

MAKING A LAZY RIVER, NOT A WHITEWATER RAPIDS

ADOPTING STREAM PROCESSING FOR INSTRUMENTATION



SEAN CRIBBS
SENIOR PRINCIPAL ENGINEER





**DIRECT
KICK**

Sunday, July 12, 2015

TEAM
CHANNEL

MATCH-UP
Kansas City @ Vancouver

TIME
9:00pm/ET

ON INSTRUMENTATION

HD not available
in all areas.

Local blackout restrictions apply. Programming subject to change.

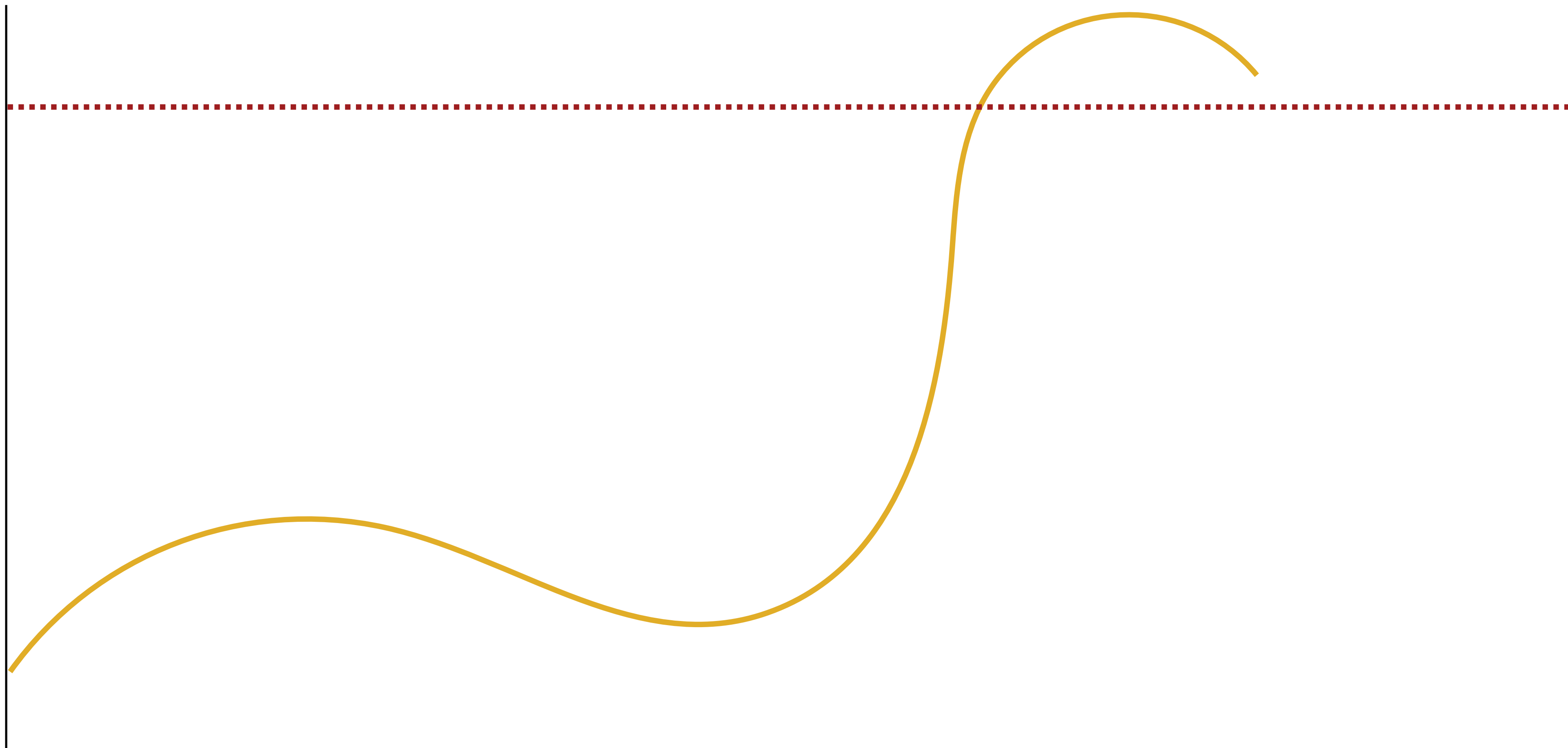
TOSHIBA

HOW ARE WE INSTRUMENTING?

