

# 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

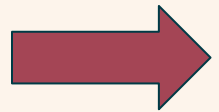
Stream  
Processing

---  
Since 2017

@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



# Before We Start...

@CAITO\_200\_OK



# Before We Start...



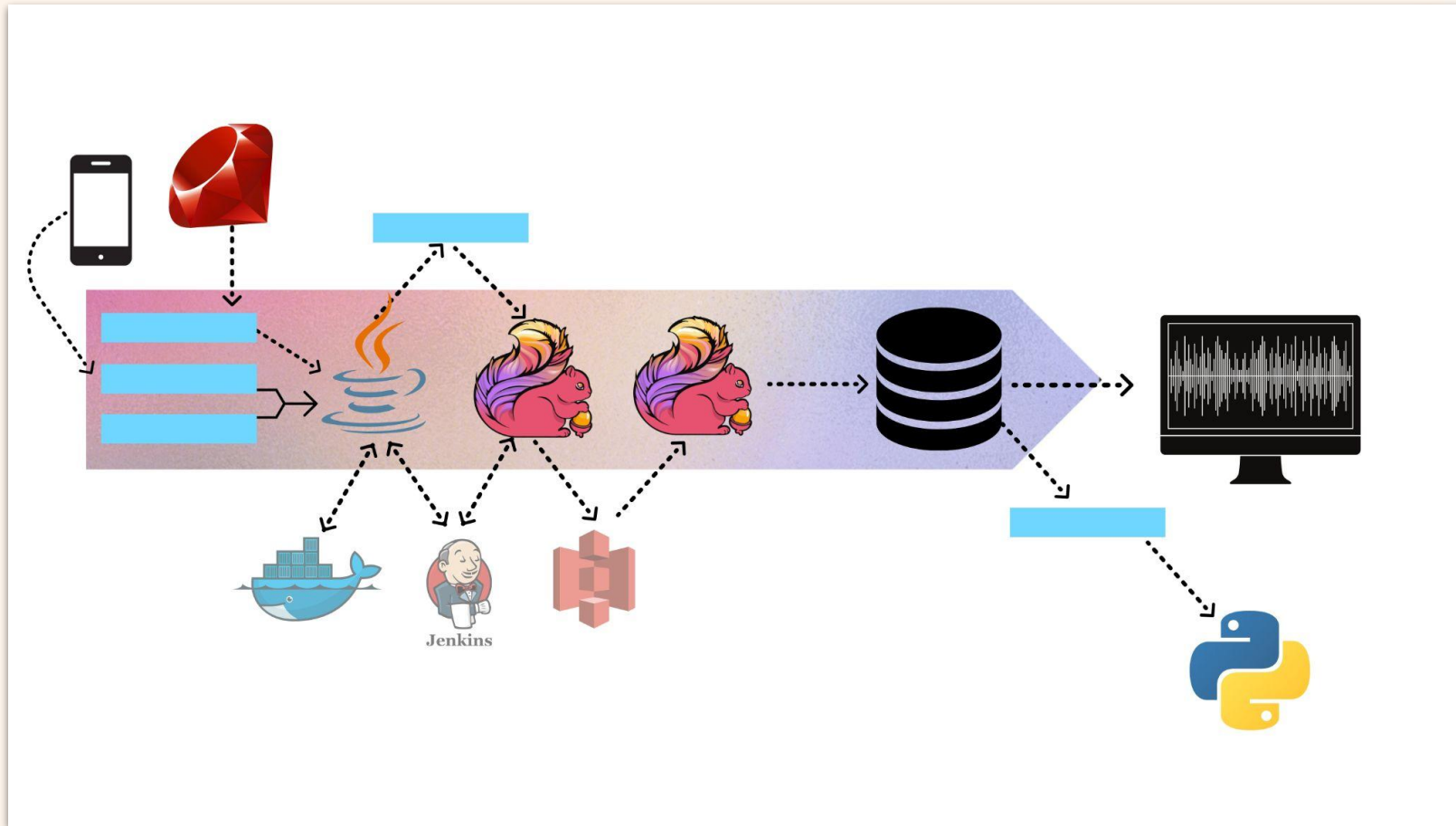
# The Challenge

@CAITO\_200\_OK

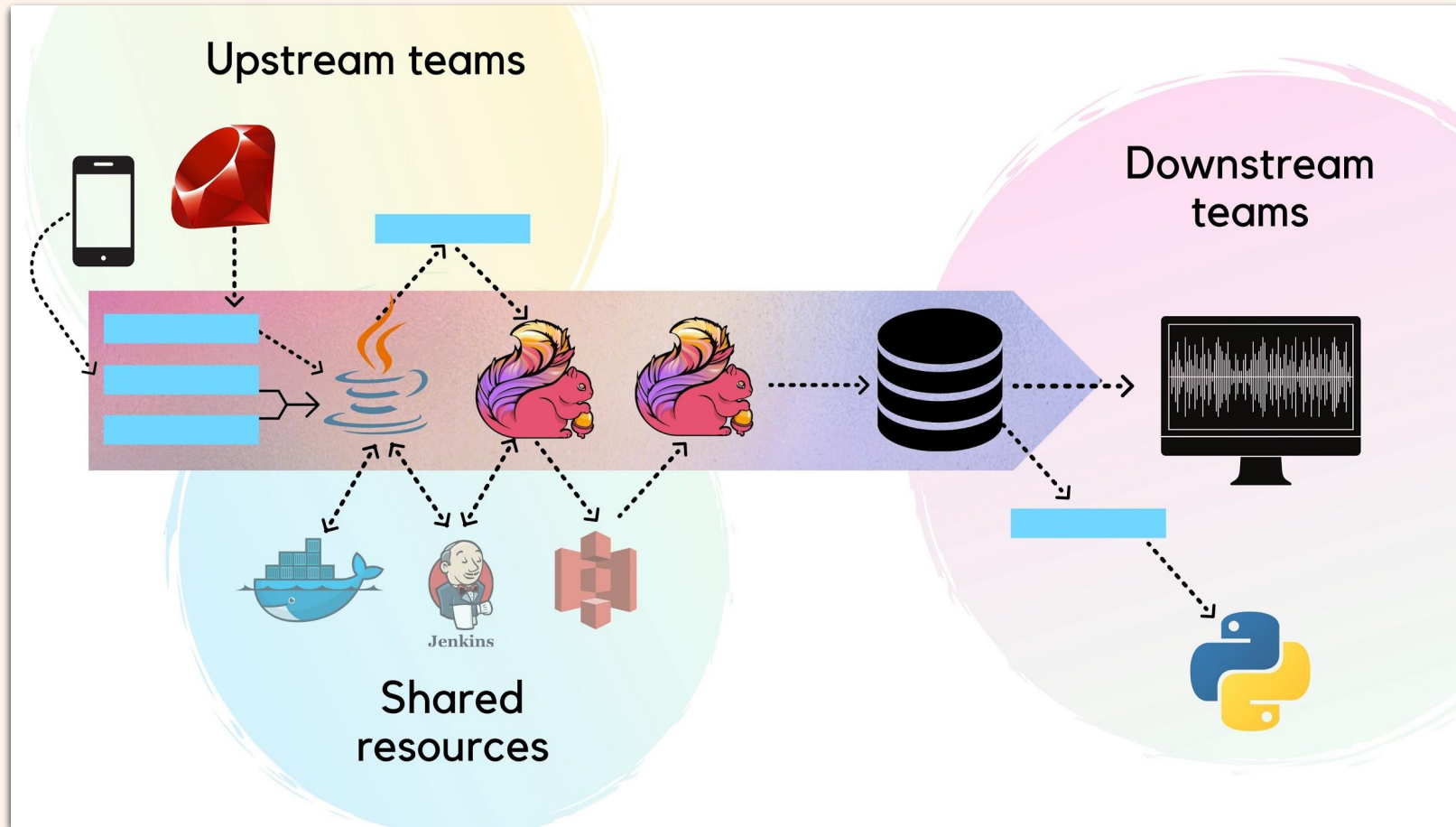




# The Challenge



# The Challenge



# "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



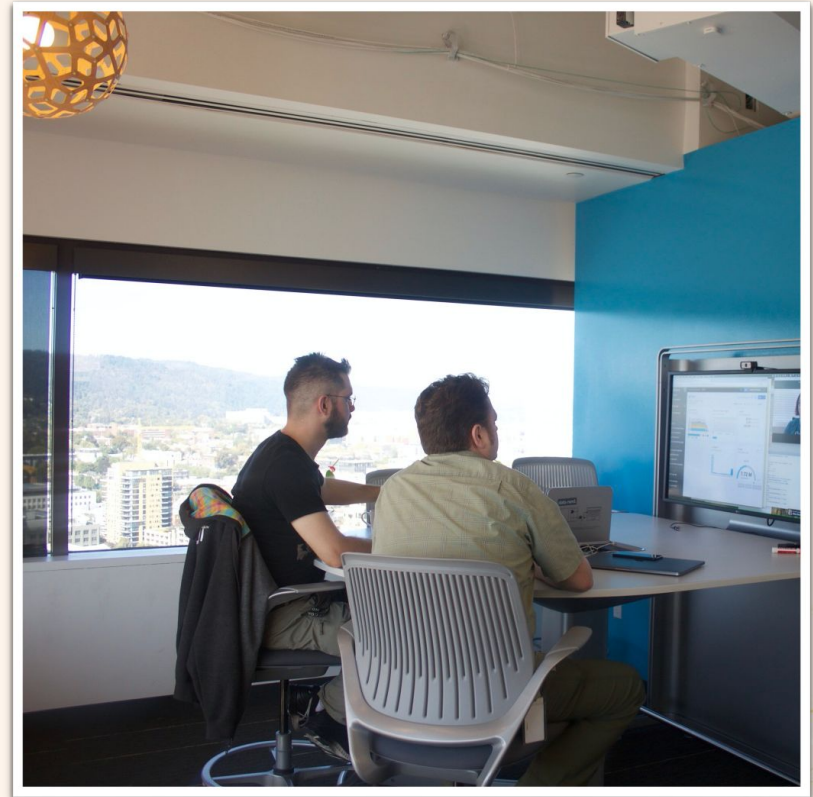
@CAITO\_200\_OK



# Metrics-Driven Metrics

The dashboard should **be**:

- Meaningful
- Iterative
- Accessible

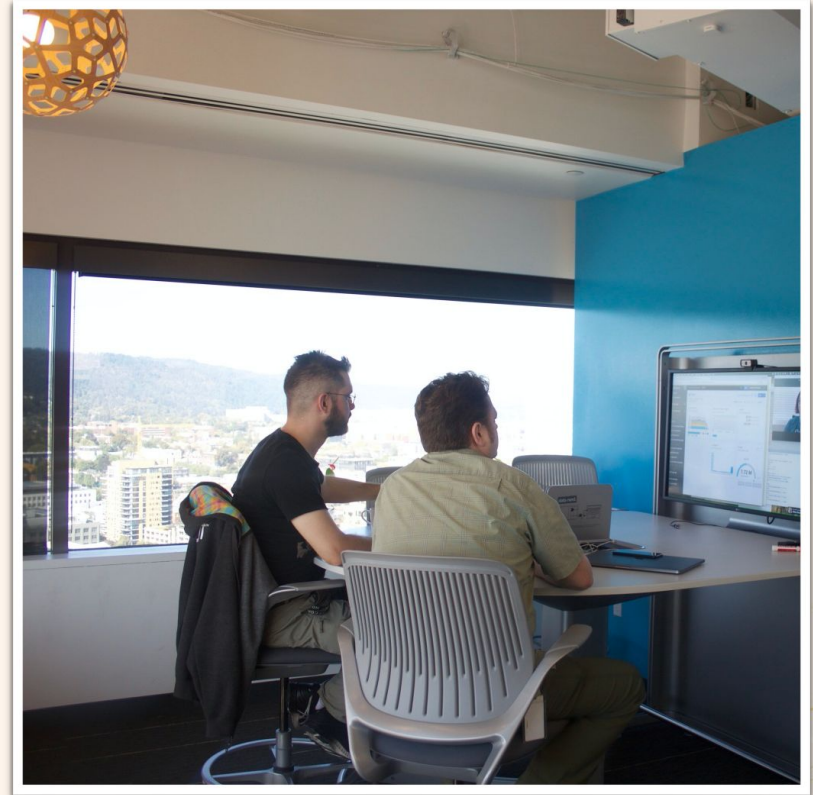




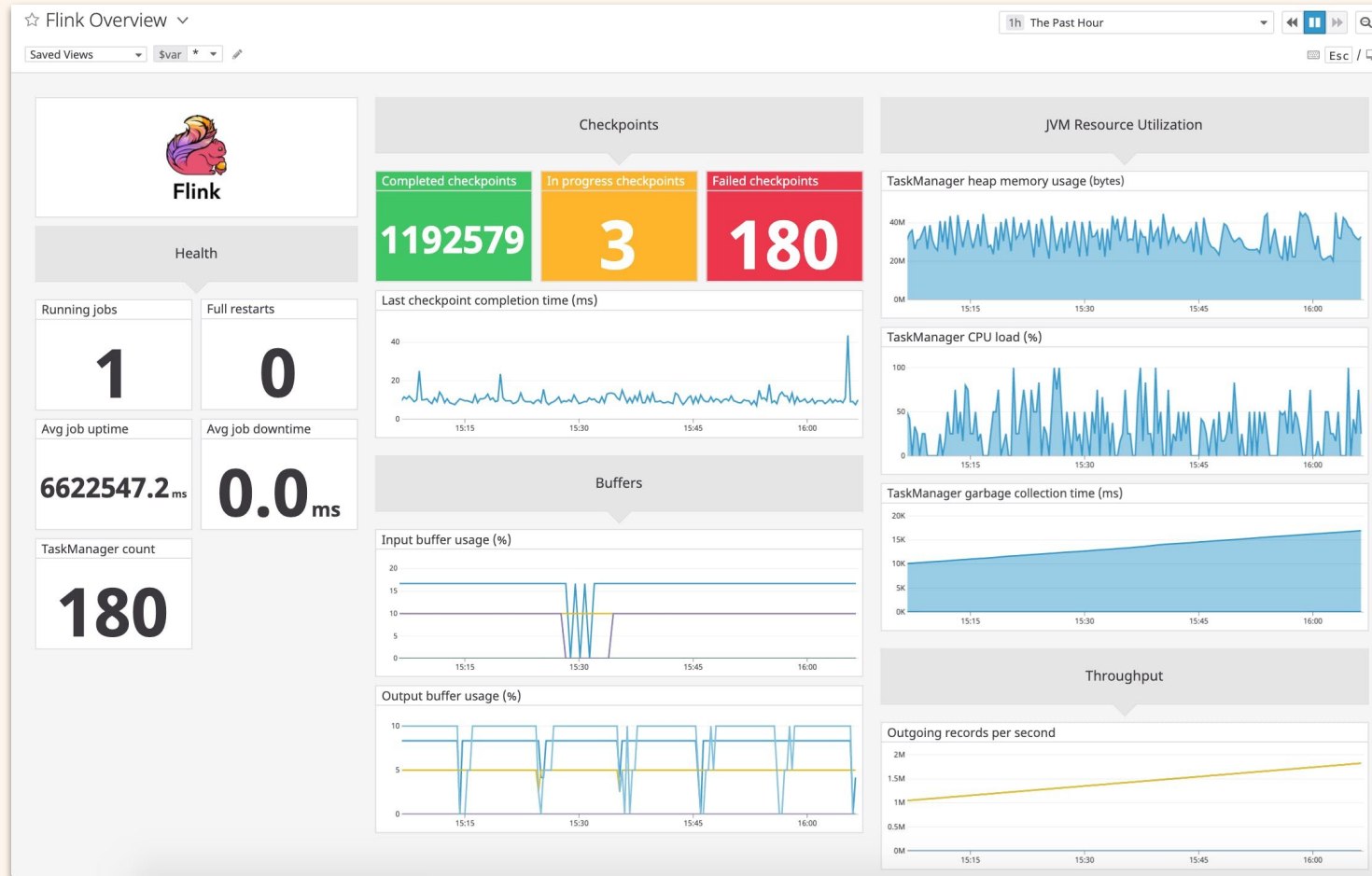
# Metrics-Driven Metrics

The dashboard should **reflect** your:

