




FROM BATCH TO STREAMS: BUILDING VALUE FROM DATA IN-MOTION

RICARDO FERREIRA



@RIFERREI

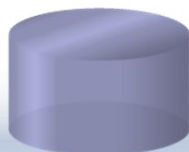
A close-up, high-contrast photograph of a man's face, likely from the movie 'The Revenant'. The man has a thick, dark beard and intense, weathered features. His expression is somber and contemplative. The lighting is dramatic, with deep shadows and bright highlights on his skin.

THE WORLD IS
NOT THE SAME
AS IT WAS

TRADITIONAL APPROACHES TO RESPONSE



DATA
CAPTURE



DATA
STORAGE



DATA
ANALYSIS




ACTION

WHAT HAS
ALREADY
HAPPENED

- EXISTING ARCHITECTURES AREN'T DESIGNED TO SENSE AND RESPOND TO REAL-TIME EVENTS
- TIME AND SPACE ANALYSIS IS TOO COMPLICATED OR NOT POSSIBLE
- LEADS TO "PAST-TENSE" DECISION MAKING

TRADITIONAL DECISION MANAGEMENT

WITH TRADITIONAL DECISION MAKING SOLUTIONS
WE CAN ONLY RESPOND TO CHANGES
IF THEY'VE ALREADY OCCURRED



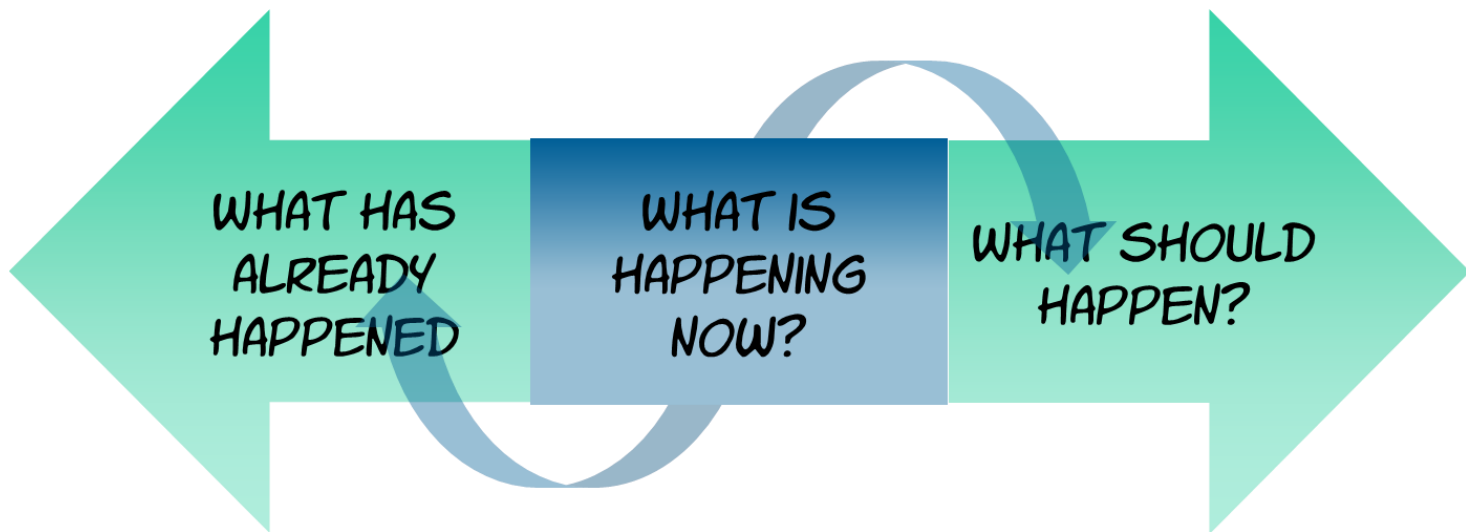
WHAT HAS
ALREADY
HAPPENED



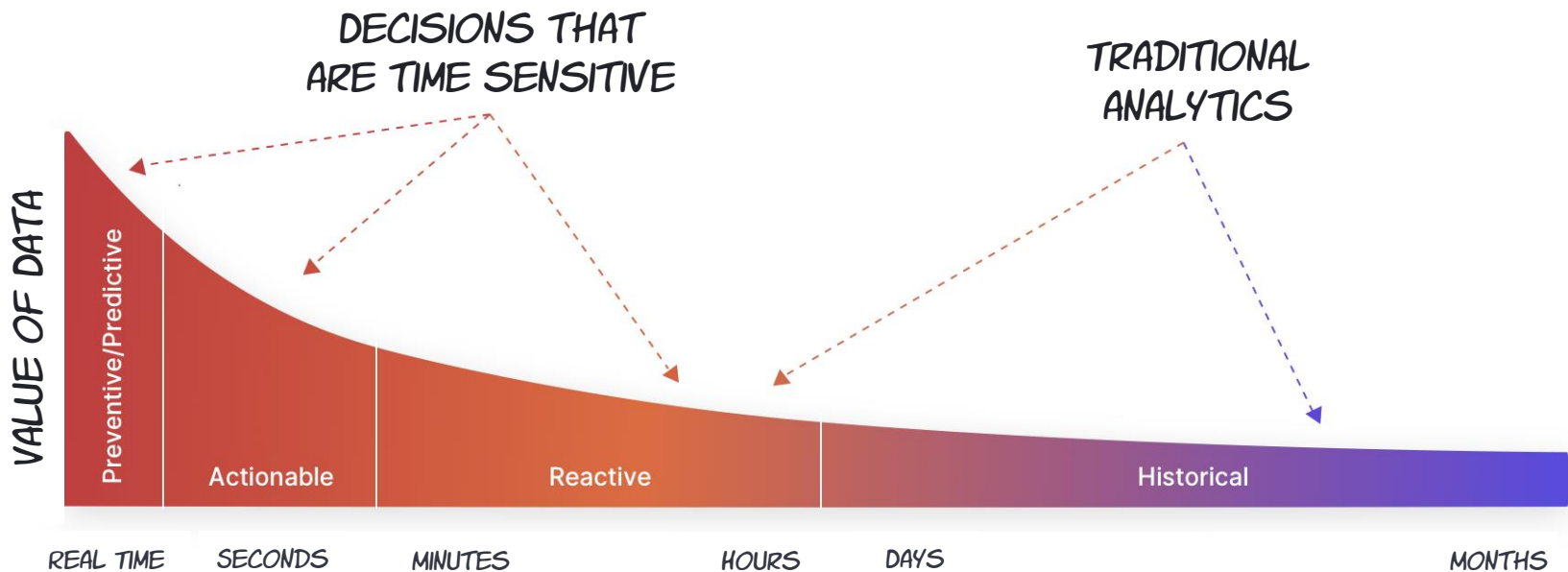
WHAT SHOULD
HAPPEN?

