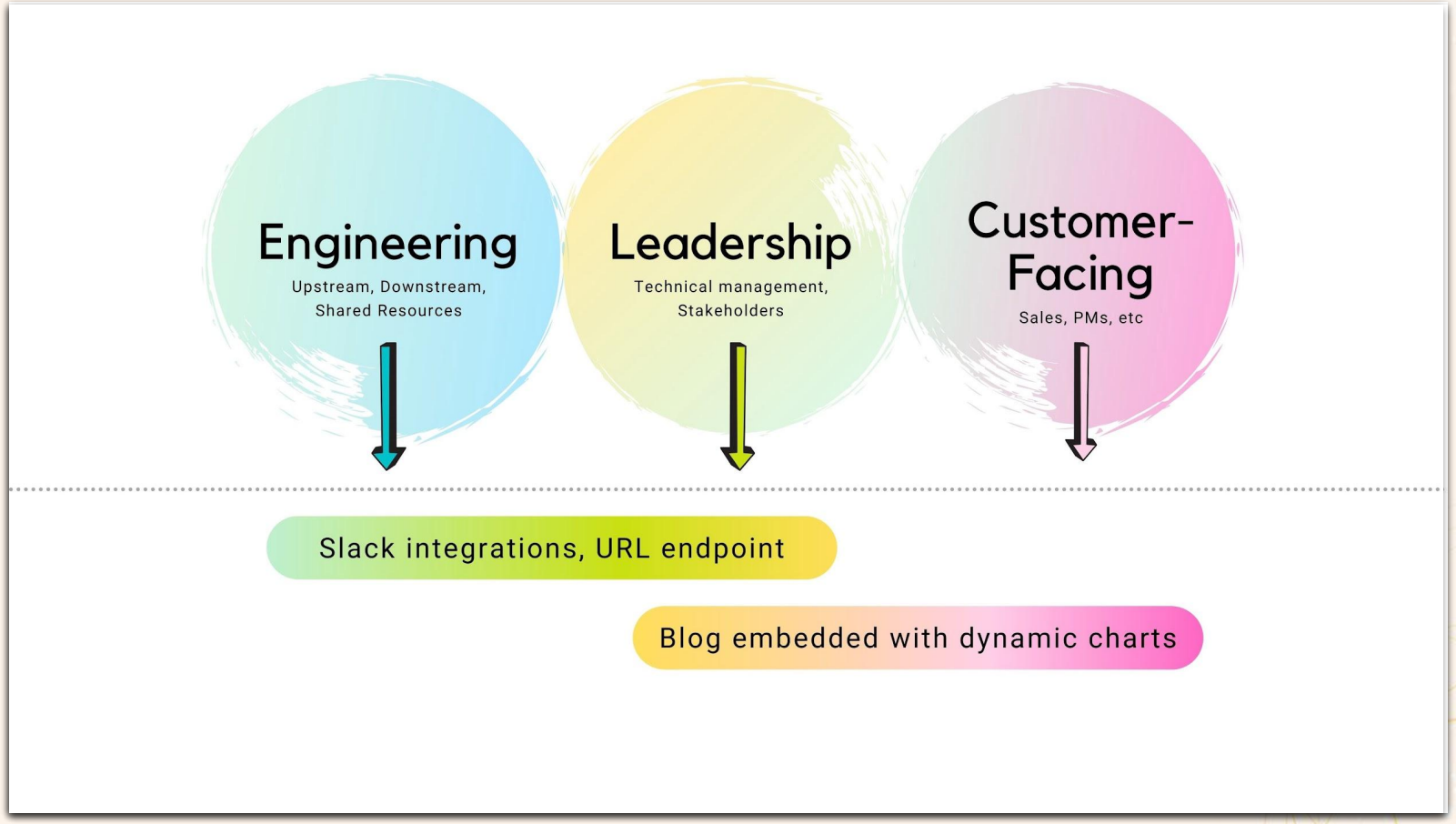
@CAITO_200_OK

Shared Language

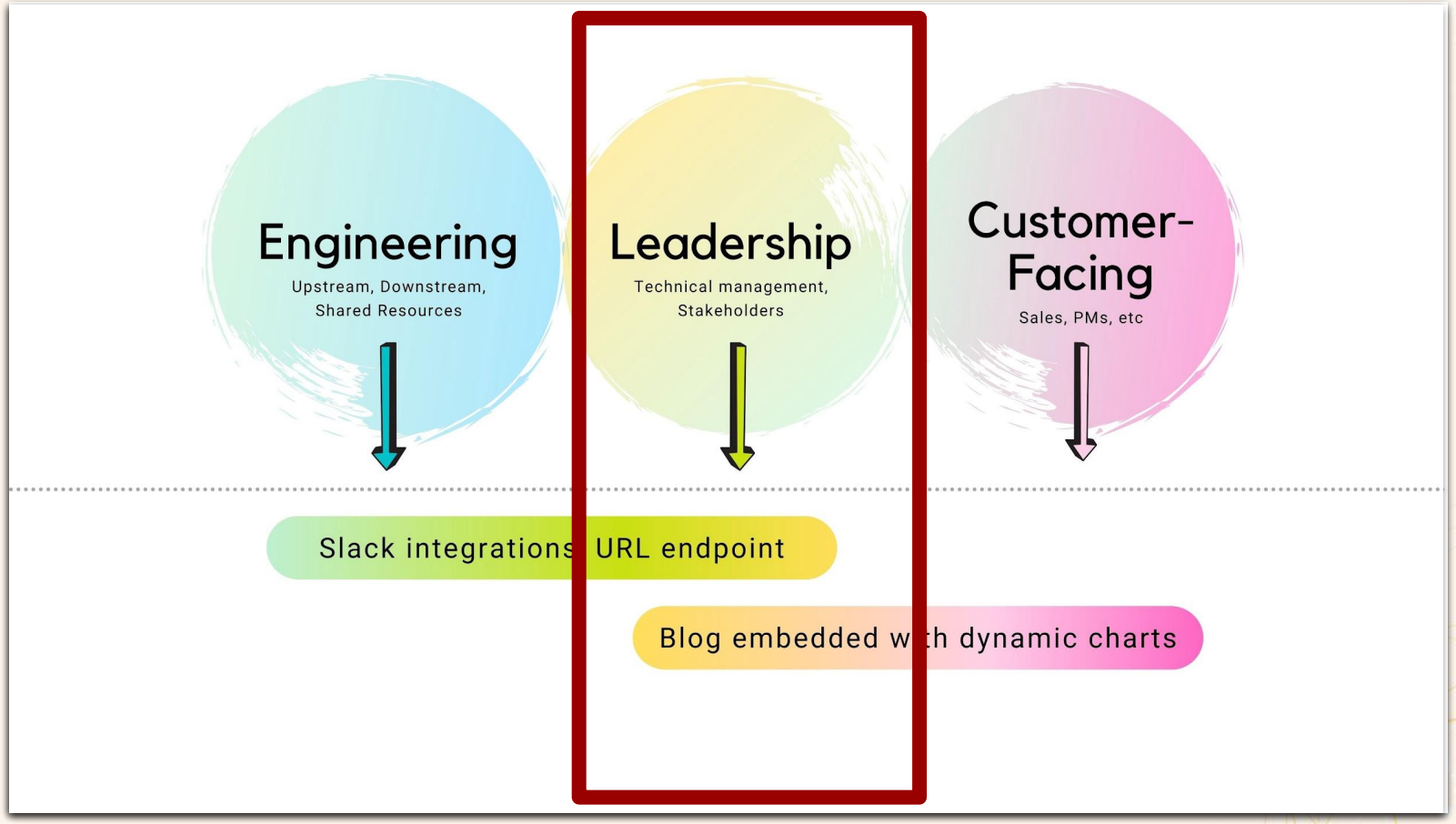


@CAITO_200_OK

Shared Language



Shared Language



@CAITO_200_OK

Conclusion

@CAITO_200_OK

Conclusion

For systems with complex integration points:

- **Metrics-Driven Metrics**: streamline technical complexity
- **Metrics as a Shared Language**: streamline interpersonal complications

Contact

- @Caito_200_OK
- caito@ververica.com
- http://caito-200-ok.com

Upcoming

- Flink Forward 2022
  - August 3, 2022
  - San Francisco

@CAITO_200_OK

Thank You!

- **O'Reilly Conference staff!!**
- "Shared Language" photo models:
  - **Zack Hobson**, CTO
  - **Cory Johannsen**, Senior Software Engineer

Quoted Contributors & Reviewers:

- **Mandy Riso** - DevOps Engineer
- **Eric Shamow** - Engineering Manager
- **Mike Hix** - Senior Software Engineer
- **Ben Ford** - Product Manager
- **Noçnica Fee** - Developer Advocate
- **Lucy Wyman** - Senior Software Engineer
- **Logan Ballard** - Site Reliability Engineer

Resources

- Prometheus push gateway - when to use it/documentation: https://prometheus.io/docs/practices/pushing/
- Prometheus push gateway on GitHub: https://github.com/prometheus/pushgateway