- Roadmap
- Highest Risk
- Most uncertain metrics



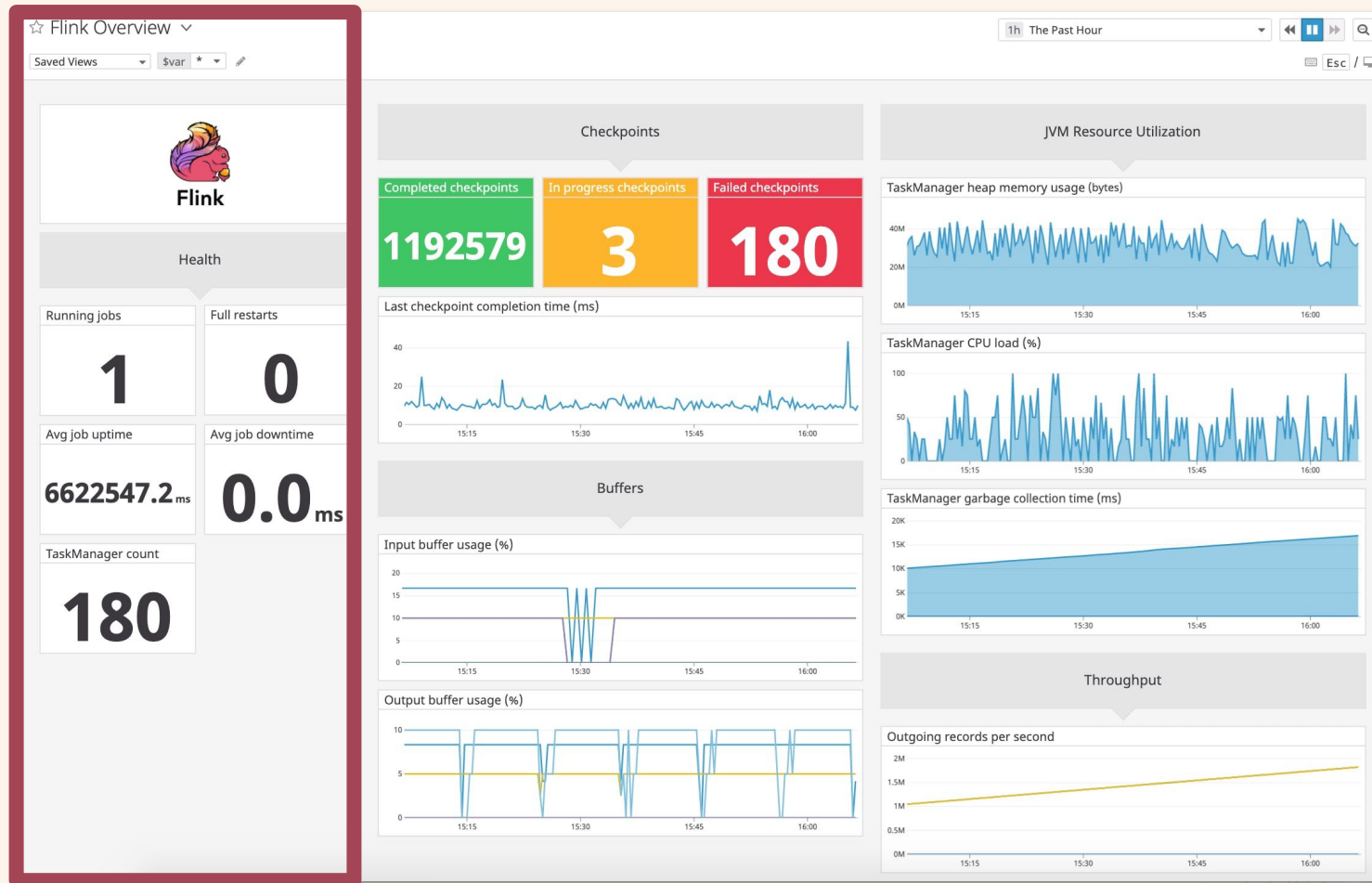
# Metrics-Driven Metrics



@CAITO\_200\_OK



# Metrics-Driven Metrics



@CAITO\_200\_OK



# Metrics-Driven Metrics

Apache Flink Dashboard

Join Rides with Fares (java RichCoFlatMap) 2018-12-06, 10:40:27 1m 40s

Overview Timeline Exceptions Configuration

```
graph LR; S1[Source: Custom Source  
Parallelism: 1] -- REBALANCE --> F[Filter  
Parallelism: 4  
Low Watermark: 1357178639999]; S2[Source: Custom Source  
Parallelism: 1] -- HASH --> C[Co-Flat Map -> Sink: Print to Std. Out  
Parallelism: 4  
Low Watermark: 1357178499999]; F -- HASH --> C;
```

Subtasks Task Metrics Watermarks Accumulators Checkpoints Back Pressure

Aggregate task statistics by TaskManager

Start Time	End Time	Duration	Name	Bytes received	Records received	Bytes sent	Records sent	Parallelism	Tasks	Status
2018-12-06, 10:40:27	2018-12-06, 10:42:08	1m 40s	Source: Custom Source	0 B	0	348 MB	1,461,469	1	1	RUNNING

Veronica Platform

Community Edition

Namespace: default

Applications

+ Create Deployment

dP Deployments

Artifacts

Secret Values

Administration

### dP Deployments

Deployment Defaults + Create Deployment

Sorting: CPU (most first) Filter: any string, or key value Status: All Status

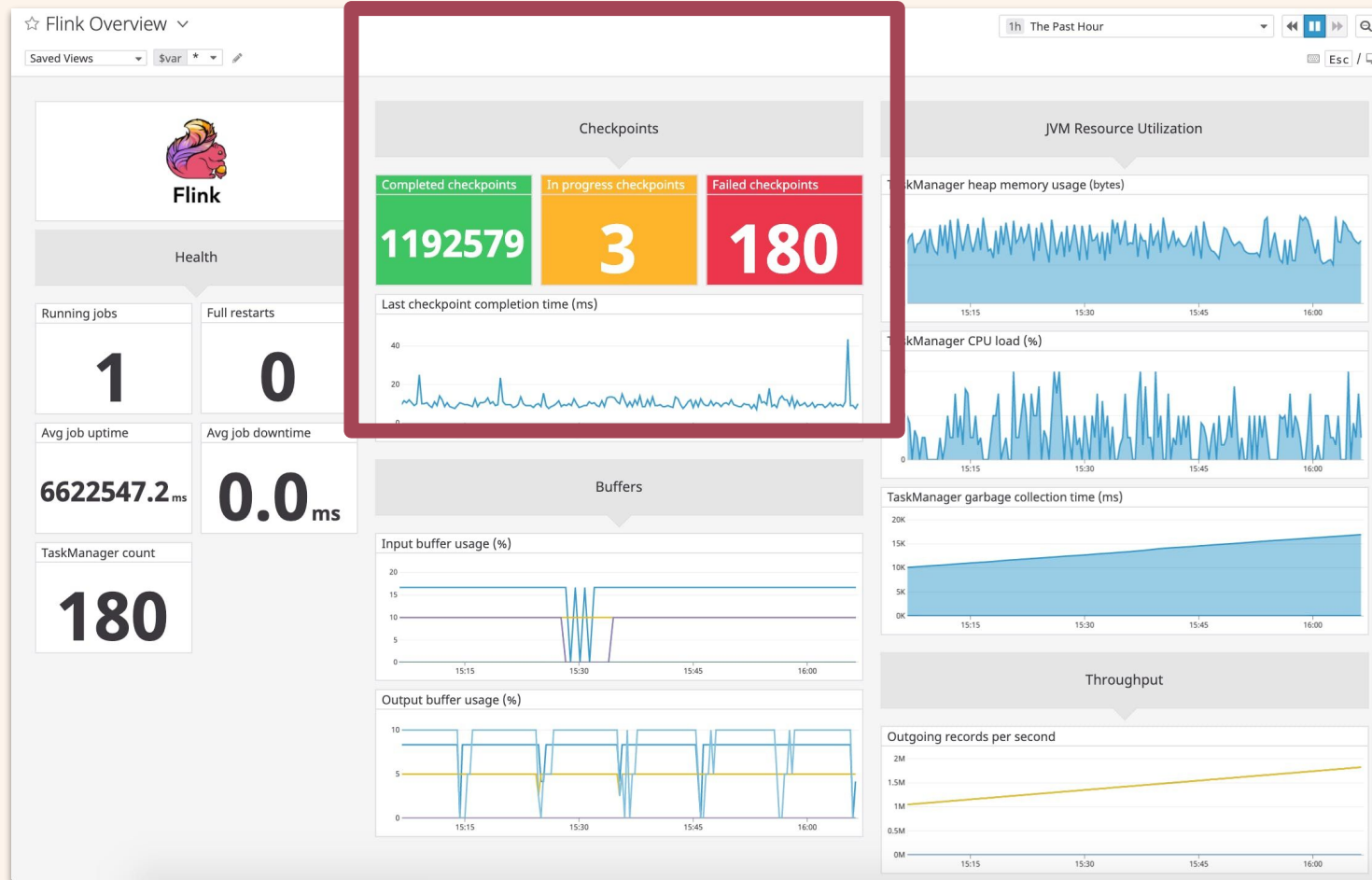
- Real Time Recommendations** Created: 2020-09-25, 13:48:40  
CPU: 3 Memory: 50 Task: 1/30 status: RUNNING Desired RUNNING Current
- S3 to ElasticSearch** Created: 2020-09-25, 14:39:28  
CPU: 2 Memory: 50 Task: 1/30 status: RUNNING Desired RUNNING Current
- Real Time Enrichment** Created: 2020-09-25, 13:52:34  
CPU: 2 Memory: 50 Task: 1/30 status: SUSPENDED Desired SUSPENDED Current
- Kafka to Parquet** Created: 2020-09-25, 13:52:18  
status: SUSPENDED RUNNING

@CAITO\_200\_OK





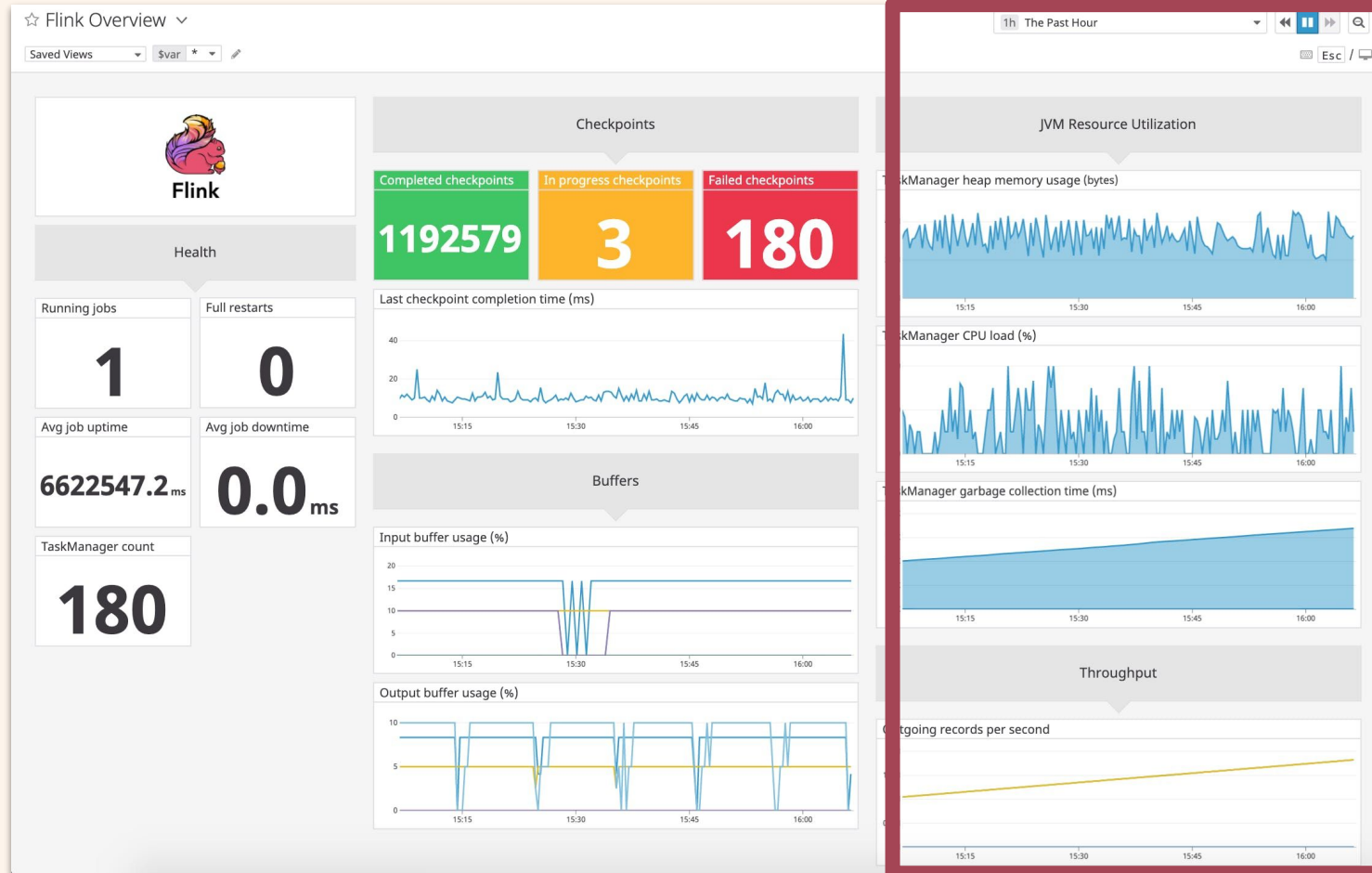
# Metrics-Driven Metrics



@CAITO\_200\_OK



# Metrics-Driven Metrics



@CAITO\_200\_OK



# Metrics-Driven Metrics

## Prometheus

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

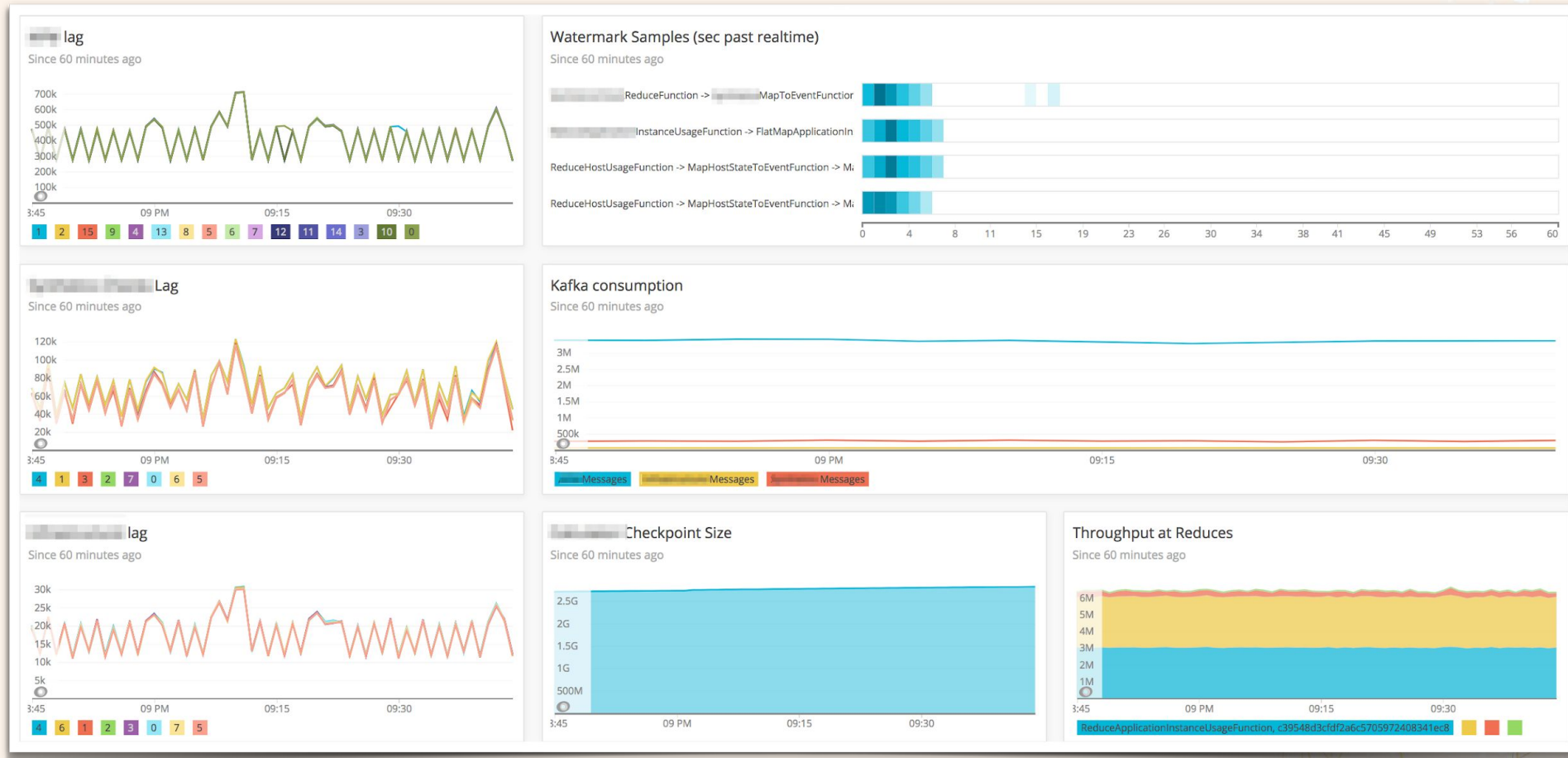
## DataDog

```
metrics.reporter.dghttp.factory.class: org.apache.flink.metrics.datadog.DatadogHttpReporterFactory
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
```