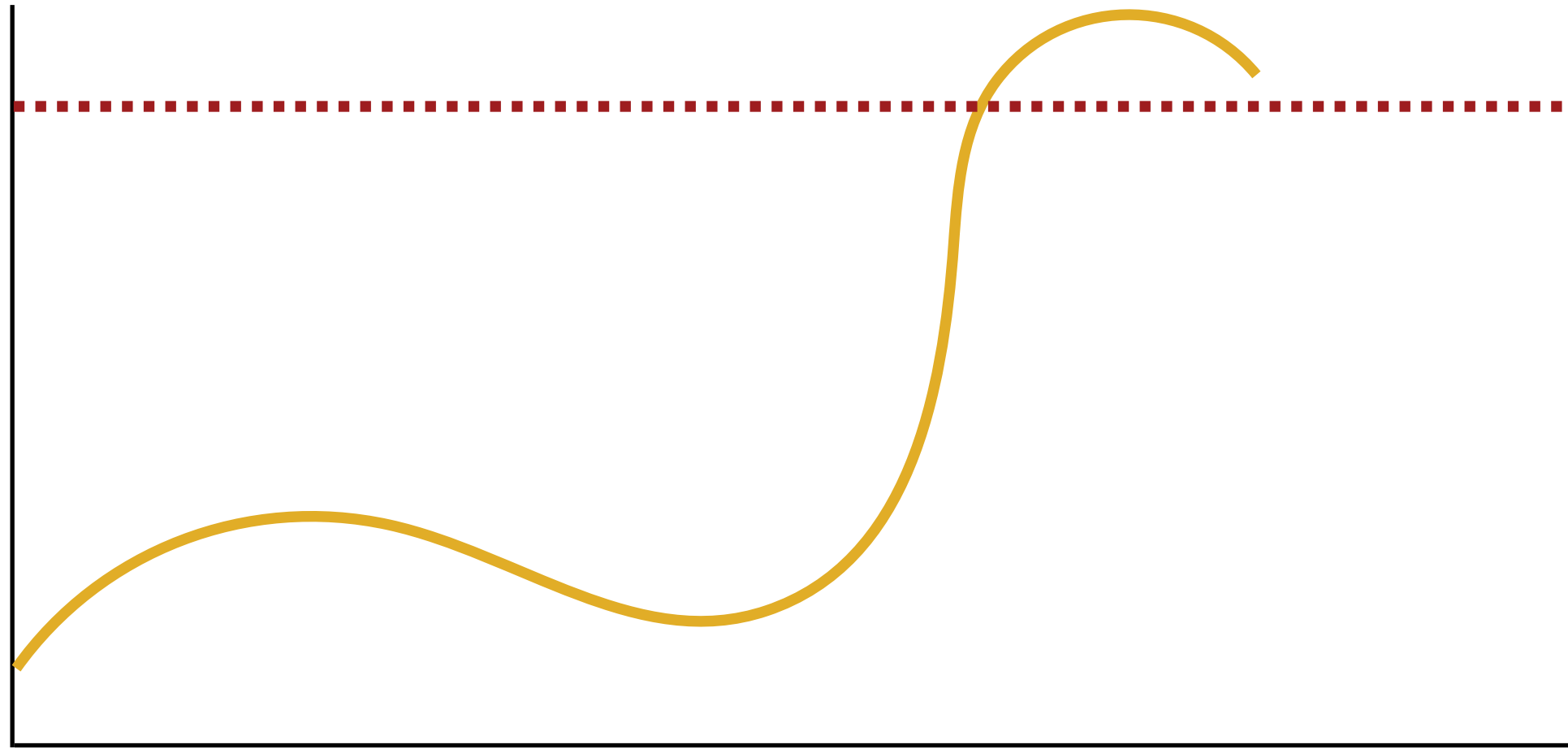
- Wide variety of products



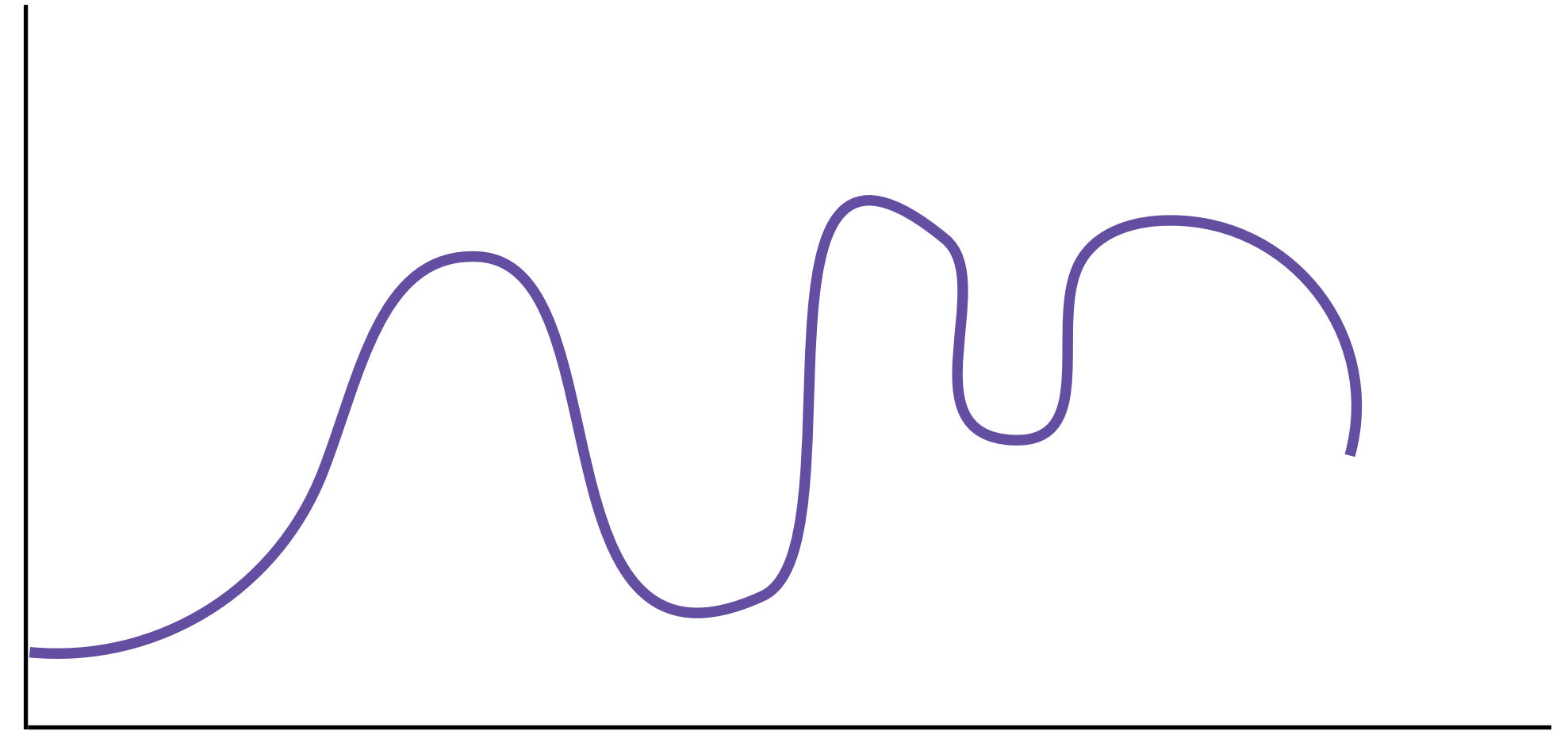
super important metric



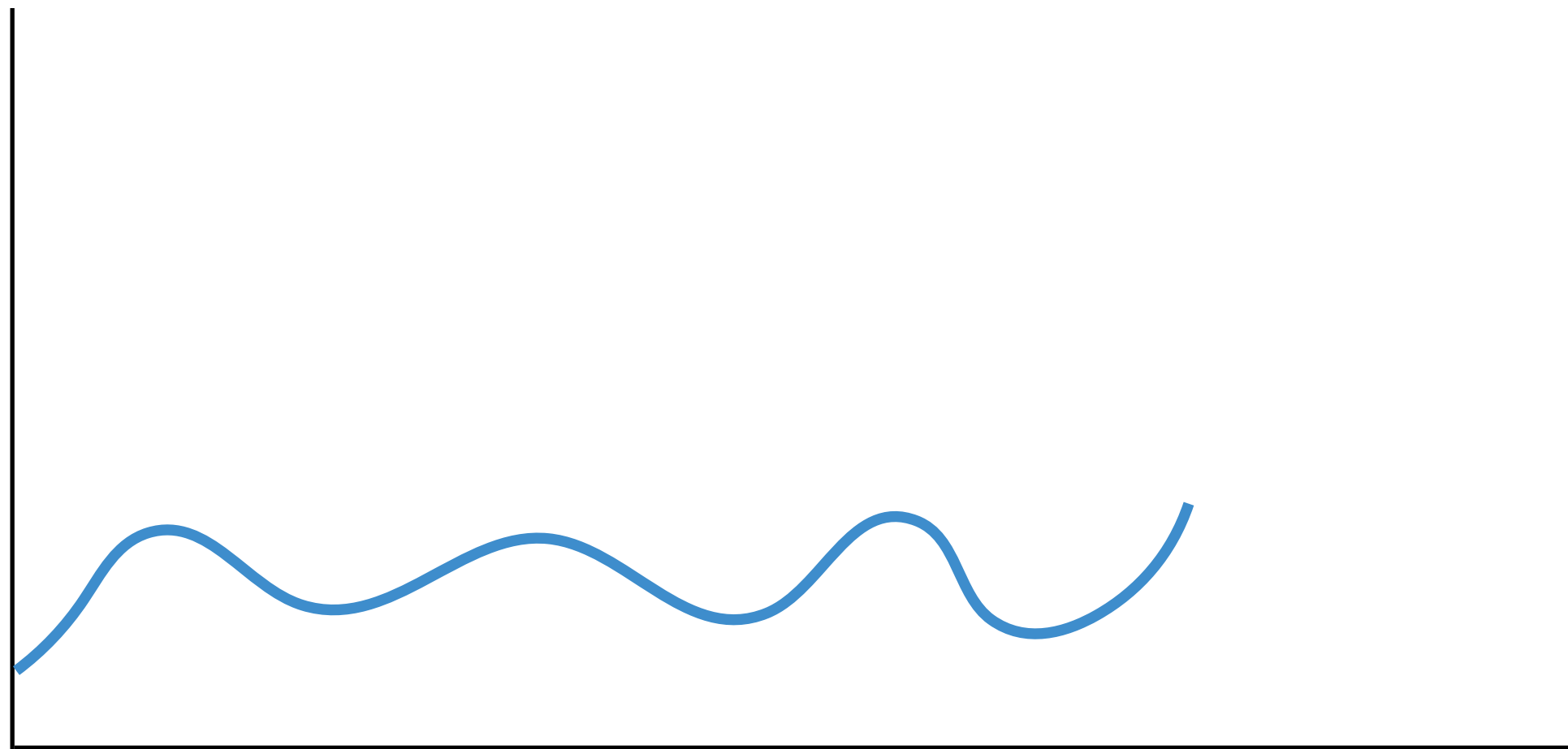
super important metric



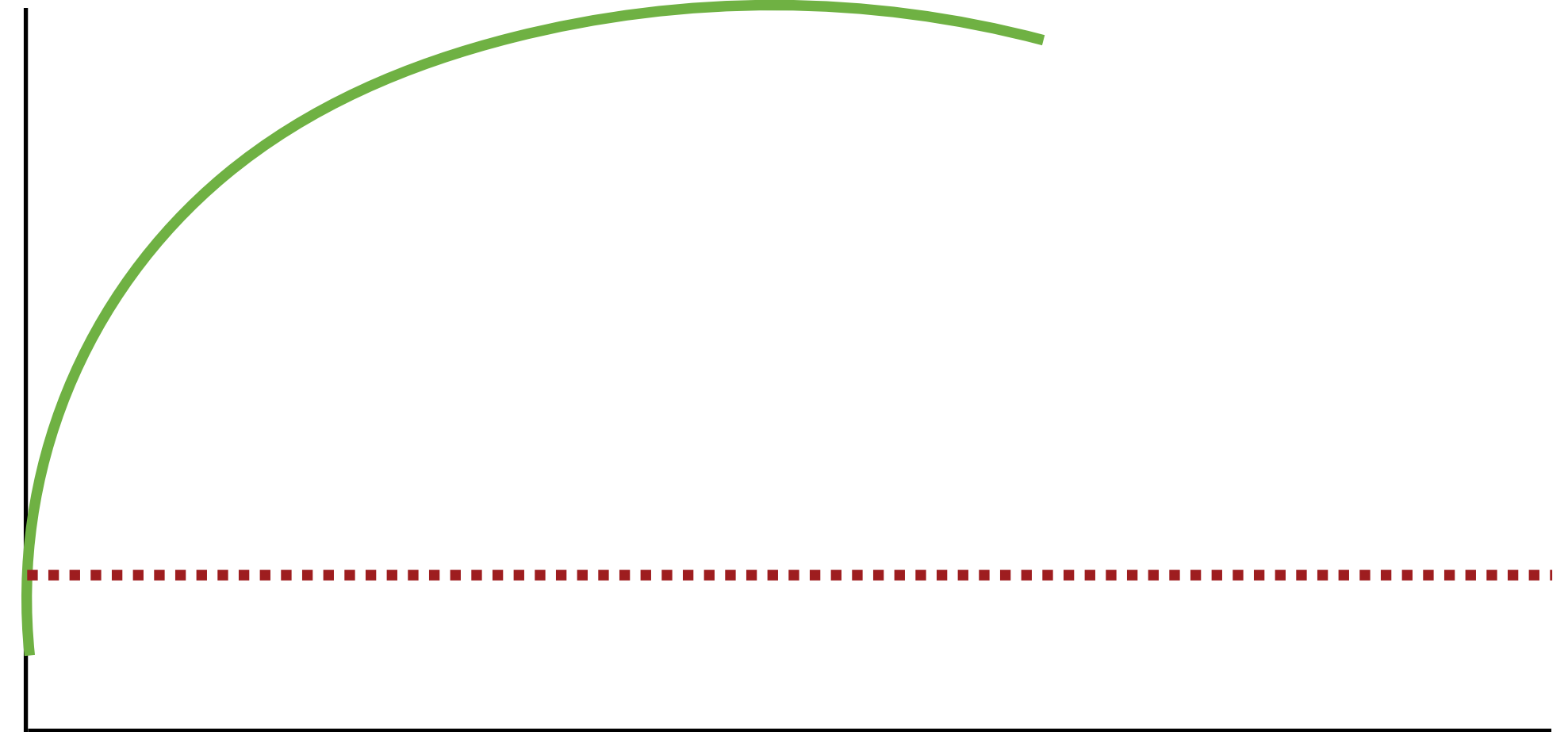
super important metric



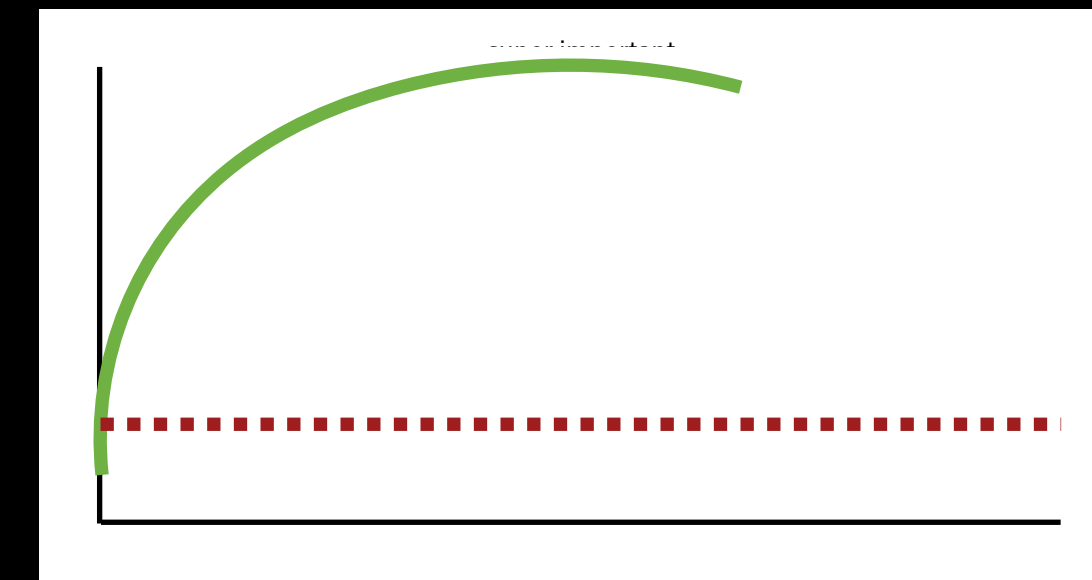
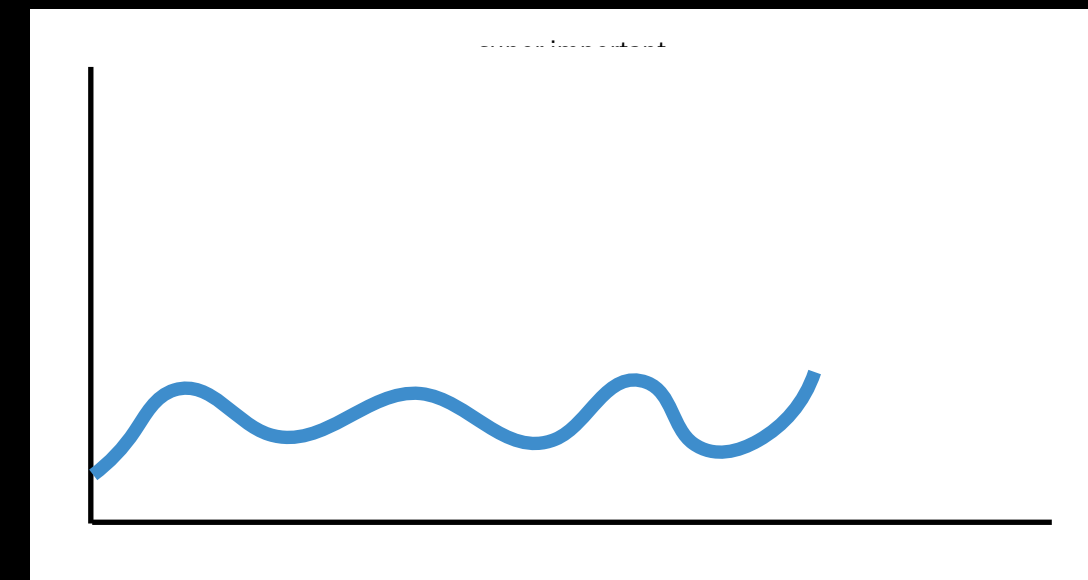
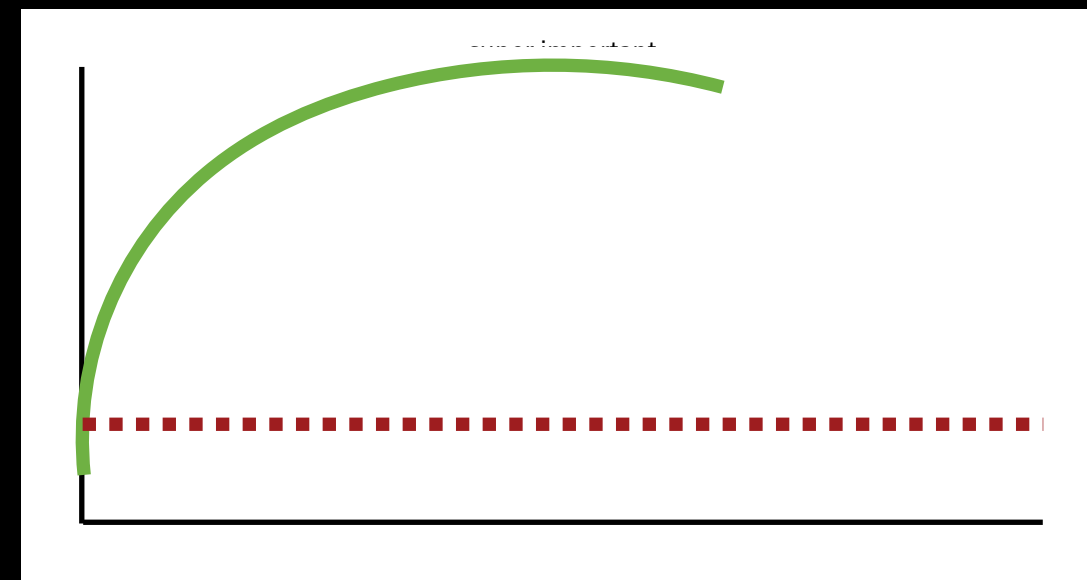
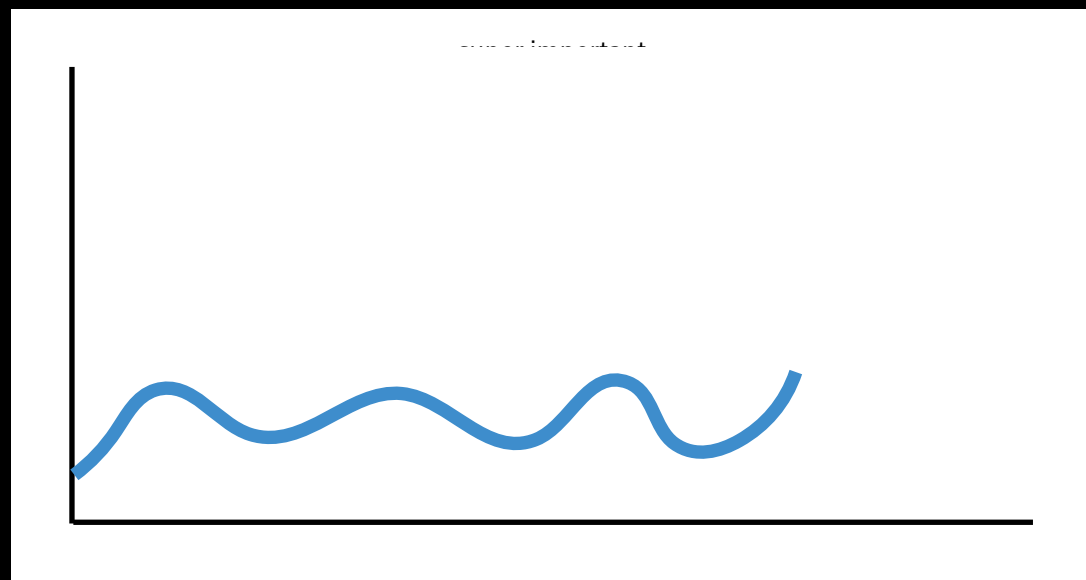
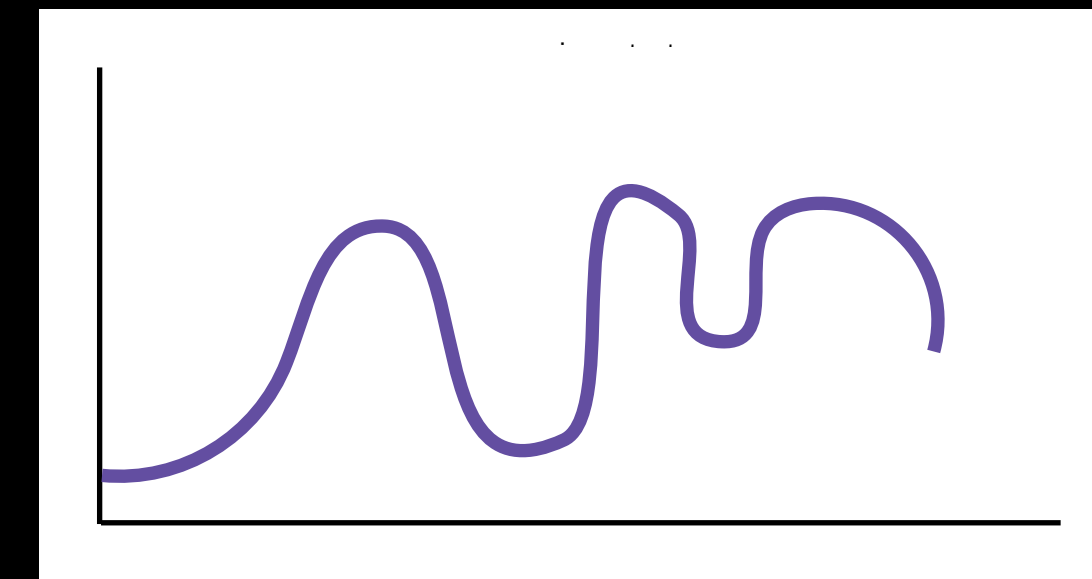
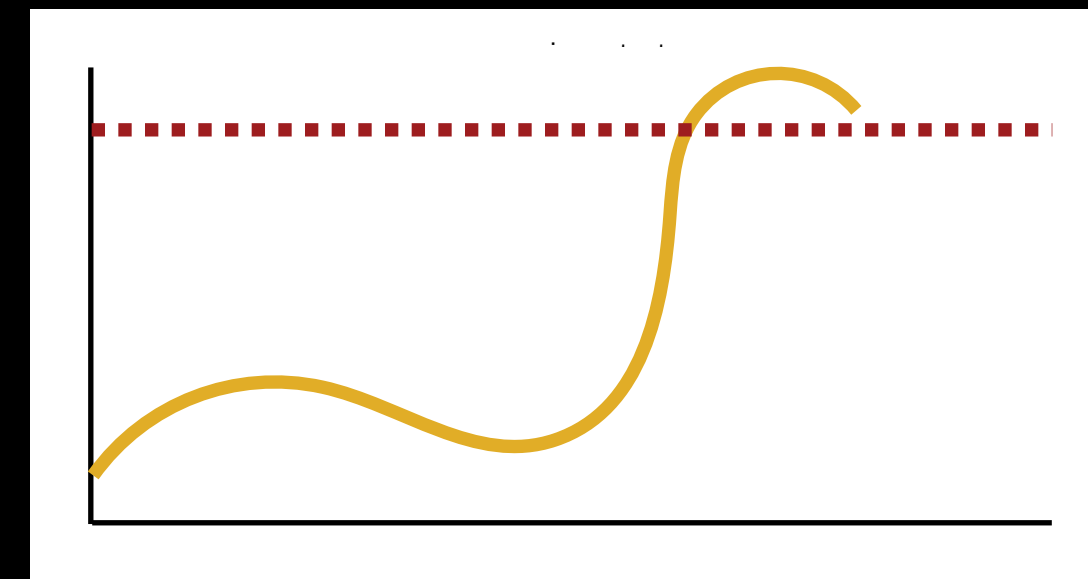
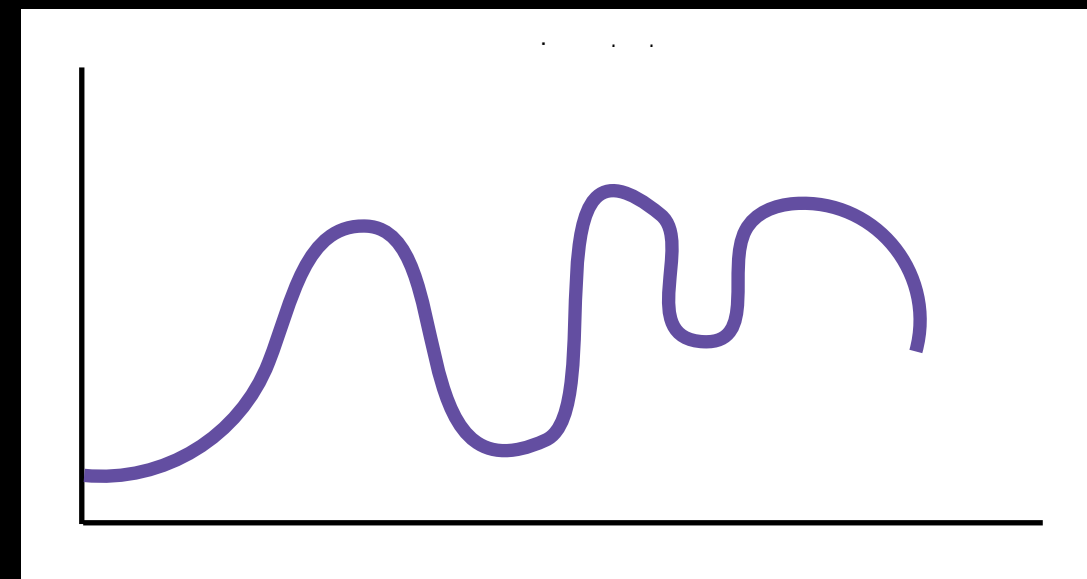
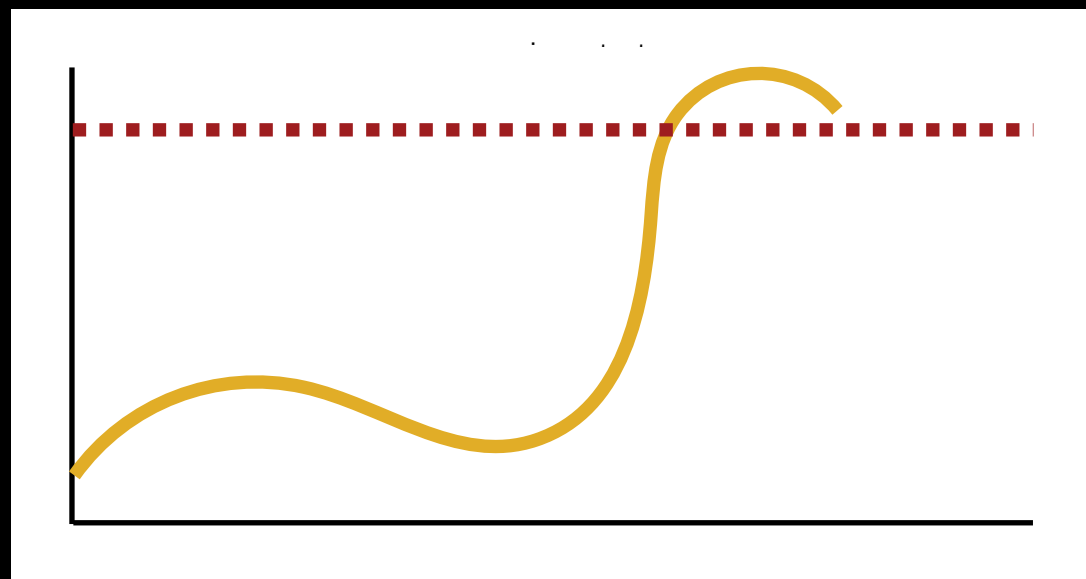
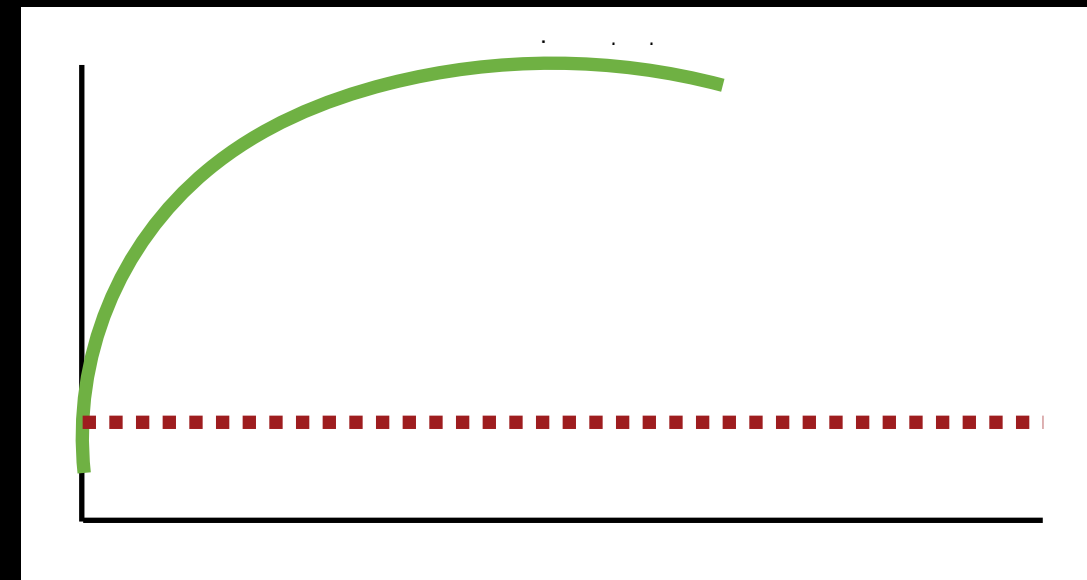
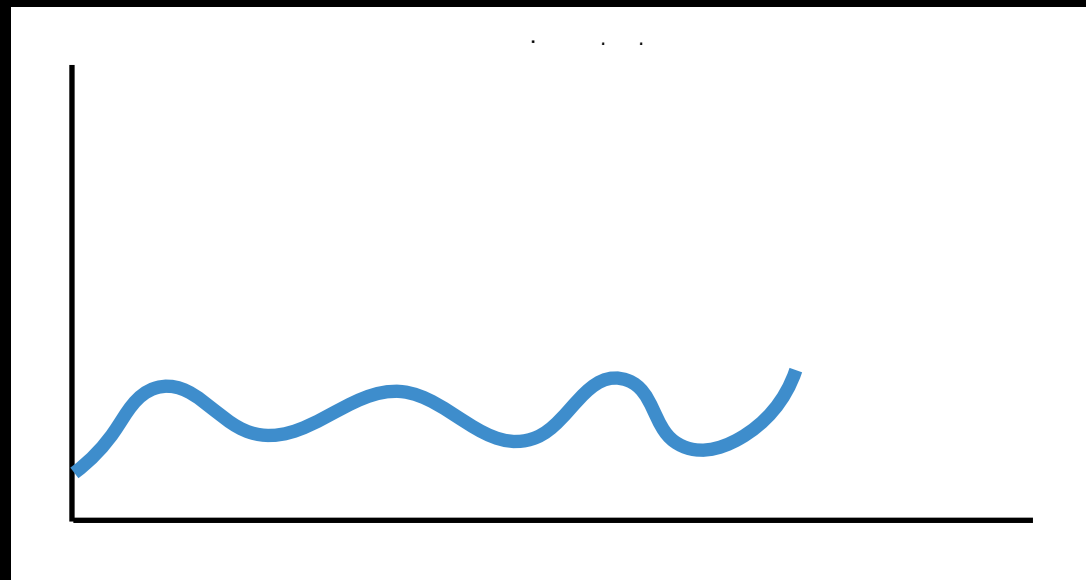
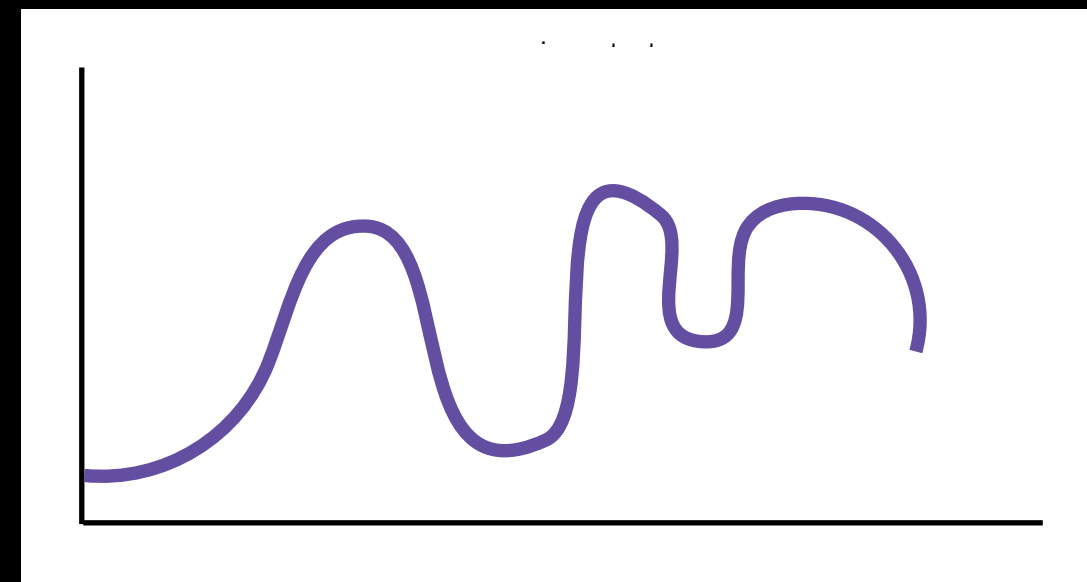
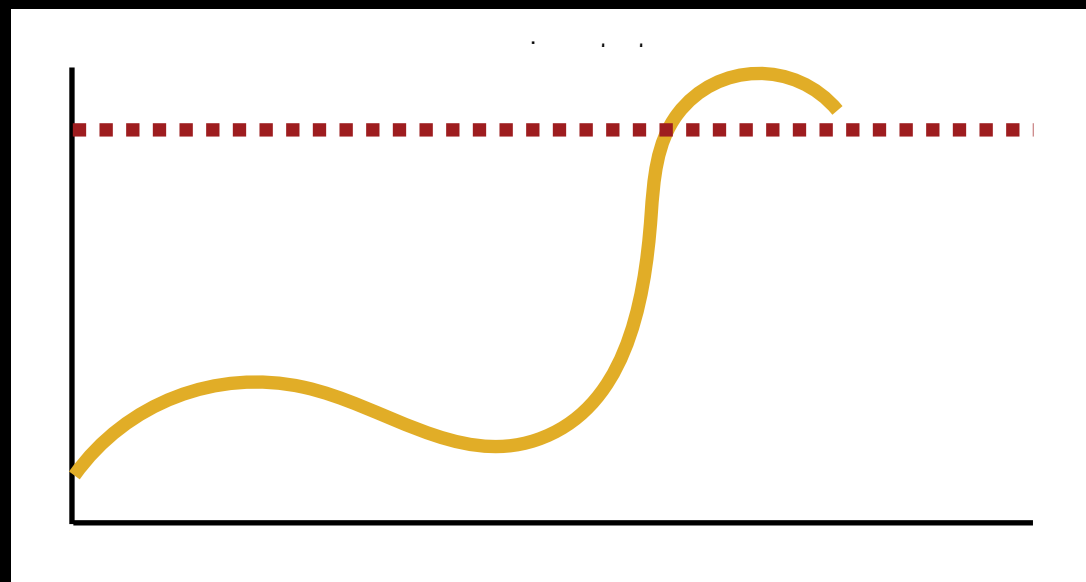
super important metric



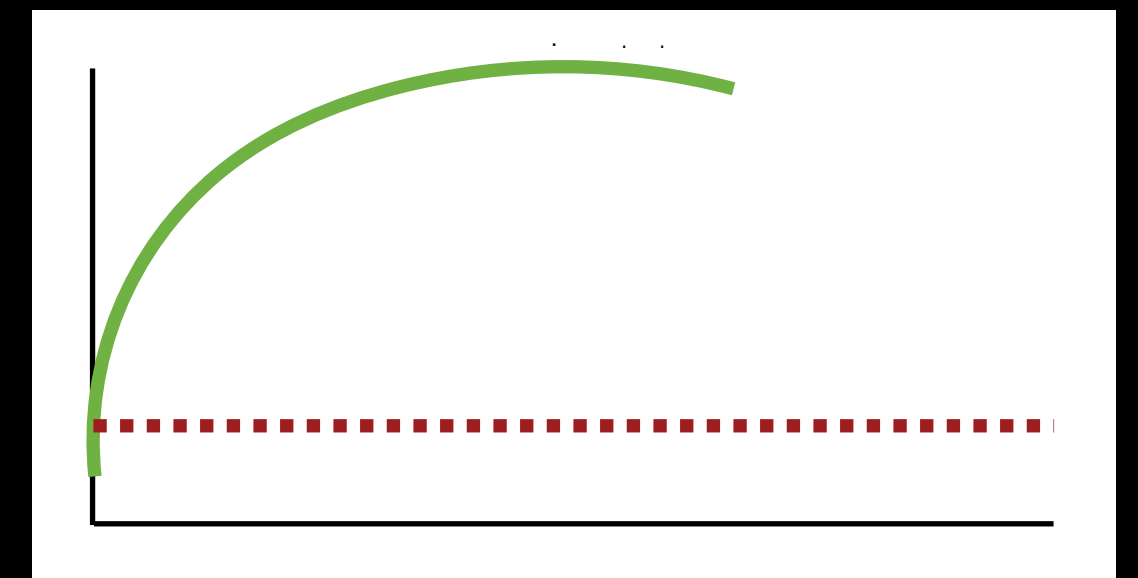
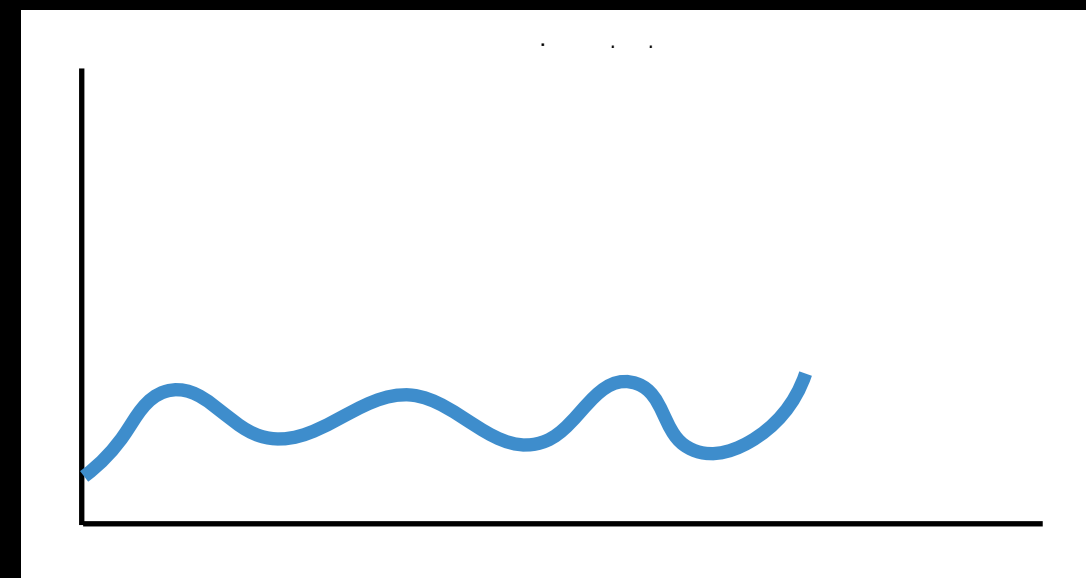
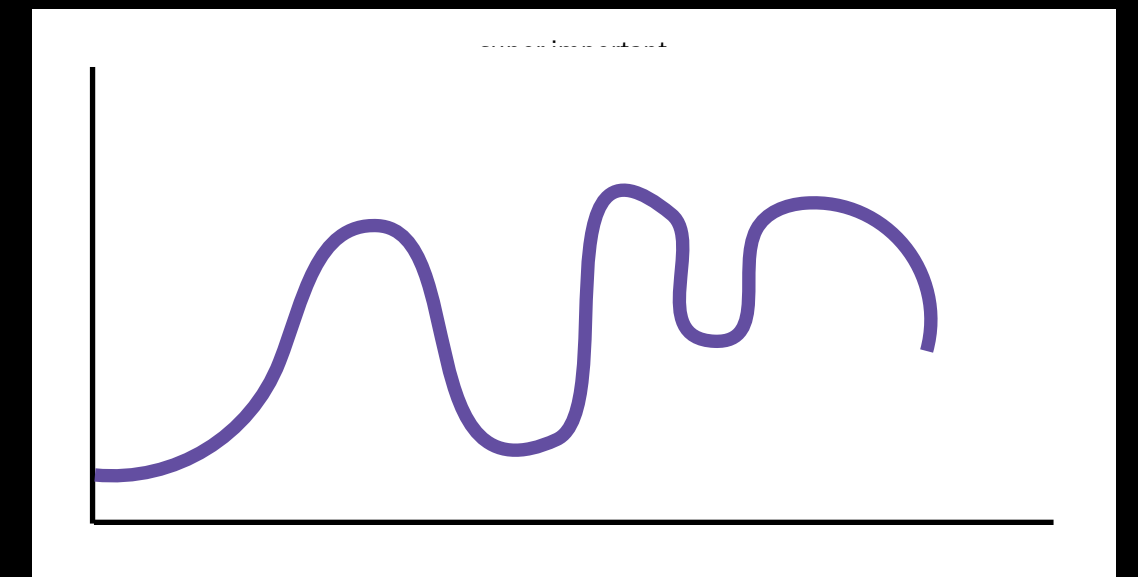
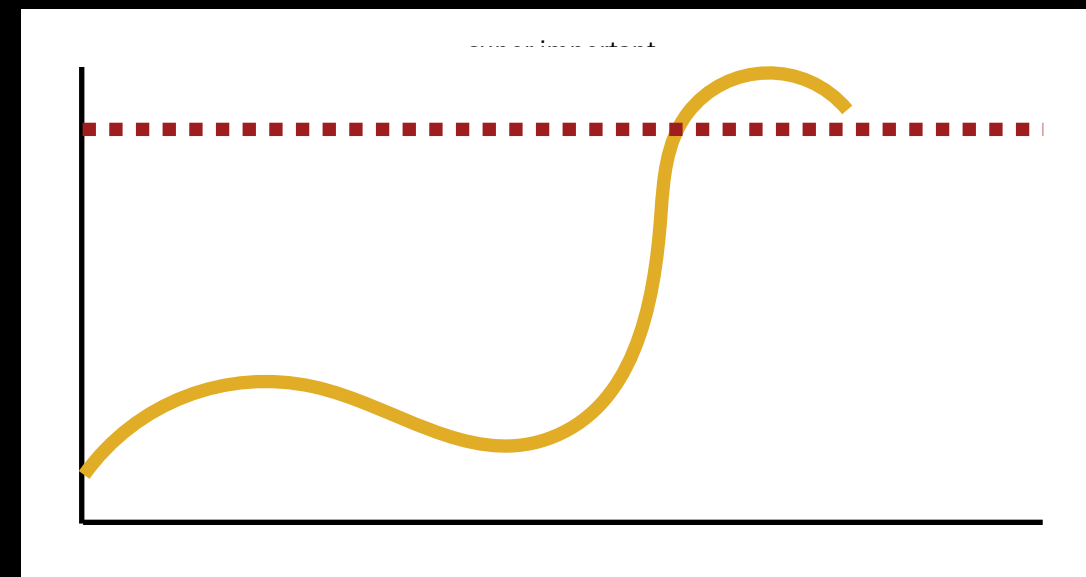
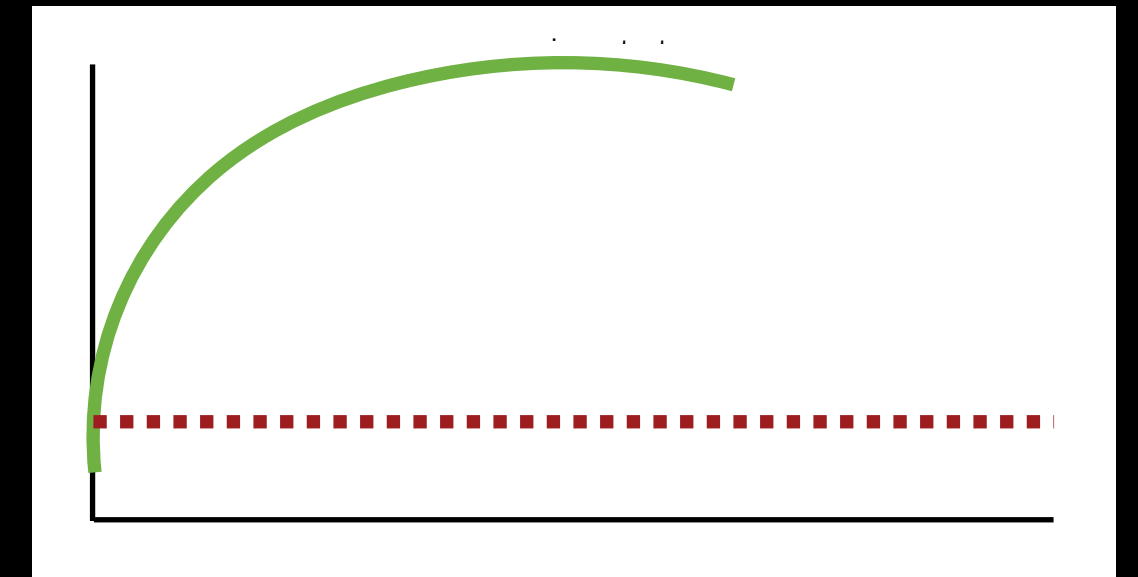
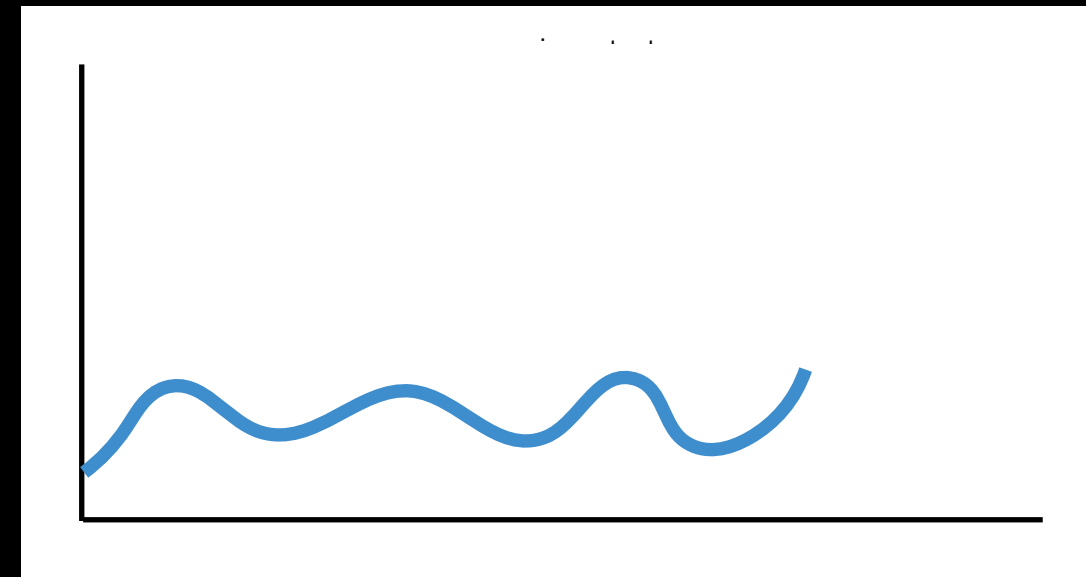
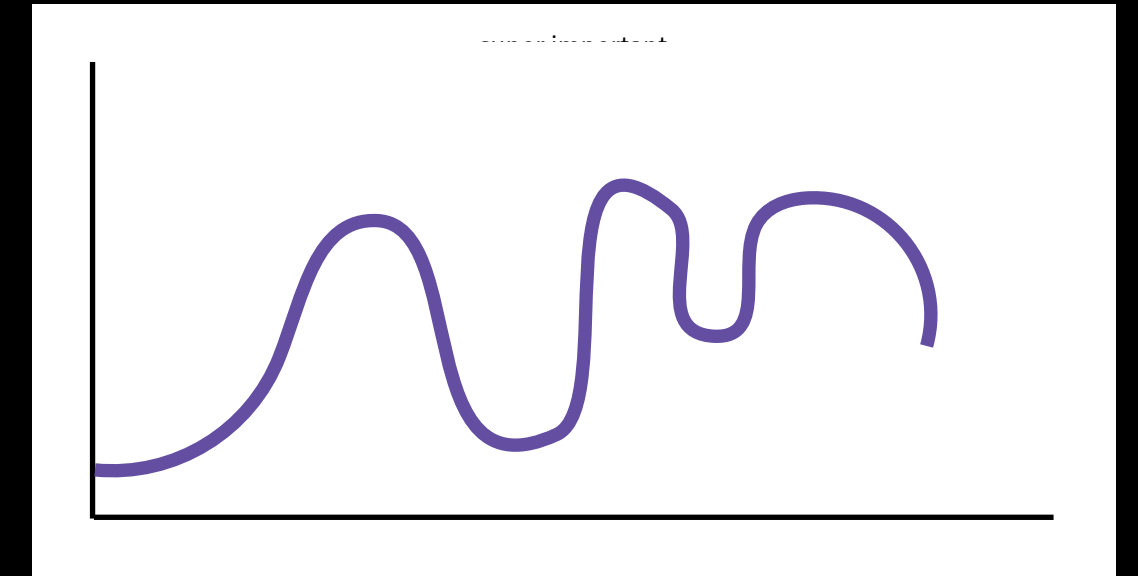
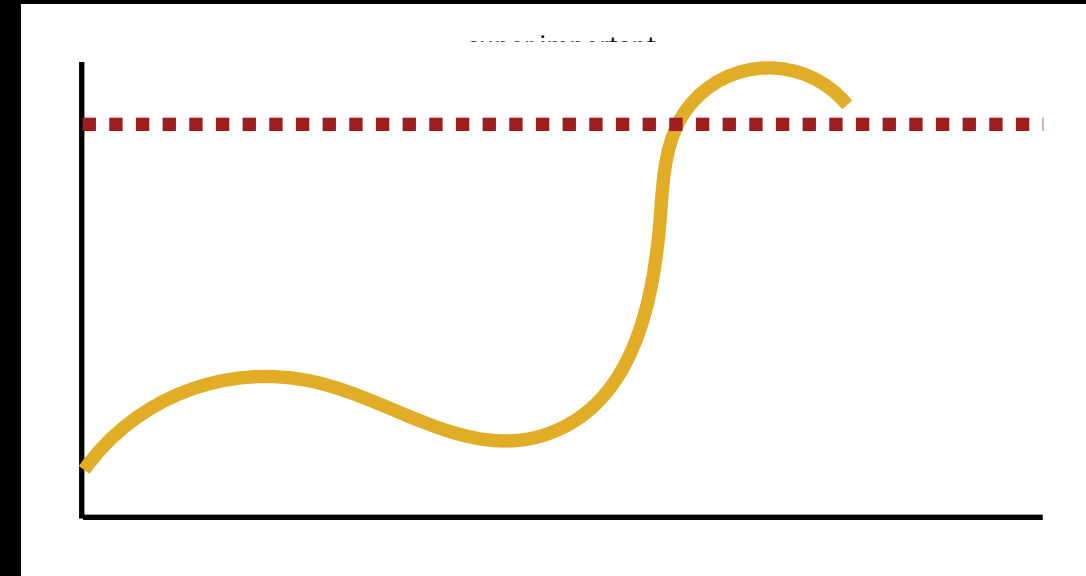
super important metric