- Prometheus Flink example: https://github.com/mbode/flink-prometheus-example
- Flink monitoring: https://flink.apache.org/news/2019/02/25/monitoring-best-practices.html

Credits - Images

- Flink logos:
  https://wints.github.io/flink-web//community.html
- Data-Driven/Aware/Informed Design:
  https://uxdesign.cc/becoming-a-data-aware-designer-1d7614ebc3ed
- Stream processing diagram:
  https://www.ververica.com/what-is-stream-processing
- 747 Airplane cockpit:
  https://www.reddit.com/r/pics/comments/5vv8qt/the_pilots_seat_and_cockpit_of_a_boeing_747/
- DataDog Flink dashboard:
  https://www.datadoghq.com/blog/monitor-apache-flink-with-datadog/
- All other photos & Images:
  Caito Scherr

Credits

- Basic Data-Driven Development principles https://www.portable.com.au/reports/principles-of-data-driven-design
- Data-driven design: https://www.springboard.com/blog/data-driven-design/
- Becoming A Data-Aware Designer, + definition image - Illustration from “Designing with Data” by King, Churchill, & Tan https://uxdesign.cc/becoming-a-data-aware-designer-1d7614ebc3ed
- https://digitalprinciples.org/wp-content/uploads/PDD_Principle-BeDataDriven_v2.pdf
- Harvard Business Review - Data Driven Culture: https://hbr.org/2020/02/10-steps-to-creating-a-data-driven-culture