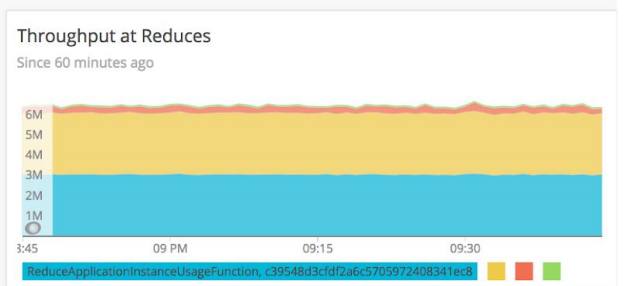
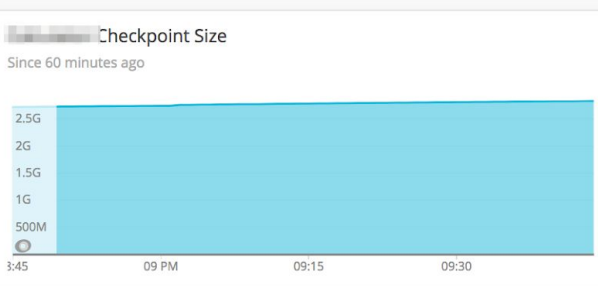
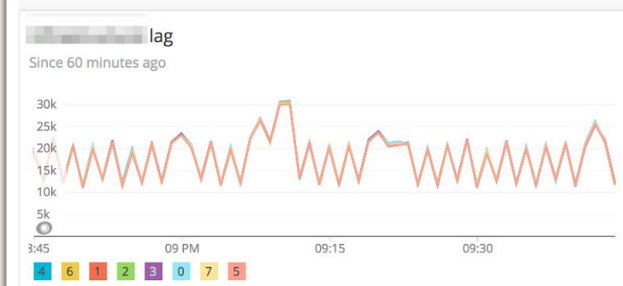
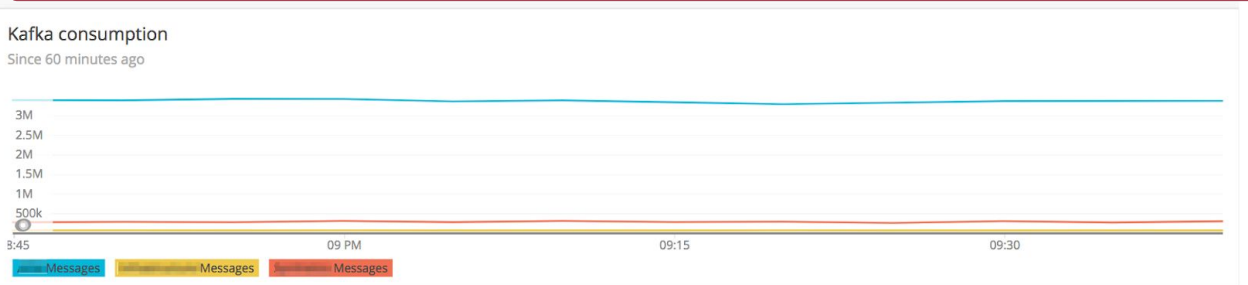
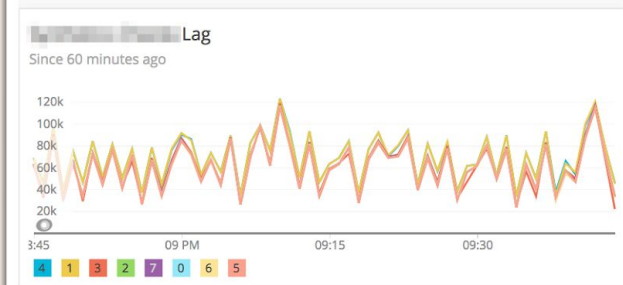
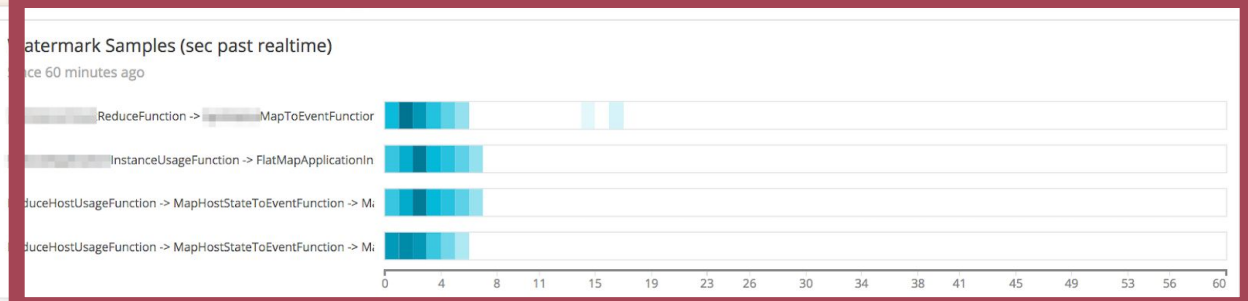
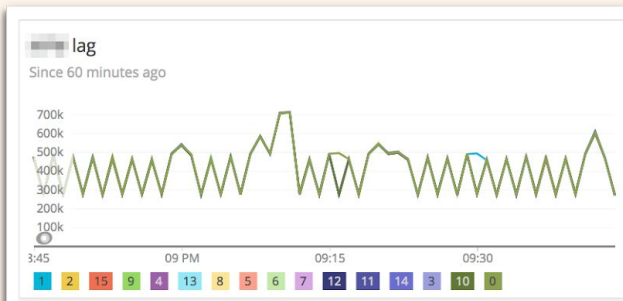