“operational visibility”

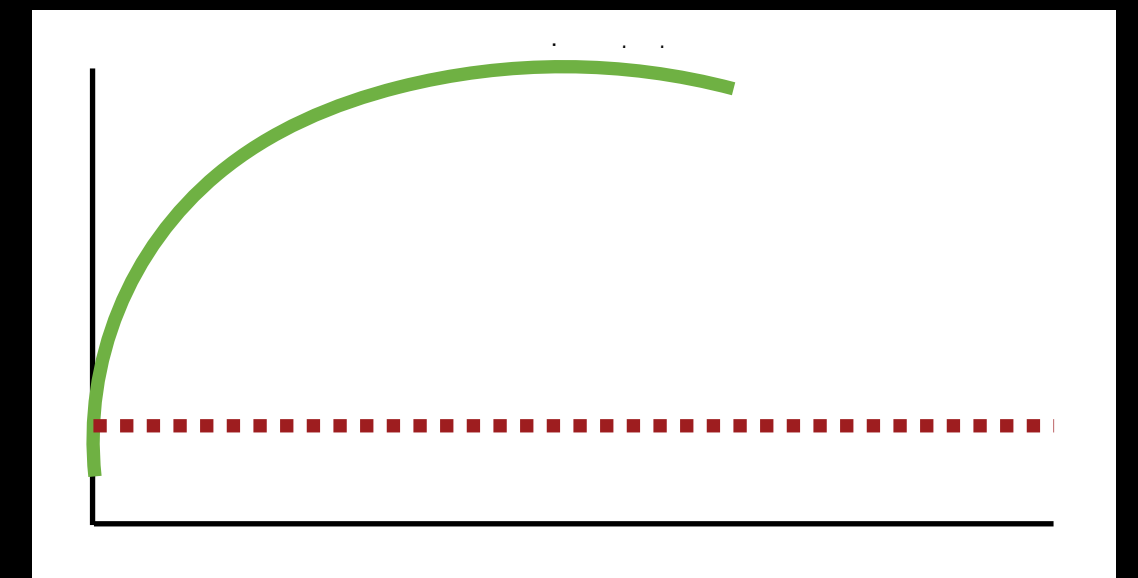
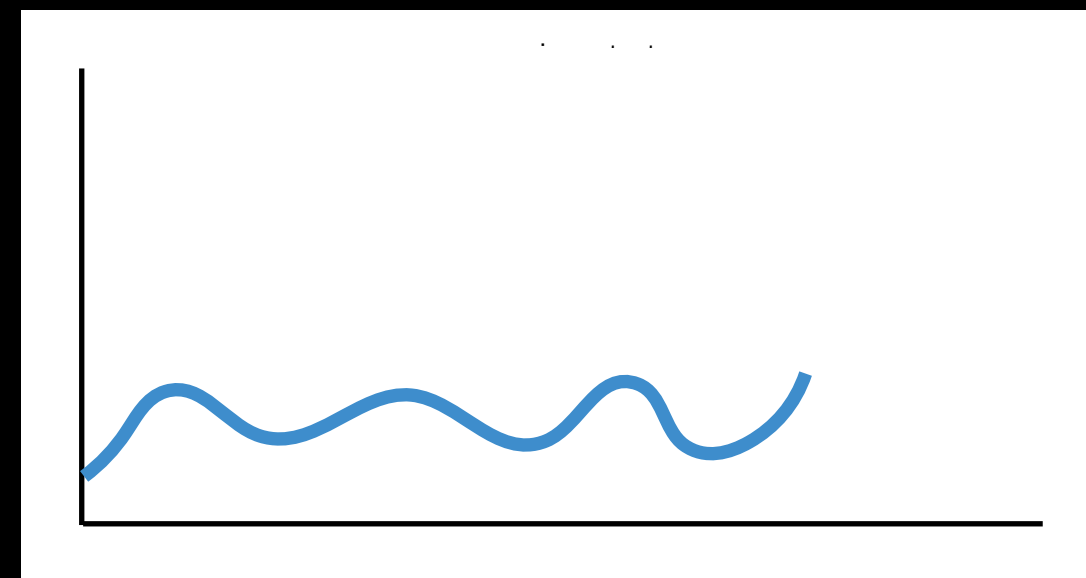
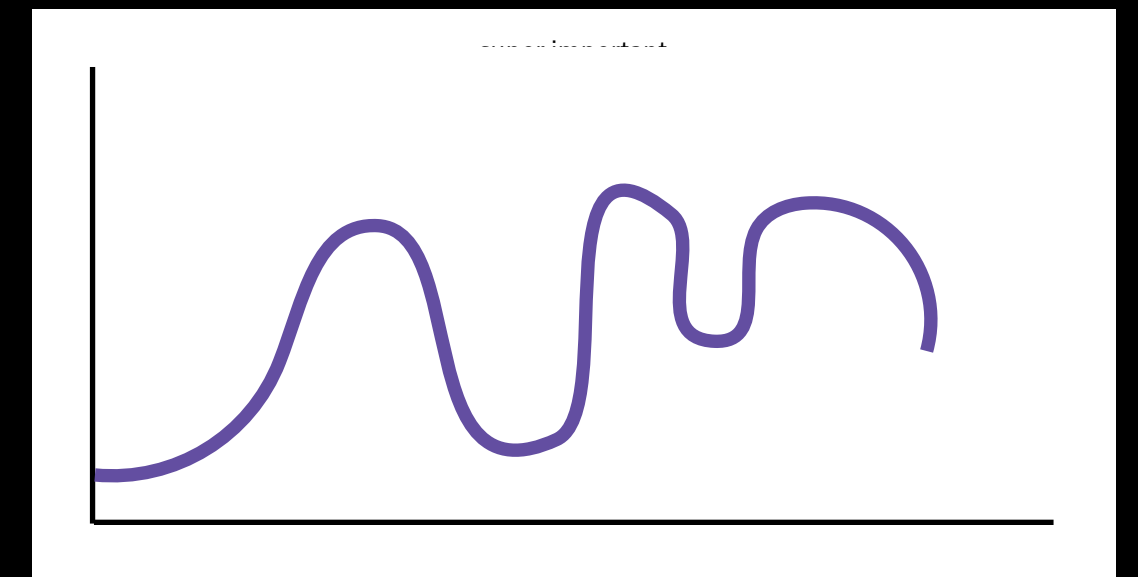
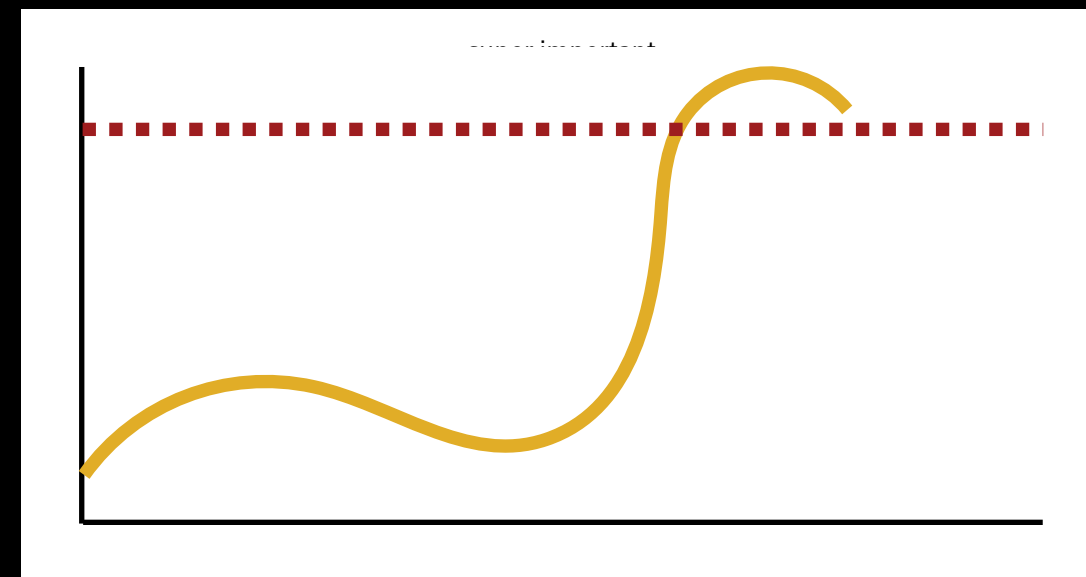
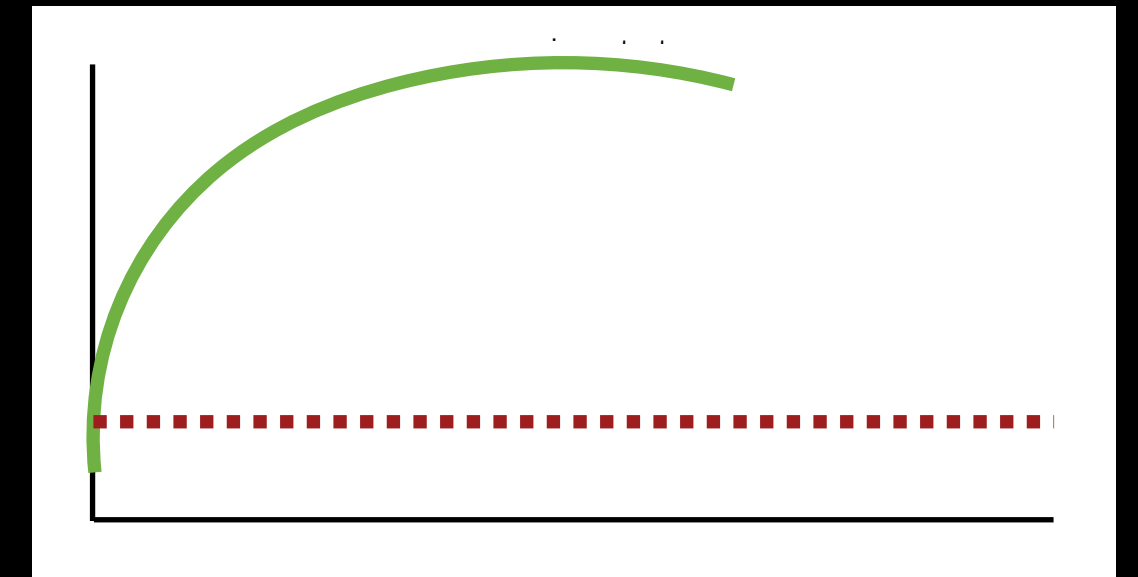
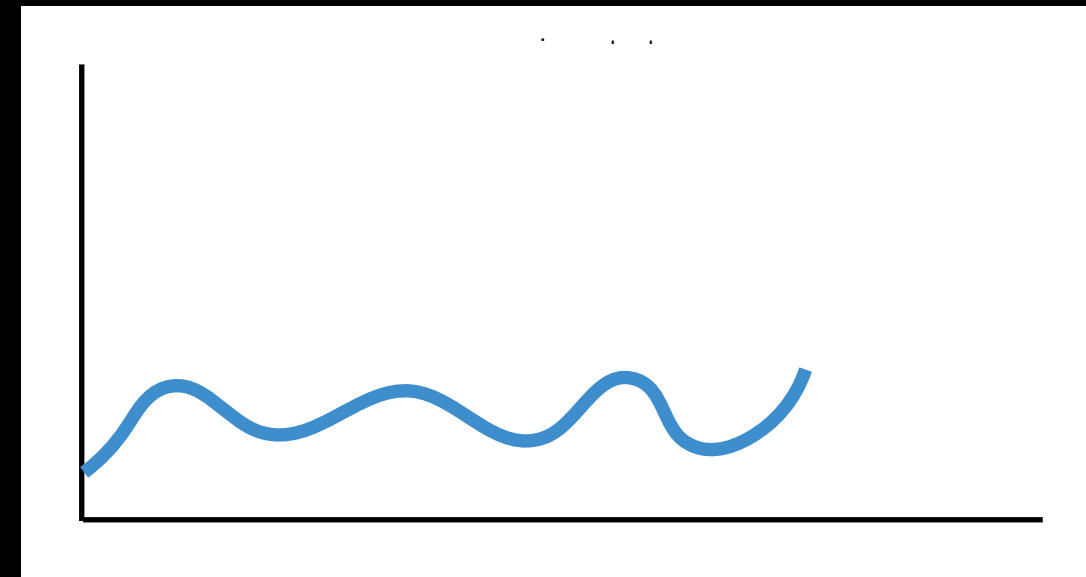
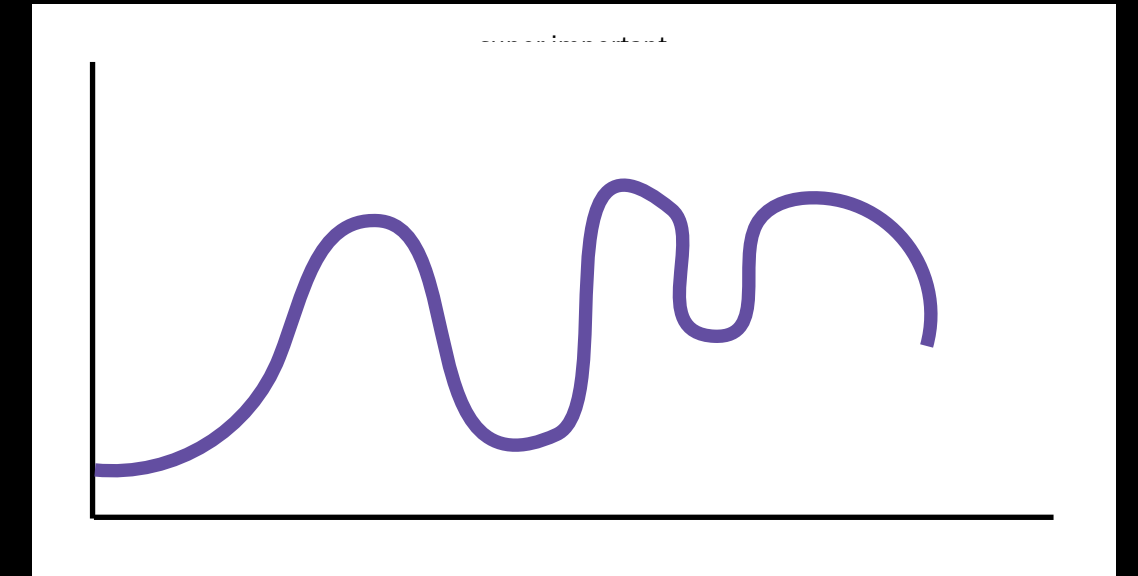
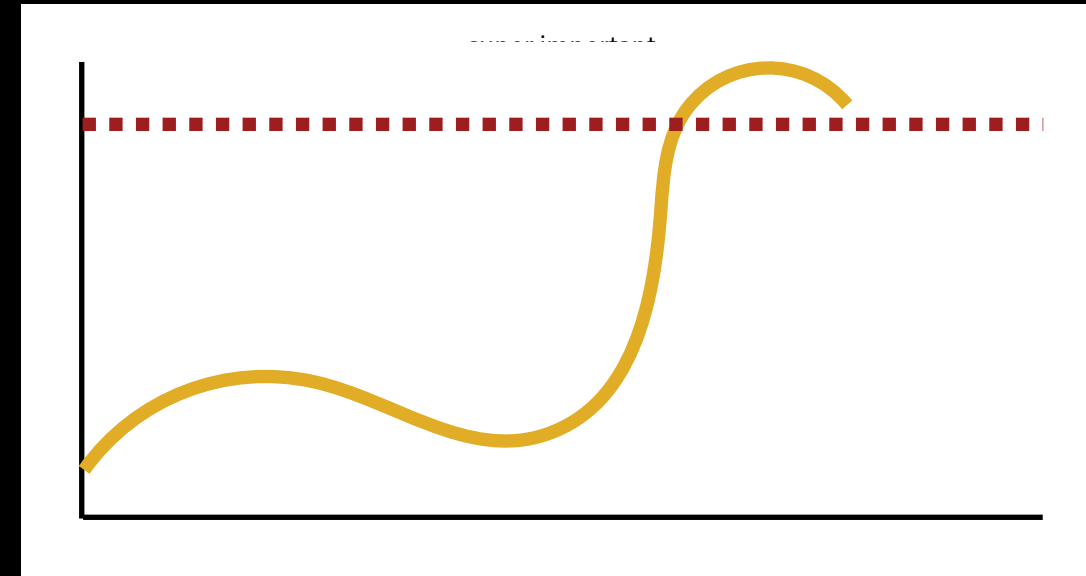


AREAS FOR IMPROVEMENT



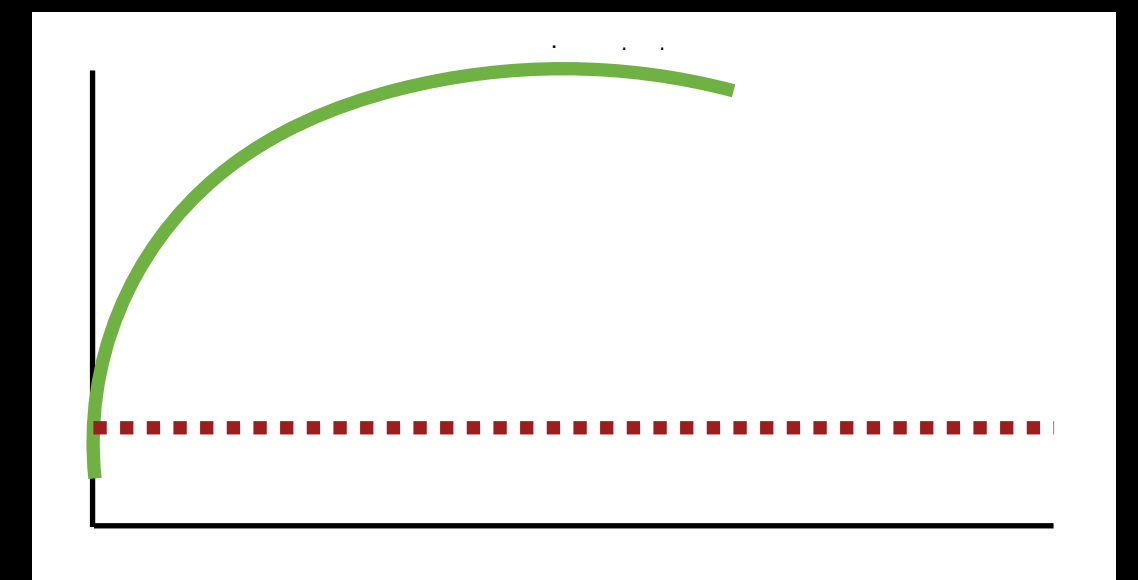
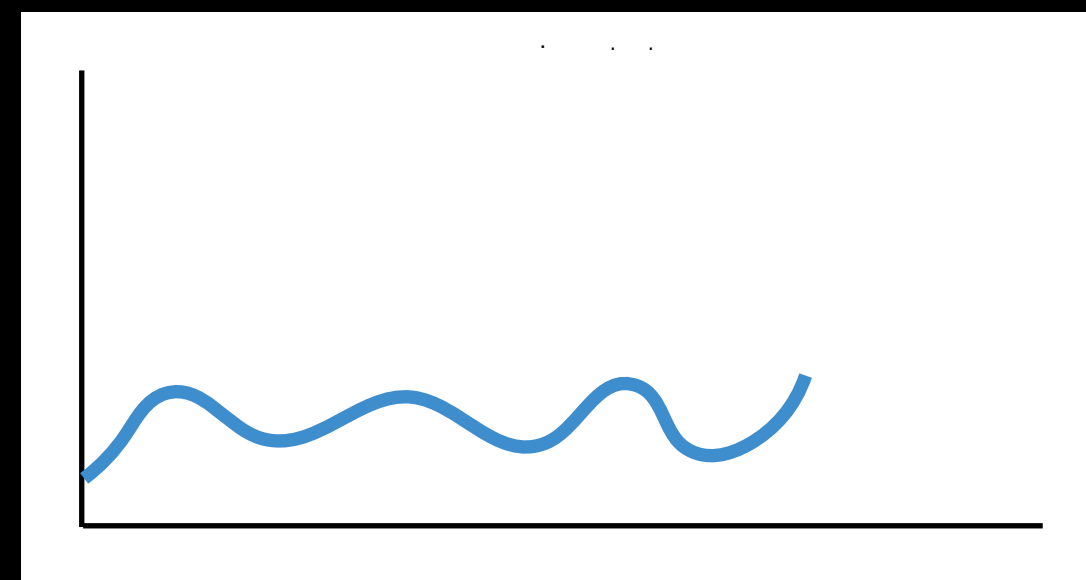
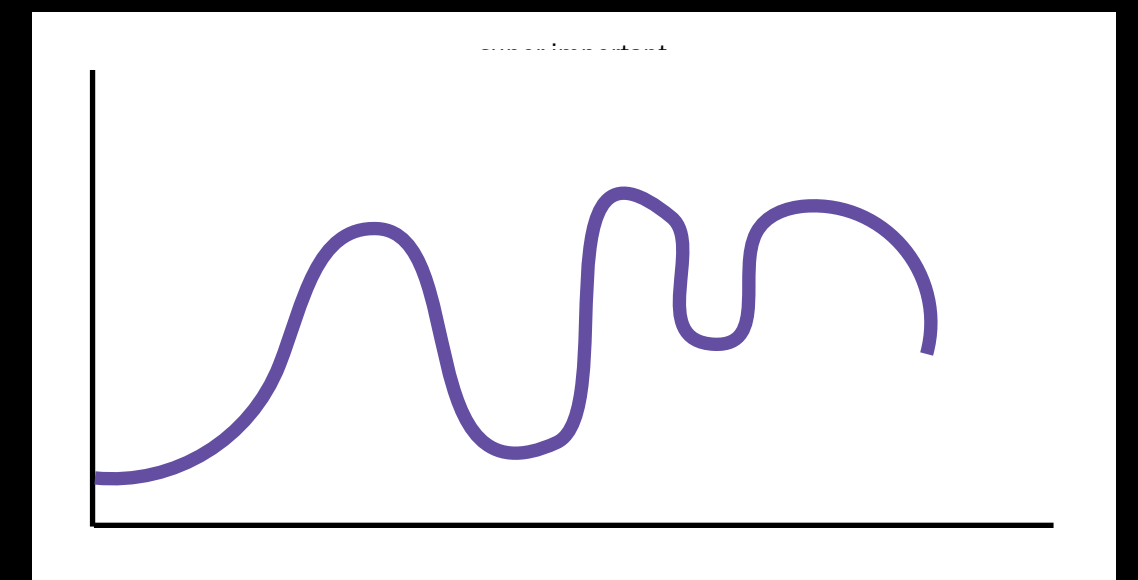
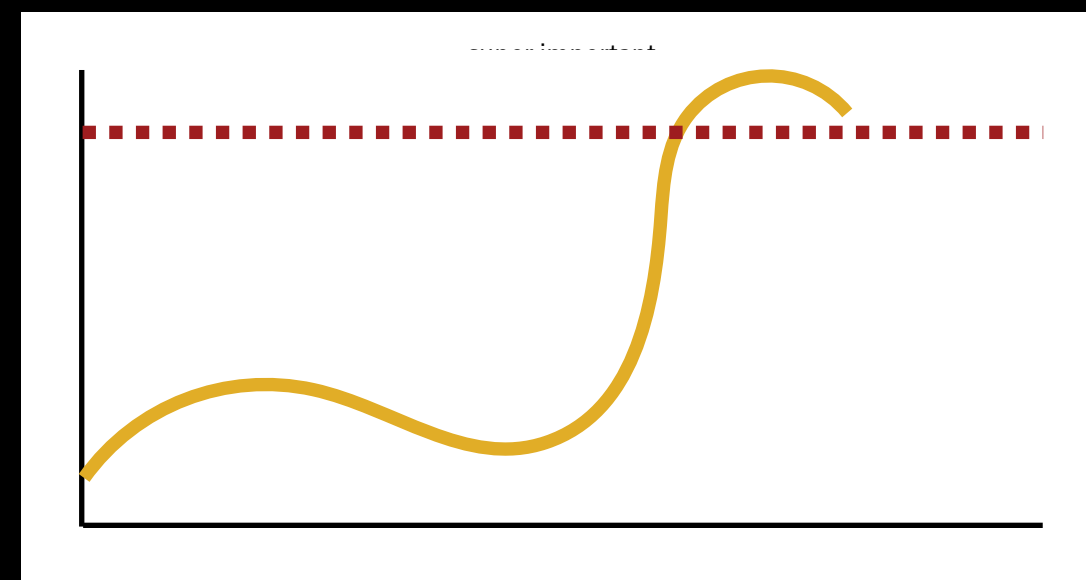
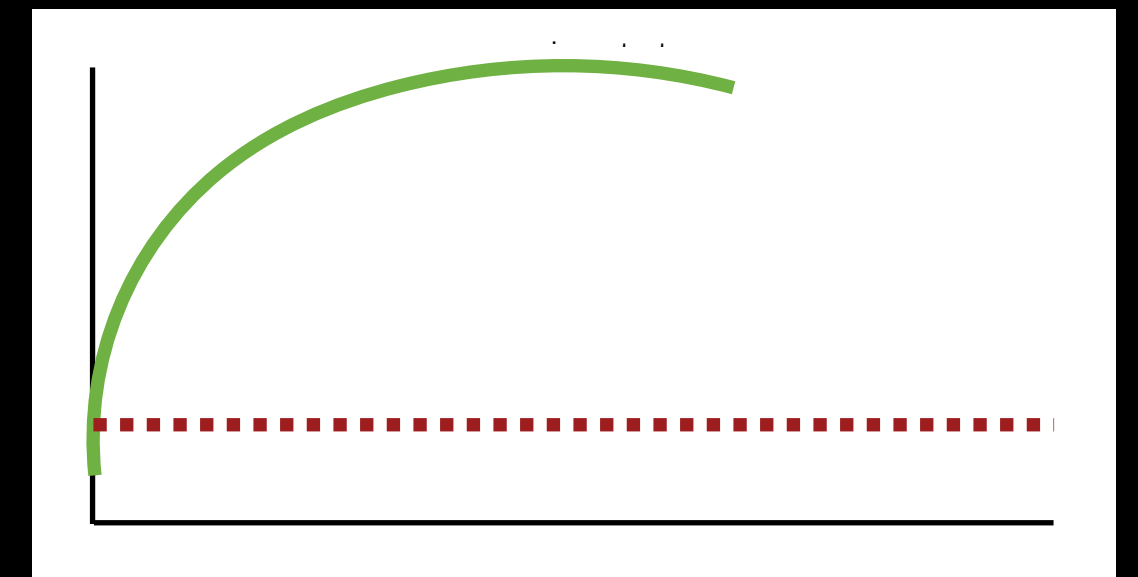
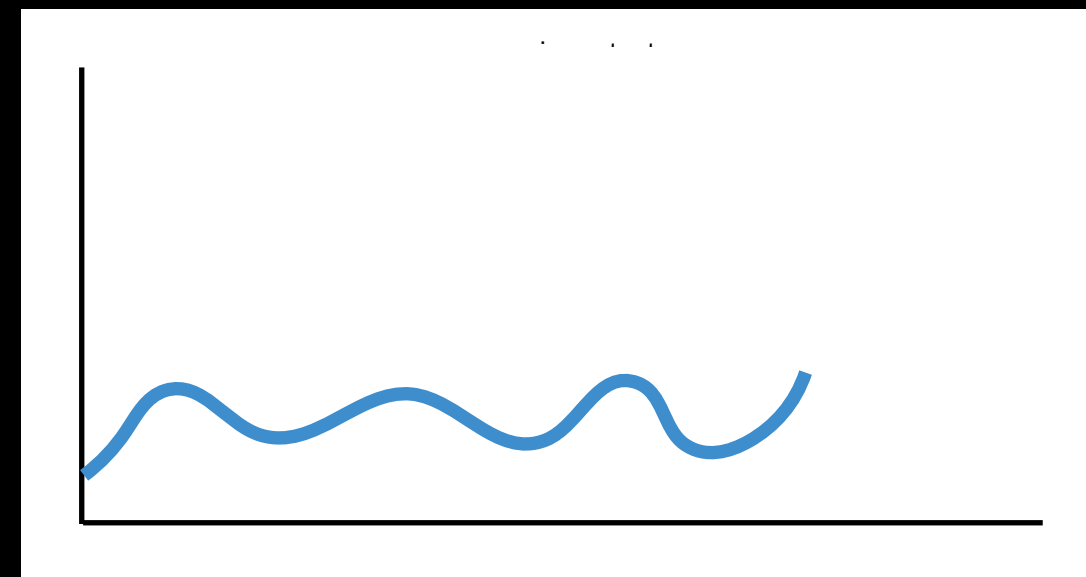
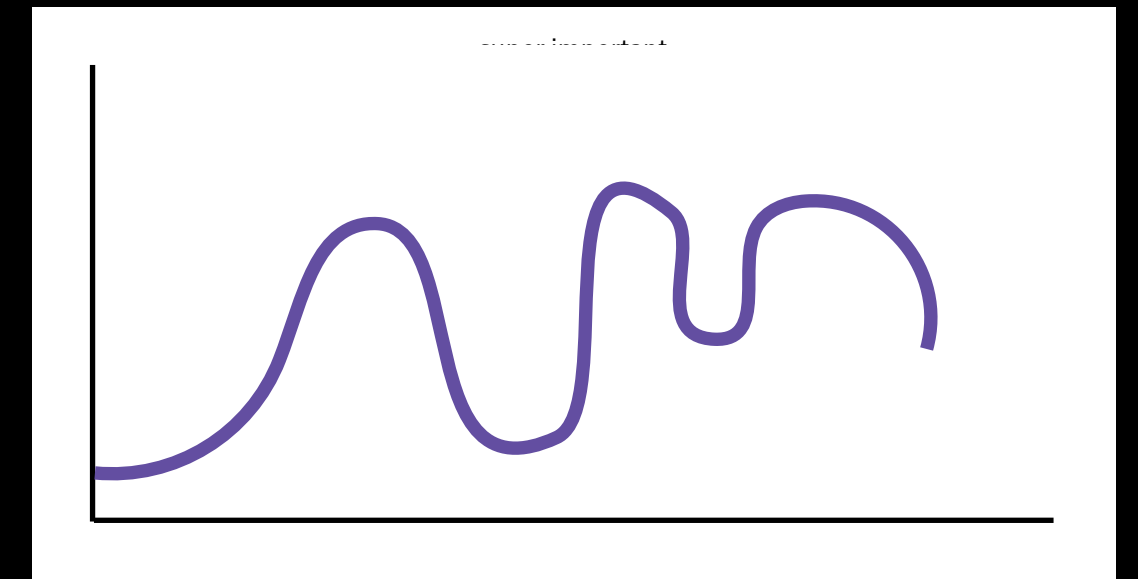
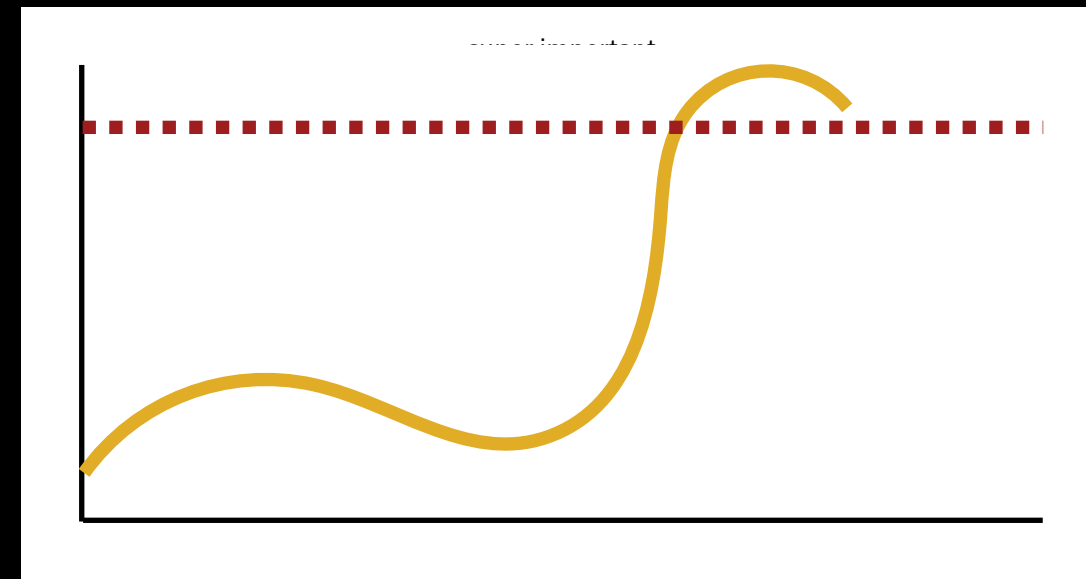
AREAS FOR IMPROVEMENT