WHY STREAMING DATA?

EVENT PROCESSING PROVIDES THE
SITUATION AWARENESS
TO KNOW WHAT'S HAPPENING NOW



WHY STREAMING DATA?



SOURCE: PERISHABLE INSIGHTS, FORRESTER



RICARDO FERREIRA

DEVELOPER ADVOCATE 🥑

❑ ELASTIC COMMUNITY TEAM 

❑ HASHICORP AMBASSADOR 

❑ BEFORE JOINING ELASTIC:

CONFLUENT, ORACLE, RED HAT

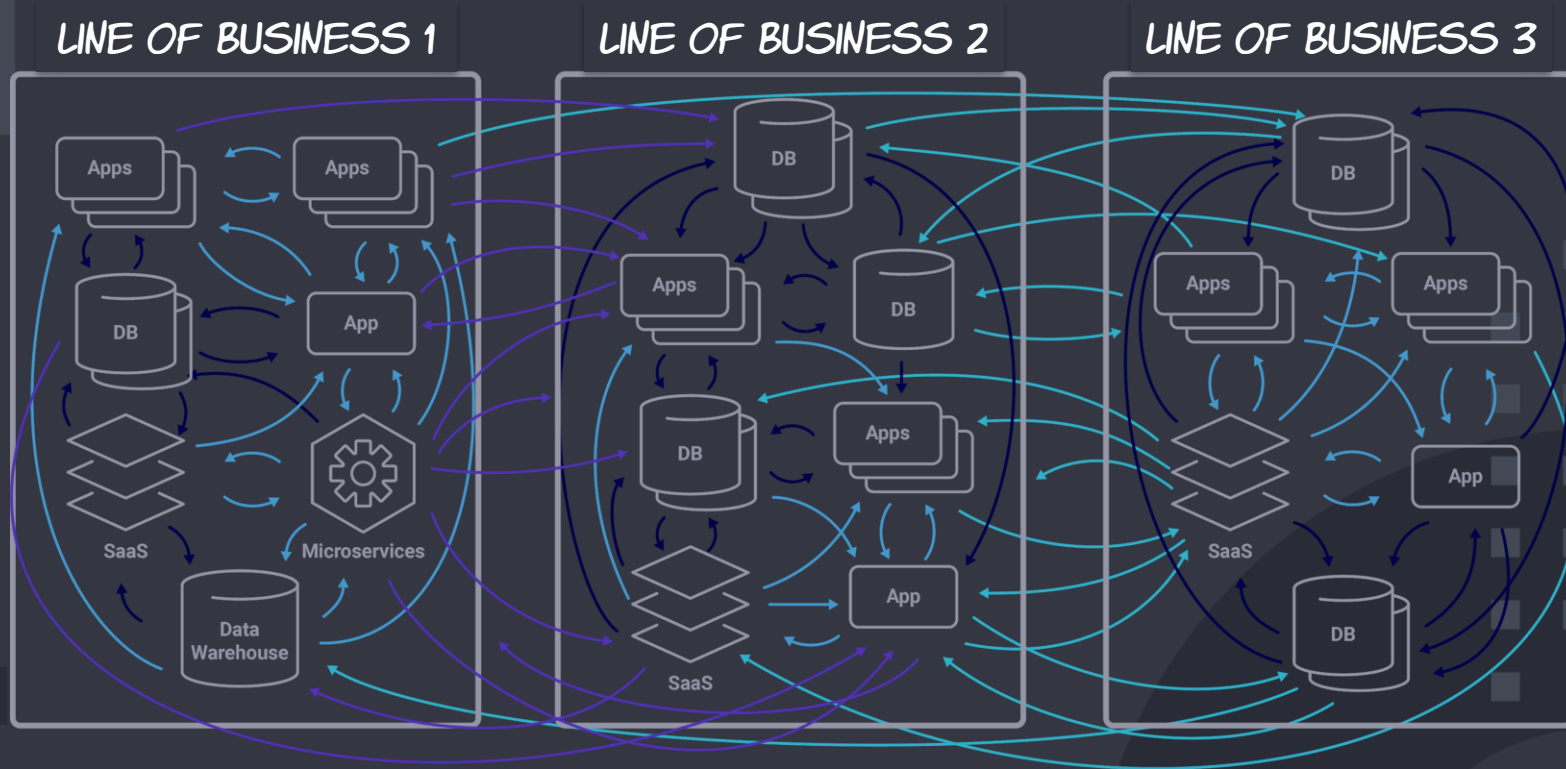
❑ DISTRIBUTED SYSTEMS, BIG DATA,
STREAMING DATA, DATABASES

❑ RIFERREI@ELASTIC.CO

❑ RIFERREI@RIFERREI.COM

THE WORLD
TODAY AS-IS

SILOED-BASED SYSTEMS ARE FRAGILE



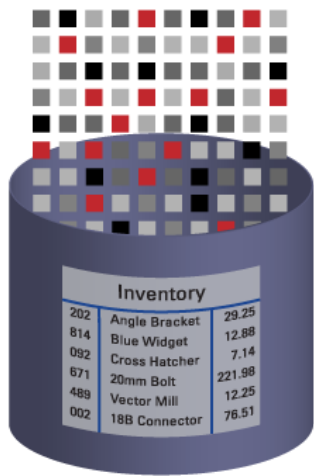
BATCH IS LATENCY BUILT-IN



PLUMBING IS HARD

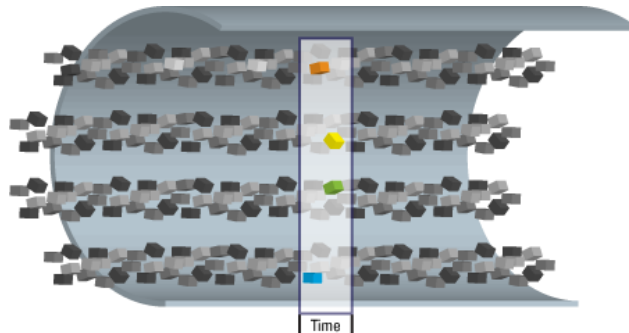


WHY STREAMING DATA?