# Metrics-Driven Metrics

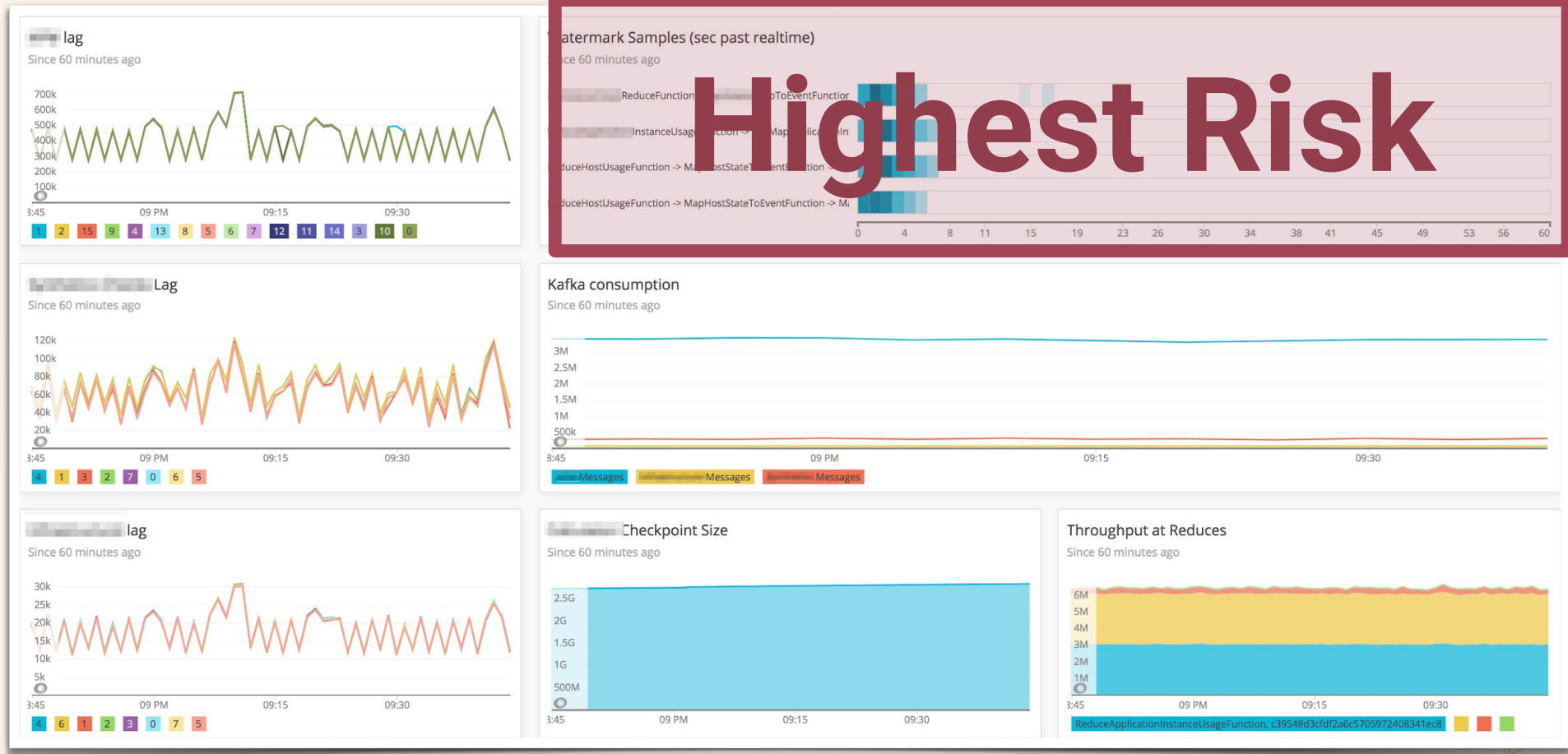




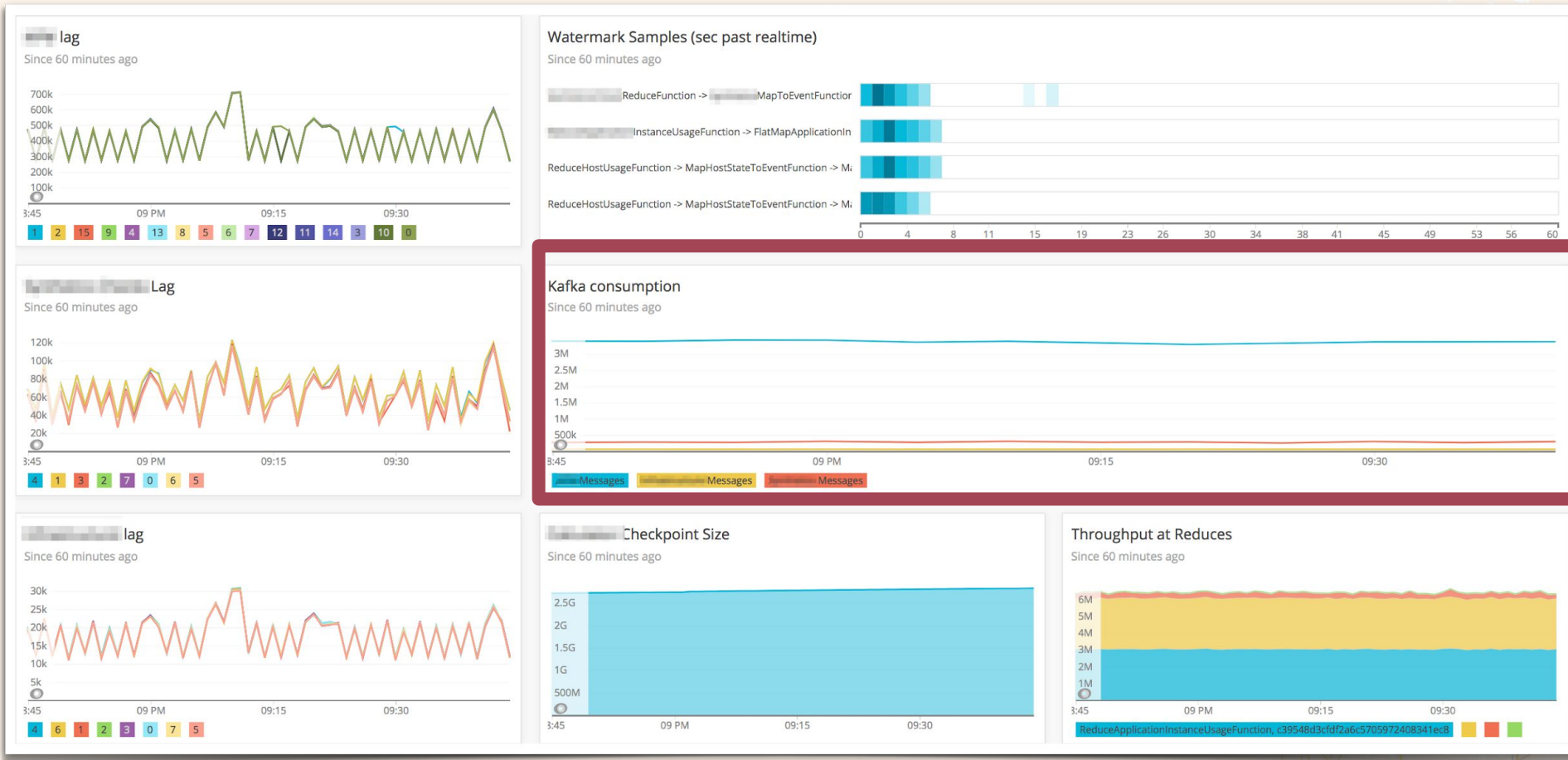
# Metrics-Driven Metrics



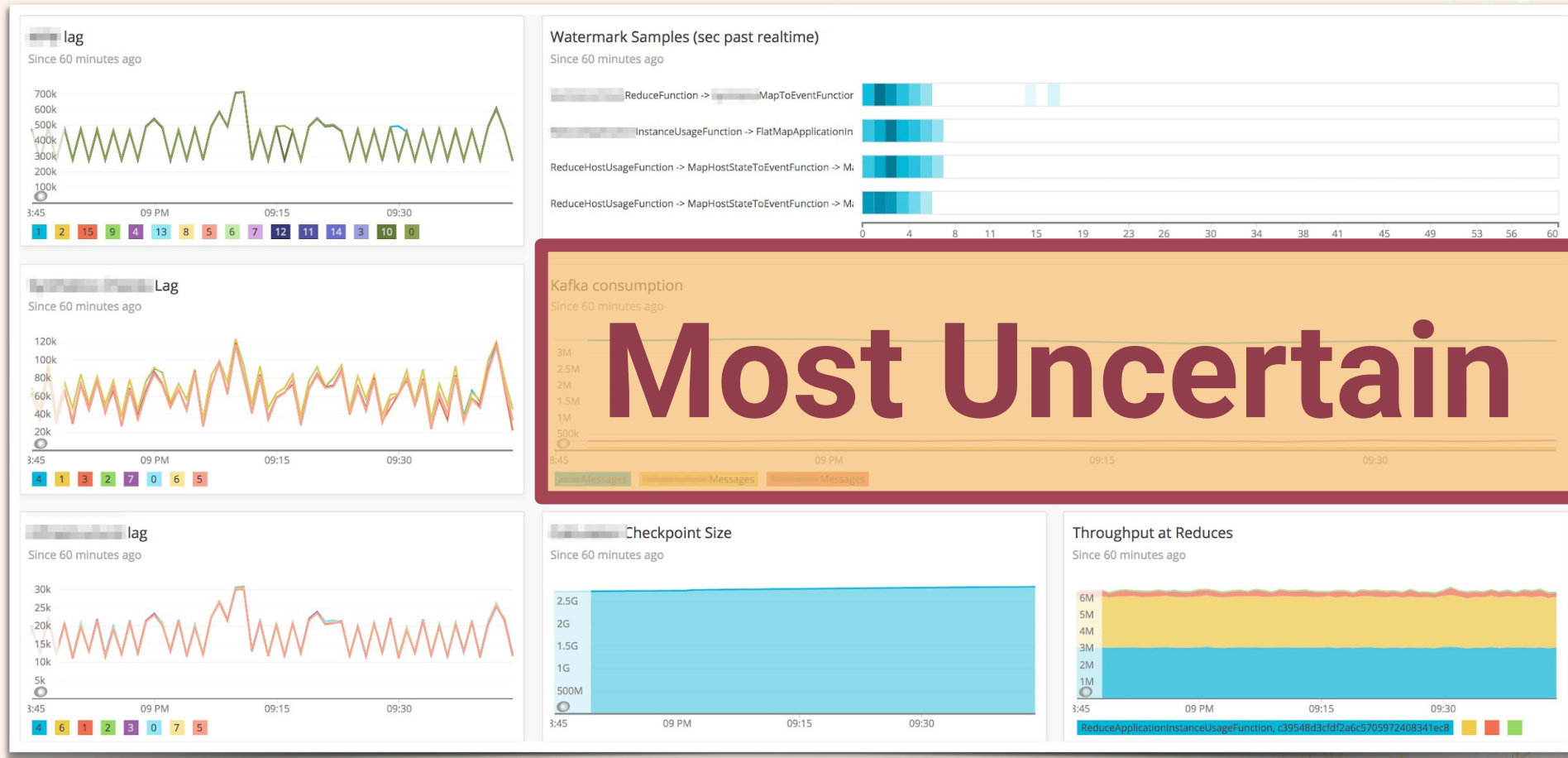
# Metrics-Driven Metrics



# Metrics-Driven Metrics

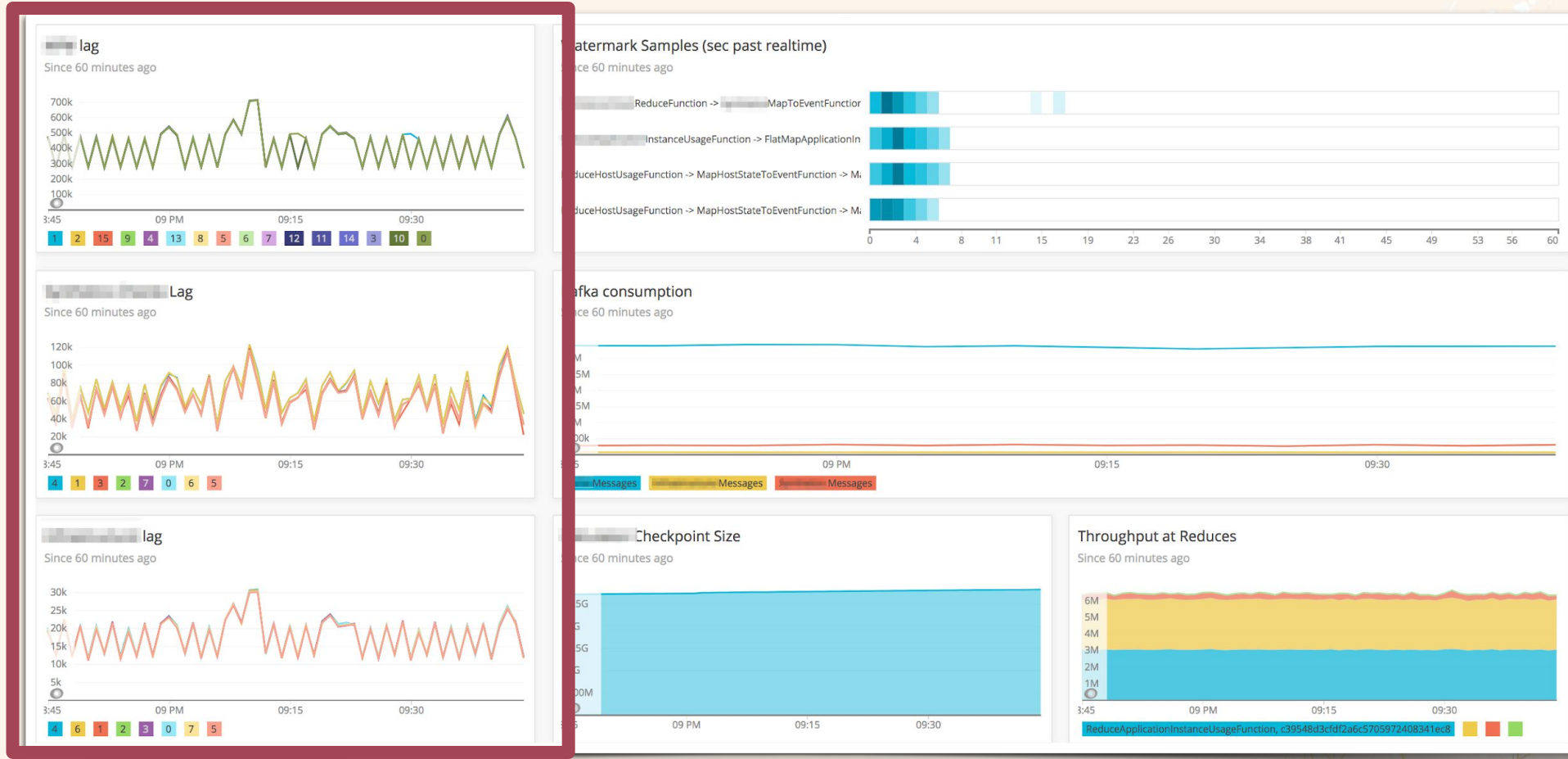


# Metrics-Driven Metrics



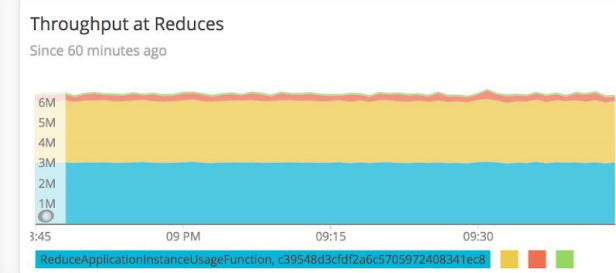
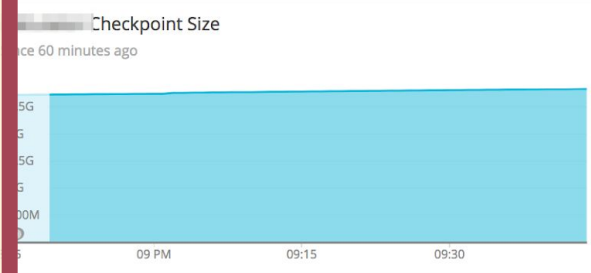
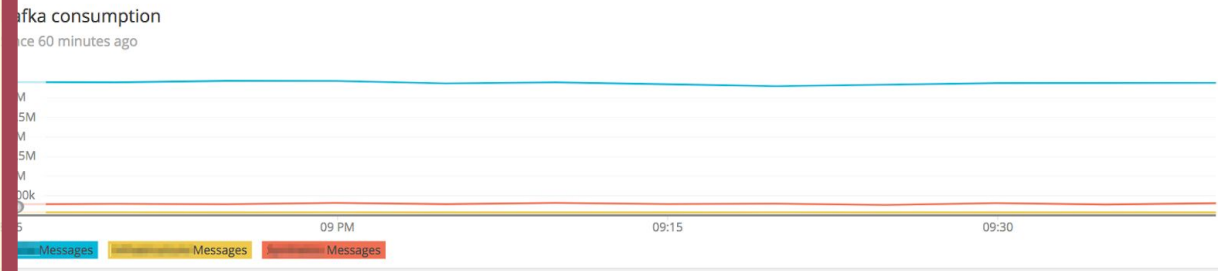
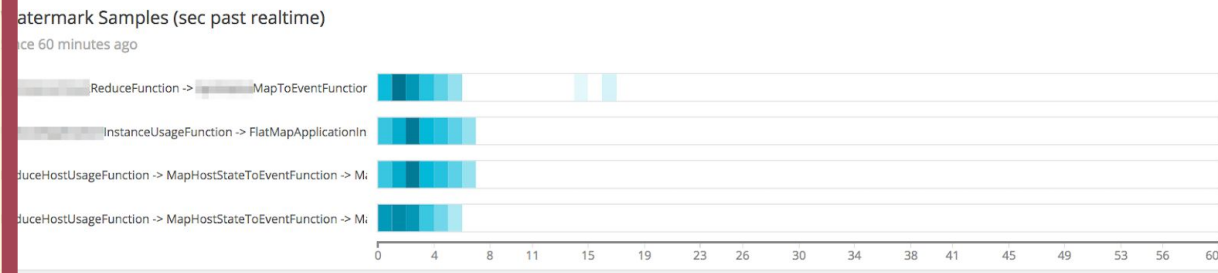
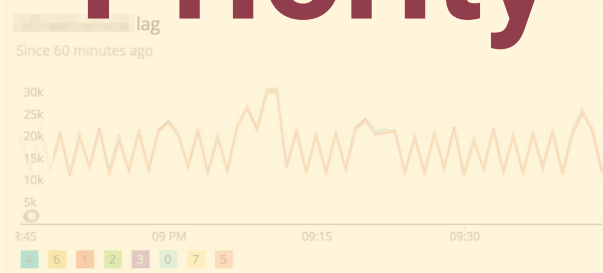