✗ Human anomaly detector



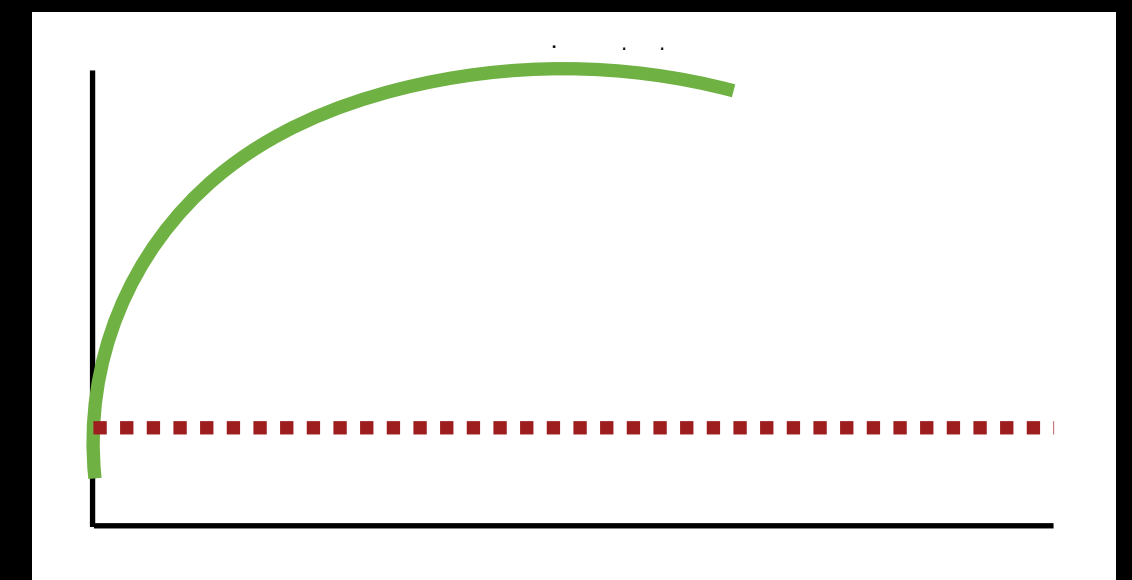
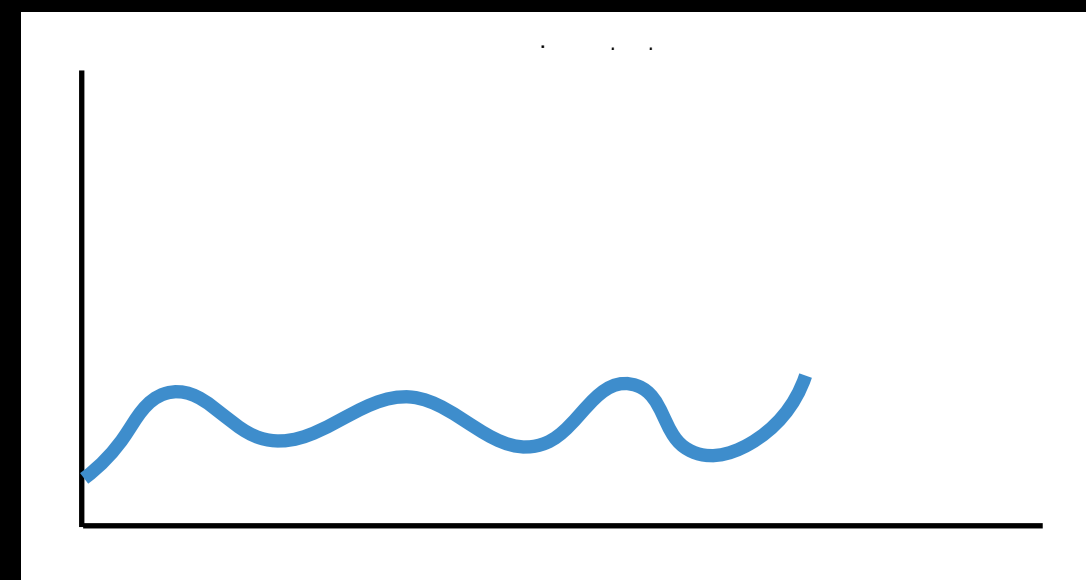
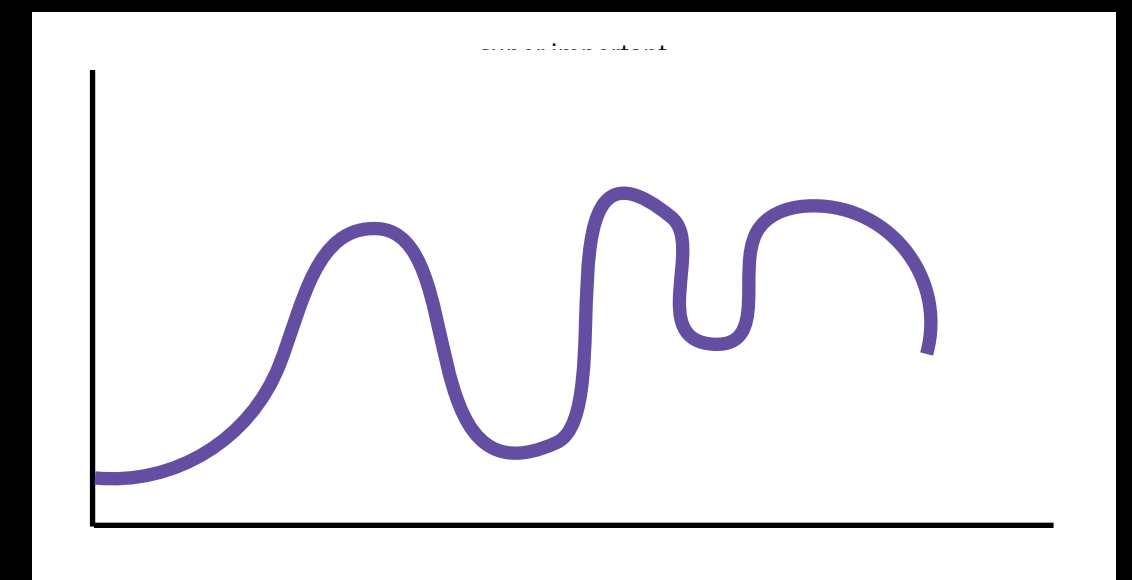
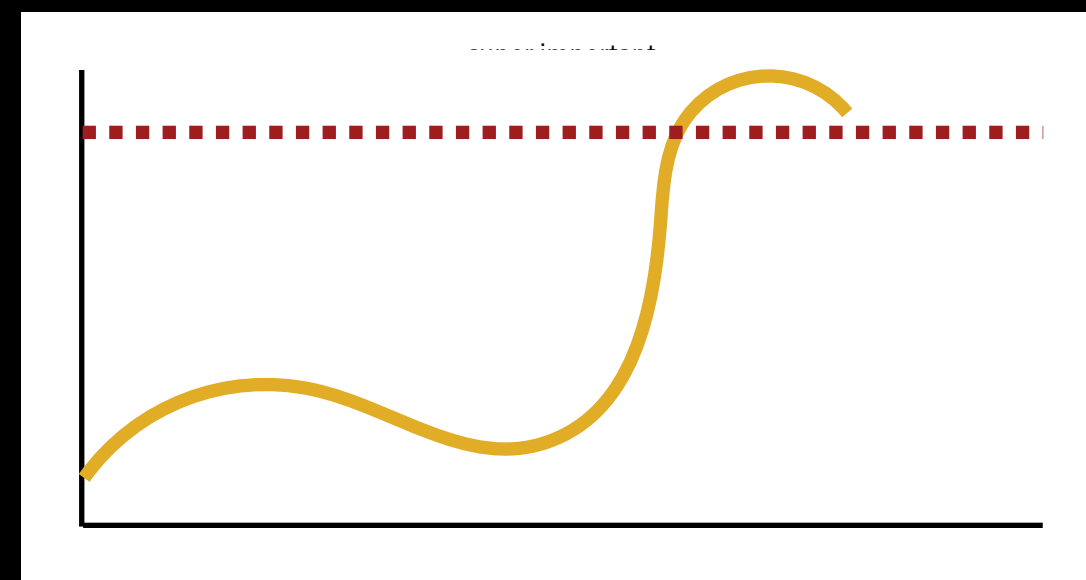
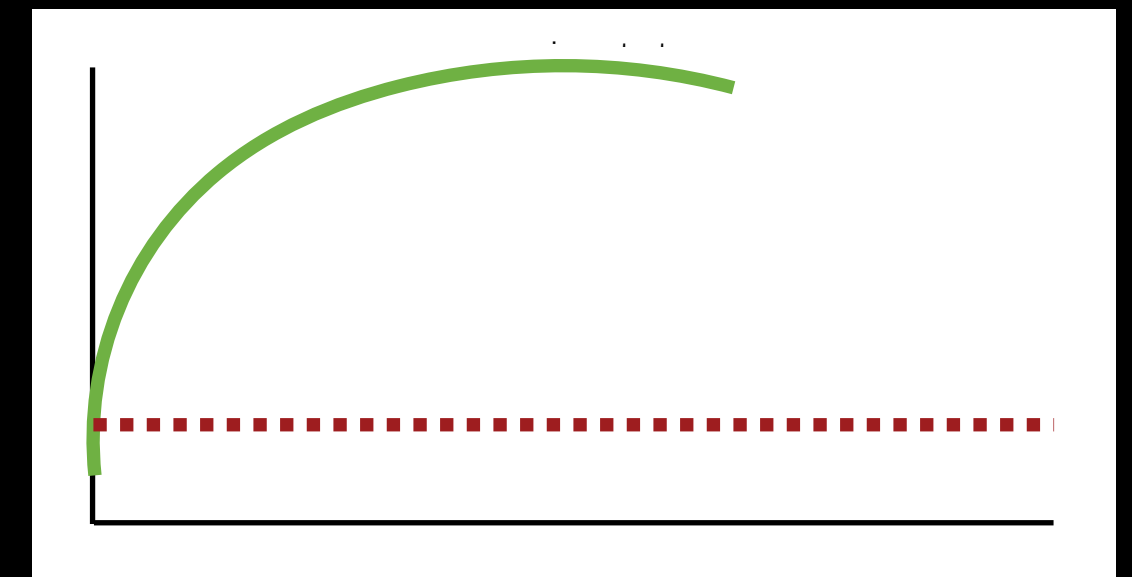
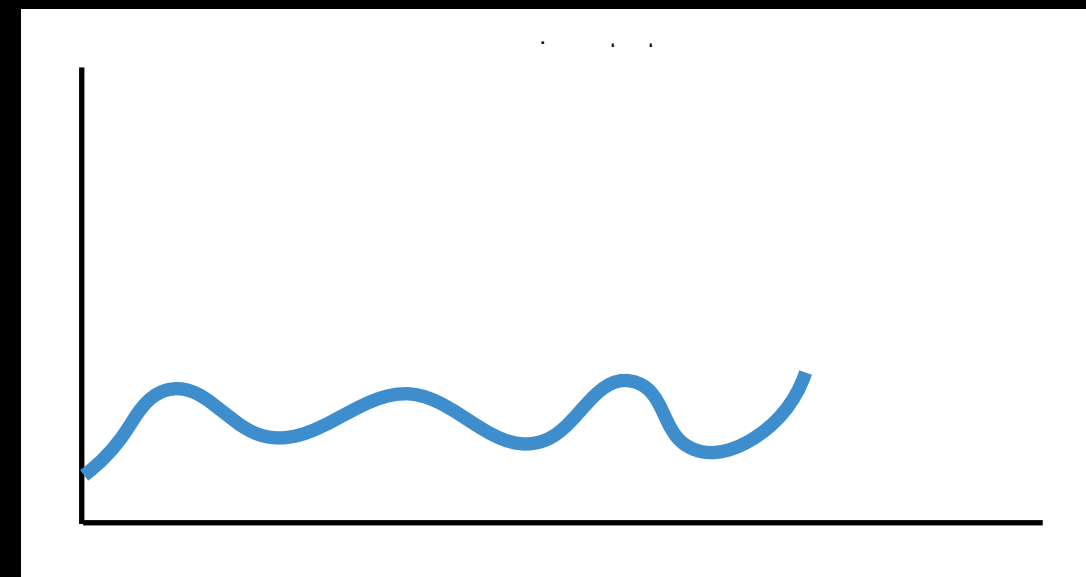
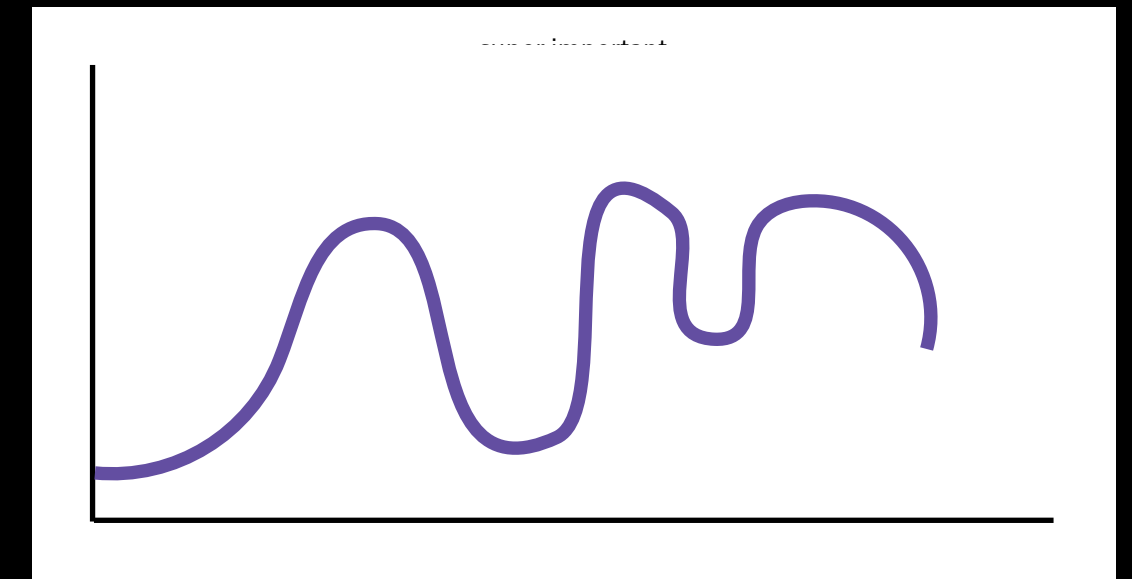
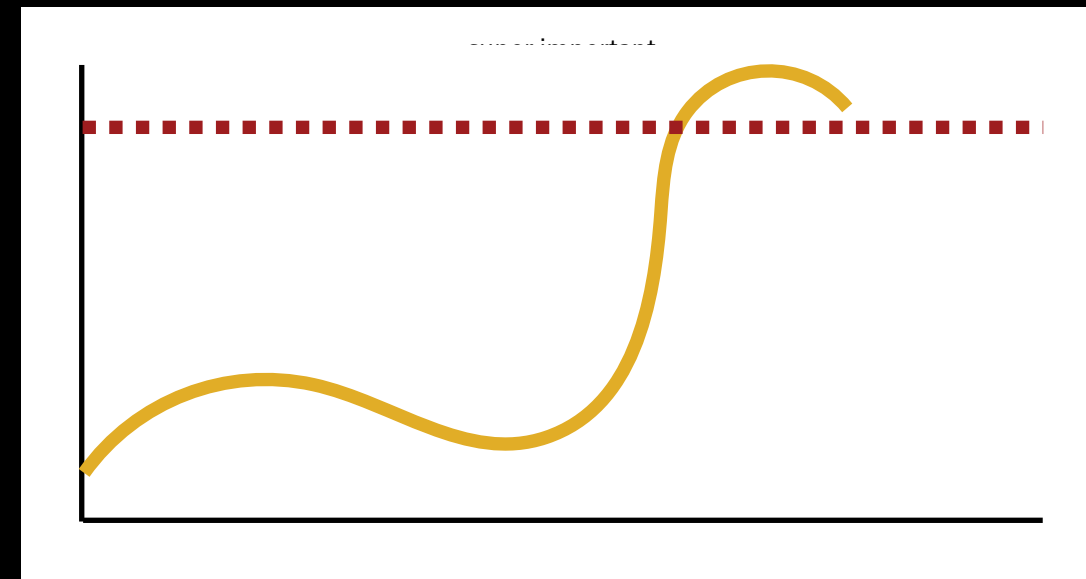
AREAS FOR IMPROVEMENT

- ✗ Human anomaly detector
- ✗ Correlation is awkward



AREAS FOR IMPROVEMENT

- ✗ Human anomaly detector
- ✗ Correlation is awkward
- ✗ Copious data, low fidelity



HOW ARE WE INSTRUMENTING?

- Wide variety of products



HOW ARE WE INSTRUMENTING?

- Wide variety of products
- Log files



HOW ARE WE INSTRUMENTING?

- Wide variety of products
- Log files
- *...more log files*





A close-up photograph of a man with dark hair, wearing a dark blue shirt, holding his hands to his head in a gesture of distress or pain. The background is a plain, light-colored wall with a corkboard visible on the right side. A red rectangular box is overlaid on the image, containing white text that reads "index.js:65 500 Server Error".

`index.js:65 500 Server Error`

A close-up photograph of a man with dark hair, wearing a blue button-down shirt. He is covering his face with both hands, with his fingers spread, suggesting a state of frustration, stress, or despair. The background is a plain, light-colored wall with a corkboard visible on the right side.

`index.js:65` **it's broke fam**

AREAS FOR IMPROVEMENT



AREAS FOR IMPROVEMENT

- ✗ Developers write bad logs



AREAS FOR IMPROVEMENT

- ✗ Developers write bad logs
- ✗ Logs lack context



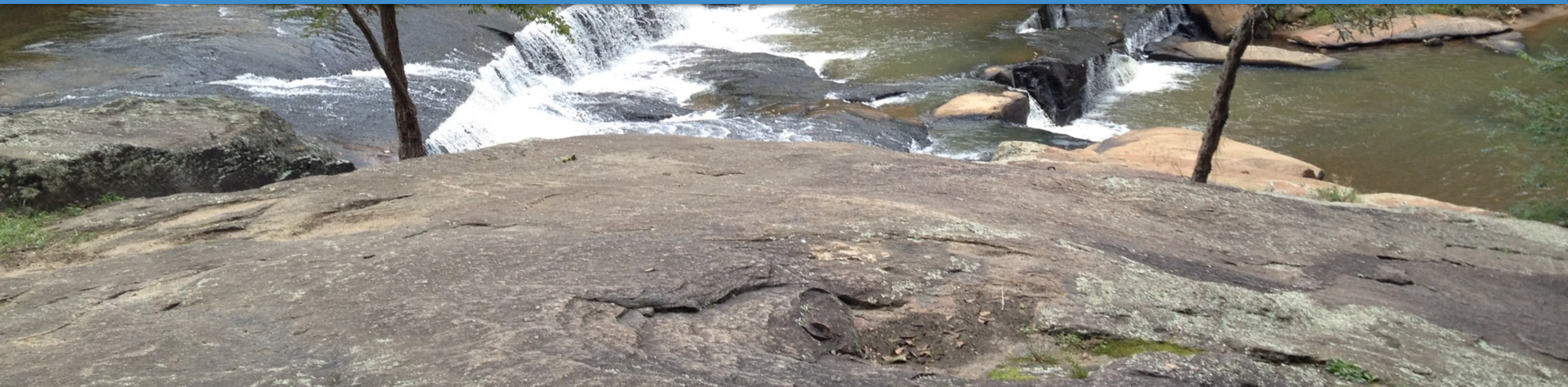
AREAS FOR IMPROVEMENT

- ✗ Developers write bad logs
- ✗ Logs lack context
- ✗ Text logs lack fidelity





WHY STREAM PROCESSING?



PROGRAMMABILITY GIVES

FULL FLEXIBILITY
OVER MONITORING
BEHAVIOR



HIGHER FIDELITY DATA GIVES A

RICH SET OF
DIMENSIONS



INCREMENTAL PROCESSING GIVES

PERFORMANCE AND
SCALE



INCREMENTAL PROCESSING GIVES

PERFORMANCE AND
SCALE

RIEMANN



<http://riemann.io>



“Sonata can capture 95% of all traffic pertaining to the query, while reducing the overall data rate by a factor of about 400 and the number of required counters by four orders of magnitude.”

NETWORK MONITORING AS A STREAMING ANALYTICS PROBLEM
GUPTA, ET AL, HOTNETS'16

CHALLENGES

CHALLENGES

✗ “You’re asking me to **program my monitoring system?**”

CHALLENGES

- ✗ “You’re asking me to **program my monitoring system?**”
- ✗ **New paradigm, new concepts:** windows, triggers, partitioning, etc

CHALLENGES

- ✗ “You’re asking me to **program my monitoring system?**”
- ✗ **New paradigm, new concepts:** windows, triggers, partitioning, etc
- ✗ Our goal is not to **make Hadoop easier/better/faster**



DRAWING INSPIRATION

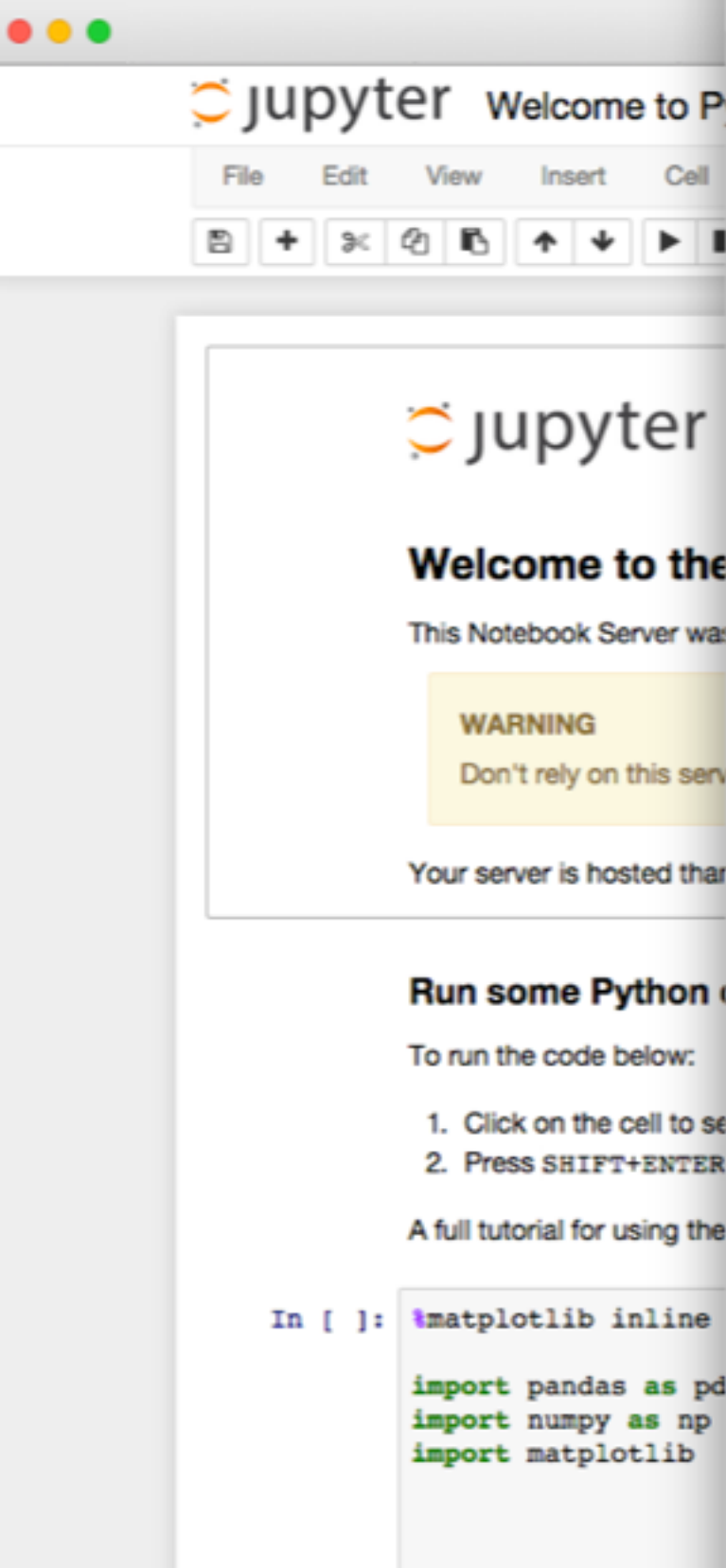
“Instead of imagining that our main task is to instruct a **computer** what to do, let us concentrate rather on explaining to **human beings** what we want a computer to do.”

