ROUTING DATA AROUND THE WORLD WITH STREAM!

Steve Koelpin
Sr. Architect, TransUnion

Don Reilly
Sr. Manager, TransUnion

STEVE KOELPIN

Sr. Architect, TransUnion

DON REILLY

Sr. Architect, TransUnion

CCOE—Admin

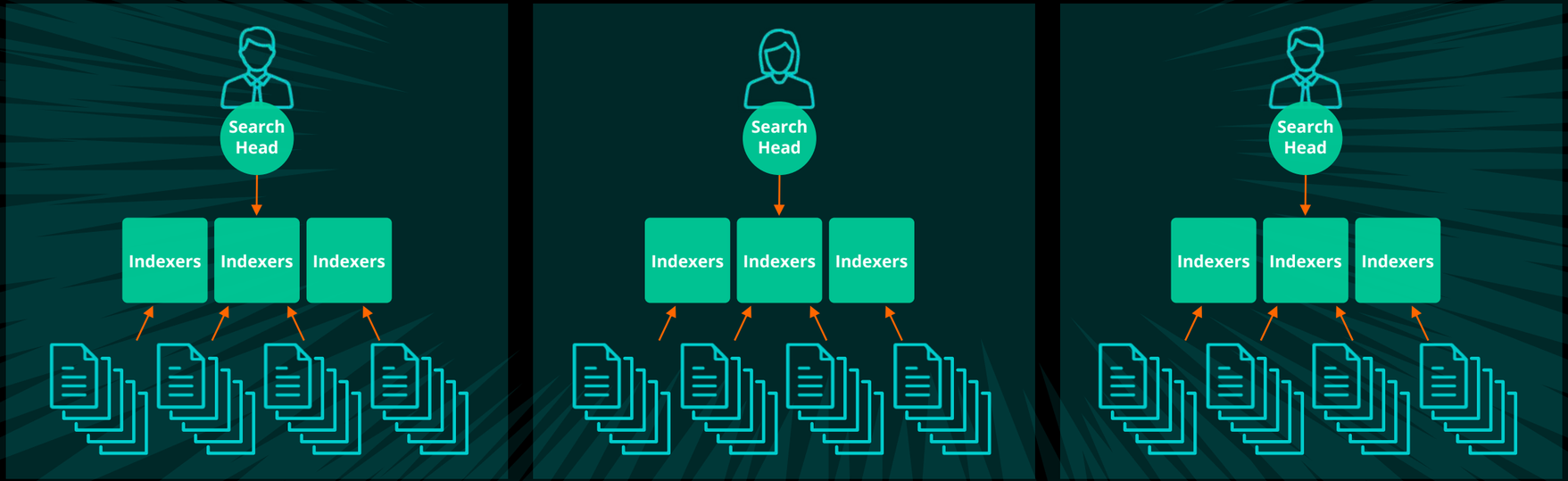
CRYPTOCON! 2023

TransUnion

The Content of this presentation is intended for general informational purposes only. TransUnion make no representation or warranty, express or implied, regarding the content of this presentation, including but not limited to warranties as to its accuracy, completeness, ownership, or fitness for a particular purpose. Your reliance on the contents of this presentation is solely at your own risk. TransUnion does not warrant, endorse, or assume liability for any of the content contained in links to third party sites included in this presentation.