# Metrics-Driven Metrics

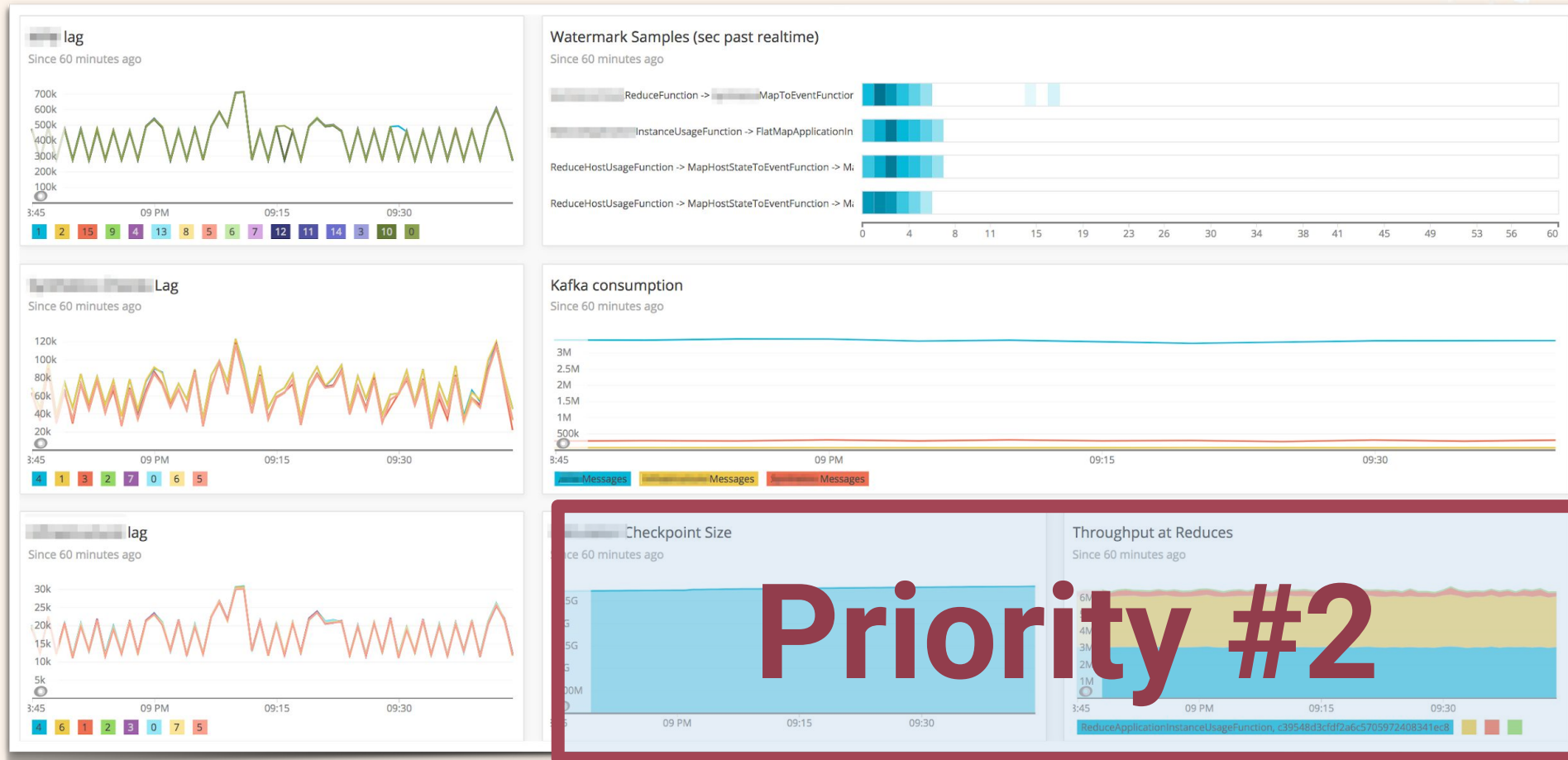


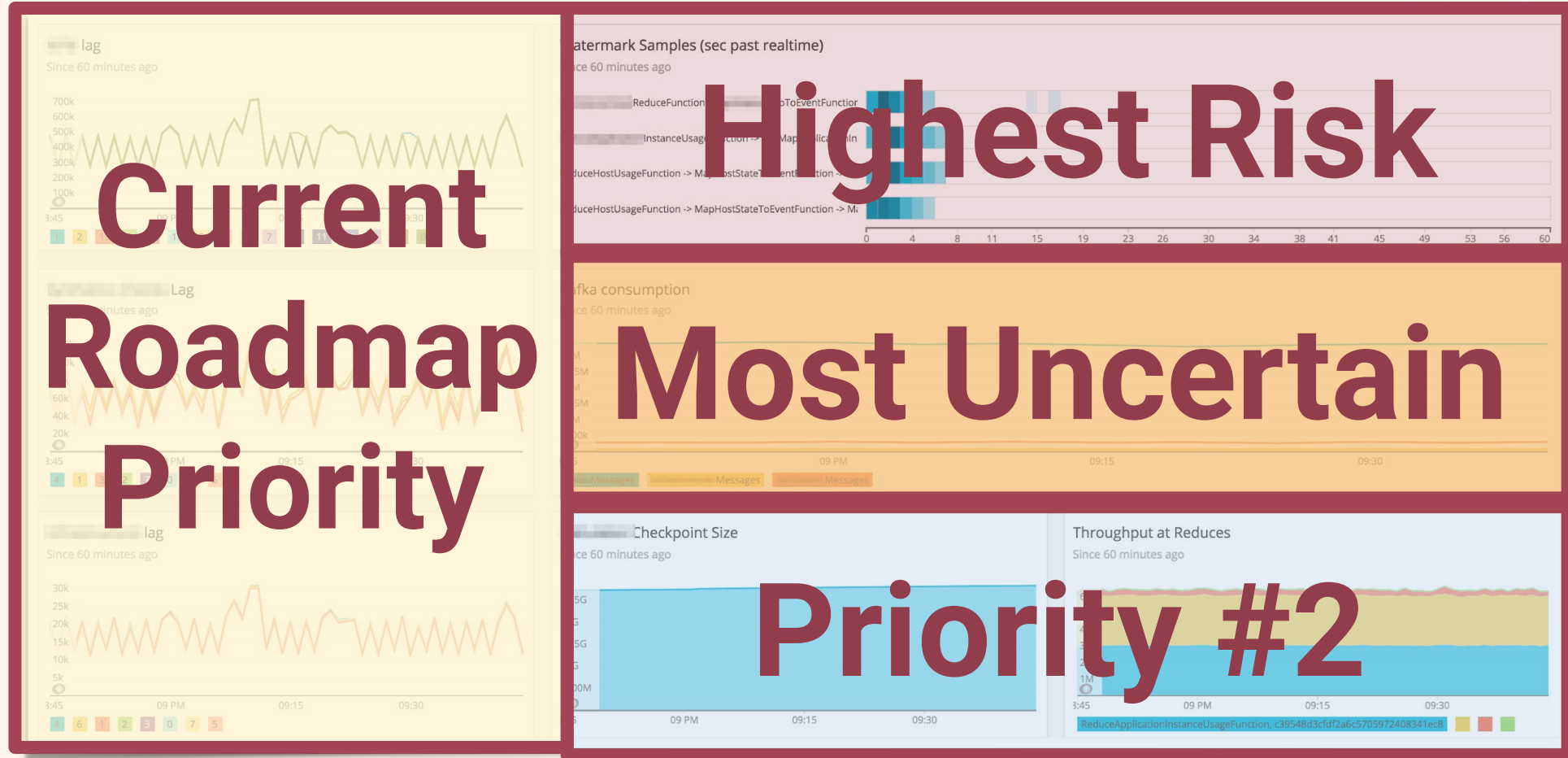
# Metrics-Driven Metrics

## Current Roadmap Priority



# Metrics-Driven Metrics







# 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



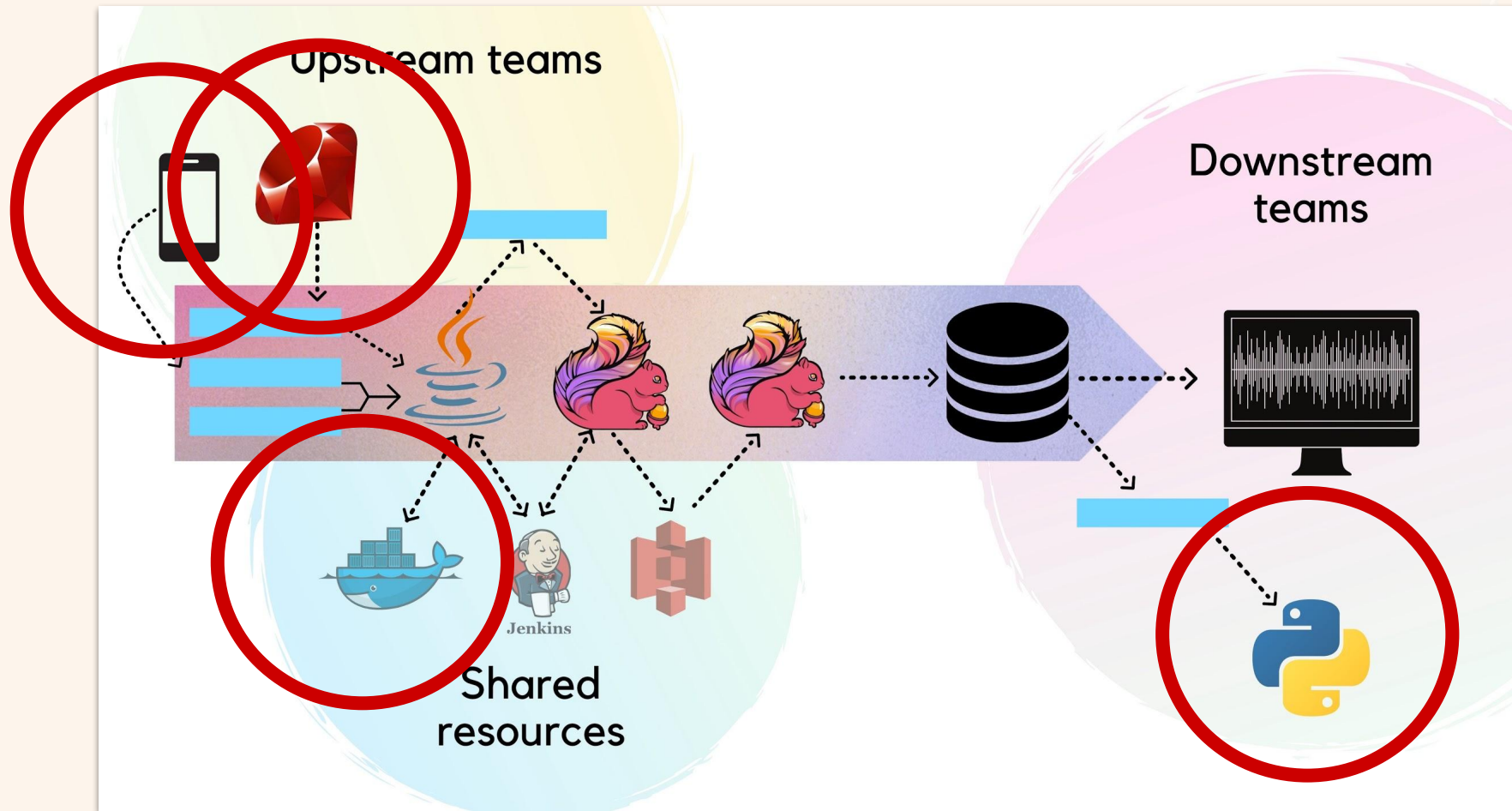








# Shared Language



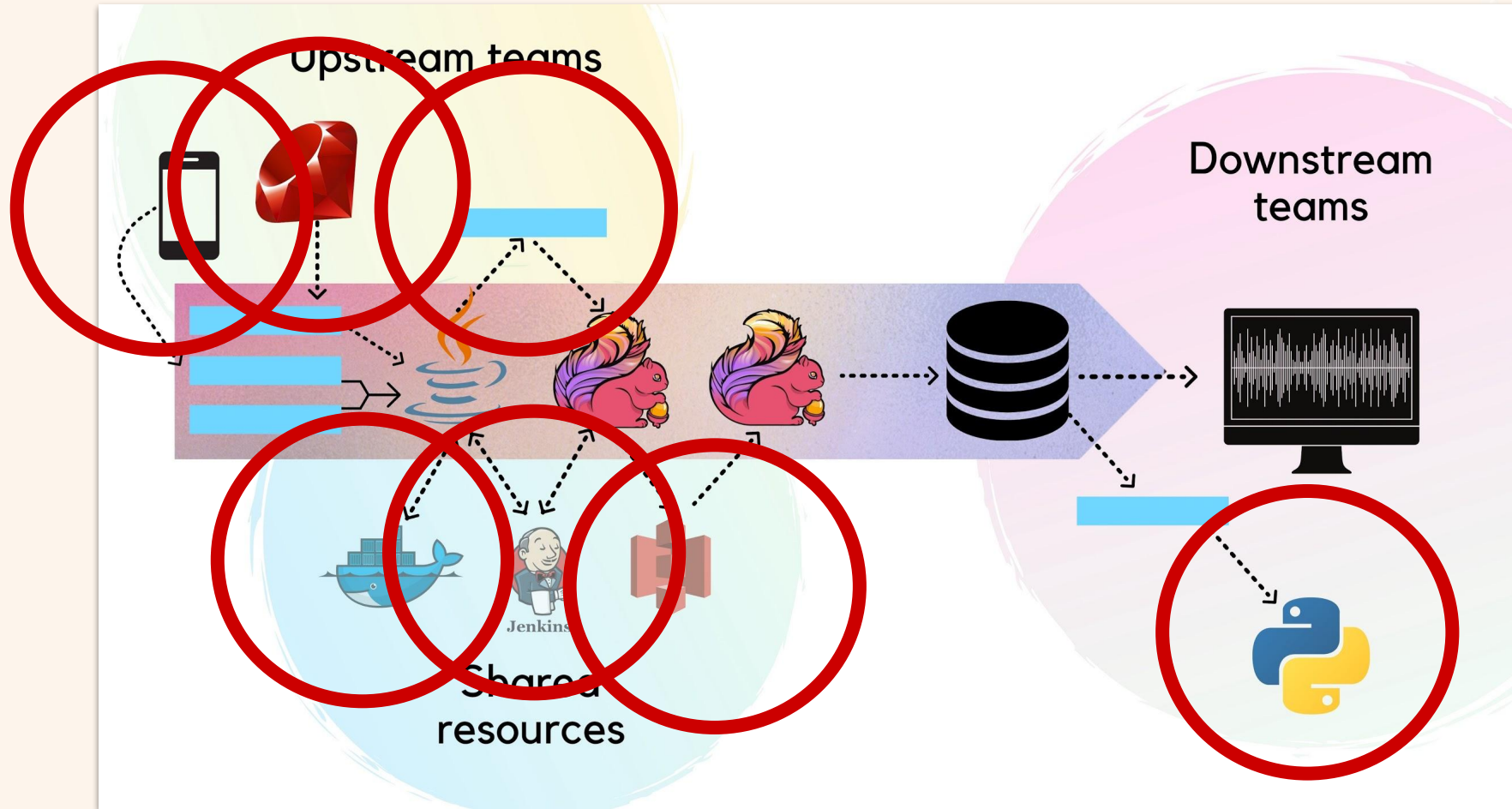




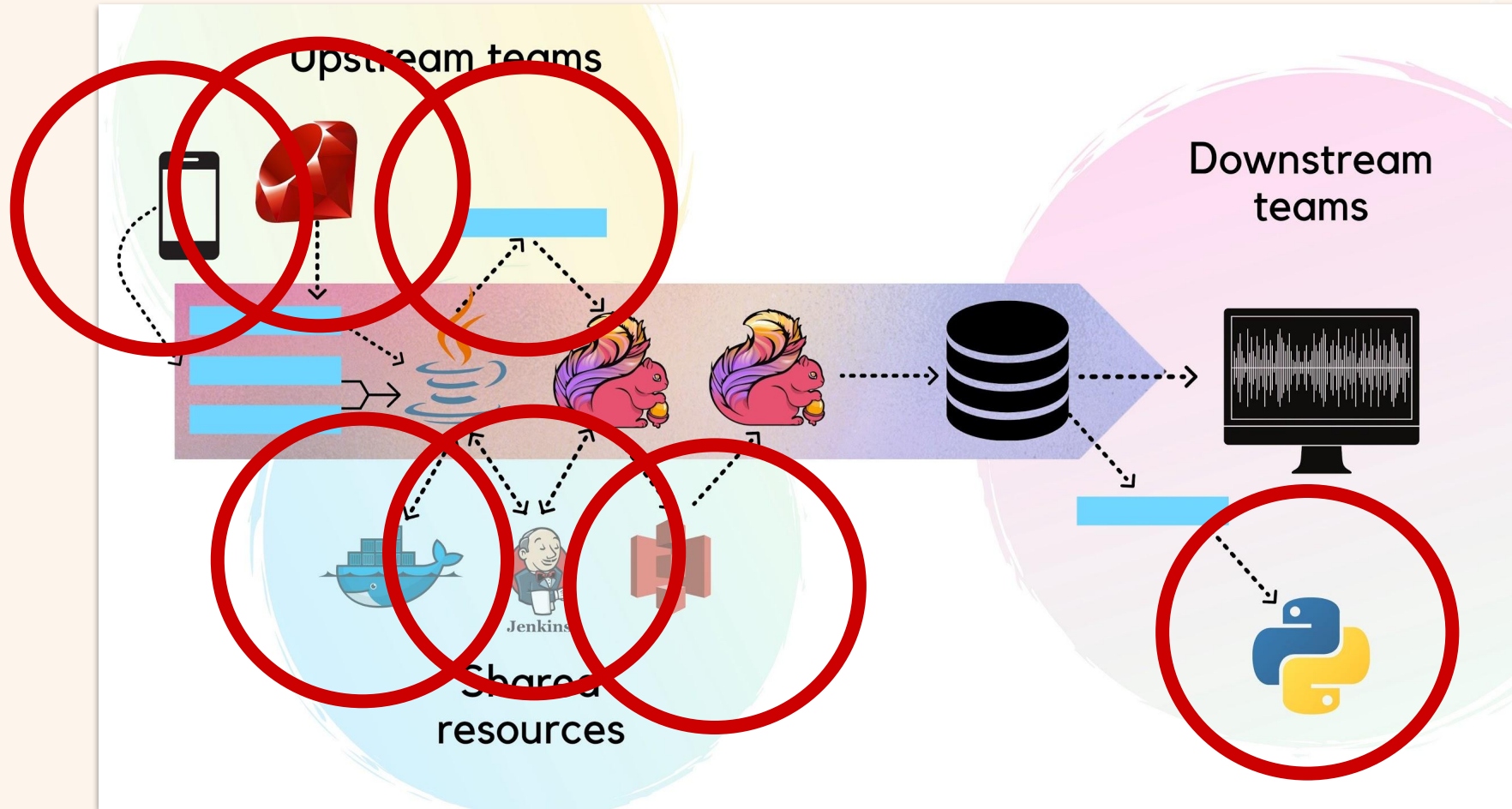




# Shared Language

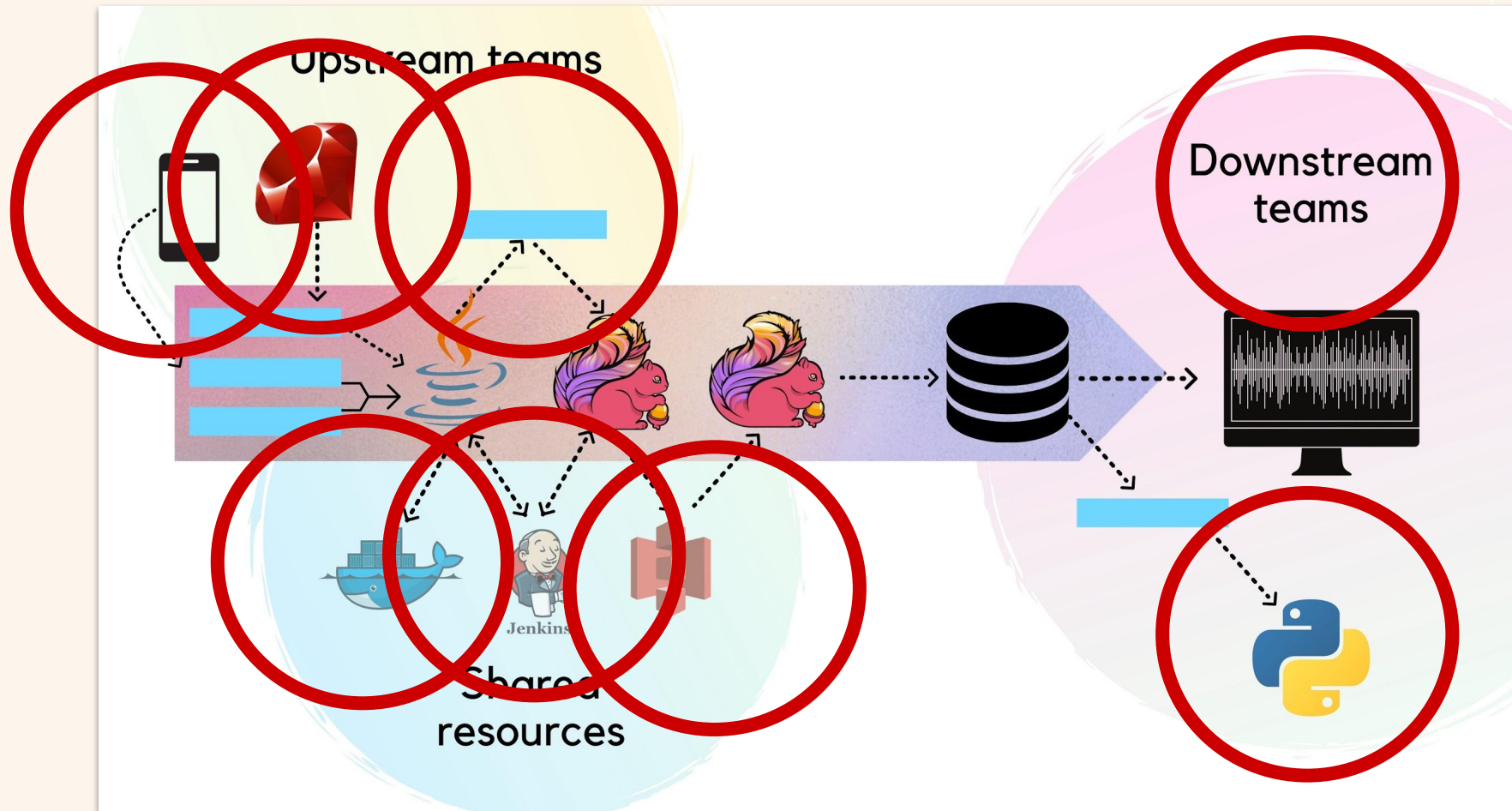


# Shared Language



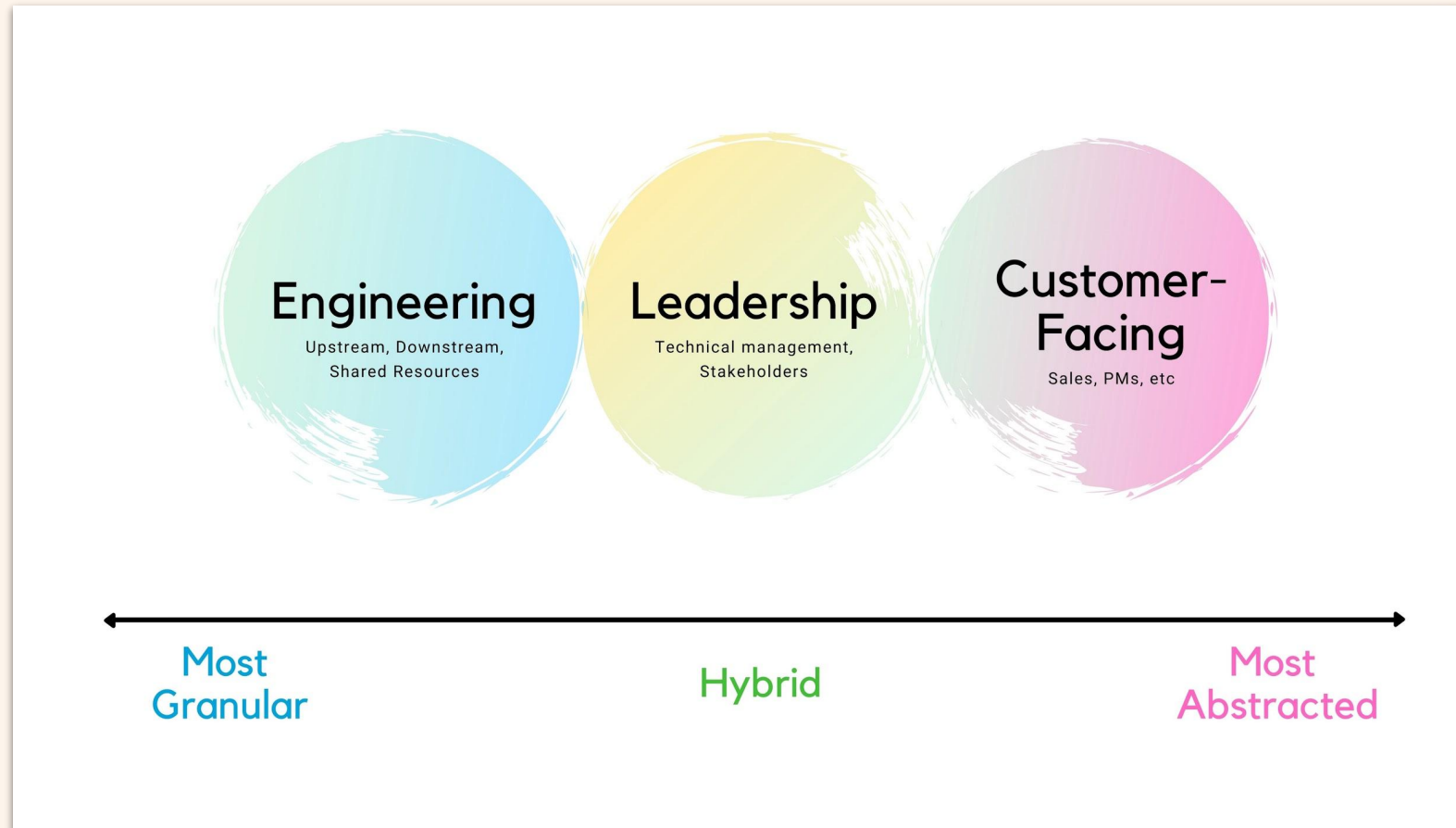


# Shared Language

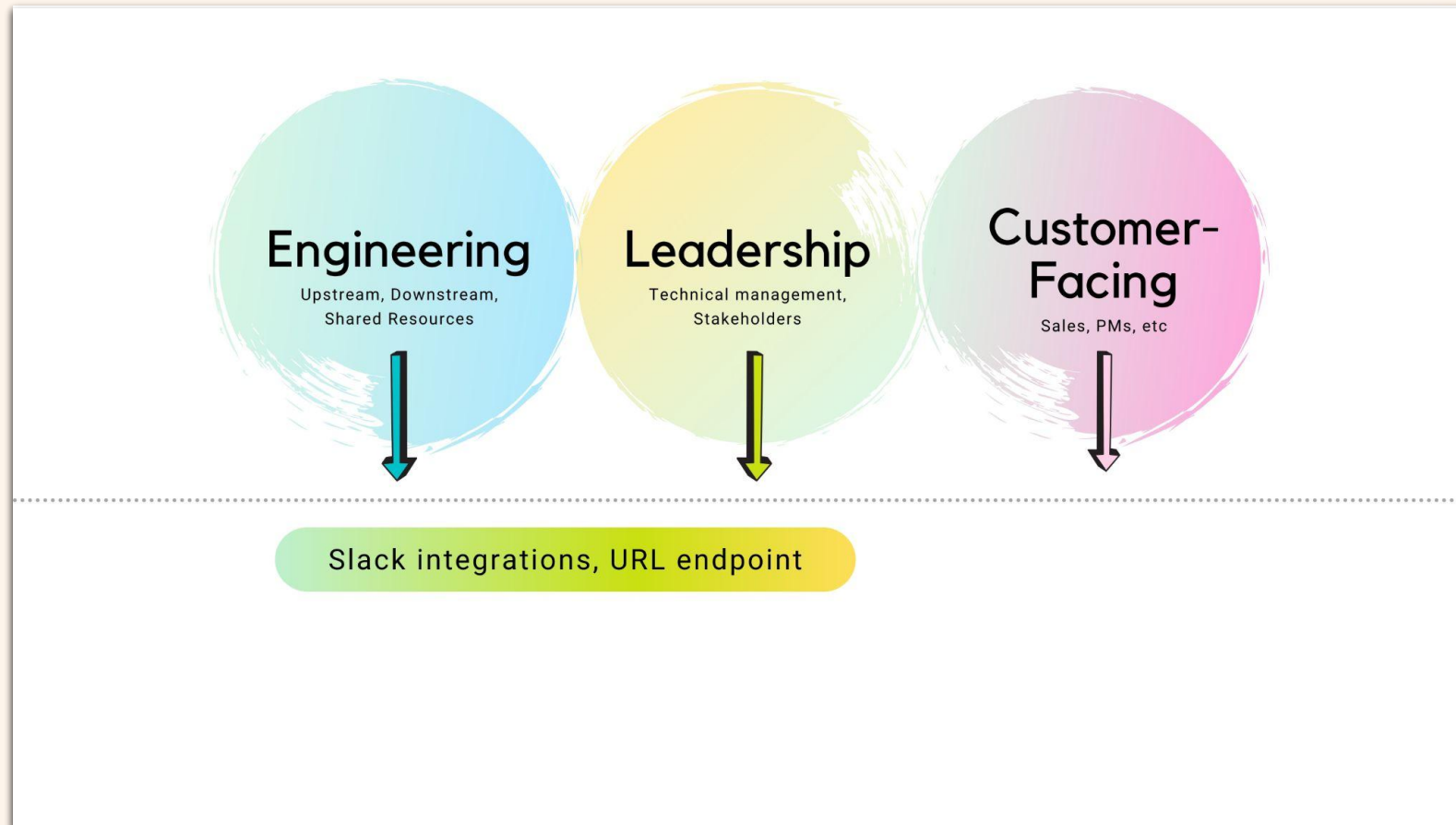




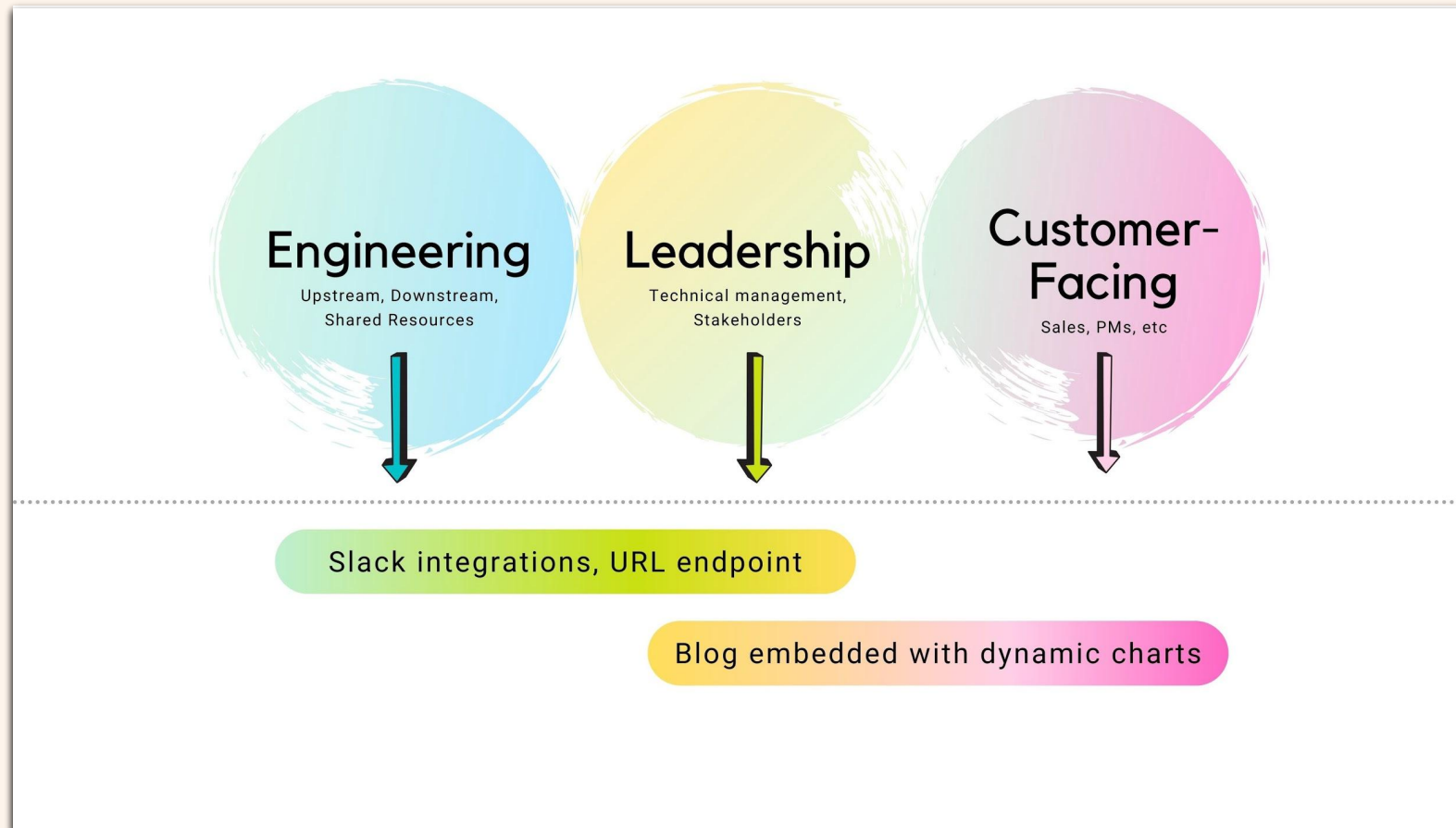