–DONALD KNUTH, 1983

INTERACTIVE REPLS WITH HISTORY AND PROSE

JUPYTER, MATHEMATICA

<http://jupyter.org/>



A screenshot of a Jupyter Notebook titled "Exploring the Lorenz System". The notebook contains the following text:

In this Notebook we explore the [Lorenz system](#) of differential equations:

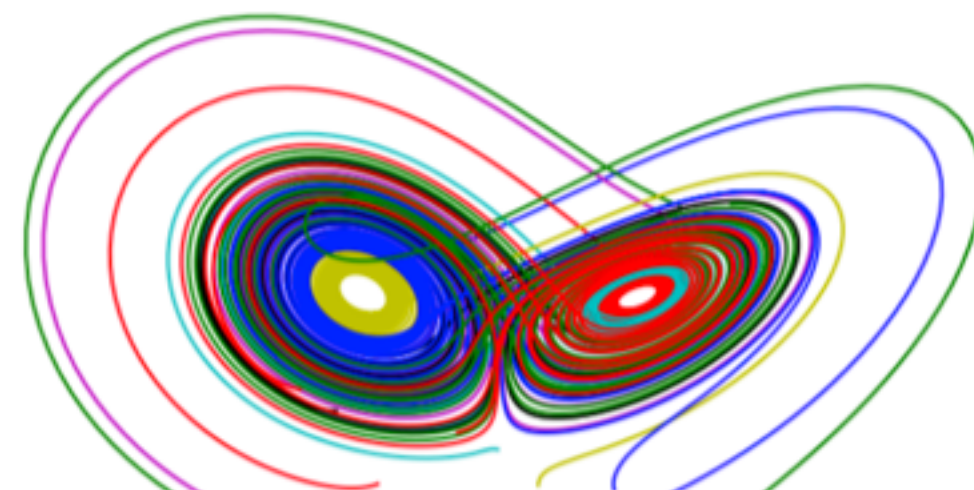
$$\begin{aligned} \dot{x} &= \sigma(y - x) \\ \dot{y} &= \rho x - y - xz \\ \dot{z} &= -\beta z + xy \end{aligned}$$

This is one of the classic systems in non-linear differential equations. It exhibits a range of complex behaviors as the parameters (σ, β, ρ) are varied, including what are known as *chaotic solutions*. The system was originally developed as a simplified mathematical model for atmospheric convection in 1963.

In [7]: `interact(Lorenz, N=fixed(10), angle=(0.,360.), sigma=(0.0,50.0), beta=(0.,5), rho=(0.0,50.0))`

The interactive interface shows sliders for the following parameters:

- angle: 308.2
- max_time: 12
- σ : 10
- β : 2.6
- ρ : 28

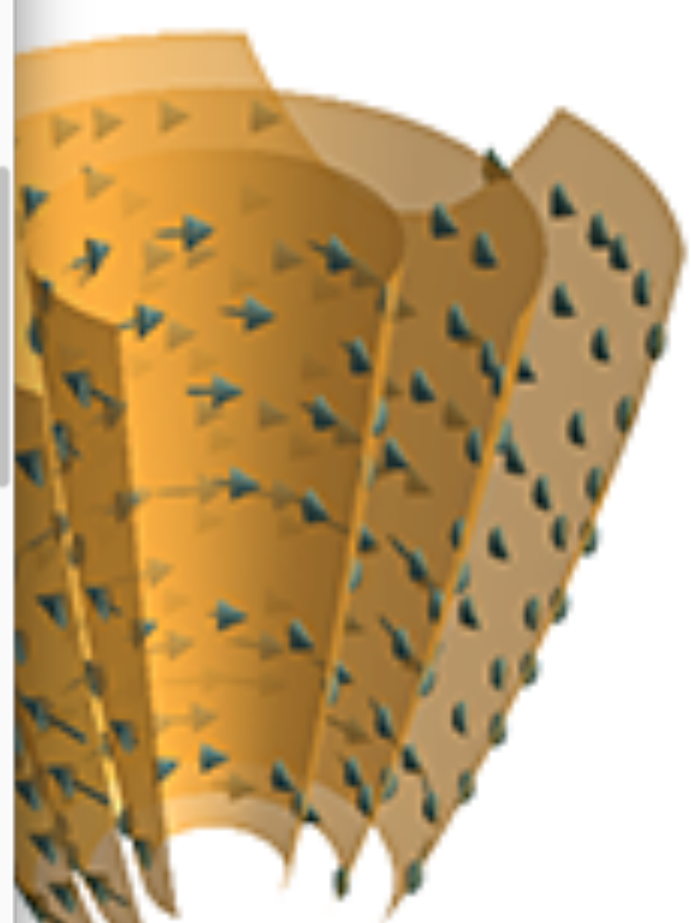


<https://www.wolfram.com/mathematica/>

A screenshot of the Wolfram Mathematica Online interface. It shows the following code:

```
Grid[Partition[Plot3D[Re[#], {x, y} ∈ D, Boxed → False, Axes → None, PlotRange → All] & /@ funs, 3]]
```

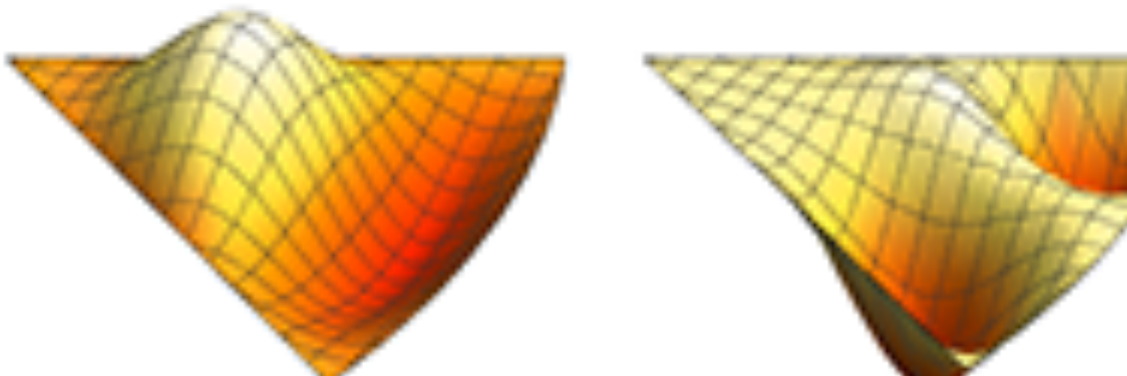
Out[5]=



A screenshot of the Wolfram Mathematica Online interface. It shows the following code:

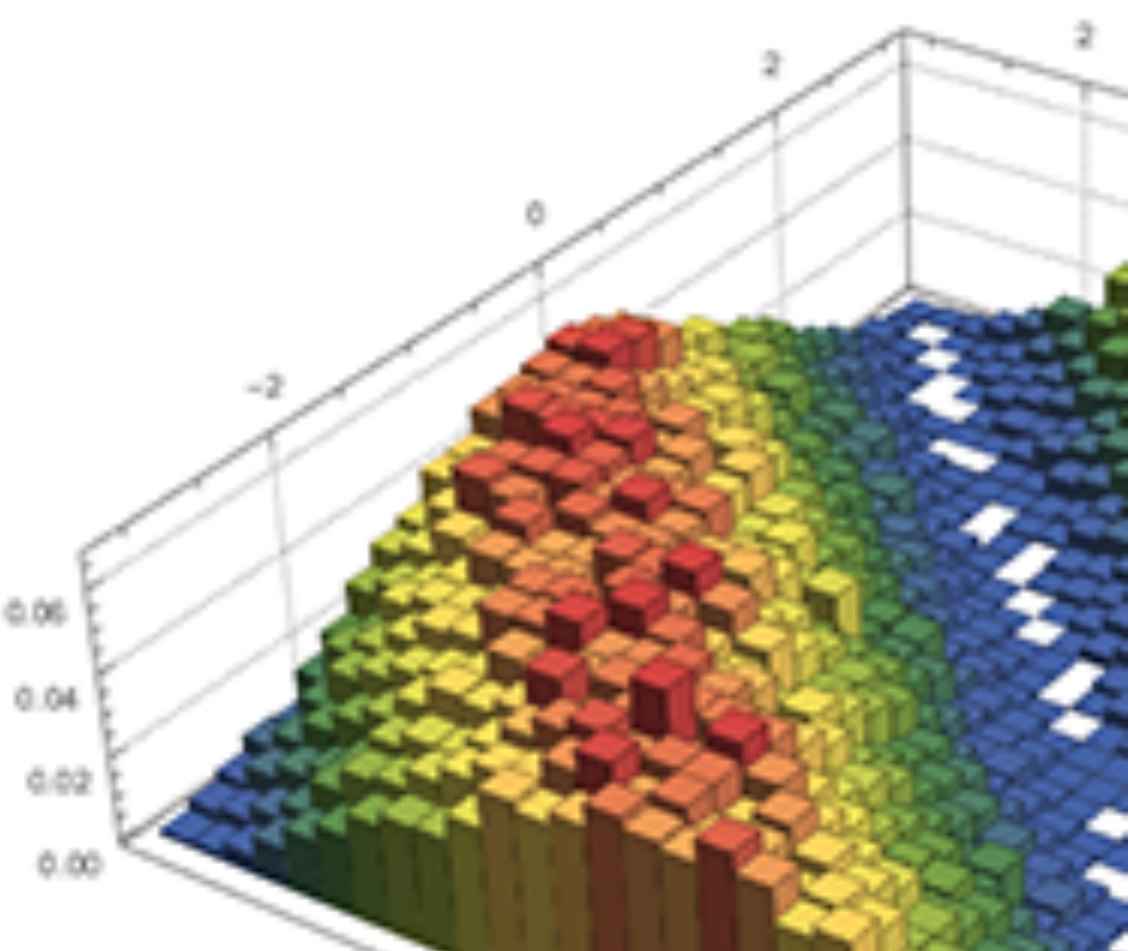
```
Grid[Partition[Plot3D[Re[#], {x, y} ∈ D, Boxed → False, Axes → None, PlotRange → All] & /@ funs, 3]]
```

Out[5]=



Out[6]=

```
evsD = MatrixPropertyDistribution[Arg[Eigenvalues[RandomSample[RandomVariate[evsD, 10^5]]], Histogram3D[evs, {-Pi, Pi}, 0.2], PDF, PlotTheme → "Scientific"]]
```



EVE

<http://play.witheve.com/#/examples/bar-graph.eve>

↑ workspaces



Pet Lengths

```
[#pet name: "dog" length: 14]  
[#pet name: "orangutan" length: 9]  
[#pet name: "lemur" length: 5]
```



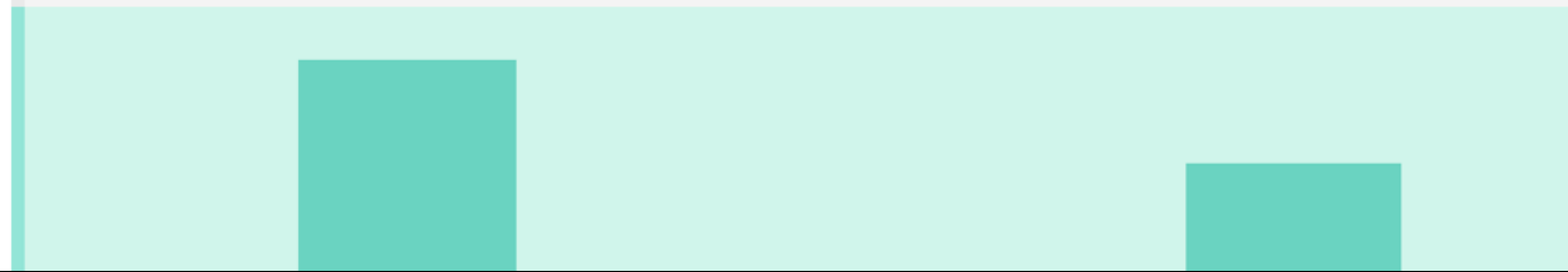
Each pet is a single bar on our graph. The bar's label is the pet's name, it's height is the pet's length. The sort property tells the bar graph to draw the bar in alphabetical order based on the pets' names.

search

```
[#pet name length]  
ix = sort[value: name]
```

bind @view

```
[#bar-graph | bar: [label: name height: length sort: ix]]
```



LITERATE PROGRAMMING BENEFITS

<https://github.com/witheve/rfcs/blob/master/proposed/syntax.md#program-structure>

LITERATE PROGRAMMING BENEFITS

- "Literate programming forces you to **consider a human audience.**"

LITERATE PROGRAMMING BENEFITS

- "Literate programming forces you to **consider a human audience.**"
- "The human brain is wired to **engage with and remember stories.**"

LITERATE PROGRAMMING BENEFITS

- "Literate programming forces you to **consider a human audience.**"
- "The human brain is wired to **engage with and remember stories.**"
- "...literate programming encourages the programmer to arrange [programs] **in a way that makes narrative sense.**"

LITERATE PROGRAMMING BENEFITS

- "Literate programming forces you to **consider a human audience.**"
- "The human brain is wired to **engage with and remember stories.**"
- "...literate programming encourages the programmer to arrange [programs] **in a way that makes narrative sense.**"
- "...you don't really understand something until you **explain it to someone else.**"

THE AHA! MOMENT

LITERATE
DASHBOARDS,
EXECUTABLE
RUNBOOKS



CREATING LITERATE DASHBOARDS

Operational Notebook

Solr Corruption / File Count Explosion

Disk Usage

Disk Usage

We want to be warned when our disk space free goes below 10% on any disk. However, disk usage for temporary files should be able to burst over the limit and fall back below without triggering an alert to our on-call engineer. Therefore, we alert on the exponentially-weighted moving average (EWMA).

```
~~~stream
where(type: ["disk", "free", "percent"])
|> by([:host, :mount])
|> ewma
|> threshold(below: 10.0)
|> forward(:on_call_alert)
|> draw(:table)
~~~
```

Possible causes

1. Log rotation on `web` hosts might be configured incorrectly, filling up the disk.
2. Stale software artifacts from builds on `ci` hosts might be consuming too much space.

Disk Usage

We want to be warned when our disk space free goes below 10% on any disk. However, disk usage for temporary files should be able to burst over the limit and fall back below without triggering an alert to our on-call engineer. Therefore, we alert on the exponentially-weighted moving average (EWMA).

	/
web-01	50.2%
web-02	46.5%
ci-01	19.3%

Possible causes

1. Log rotation on `web` hosts might be configured incorrectly, filling up the disk.
2. Stale software artifacts from builds on `ci` hosts might be consuming too much space.