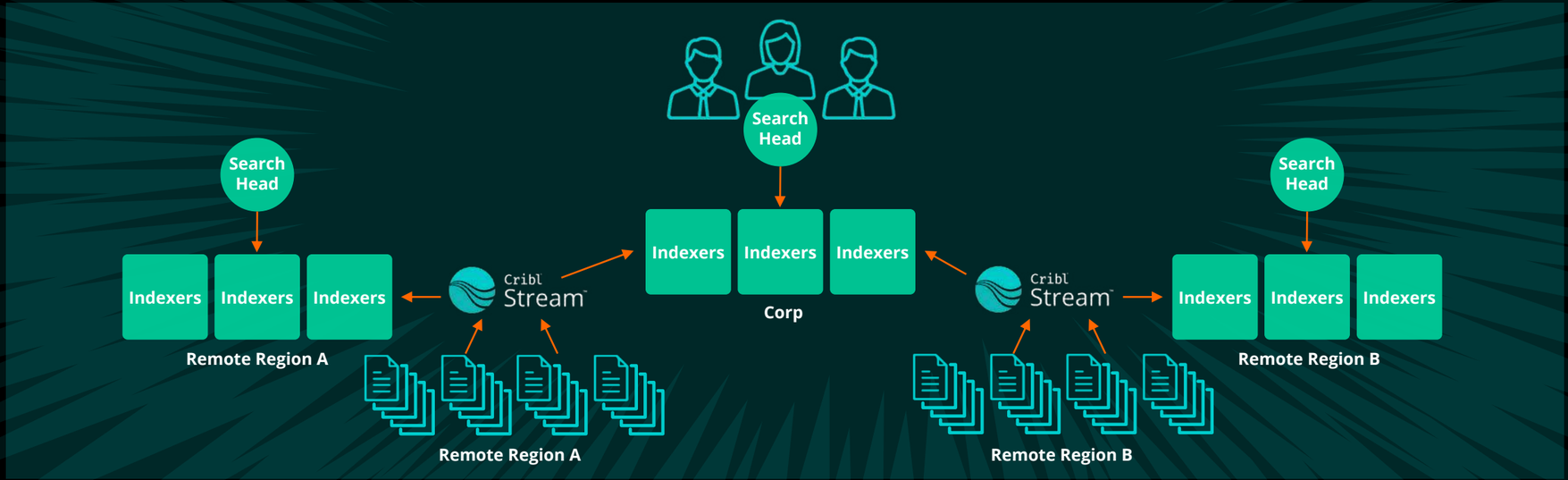
PROBLEM STATEMENT

We operate in **over 30 countries** around the world and lack visibility in the form of a single centralized view for our stakeholders and operators.



SOLUTION

Use Cribl Stream to aggregate and route data from remote regions into a centralized location.



TransUnion

AGGREGATING DATA ON THE EDGE

Send this...

Instead of sending this...

SIZE REDUCTION BY AGGREGATING

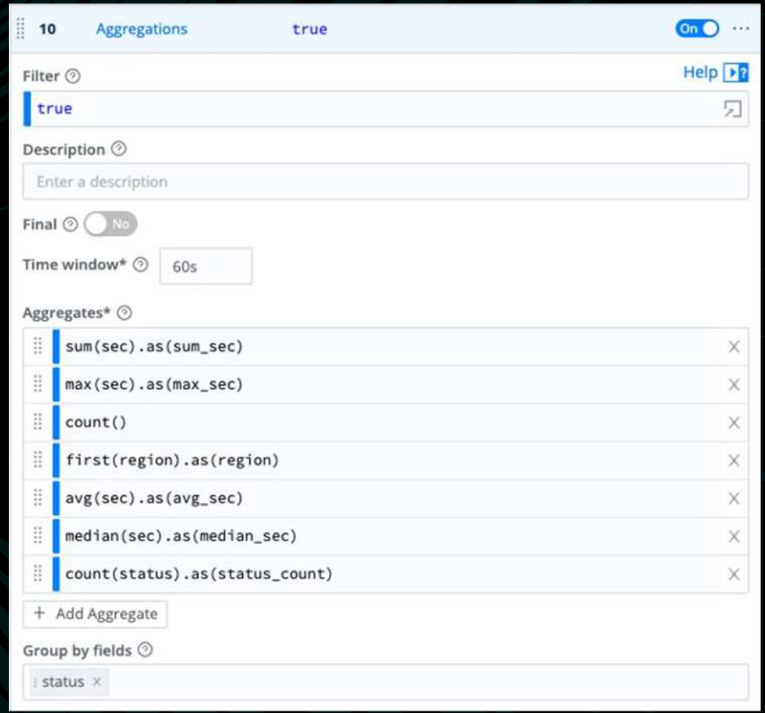
	_raw Length	Full Event Length	Number of Fields	Number of Events
IN	73.82KB	120.58KB	19	100
OUT	8.00B	1.38KB	16	4
DIFF	↓ -99.99%	↓ -98.85%	↓ -15.79%	↓ -96.00%

HOW AGGREGATION WORKS

Aggregating in Splunk

```
| bin _time span=60s
| stats max(resp_time) by _time, status
```