# Shared Language



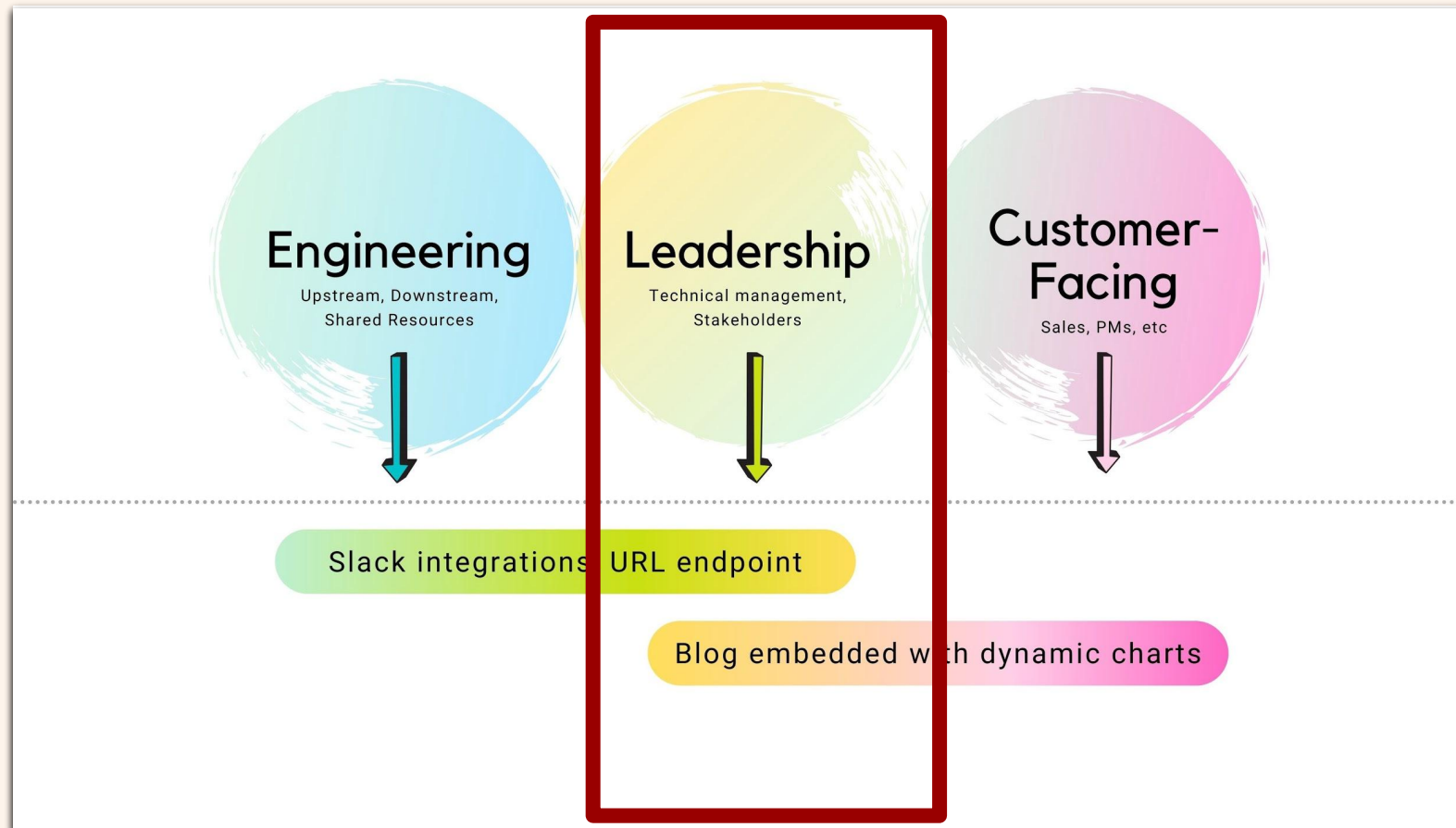
# Shared Language



# Shared Language



# Shared Language





# 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



# Introduction

*Twitter*

**Caito\_200\_OK**

*Content*

**<https://medium.com/@caito>**

**<http://caito-200-ok.com/>**

*Email*

**Caito@ververica.com**

@CAITO\_200\_OK



# Thank You!



- **YOW! 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/docuemntation:  
<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/](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](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>