TRADITIONAL ANALYTICS

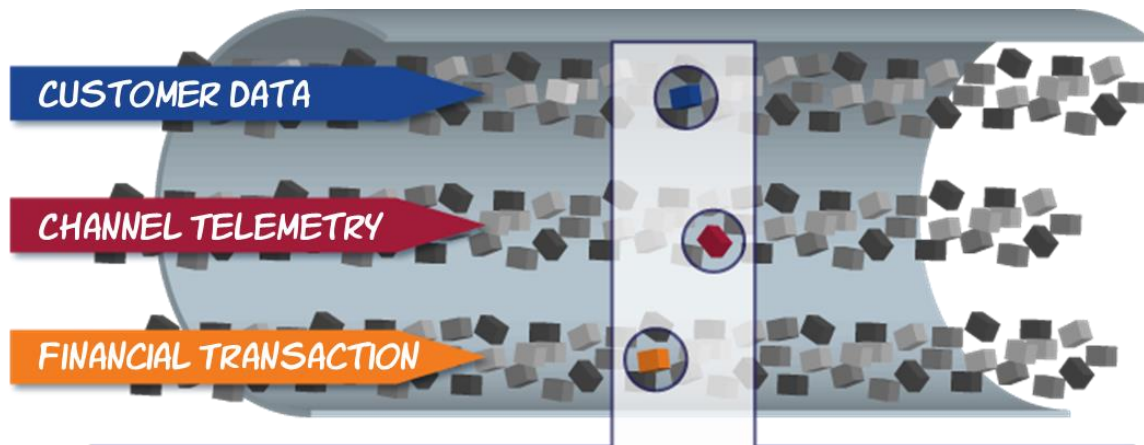
- DATA GOES INTO THE DATABASE
- ANALYZED AFTER IT IS STORED



STREAMING DATA

- EVENTS GO THROUGH
- ANALYZED AS THEY HAPPEN

STREAMING DATA = BETTER EXPERIENCE

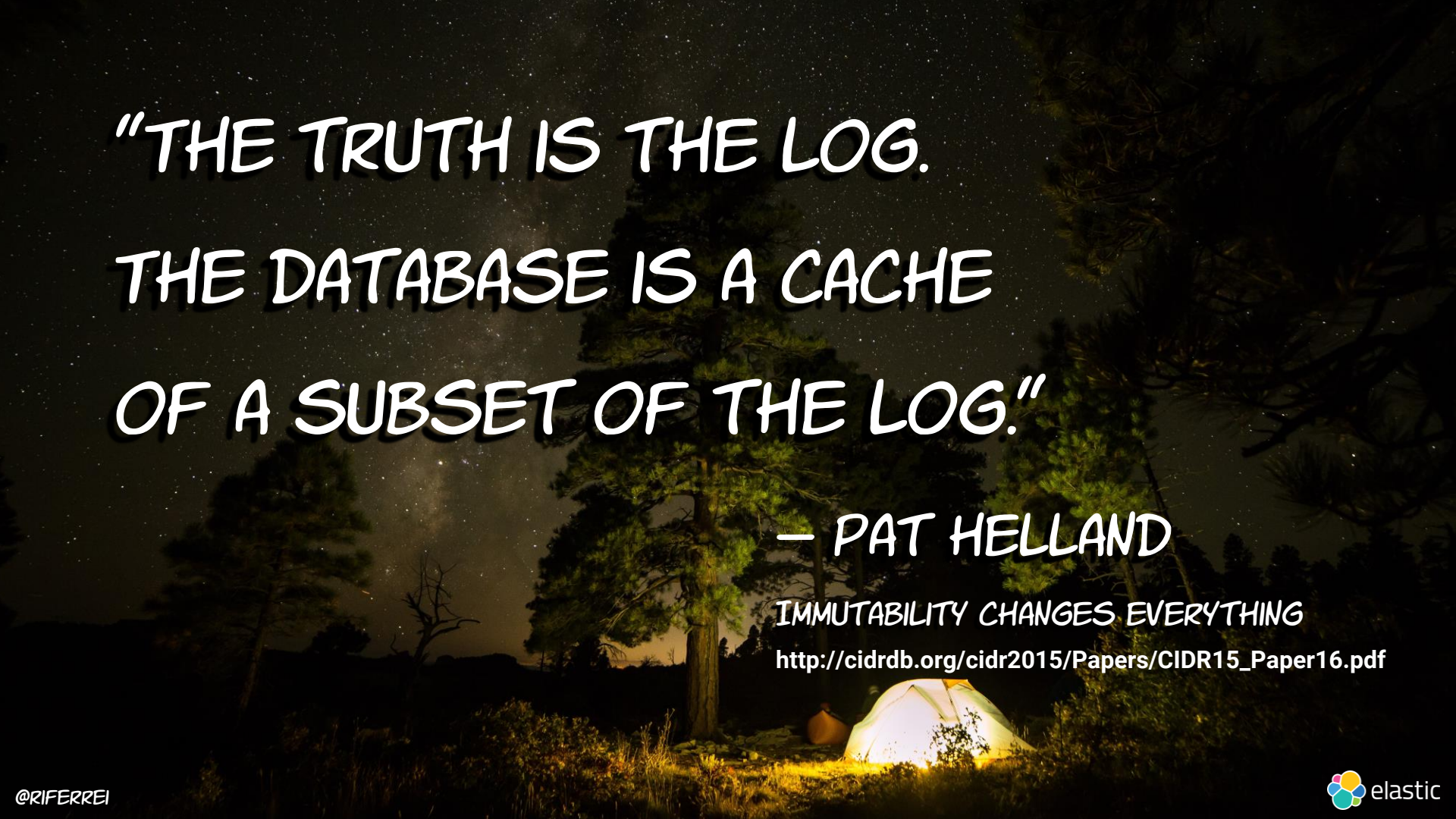


IF THE CHANNEL DEVICE IS NOT AVAILABLE AND THE CUSTOMER IS TRYING TO COMPLETE AN ONLINE TRANSACTION, THEN RETRIEVE A LIST OF OTHER DEVICE OPTIONS AND PRESENT IT BACK TO THE CUSTOMER AS QUICKLY AS POSSIBLE.