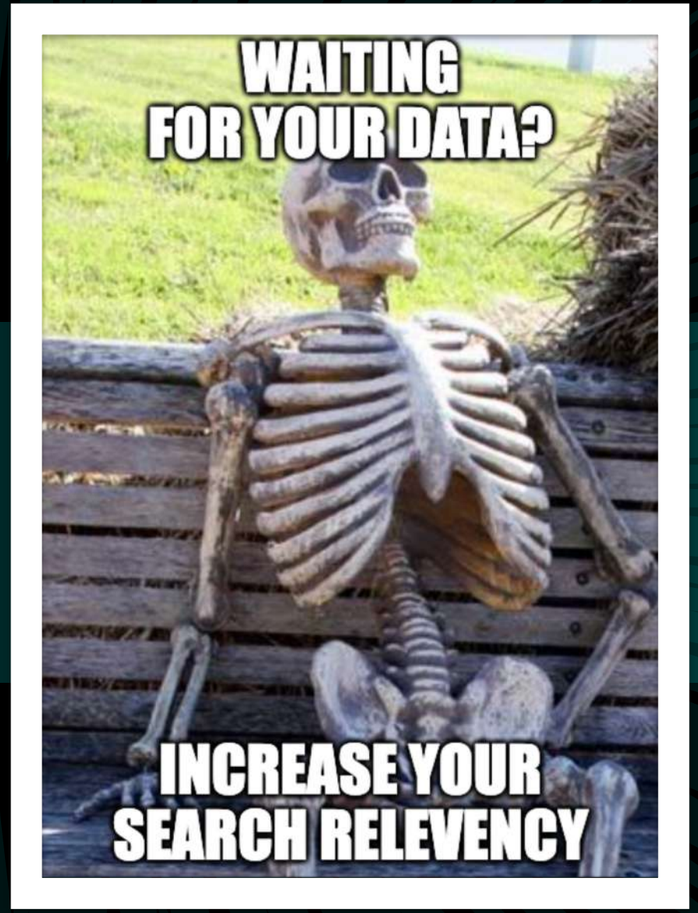
Aggregating in Stream



SEARCH RELEVANCE

Search Raw Events?

Index Relevant Data Points.



TO PUSH OR TO PULL

Distributed Search Against Data at Rest

PULL

Aggregating and Routing Data

PUSH

PULLING DATA

- Relatively easy to implement
- Don't need to understand format of data upfront
- Better suited for adhoc searching

PROS

- Increased network load
- Impossible to achieve 100% Search Relevance
- Unpredictable search run times

CONS

PUSHING DATA

- Better suited for recurrent searching
- Uniform data structures
- 100% search relevancy
- Predictable search run times
- Reduced network load

PROS

- Requires understanding of data usage
- More planning required for implementation
- Changes will not work retroactively
- Slight increase of Ingest & Storage

CONS

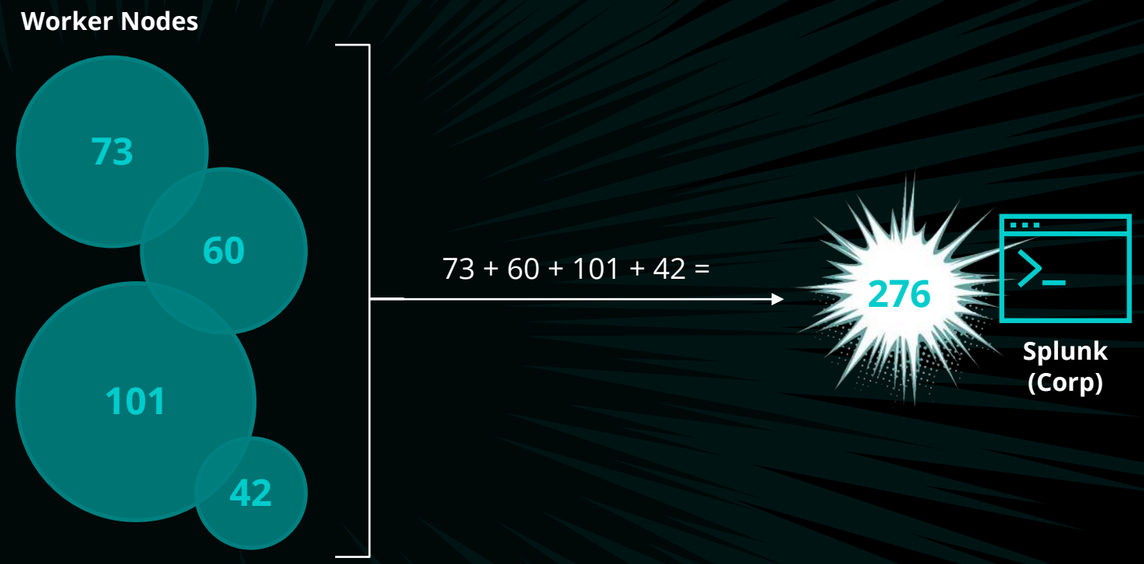
WORKER NODES HAVE THEIR INDEPENDENCE

Problem

We're aggregating on one of many nodes per region

Solution

Use a second layer of aggregation (i.e., scheduled search) to further aggregate and push to a summary index



Average Formula = $\frac{\text{Total Sum of All Numbers}}{\text{Number of Items in the Set}}$