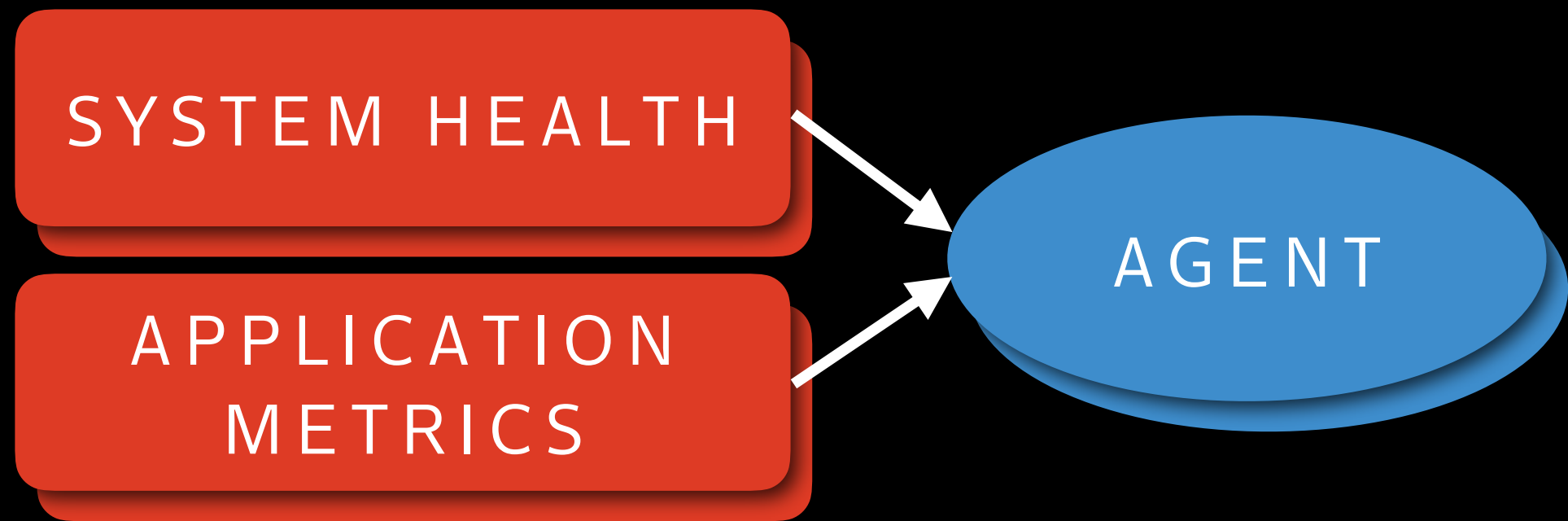
UNDER THE HOOD

OPERATIONAL VISIBILITY PROJECT

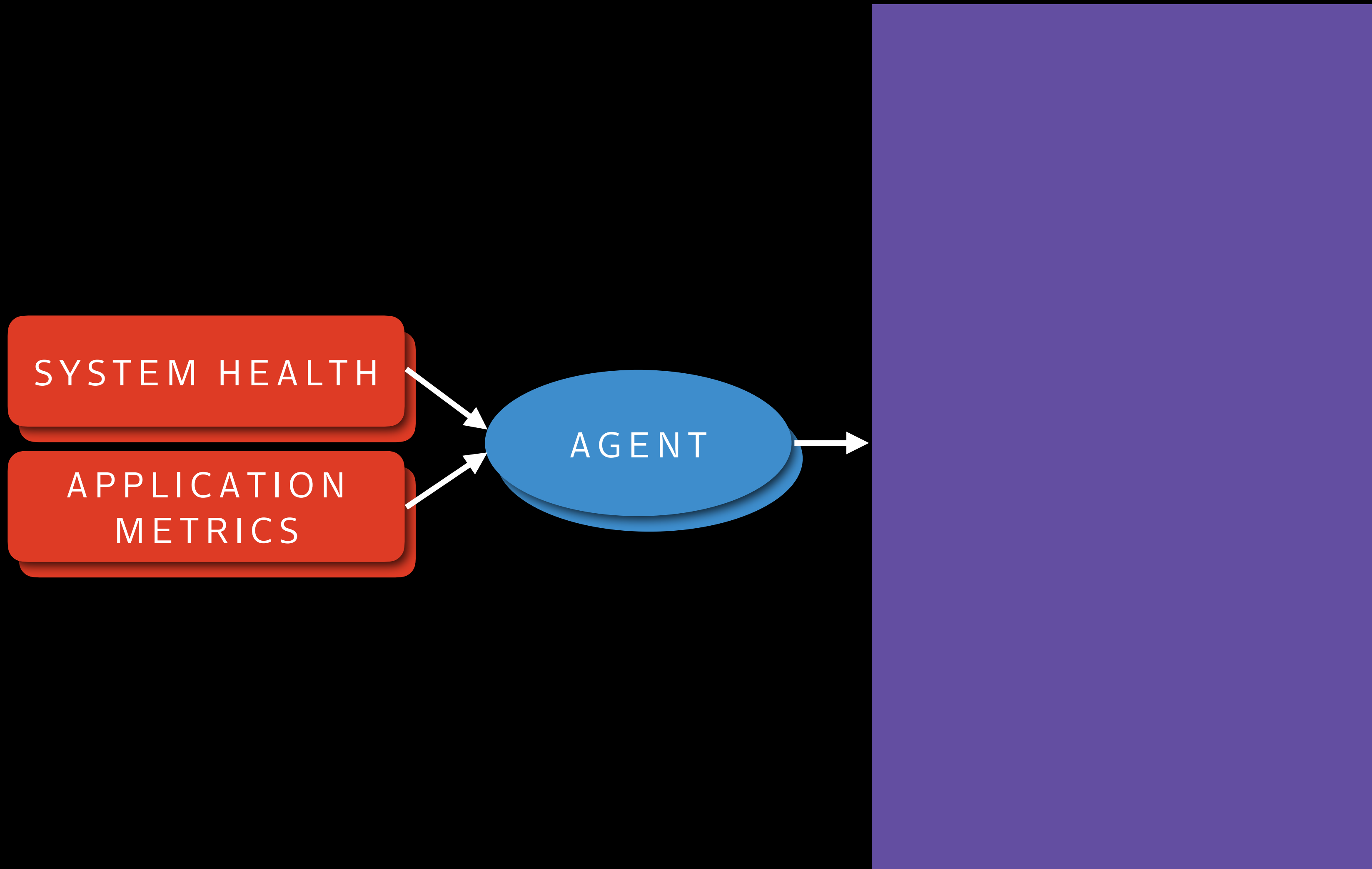
SYSTEM HEALTH

APPLICATION
METRICS

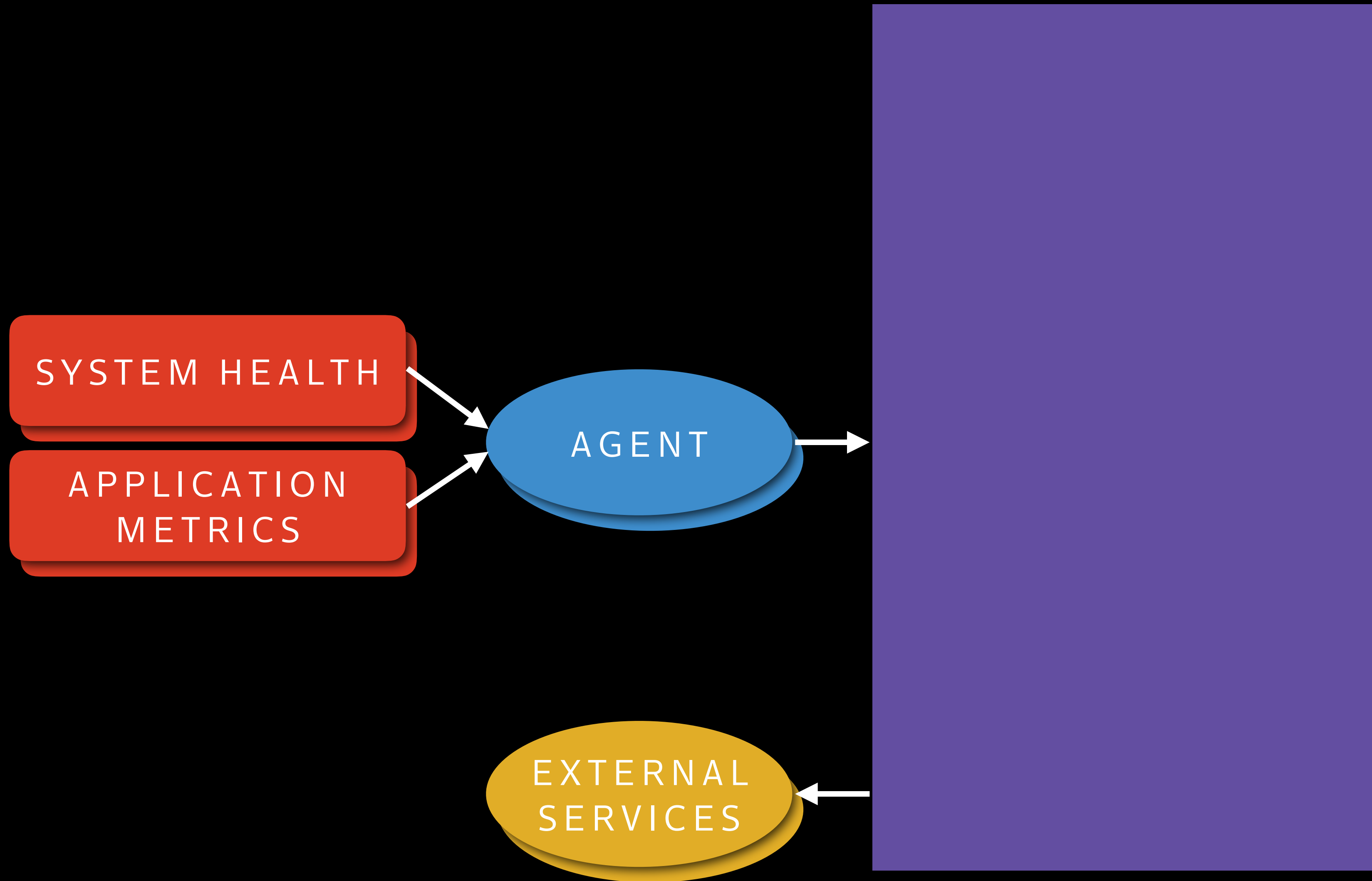
OPERATIONAL VISIBILITY PROJECT



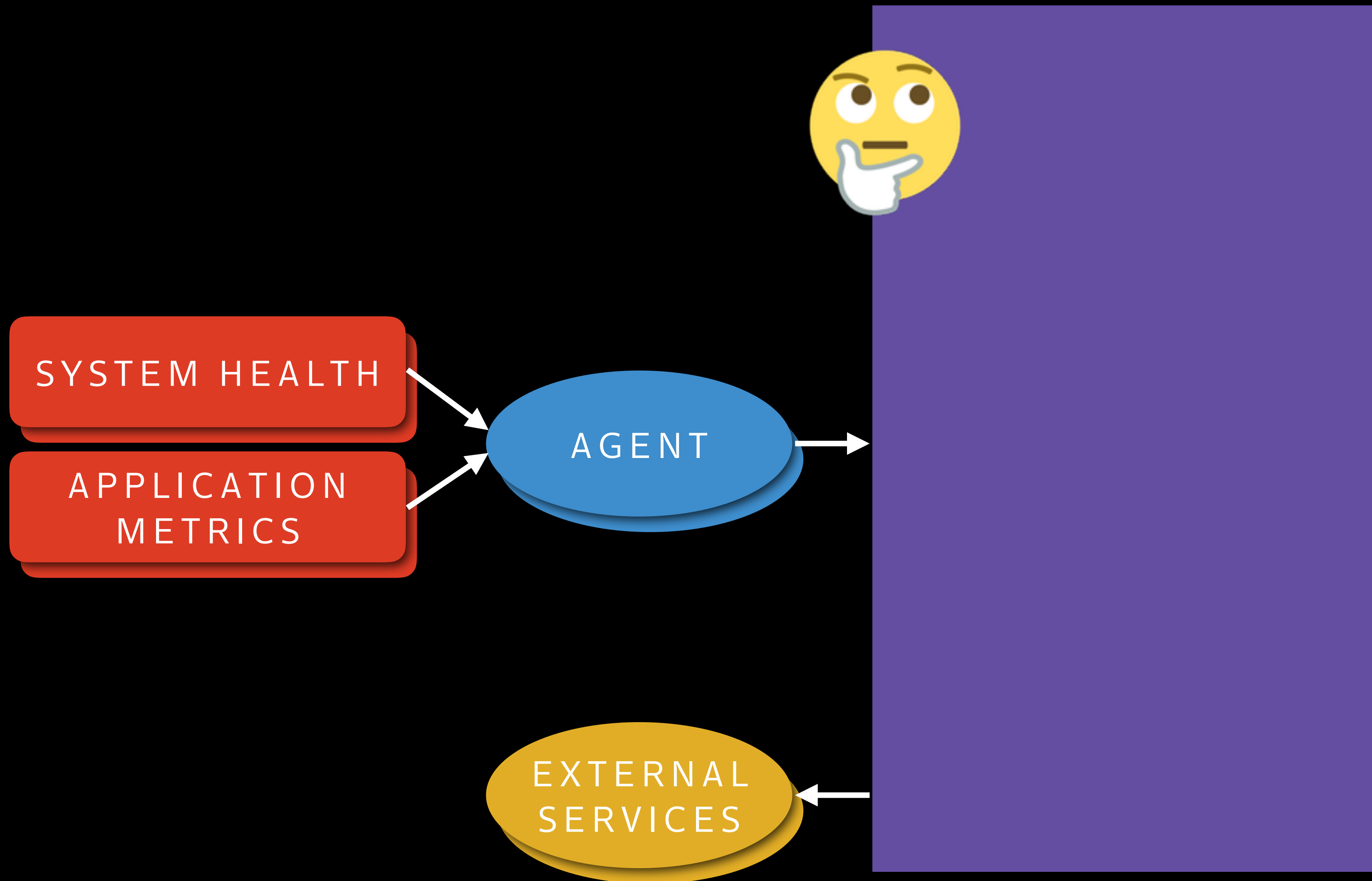
OPERATIONAL VISIBILITY PROJECT



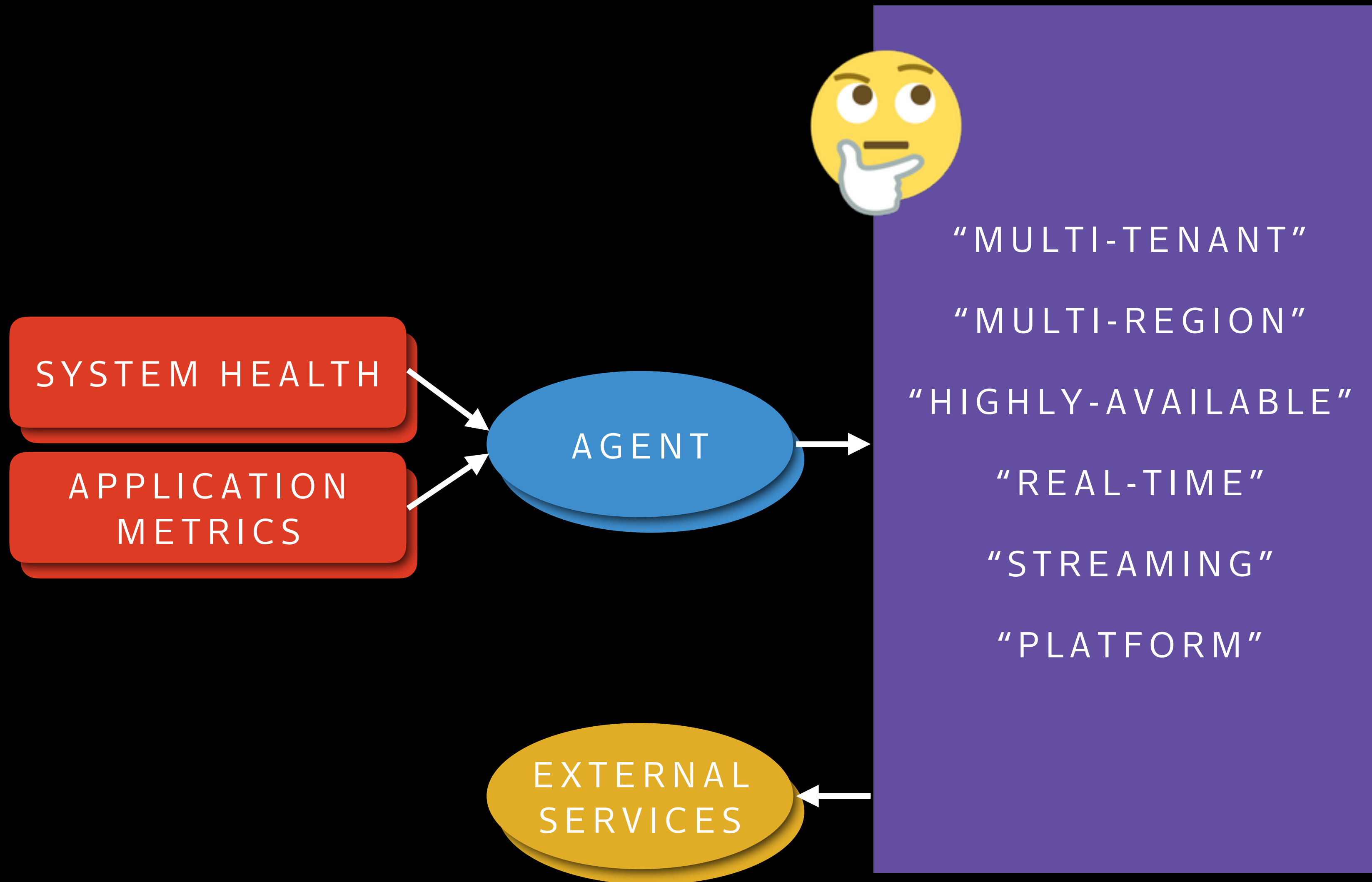
OPERATIONAL VISIBILITY PROJECT



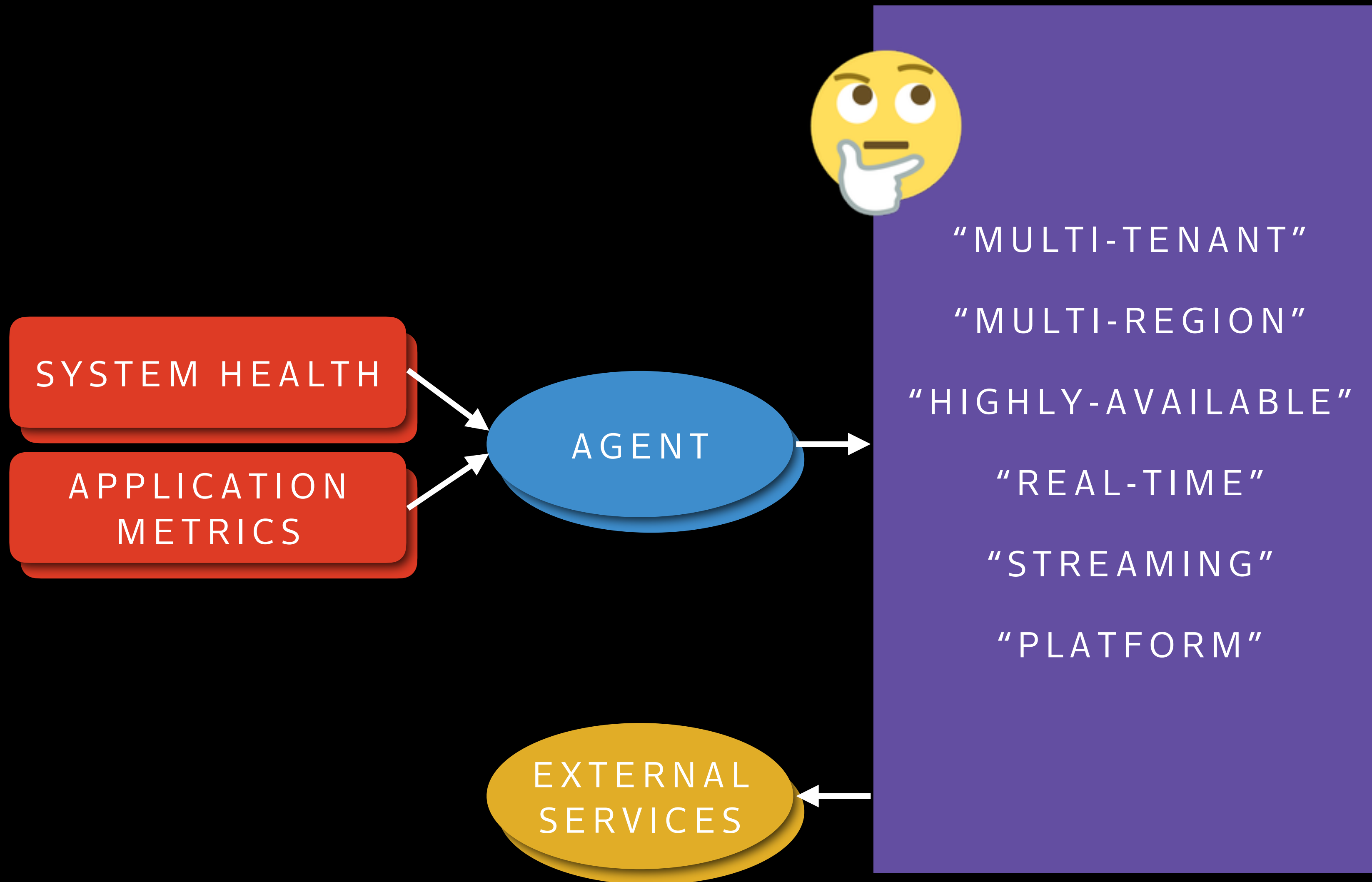
OPERATIONAL VISIBILITY PROJECT



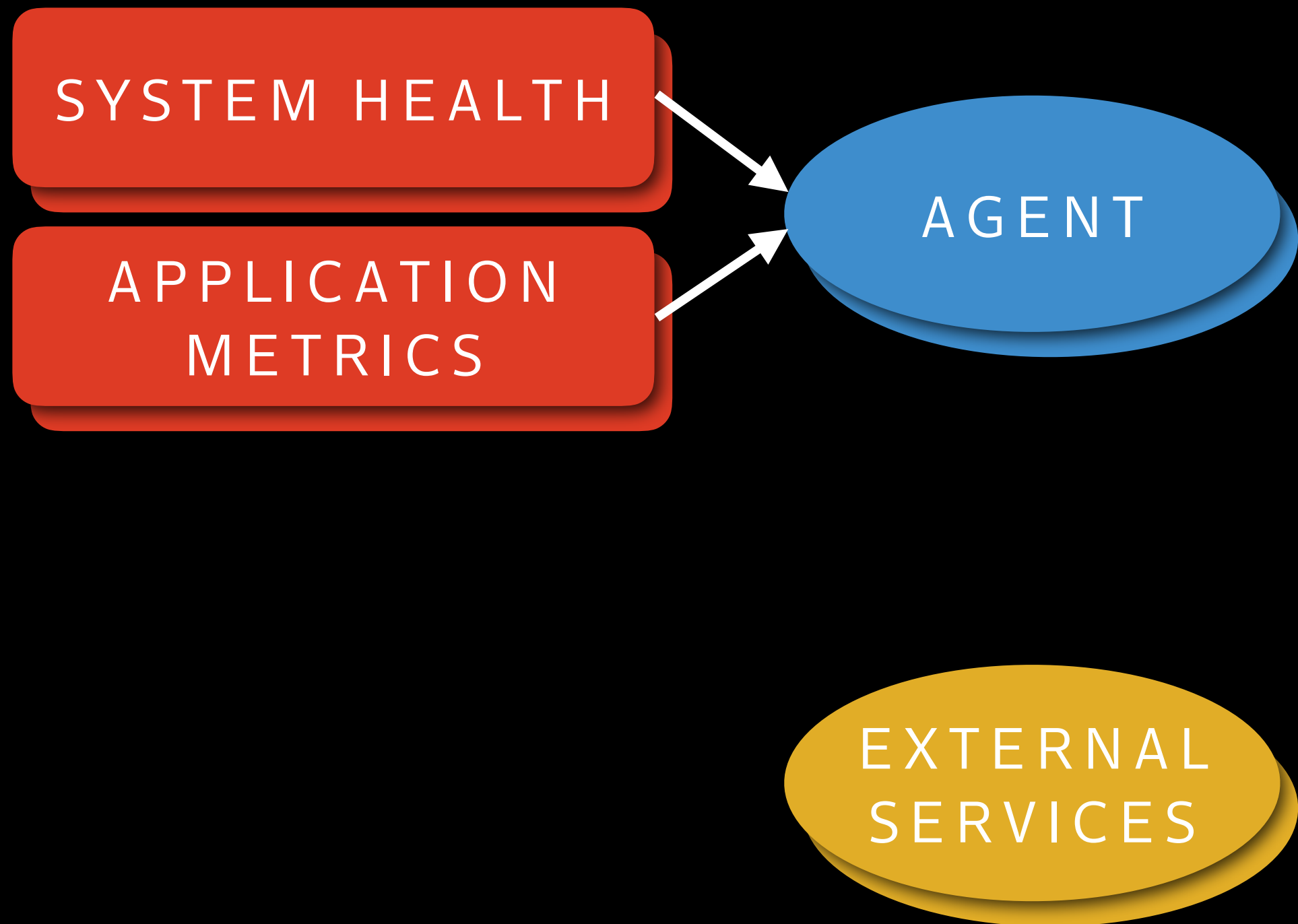
OPERATIONAL VISIBILITY PROJECT



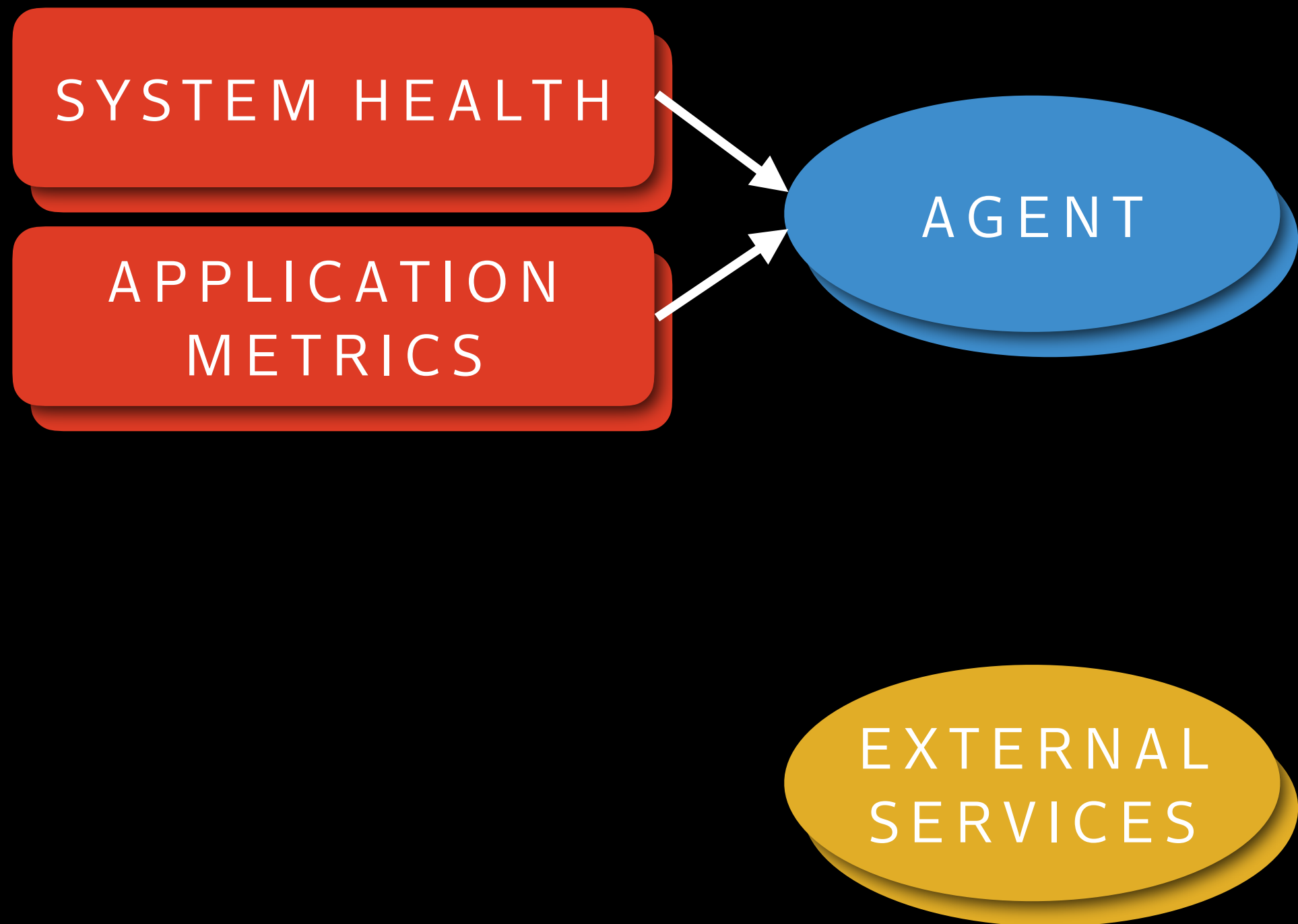
OPERATIONAL VISIBILITY PROJECT



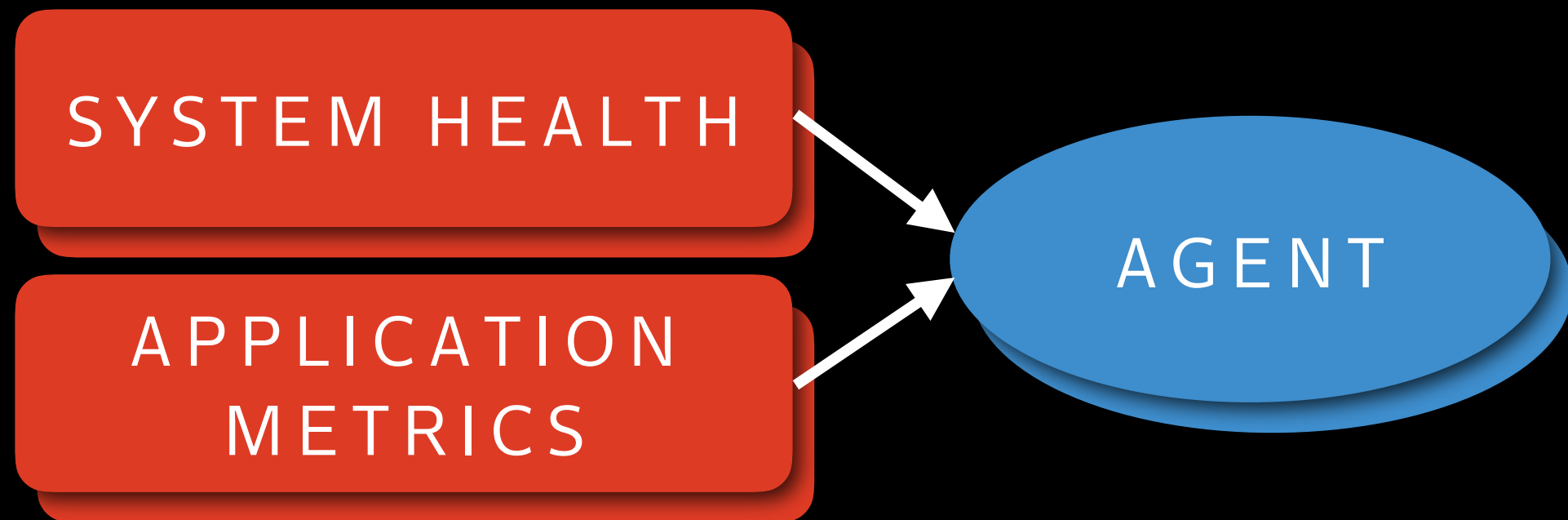
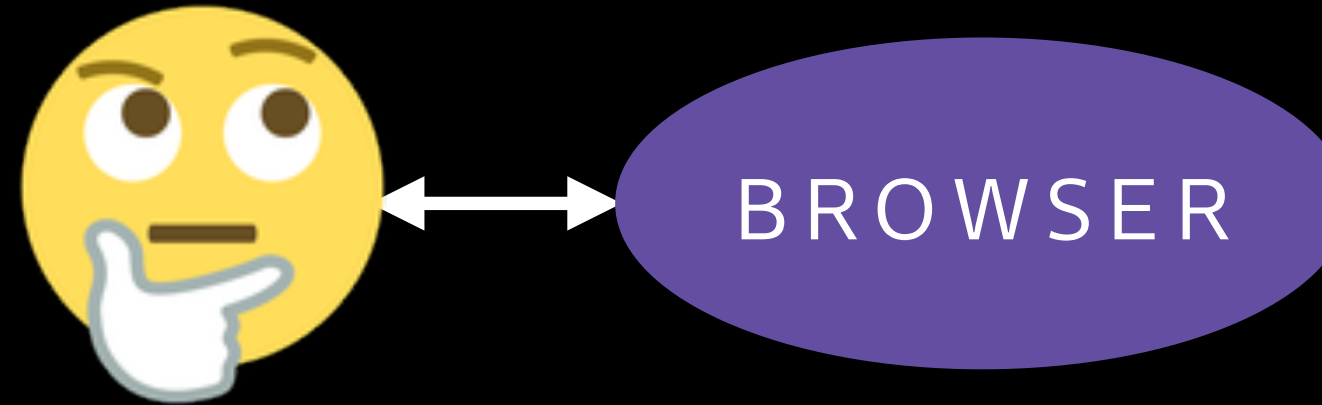
OPERATIONAL VISIBILITY PROJECT