STREAMING DATA JOURNEY

COMMIT LOG



A night scene with a tent and trees under a starry sky. The background is a dark, starry night sky with the Milky Way visible. In the foreground, there are silhouettes of trees and a small, illuminated tent on the right side. The text is overlaid on the sky and trees.

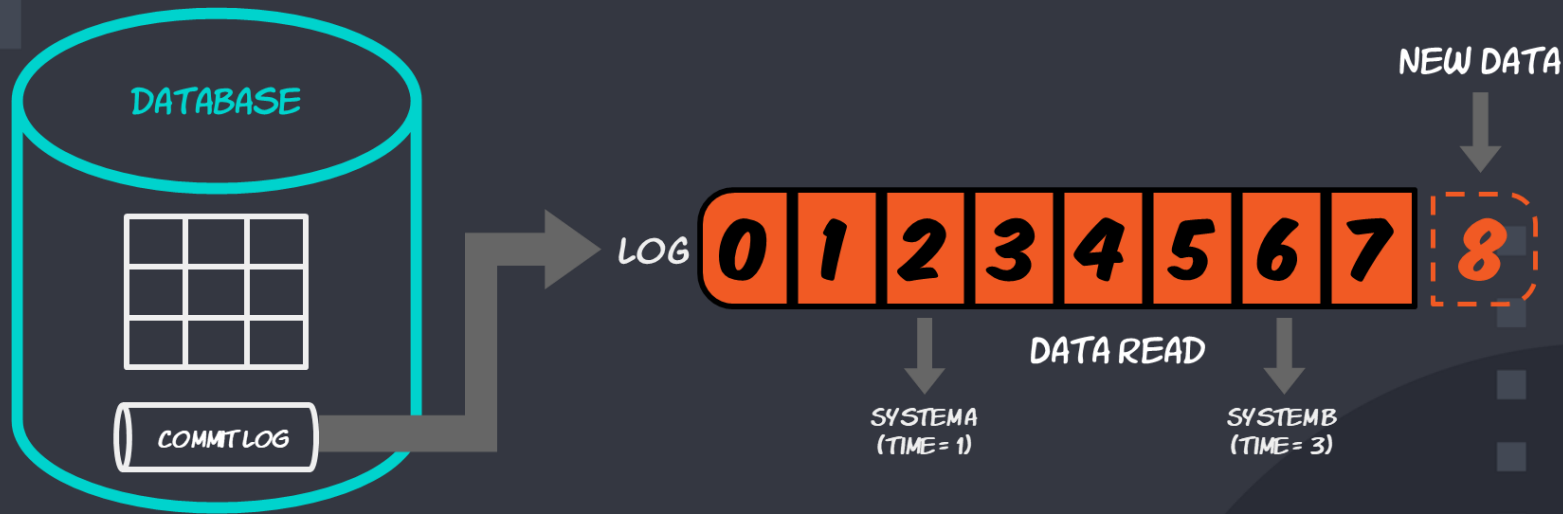
"THE TRUTH IS THE LOG.
THE DATABASE IS A CACHE
OF A SUBSET OF THE LOG."