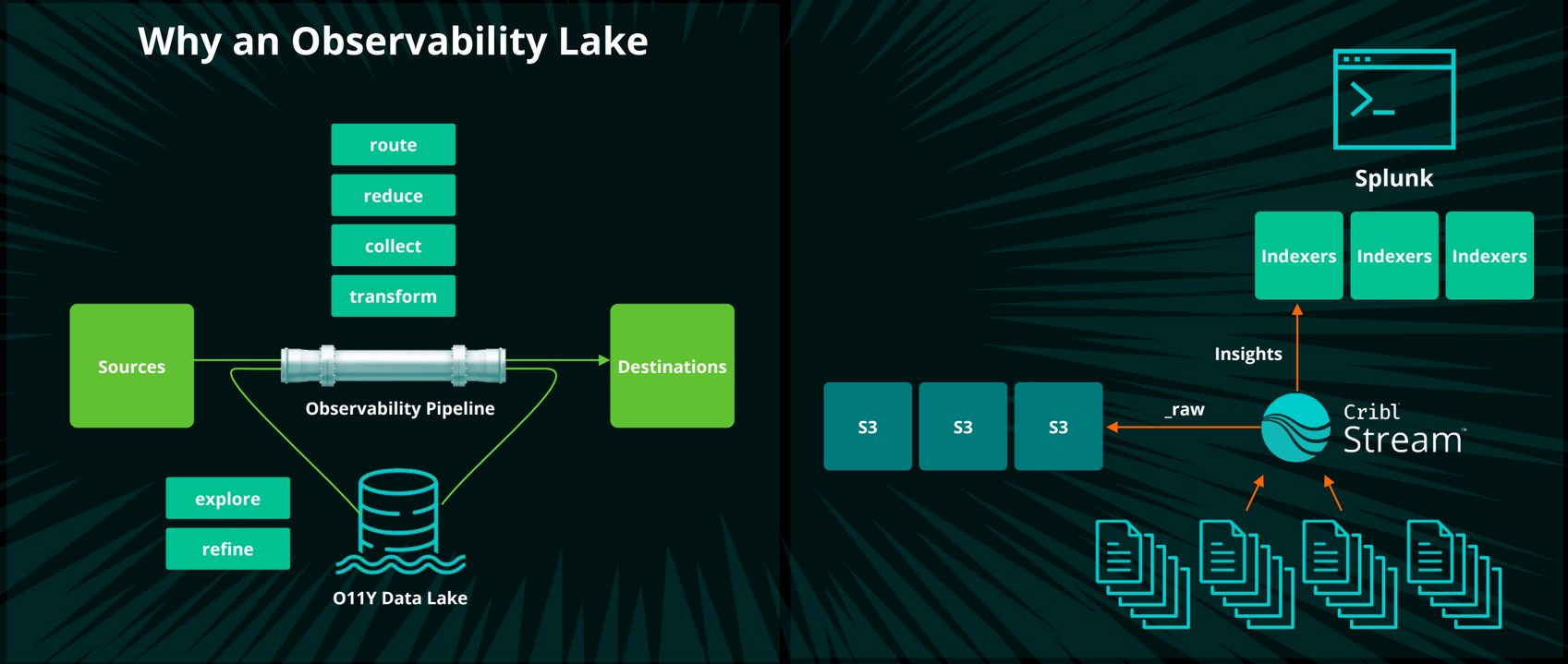

DecisionEdge Global Health
Decision Edge Global View
Select App
United States
USA
India
South Africa
UK
Hong Kong
Philippines
Columbia
Canada
DR
100
Volume
Response Time
Error Rate
Time
13:29:11
Past Day
Changes
0
Past Day
Incidents
55
Reponse Time usa
Lower
Upper
Baseline
Actual
Trending HealthScores
Errors
Volume
Response_Time
ServiceHealthScore
Error Count usa
2.956030714032111
0.23
Notifications
Incident Number
Description
AG
Urgency
TimeOfEvent
Volume usa

TransUnion

FUTURE STATE

Ingest the insights and store the _raw

- Significantly reduced license costs
- Fast search-time query speeds
- Lower storage costs



PUSHING DATA

- Better suited for recurrent searching
- Uniform data structures
- 100% search relevancy
- Predictable search run times
- Reduced network load
- Massive decrease of Storage & License!

PROS

- Requires understanding of data usage
- More planning required for implementation
- Changes will not work retroactively
- ~~Slight increase of Ingest & Storage~~

CONS

RESOURCES

Aggregating Data https://cribl.io/blog/search-observability-data/

Cribl Search https://cribl.io/blog/search-observability-data/

O11Y Data Lake https://cribl.io/blog/observability-lake/

THANK YOU.

CRYPTOCON! 2023

TransUnion