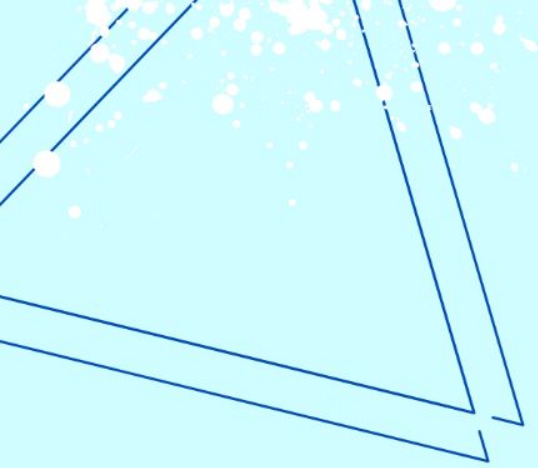




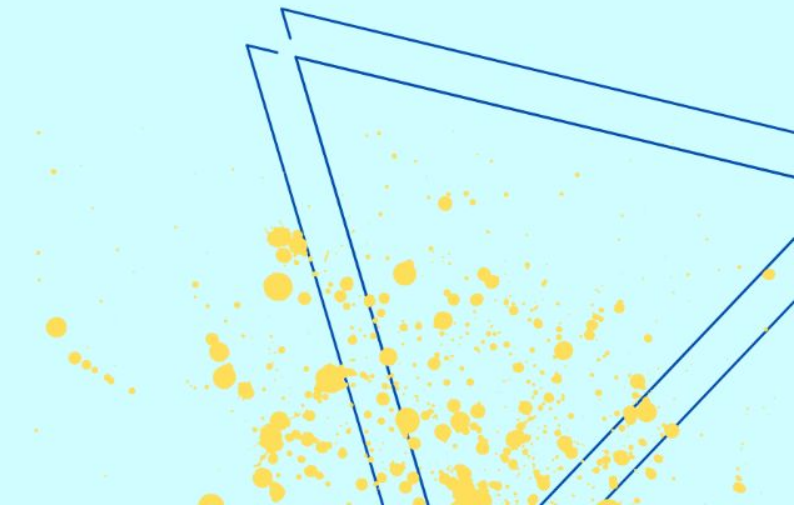
ververica

---





@CAITO-200\_OK

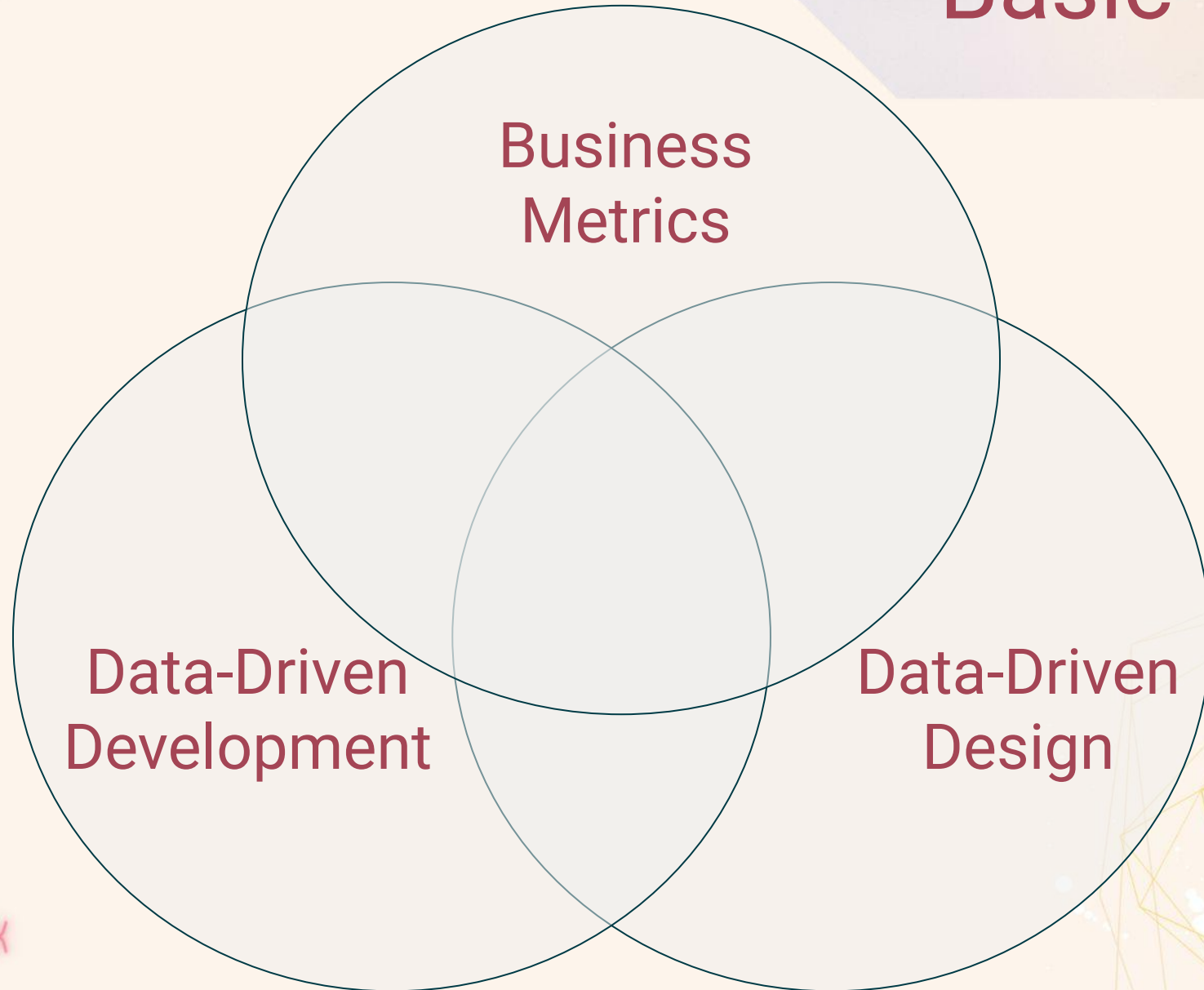


# Basic Principles

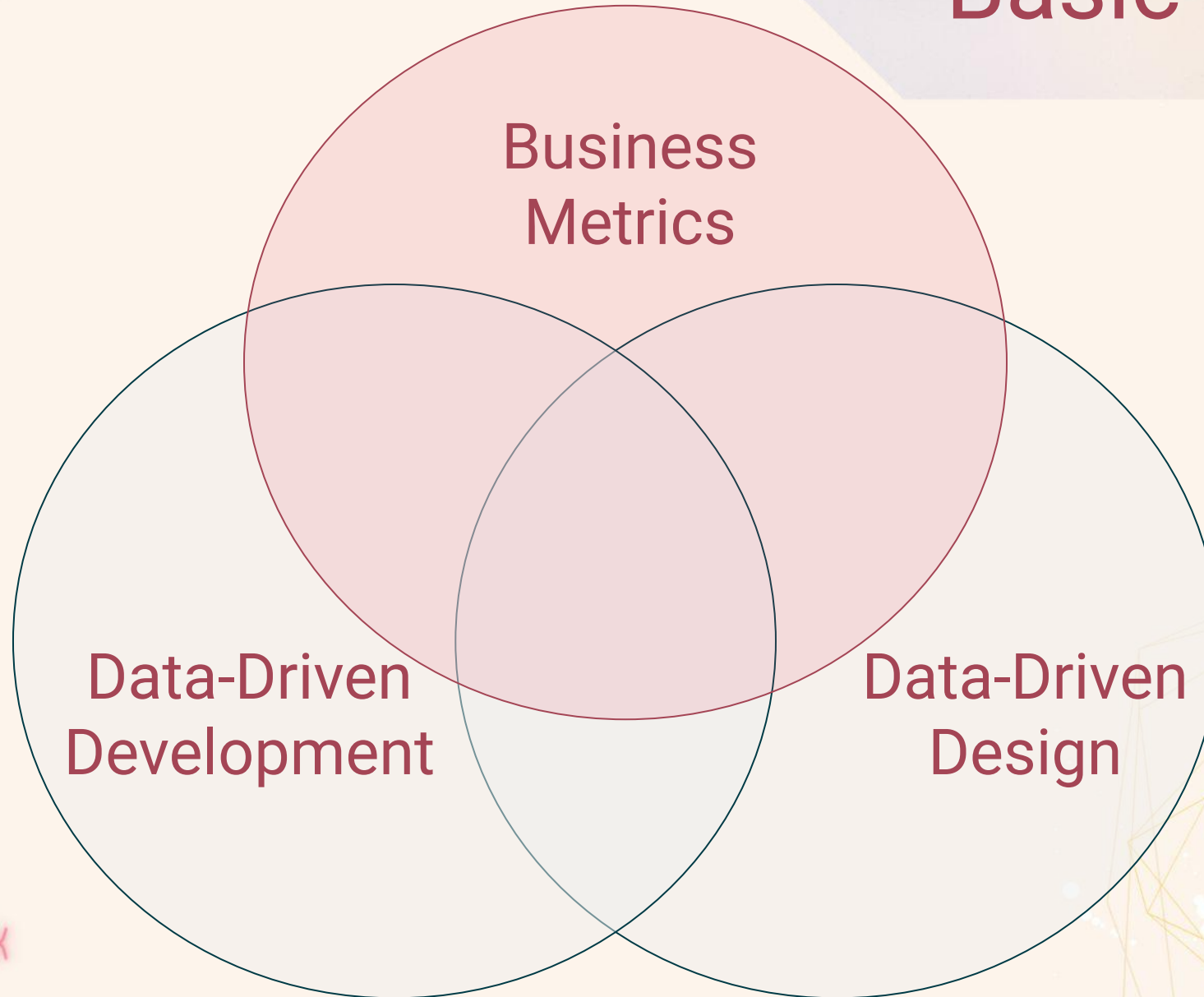
@CAITO\_200\_OK



# Basic Principles

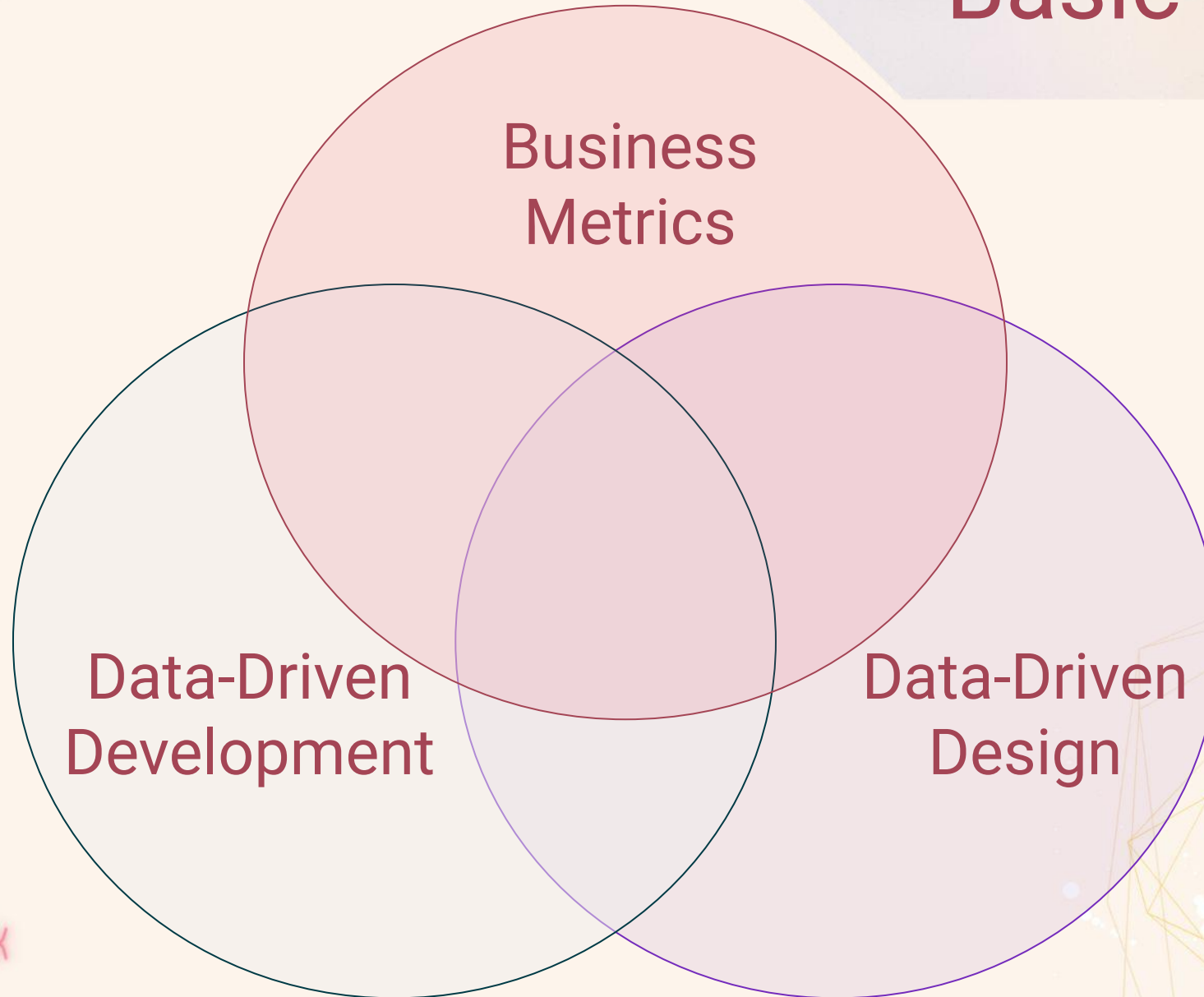


# Basic Principles

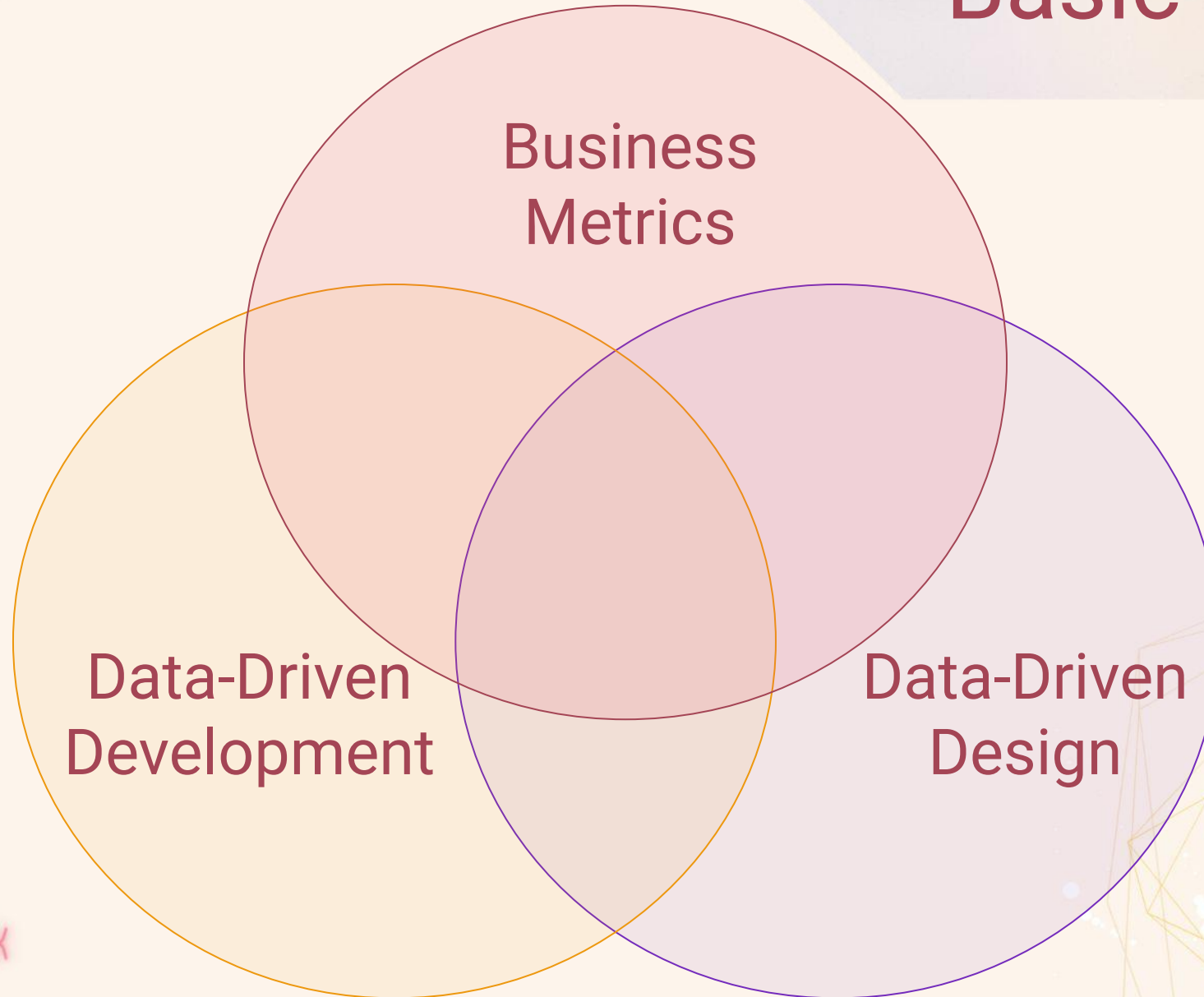




# Basic Principles



# Basic Principles



# CONCLUSION

Audience	Tactic	Tools
You/Your Team	<ul style="list-style-type: none"><li>• Dashboard should represent your priorities &amp; values</li><li>• Bring Development + Operations together</li></ul>	<b>Analytics platforms</b> <ul style="list-style-type: none"><li>• Flink UI, Ververica Platform</li><li>• Data Dog</li><li>• New Relic</li><li>• Grafana/Prometheus...</li></ul>
Upstream + Downstream Teams	<ul style="list-style-type: none"><li>• Understand each other's "normal"</li><li>• Clear ownership for integration points</li></ul>	<b>Analytics + chat integrations</b> <ul style="list-style-type: none"><li>• Slack + PagerDuty</li><li>• Slack + Jenkins...</li></ul>
Leadership + Stakeholders	<ul style="list-style-type: none"><li>• Business metrics</li><li>• Pivot between summary &amp; depth</li><li>• Use tools they're familiar with</li></ul>	<b>Automated, non-engineering spaces</b> <ul style="list-style-type: none"><li>• URL endpoints</li><li>• Internal wiki/blog (with embedded, automated output)</li></ul>
Customers	<ul style="list-style-type: none"><li>• More manual approach</li></ul>	<b>Manual approach</b> <ul style="list-style-type: none"><li>• Your company's customer-facing teams</li></ul>



# BASIC PRINCIPLES



**Quantitative**

**Qualitative**





# BASIC PRINCIPLES

Data-Driven Design



**Quantitative**

**Qualitative**



# BASIC PRINCIPLES

Data-Driven Design

Data-Informed Design



**Quantitative**

**Qualitative**



# BASIC PRINCIPLES

Data-Driven Design