— PAT HELLAND

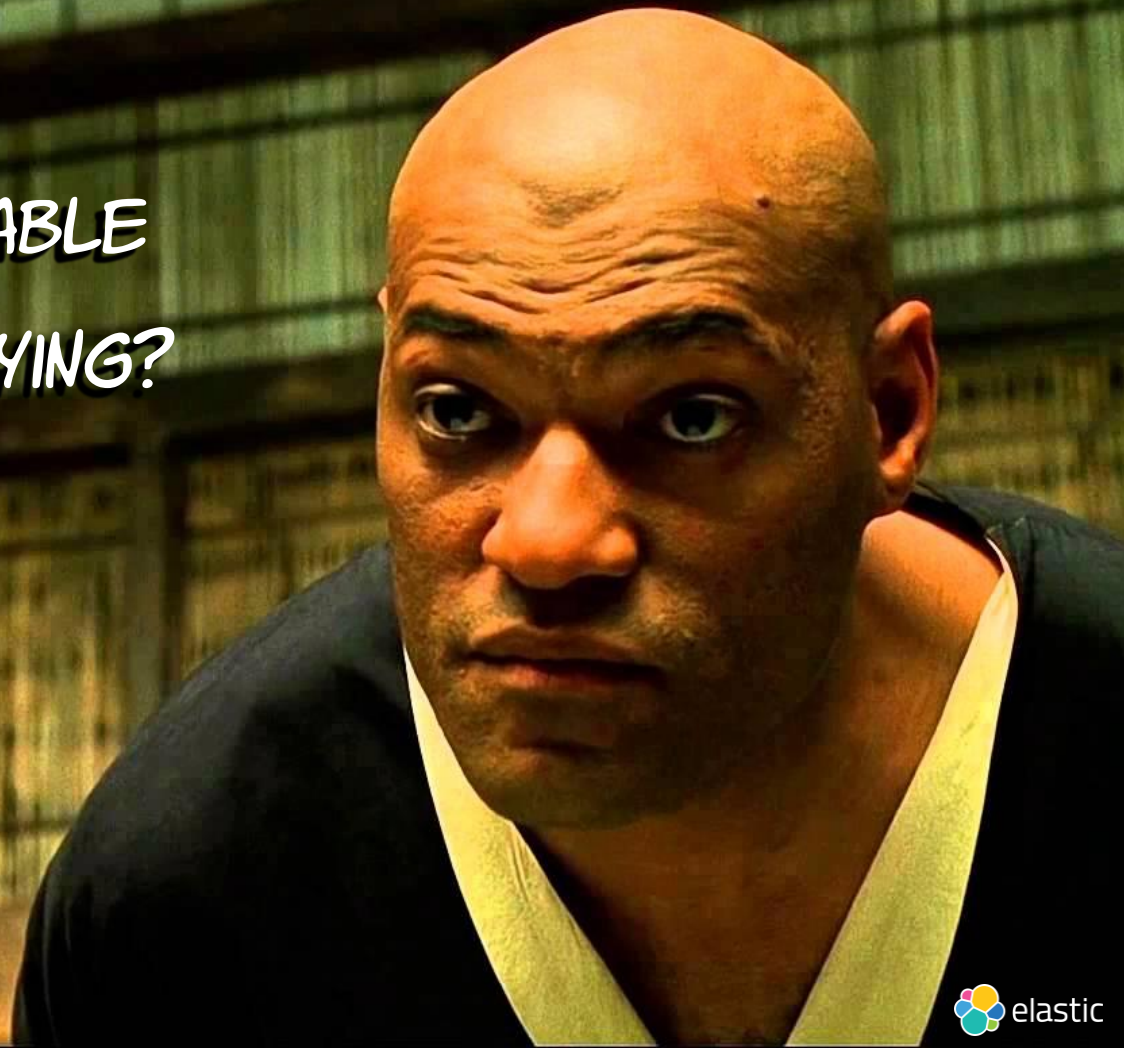
IMMUTABILITY CHANGES EVERYTHING

http://cidrdb.org/cidr2015/Papers/CIDR15_Paper16.pdf

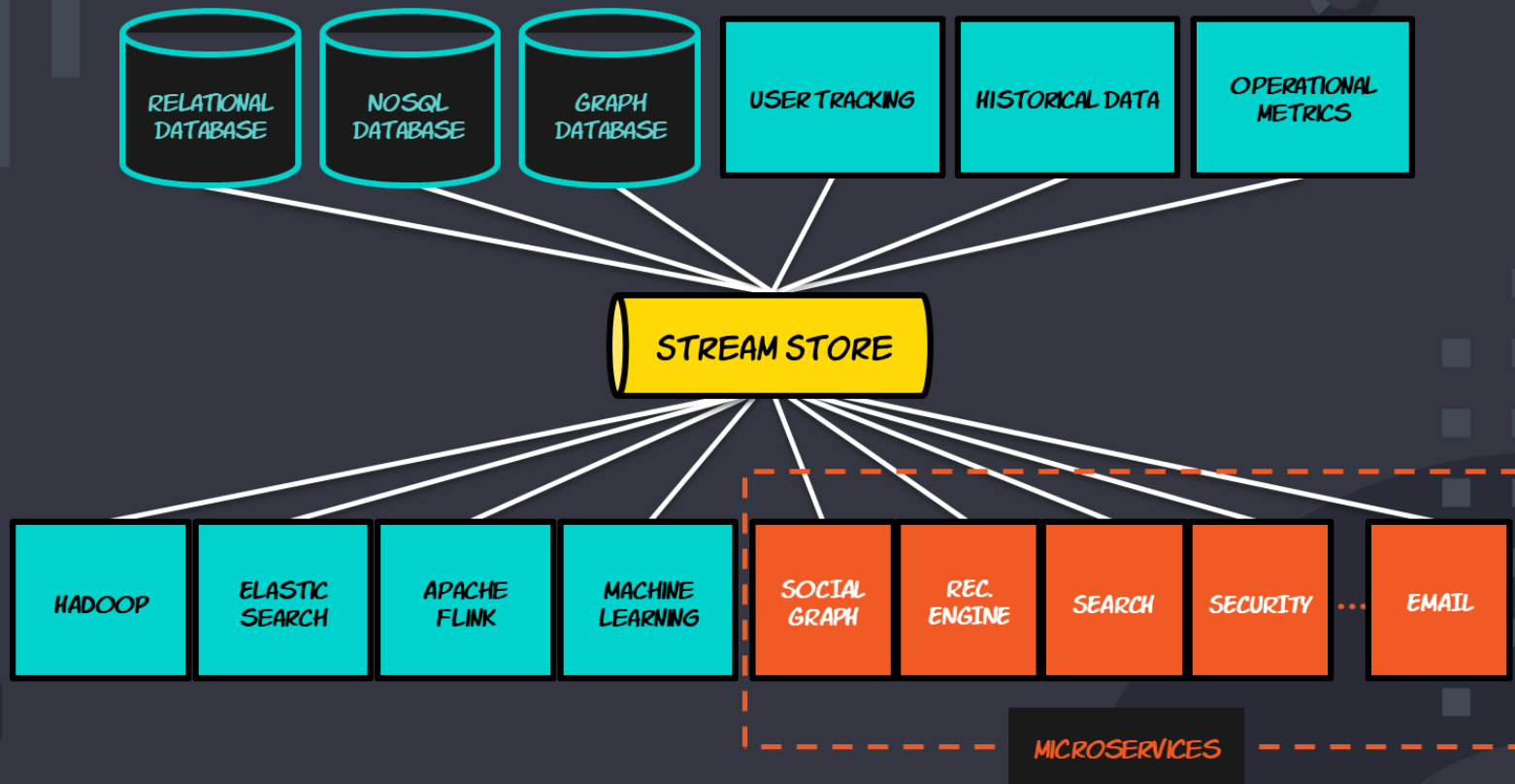
EVERYTHING IS A DATA STREAM



DO YOU THINK IS A TABLE
THAT YOU ARE QUERYING?