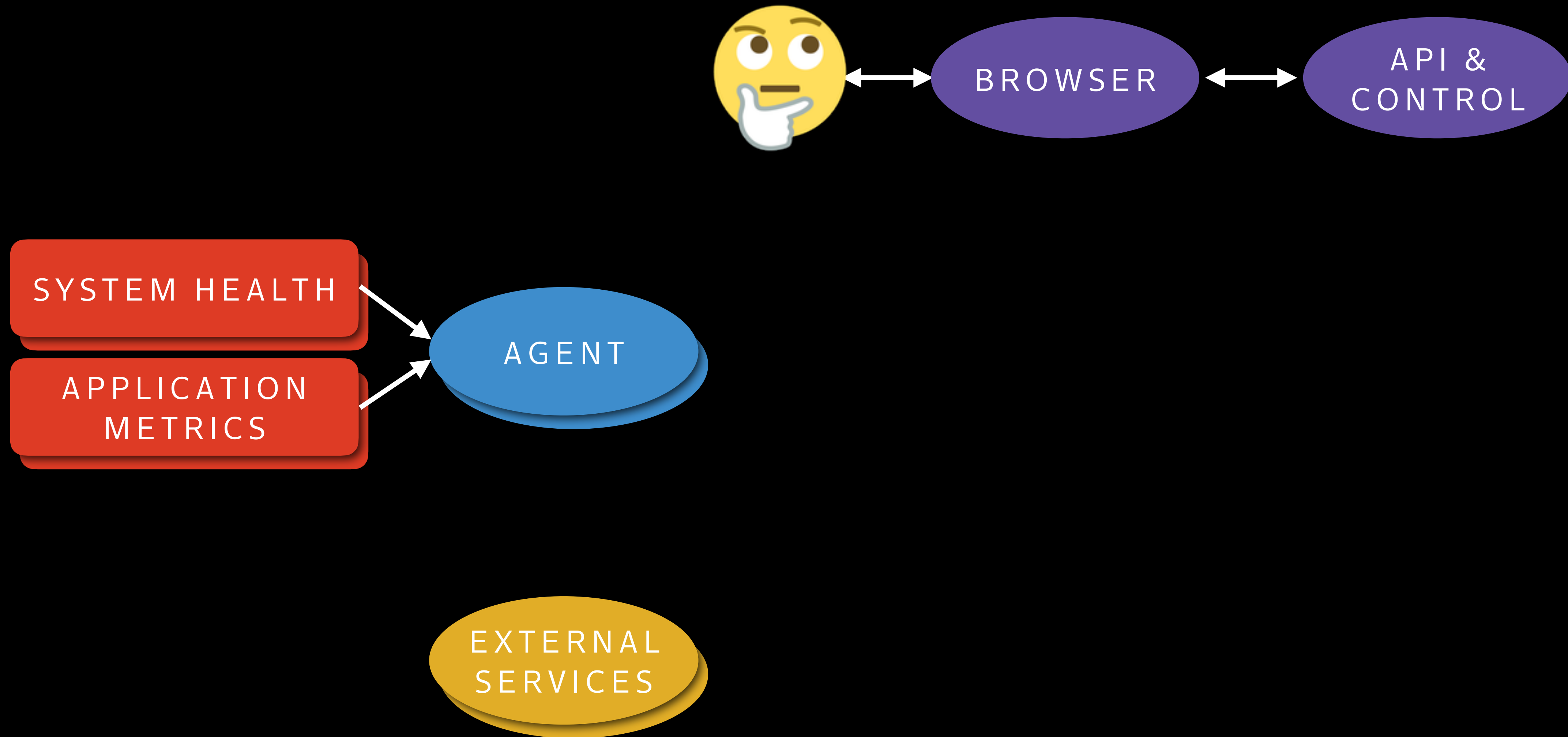
OPERATIONAL VISIBILITY PROJECT



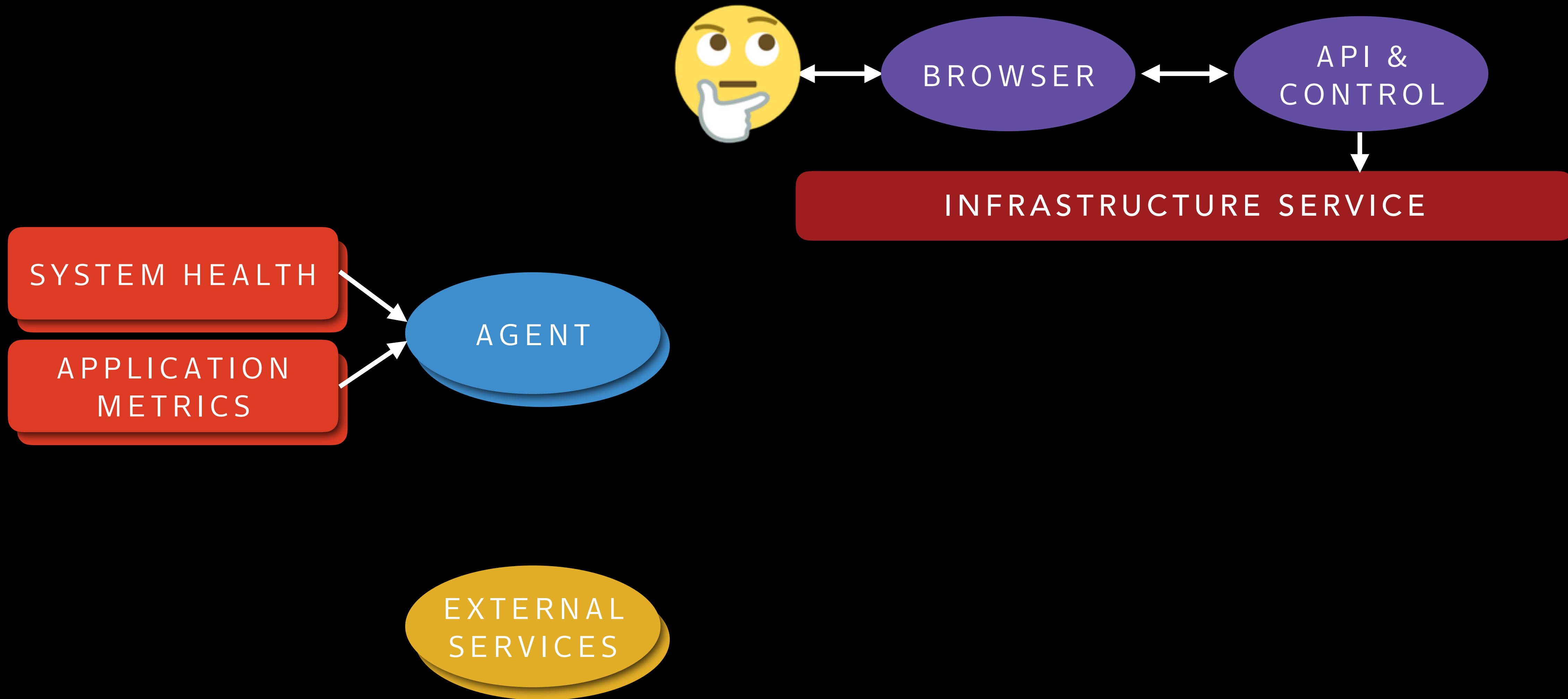
OPERATIONAL VISIBILITY PROJECT



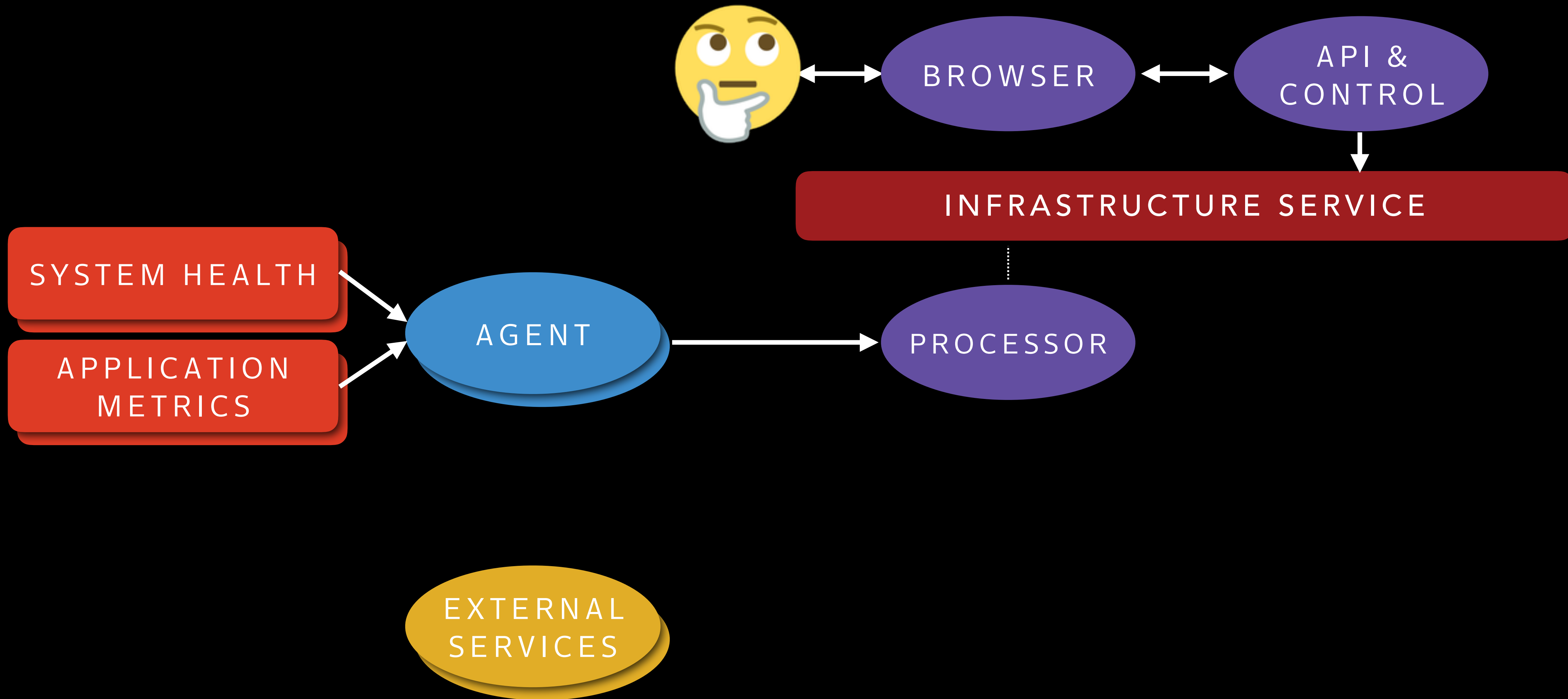
OPERATIONAL VISIBILITY PROJECT



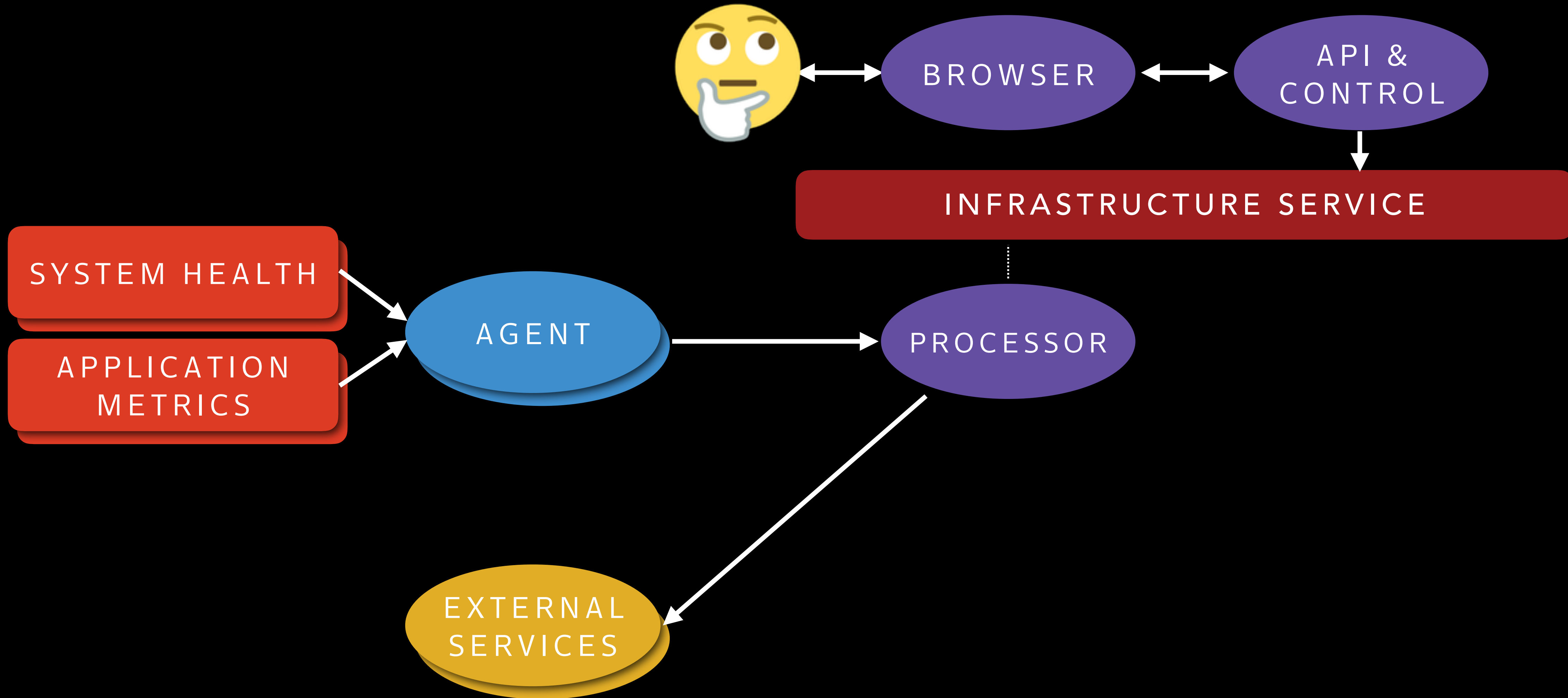
OPERATIONAL VISIBILITY PROJECT



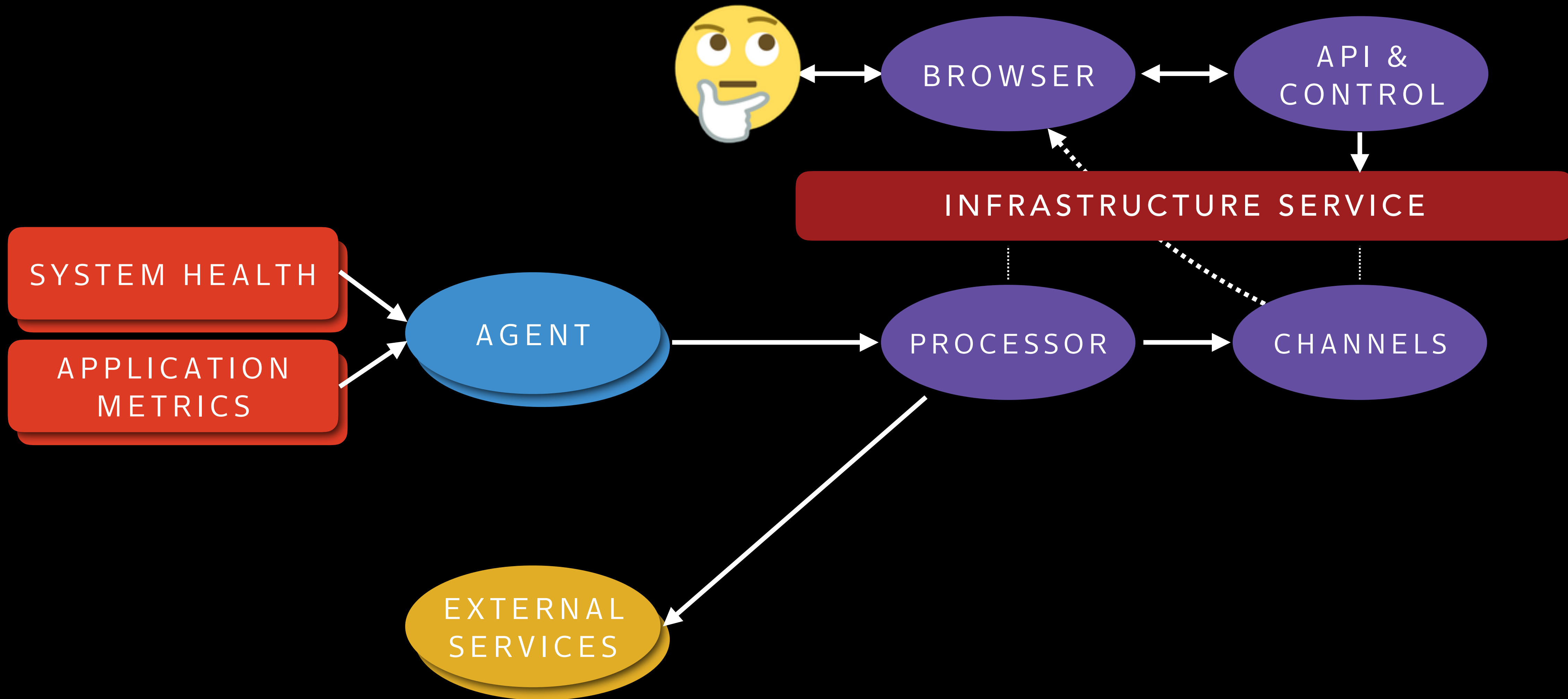
OPERATIONAL VISIBILITY PROJECT



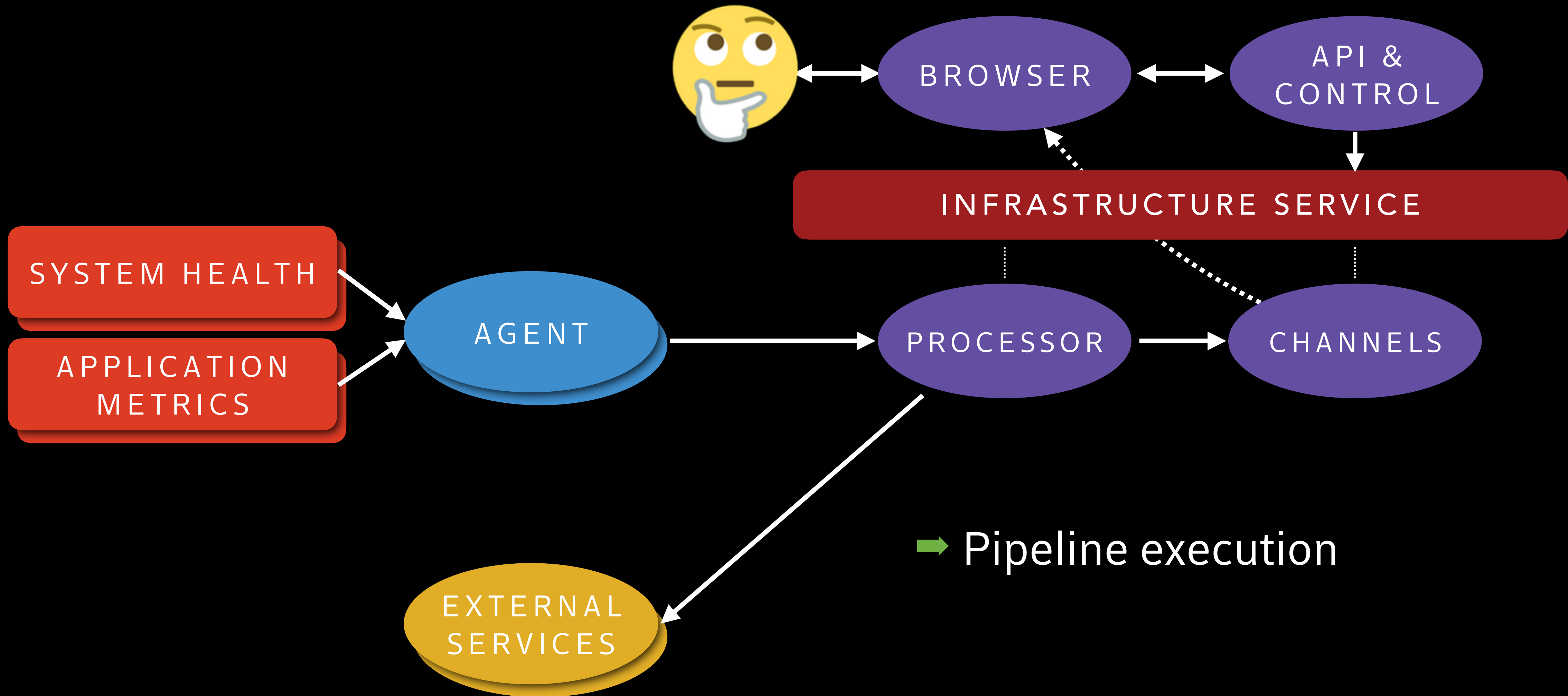
OPERATIONAL VISIBILITY PROJECT



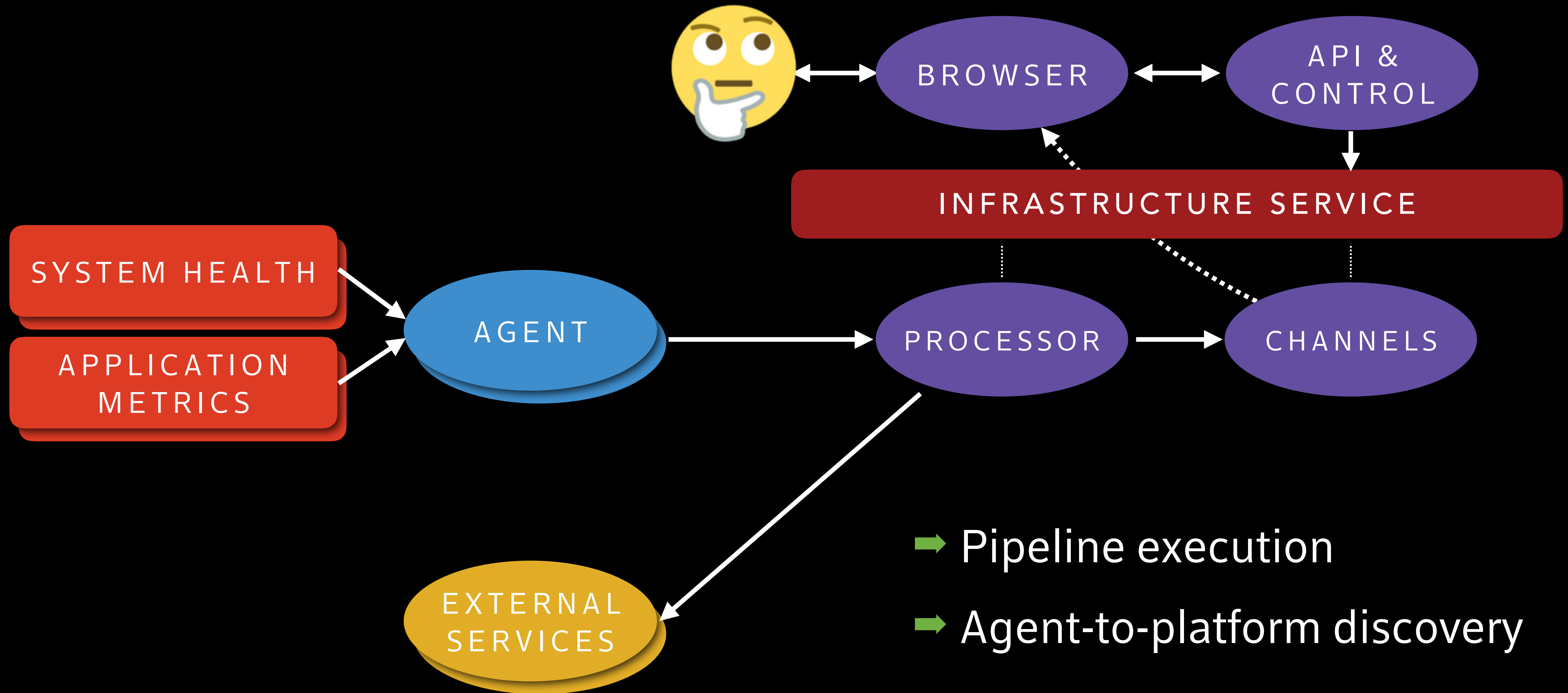
OPERATIONAL VISIBILITY PROJECT



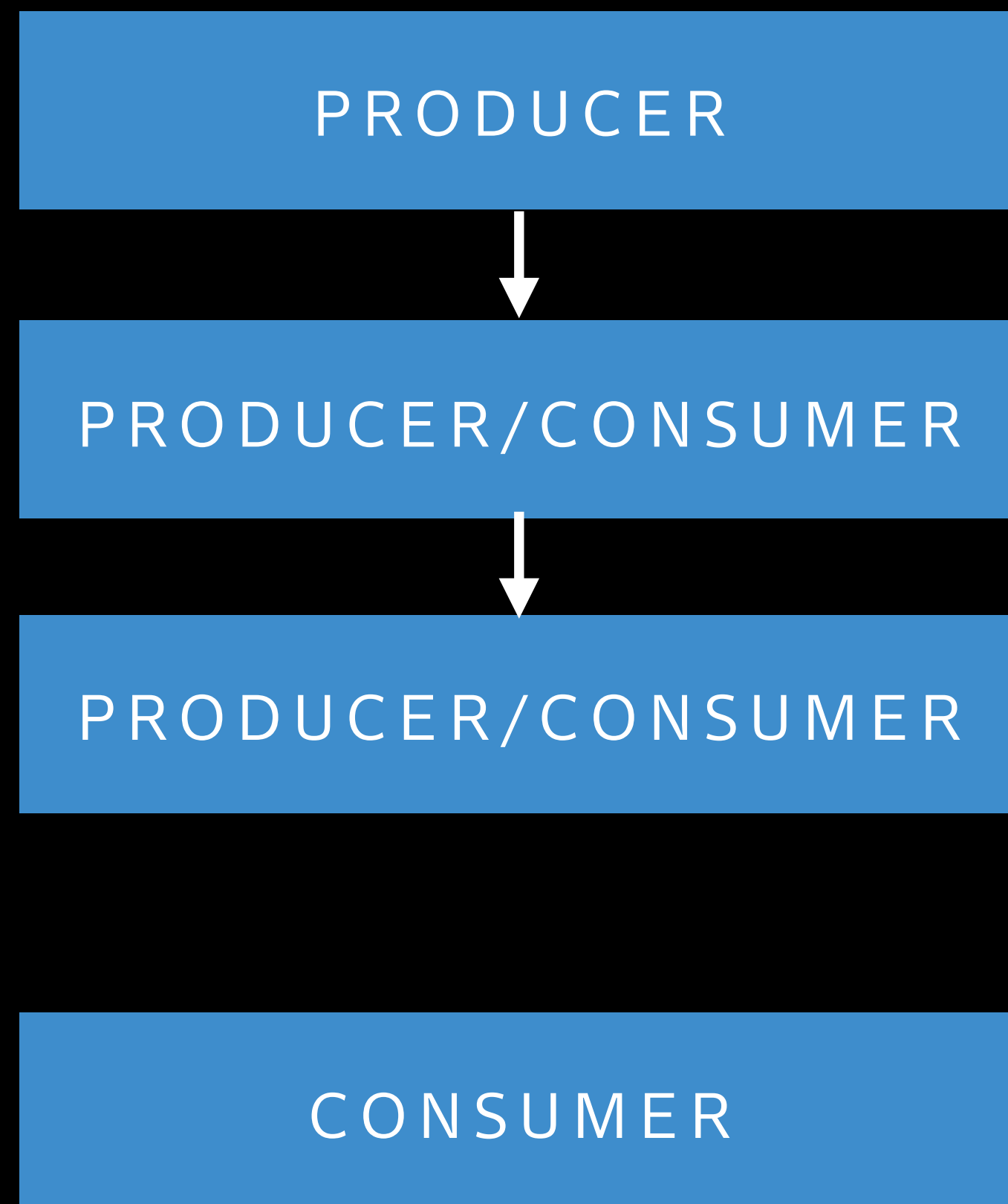
OPERATIONAL VISIBILITY PROJECT



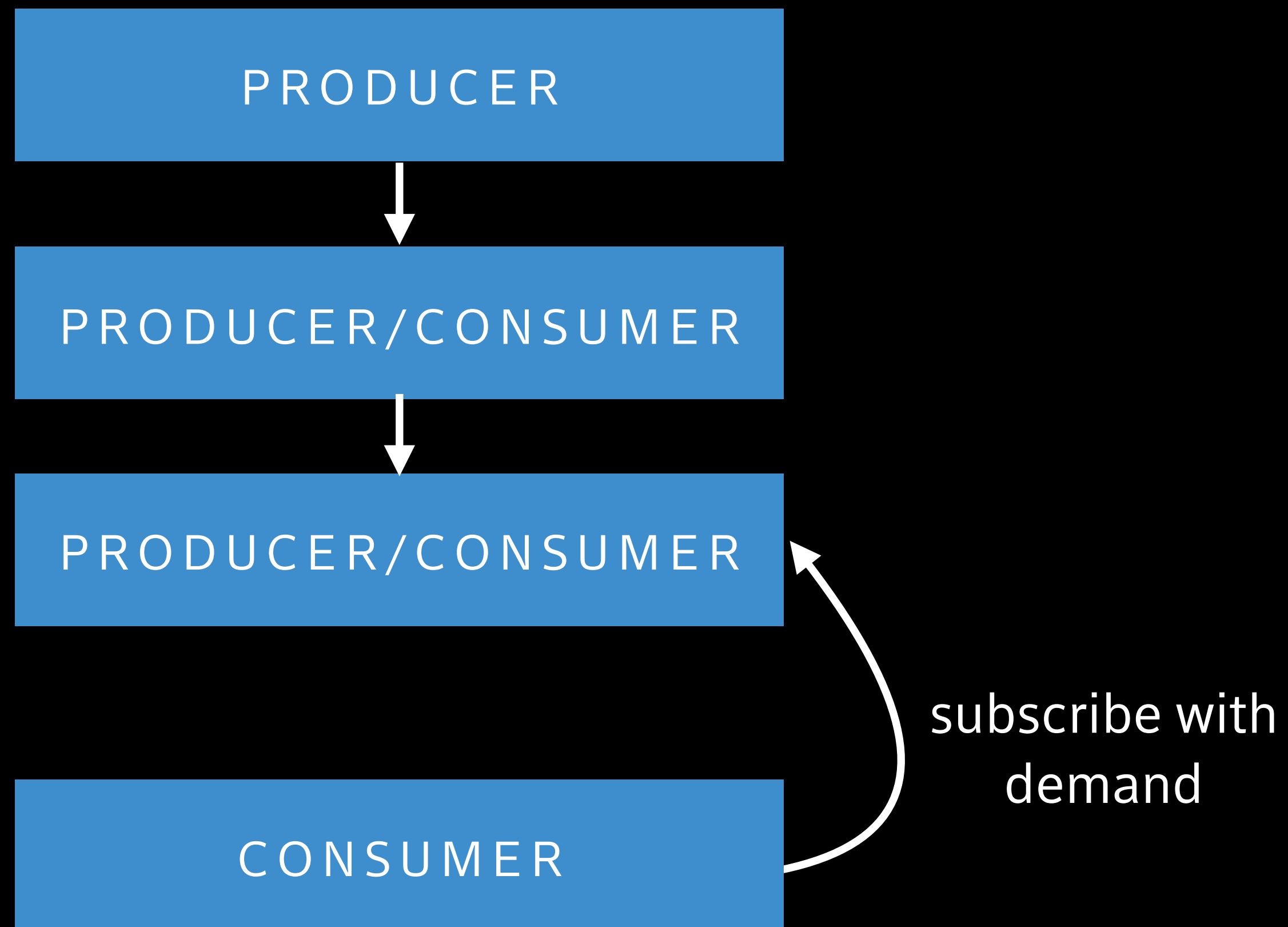
OPERATIONAL VISIBILITY PROJECT



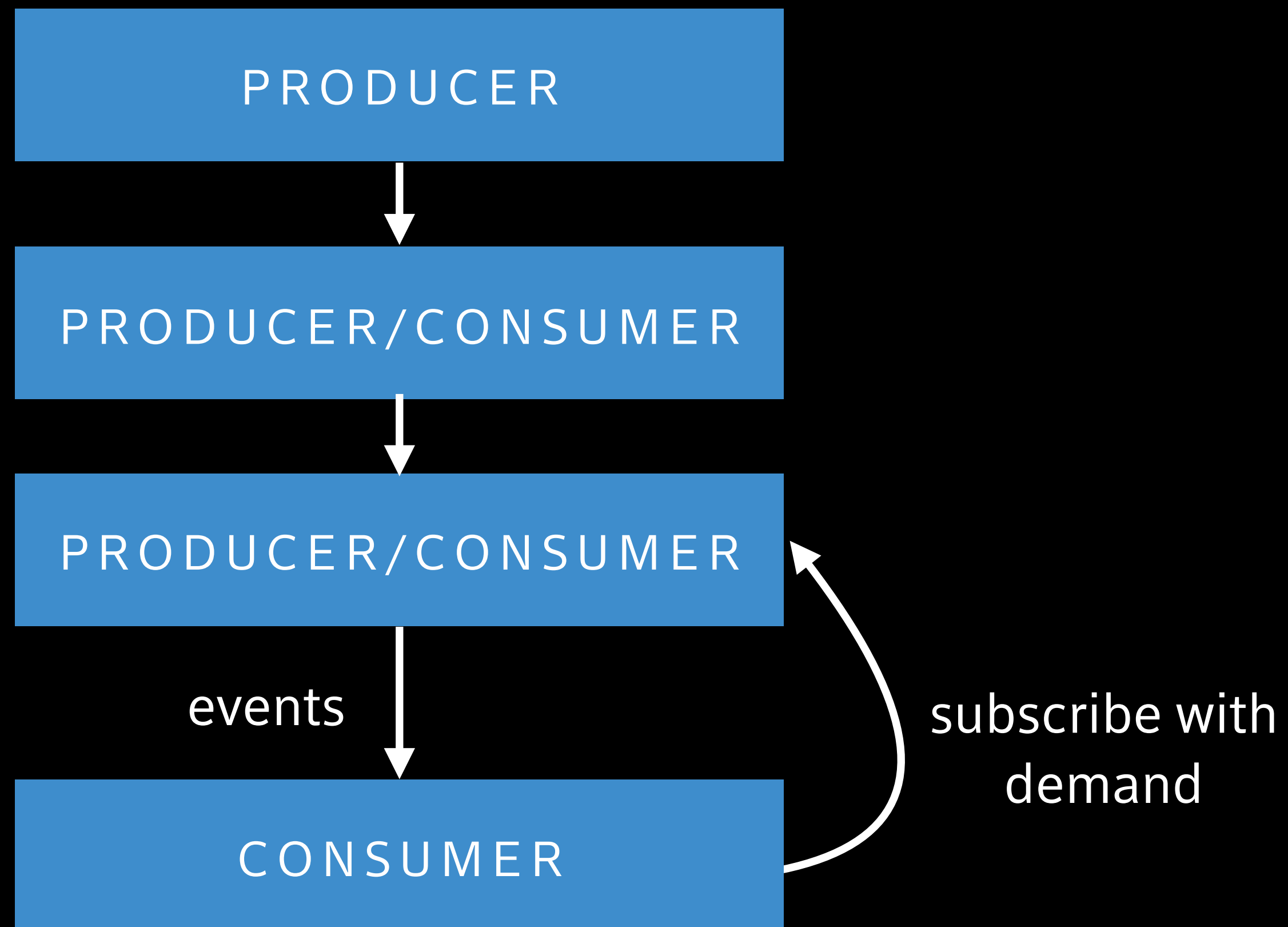
ELIXIR GENSTAGE & FLOW



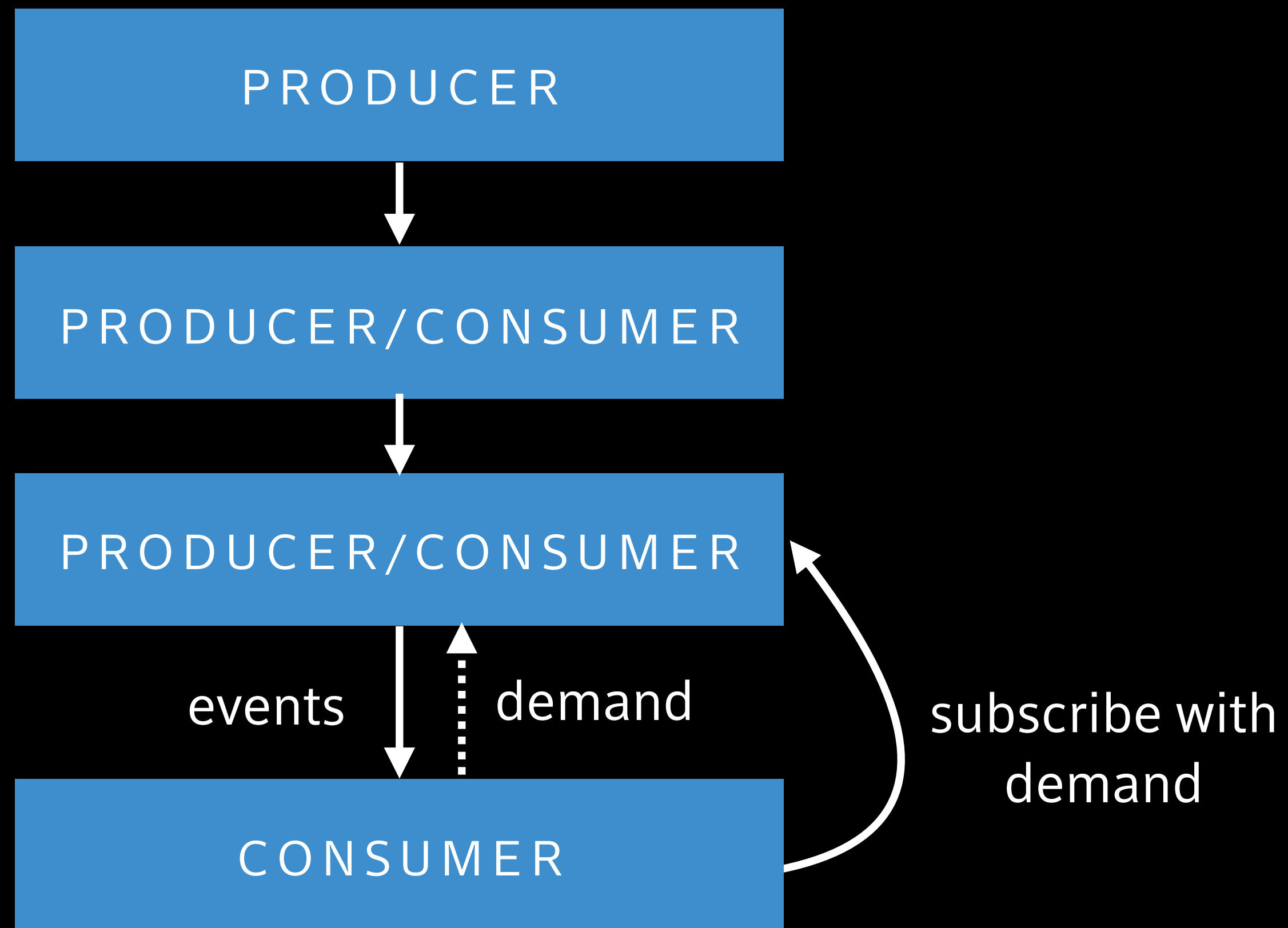
ELIXIR GENSTAGE & FLOW



ELIXIR GENSTAGE & FLOW

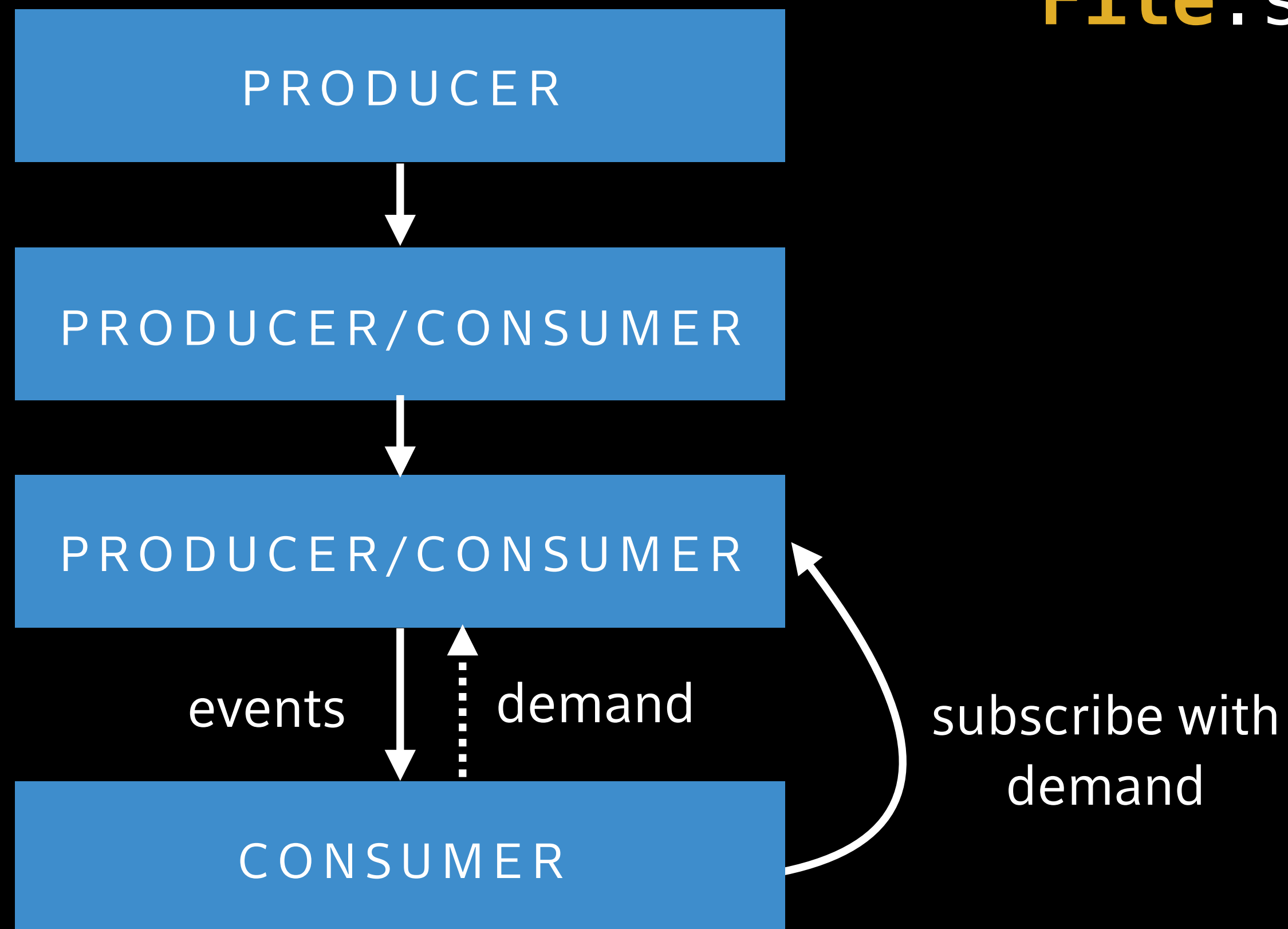


ELIXIR GENSTAGE & FLOW

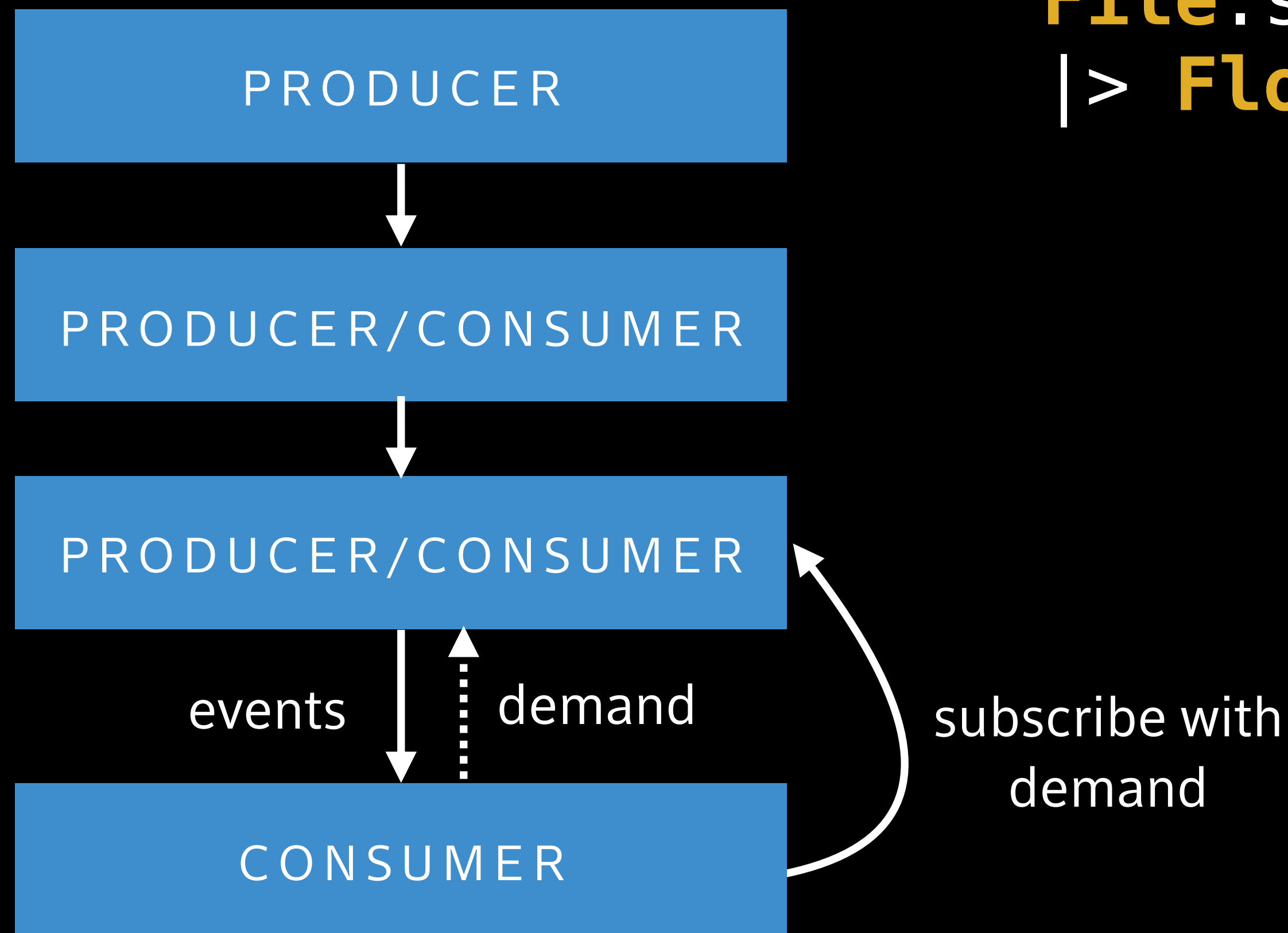


ELIXIR GENSTAGE & FLOW

```
File.stream!("/tmp/words.txt")
```

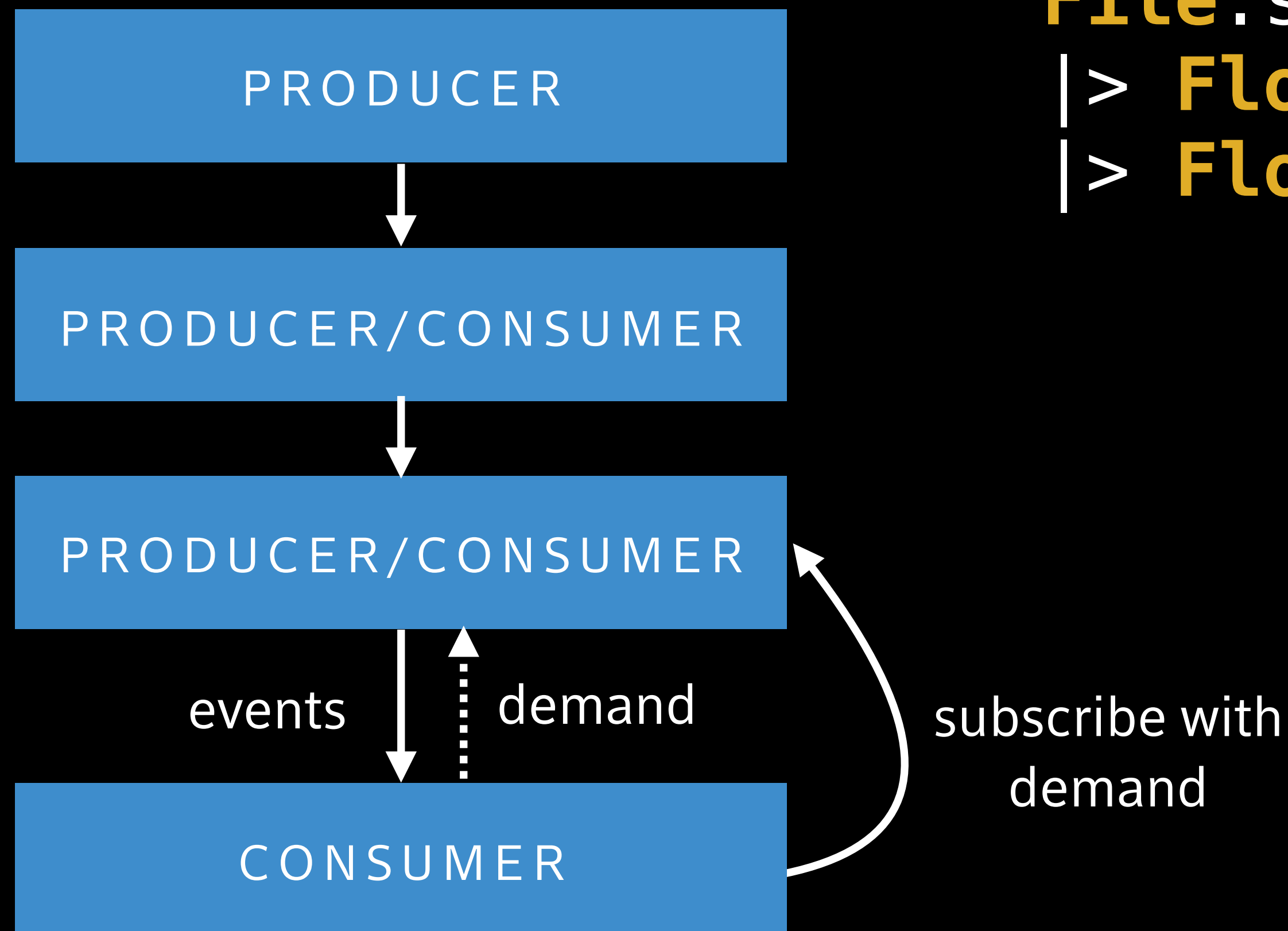


ELIXIR GENSTAGE & FLOW



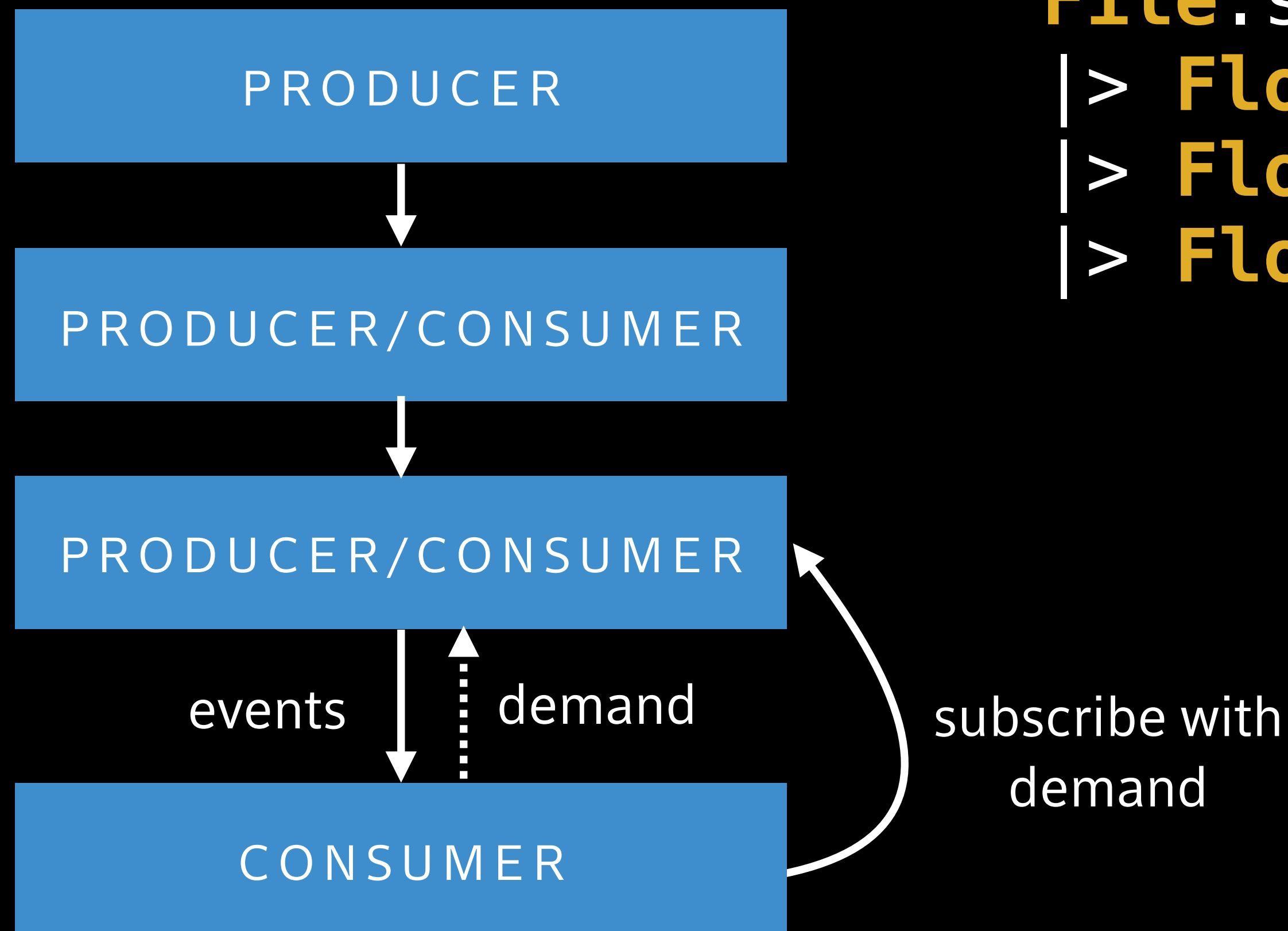
```
File.stream!("/tmp/words.txt")  
|> Flow.from_enumerable()
```

ELIXIR GENSTAGE & FLOW



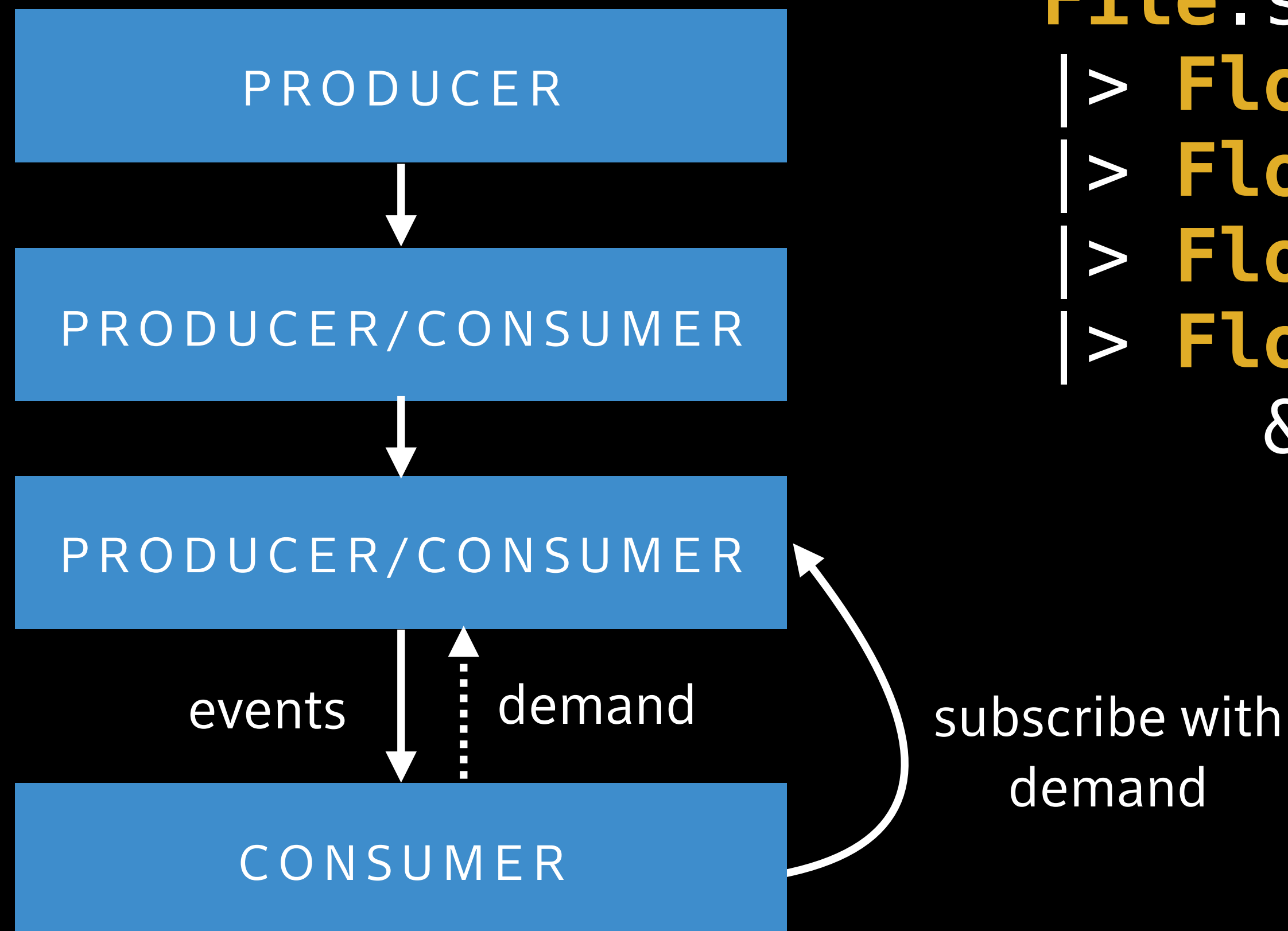
```
File.stream!("/tmp/words.txt")  
|> Flow.from_enumerable()  
|> Flow.flat_map(&String.split/1)
```

ELIXIR GENSTAGE & FLOW



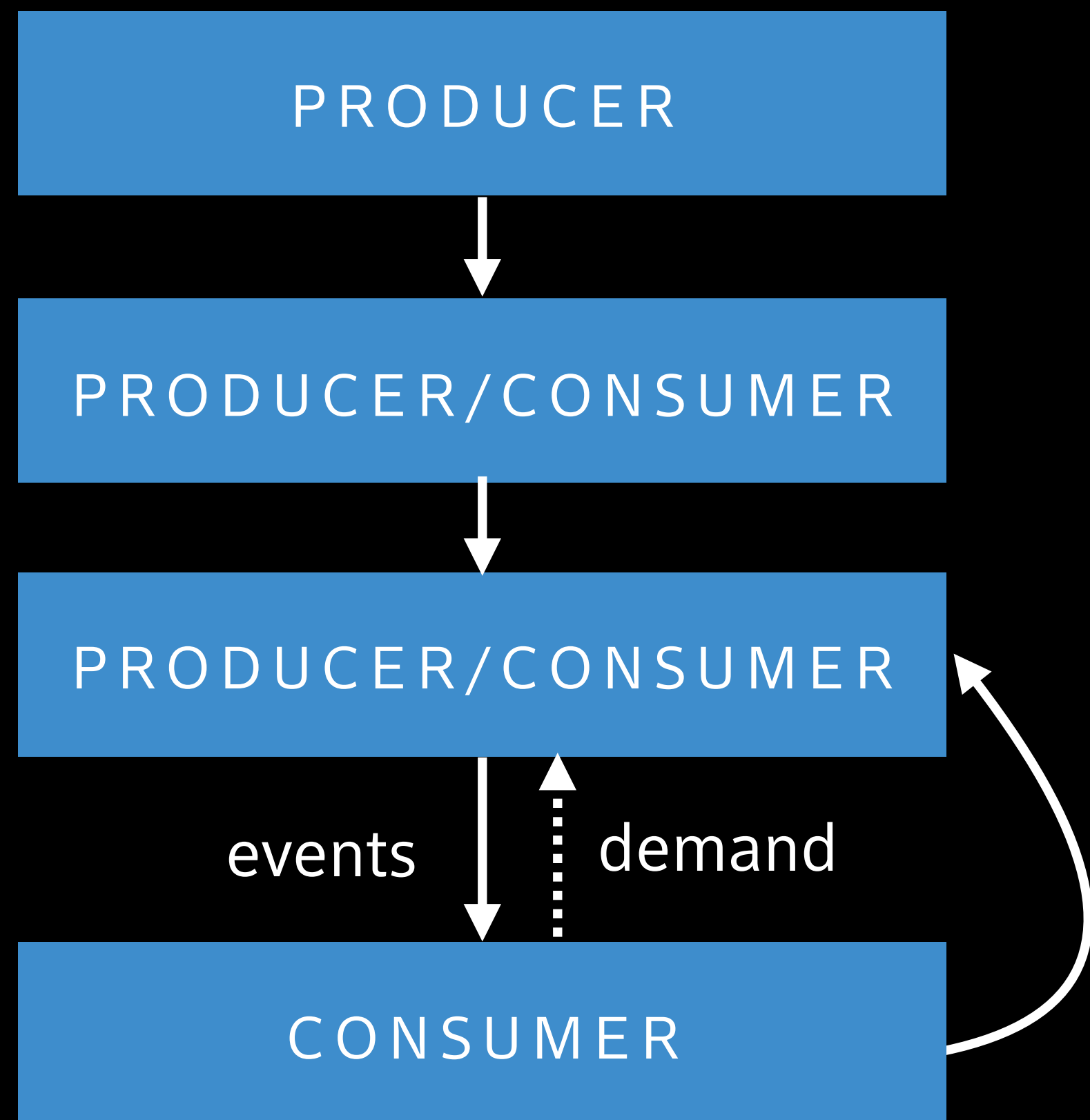
```
File.stream!("/tmp/words.txt")  
|> Flow.from_enumerable()  
|> Flow.flat_map(&String.split/1)  
|> Flow.partition()
```

ELIXIR GENSTAGE & FLOW



```
File.stream!("/tmp/words.txt")  
|> Flow.from_enumerable()  
|> Flow.flat_map(&String.split/1)  
|> Flow.partition()  
|> Flow.reduce(&Map.new/0,  
              &(Map.update(&2, &1, 0,  
                           fn c -> c+1 end)))
```

ELIXIR GENSTAGE & FLOW



```
File.stream!("/tmp/words.txt")  
|> Flow.from_enumerable()  
|> Flow.flat_map(&String.split/1)  
|> Flow.partition()  
|> Flow.reduce(&Map.new/0,  
              &(Map.update(&2, &1, 0,  
                          fn c -> c+1 end)))  
|> Enum.to_list()
```

subscribe with
demand

SEEKING DOMAIN-SPECIFIC ABSTRACTIONS

METRICS-FOCUSED STREAM COMBINATORS

SEEKING DOMAIN-SPECIFIC ABSTRACTIONS

METRICS-FOCUSED STREAM COMBINATORS

```
where(type: ["disk", "free", "percent"])
```

SEEKING DOMAIN-SPECIFIC ABSTRACTIONS

METRICS-FOCUSED STREAM COMBINATORS

```
where(type: ["disk", "free", "percent"])  
|> by([:host, :mount])
```


SEEKING DOMAIN-SPECIFIC ABSTRACTIONS

METRICS-FOCUSED STREAM COMBINATORS

```
where(type: ["disk", "free", "percent"])  
|> by([:host, :mount])  
|> ewma
```

SEEKING DOMAIN-SPECIFIC ABSTRACTIONS

METRICS-FOCUSED STREAM COMBINATORS

```
where(type: ["disk", "free", "percent"])  
|> by([:host, :mount])  
|> ewma  
|> threshold(below: 10.0)
```

SEEKING DOMAIN-SPECIFIC ABSTRACTIONS

METRICS-FOCUSED STREAM COMBINATORS

```
where(type: ["disk", "free", "percent"])  
|> by([:host, :mount])  
|> ewma  
|> threshold(below: 10.0)  
|> forward(:on_call_alert)
```

METRICS-FOCUSED STREAM COMBINATORS

```
where(type: ["disk", "free", "percent"])  
|> by([:host, :mount])  
|> ewma  