Data-Informed Design

Data-Aware Design

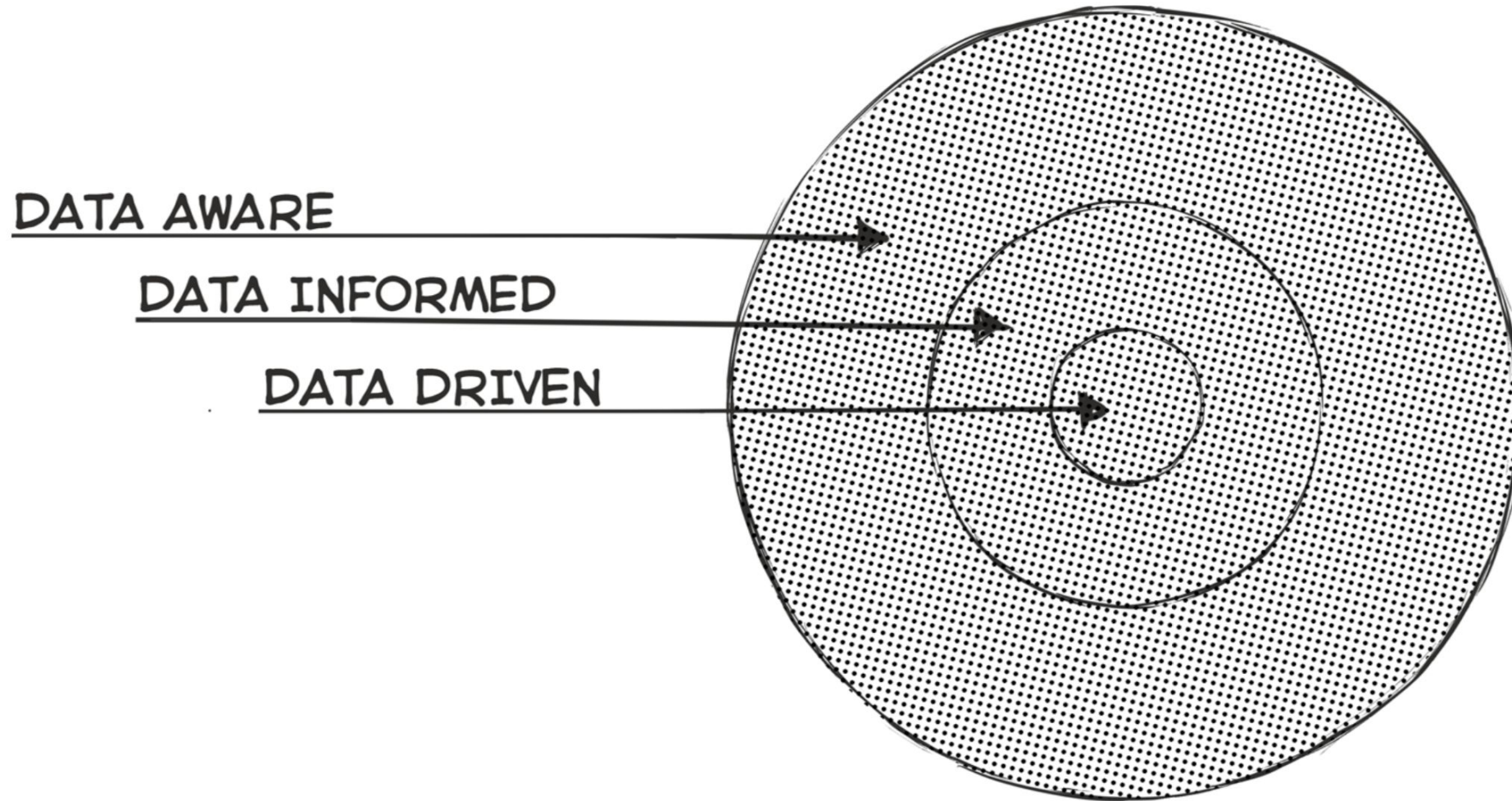


**Quantitative**

**Qualitative**



# BASIC PRINCIPLES





# THE SHARED LANGUAGE OF METRICS

# A SHARED LANGUAGE

Even the best data-driven development  
**fails easily** when it's **only built for yourself.**

# Metrics as a Shared Language

Engineering Team

Up + Downstream Teams

Leadership/Stakeholders

Customers

Most  
Technical



Most  
Abstract



# Metrics as a Shared Language

Engineering Team

Up + Downstream Teams

Leadership/Stakeholders

Customers

Most  
Technical



Most  
Abstract





# Metrics as a Shared Language

Engineering Team

Up + Downstream Teams

Leadership/Stakeholders

Customers

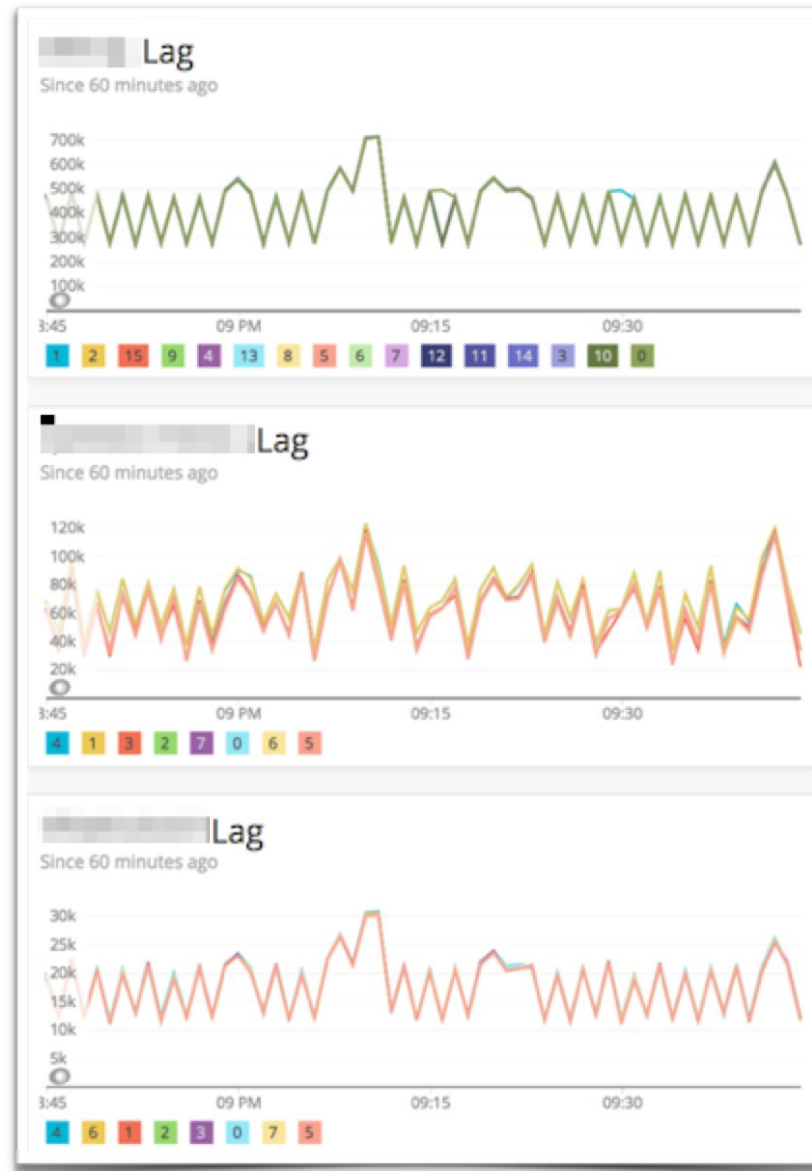
Most  
Technical



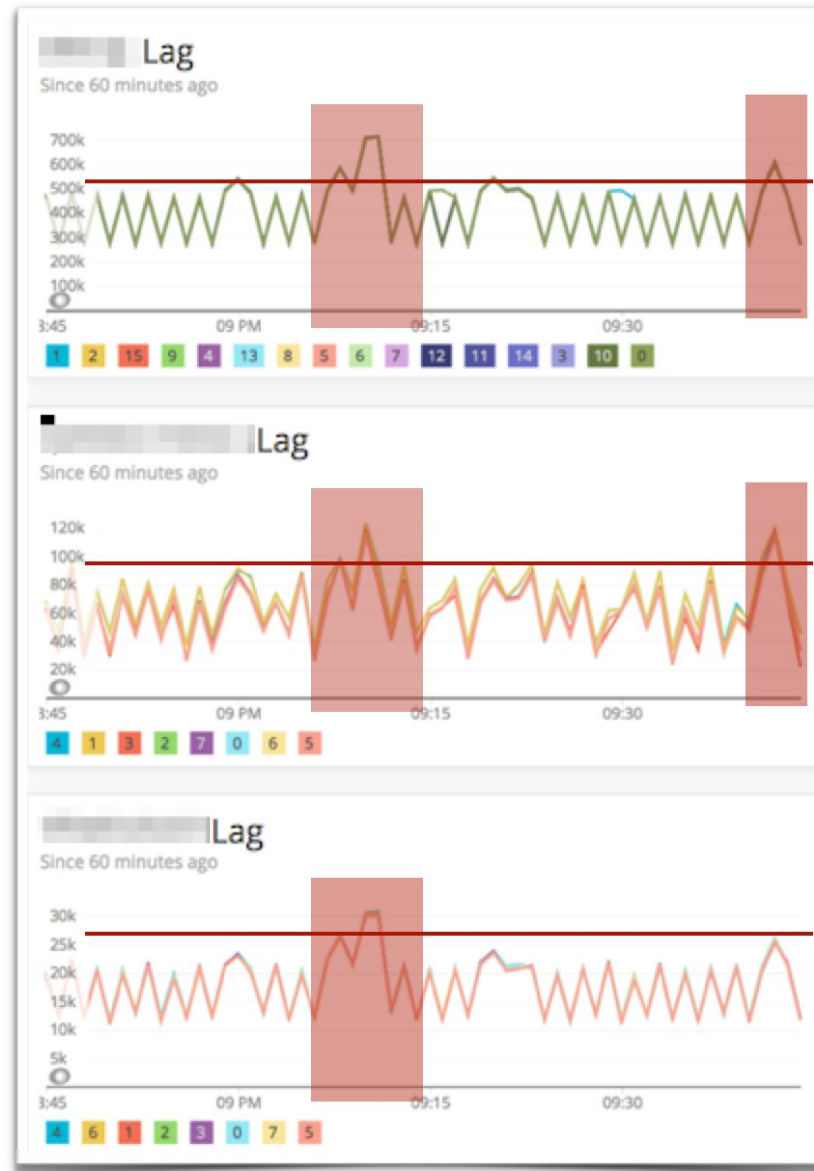
Most  
Abstract



# A SHARED LANGUAGE



# A SHARED LANGUAGE



# A SHARED LANGUAGE



# Metrics as a Shared Language

Engineering Team

Up + Downstream Teams

Leadership/Stakeholders

Customers

Most  
Technical



Most  
Abstract





# Metrics as a Shared Language

Engineering Team

Up + Downstream Teams

Leadership/Stakeholders

Customers

Most  
Technical



Most  
Abstract



# Metrics as a Shared Language

Engineering Team

Up + Downstream Teams

Leadership/Stakeholders

Customers

Most  
Technical



Most  
Abstract



# Summary

Concept	Principles	Implementation	Benefit
<b>Metrics-Driven-Metrics cycle</b>	Meaningful, iterative, accessible	High risk, most uncertain, current roadmap priority	Incident response,
<b>Metrics as a shared language</b>	Use tools familiar to your audience	Small modular units	Automate people & process challenges



# Before We Start...

## Flink's REST API integration

```
http://hostname:8081/jobmanager/metrics  
    /taskmanagers/<taskmanagerid>/metrics  
    /taskmanagers/metrics  
    /jobs/metrics?jobs=D,E,F  
    ...
```