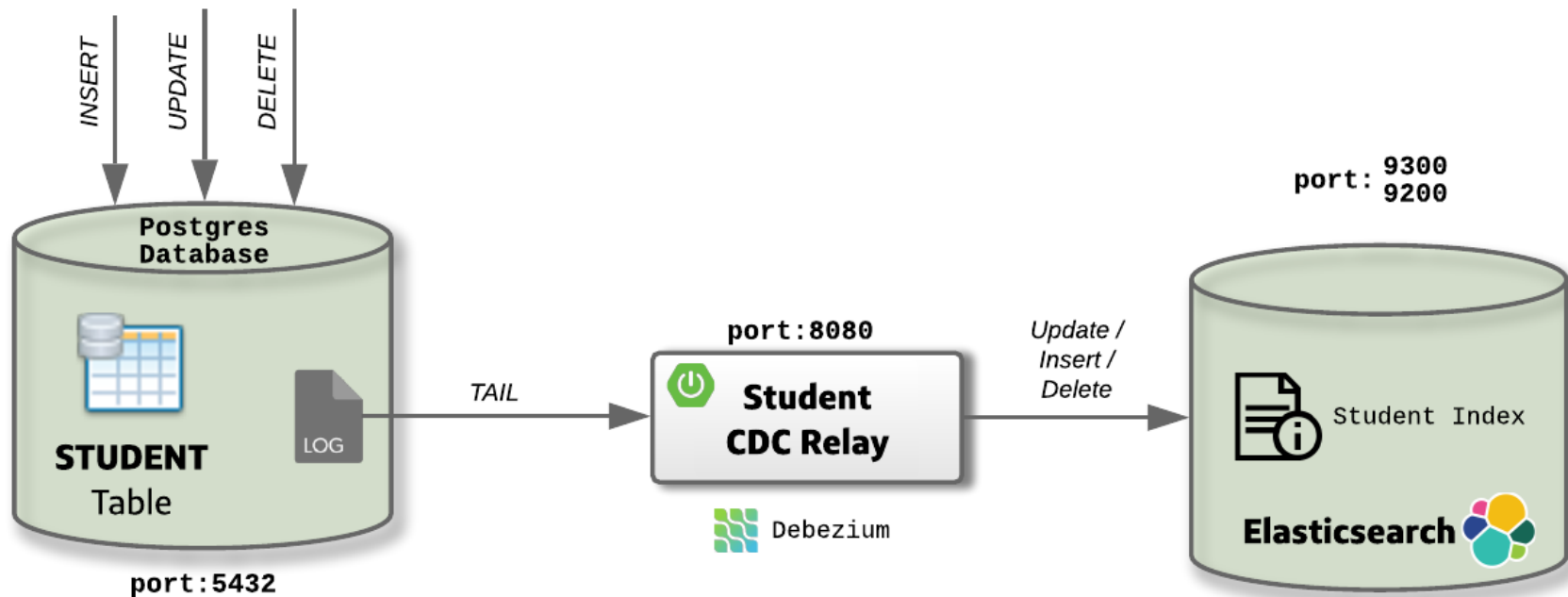


MEET THE STREAM STORE



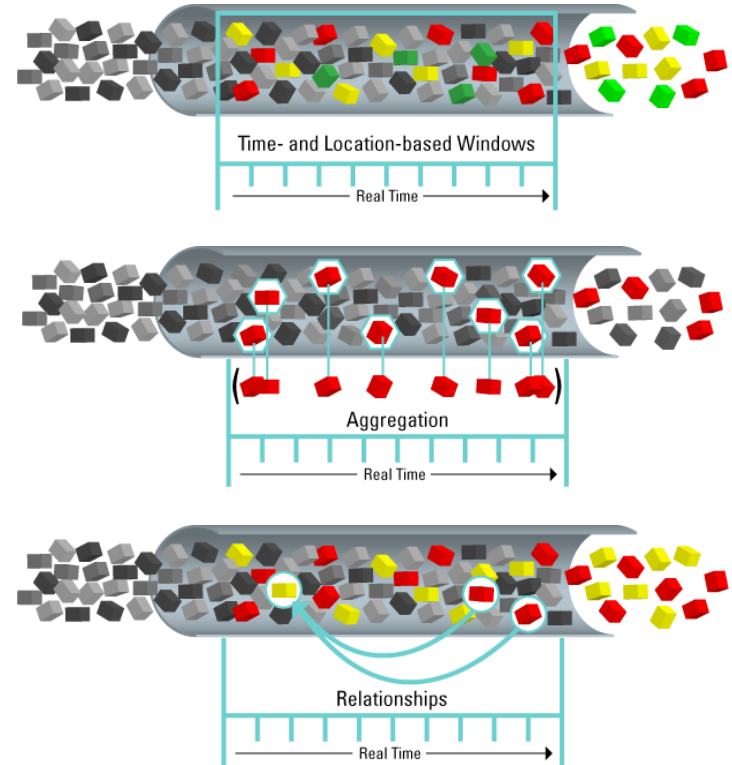
LESSONS LEARNED

TRANSFORM TRANSACT DATA TO STREAMS

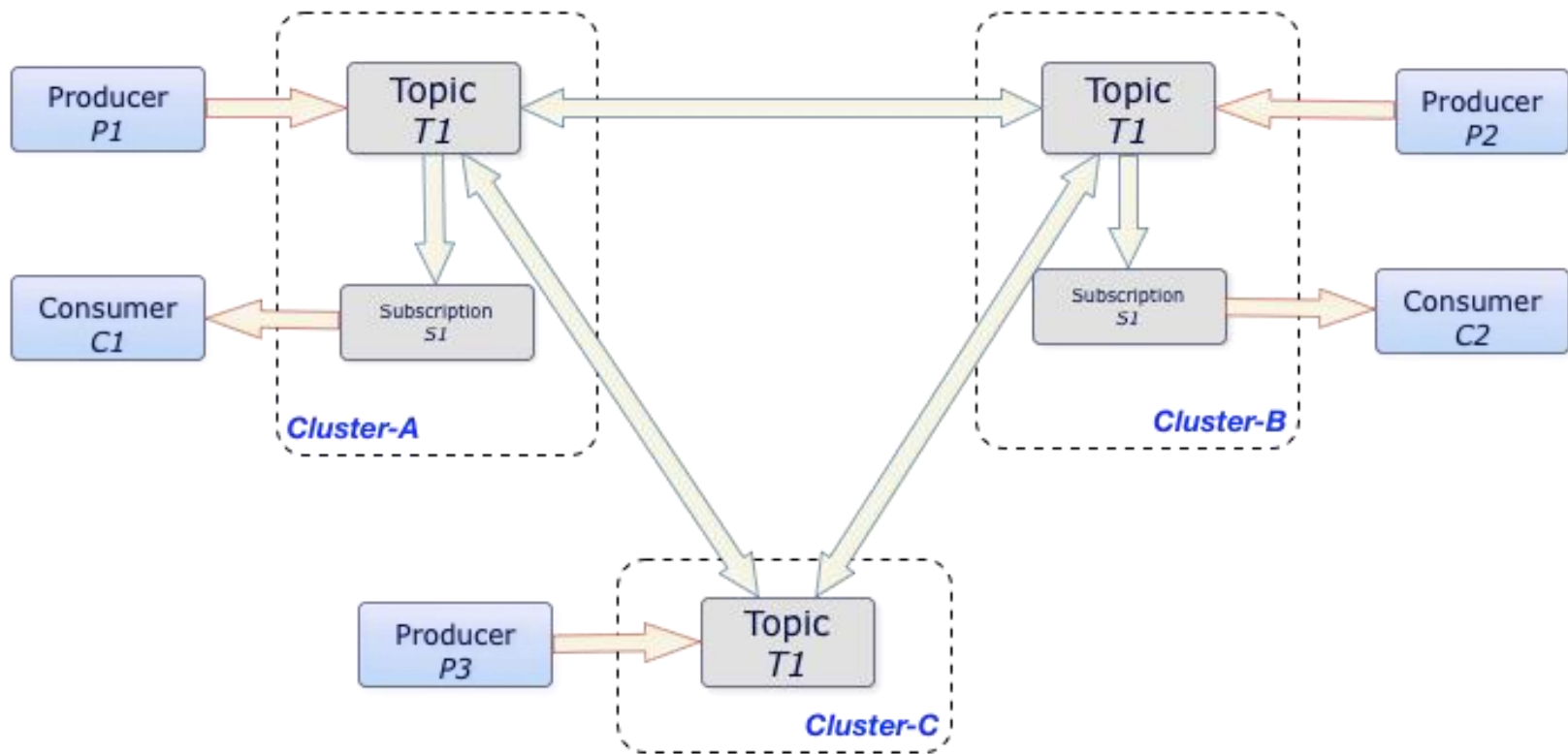


LEVERAGE DIFFERENT CONSTRUCTS

- TIME- AND TEMPORAL WINDOWS
 - WITHIN, NEAR, ETC. BASED IN REAL-TIME CONTEXT
- AGGREGATION
 - ACCUMULATION OF VALUES OR QUANTITY
 - SUM, AVERAGE, MIN, MAX, ETC.
- RELATIONSHIPS
 - EVENT A CAUSED EVENT B TO HAPPEN (CAUSALITY)