|> threshold(below: 10.0)  
|> forward(:on_call_alert)  
|> draw(:table)
```

AUTOMATIC WRITE-ATTENUATION BY

SEGMENTING PIPELINES FOR RE-USE

AUTOMATIC WRITE-ATTENUATION BY

SEGMENTING PIPELINES FOR RE-USE

```
where(type: ["disk", "free", "percent"])  
|> by([:host, :mount])  
|> ewma  
|> threshold(below: 10.0)  
|> forward(:on_call_alert)  
|> draw(:table)
```

AUTOMATIC WRITE-ATTENUATION BY

SEGMENTING PIPELINES FOR RE-USE

```
where(type: ["disk", "free", "percent"])
|> by([:host, :mount])
|> ewma
|> threshold(below: 10.0)
|> forward(:on_call_alert)
|> draw(:table)
```

AUTOMATIC WRITE-ATTENUATION BY

SEGMENTING PIPELINES FOR RE-USE

```
where(type: ["disk", "free", "percent"])
```

```
|> by([:host, :mount])
```

```
|> ewma
```

```
|> threshold(below: 10.0)
```

```
|> forward(:on_call_alert)
```

```
|> draw(:table)
```

```
where(type: ["disk", "free", "percent"])
```

```
|> by([:host, :mount])
```

```
|> ewma
```

```
|> history(minutes: 30)
```


AUTOMATIC WRITE-ATTENUATION BY

SEGMENTING PIPELINES FOR RE-USE

```
where(type: ["disk", "free", "percent"])
```

```
|> by([:host, :mount])
```

```
|> ewma
```

```
|> threshold(below: 10.0)
```

```
|> forward(:on_call_alert)
```

```
|> draw(:table)
```

```
where(type: ["disk", "free", "percent"])
```

```
|> by([:host, :mount])
```

```
|> ewma
```

```
|> history(minutes: 30)
```

AUTOMATIC WRITE-ATTENUATION BY

SEGMENTING PIPELINES FOR RE-USE

```
where(type: ["disk", "free", "percent"])
```

```
|> by([:host, :mount])
```

```
|> ewma
```

```
|> threshold(below: 10.0)
```

```
|> forward(:on_call_alert)
```

```
|> draw(:table)
```



```
|> history(minutes: 30)
```



LESSONS LEARNED

RETROSPECTIVE

RETROSPECTIVE

- ✓ Literate programs make for **good collaboration**

RETROSPECTIVE

- ✓ Literate programs make for **good collaboration**
- ✗ Our vision is **ahead of the organization**

RETROSPECTIVE

- ✓ Literate programs make for **good collaboration**
- ✗ Our vision is **ahead of the organization**
- ✗ Stream processing is **just the means, not the end**

RETROSPECTIVE

- ✓ Literate programs make for **good collaboration**
- ✗ Our vision is **ahead of the organization**
- ✗ Stream processing is **just the means, not the end**
- ✗ Much more **research is needed**

REFERENCES

<https://git.io/vQIgp>



extra slides

WHY ELIXIR?



WHY ELIXIR?

✓ Familiarity



WHY ELIXIR?

- ✓ Familiarity
- ✓ Meta-programming



WHY ELIXIR?

- ✓ Familiarity
- ✓ Meta-programming
- ✓ Transparent, low-latency runtime



WHY ELIXIR?

- ✓ Familiarity
- ✓ Meta-programming
- ✓ Transparent, low-latency runtime
- ✓ Our ops time/budget is small



WHY ELIXIR?

- ✓ Familiarity
- ✓ Meta-programming
- ✓ Transparent, low-latency runtime
- ✓ Our ops time/budget is small
- ✓ *Generic Erlang/Elixir pitch, "let it crash", etc*



“For small inputs ($\leq 0.5\text{GB}$), the Metis single-machine MapReduce system performs best. This matters, as small inputs are common in practice: **40–80% of Cloudera customers’ MapReduce jobs and 70% of jobs in a Facebook trace have $\leq 1\text{GB}$ of input.**”

MUSKETEER: ALL FOR ONE, ONE FOR ALL IN DATA PROCESSING SYSTEMS
GOG, ET AL, EUROSYS '15