GEO REPLICATION AND MULTI-TENANCY



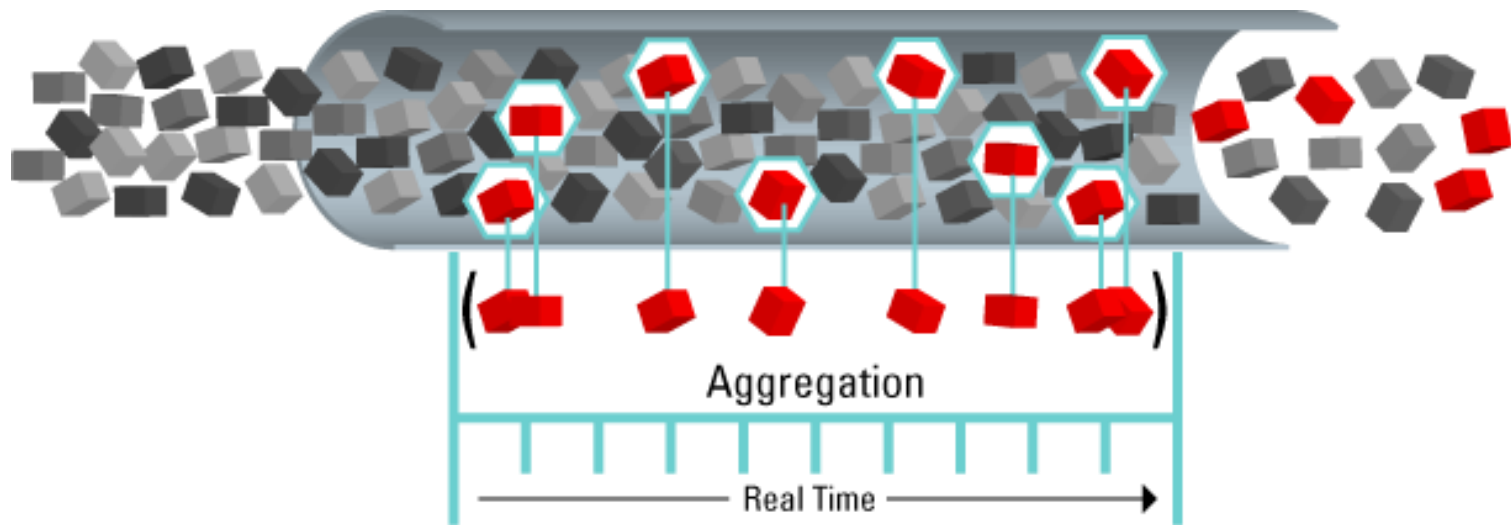
BUILD VERSUS BUY ANALYSIS

- INFRASTRUCTURE OWNERSHIP
- DATA GRAVITY EVALUATION
- PRODUCT COMPLETENESS

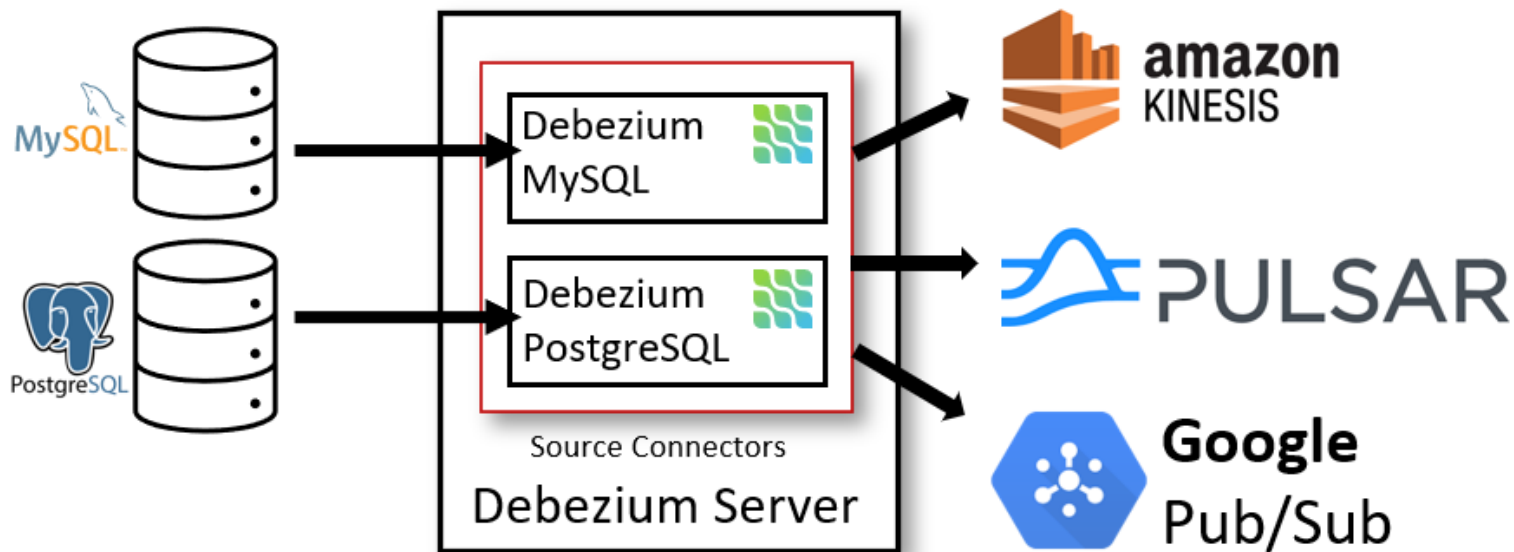


WHAT TRAPS TO AVOID

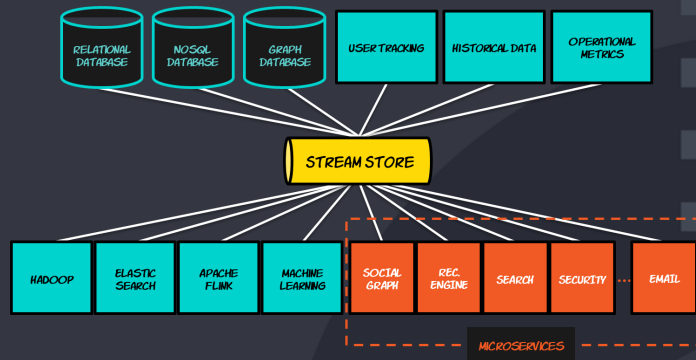
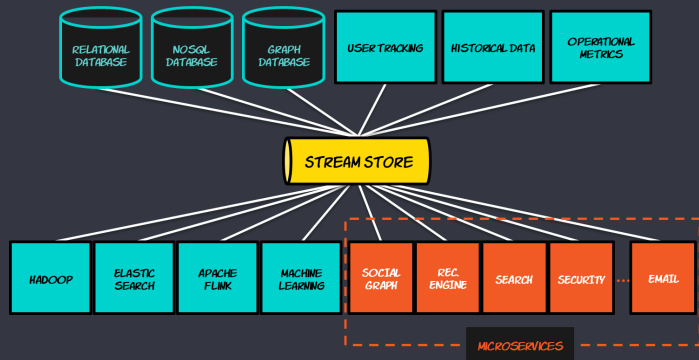
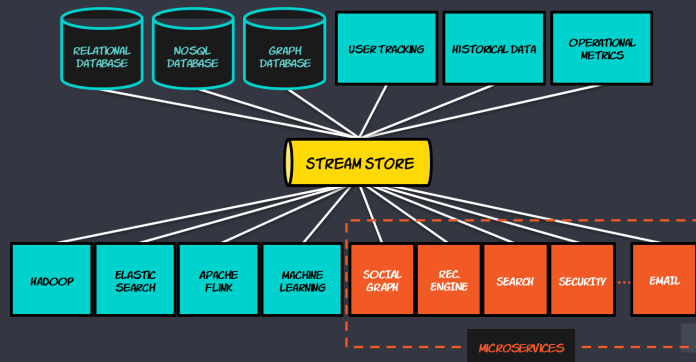
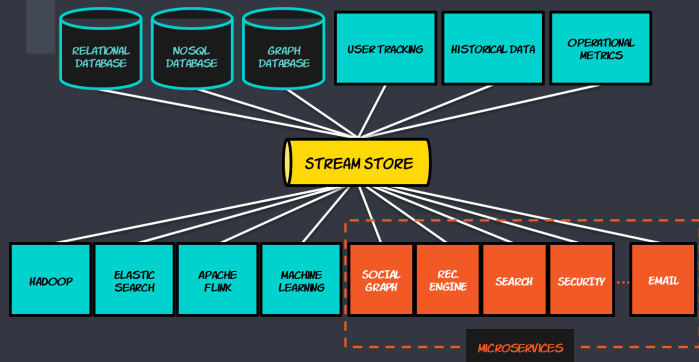
USING STREAMS ONLY TO AGGREGATE DATA



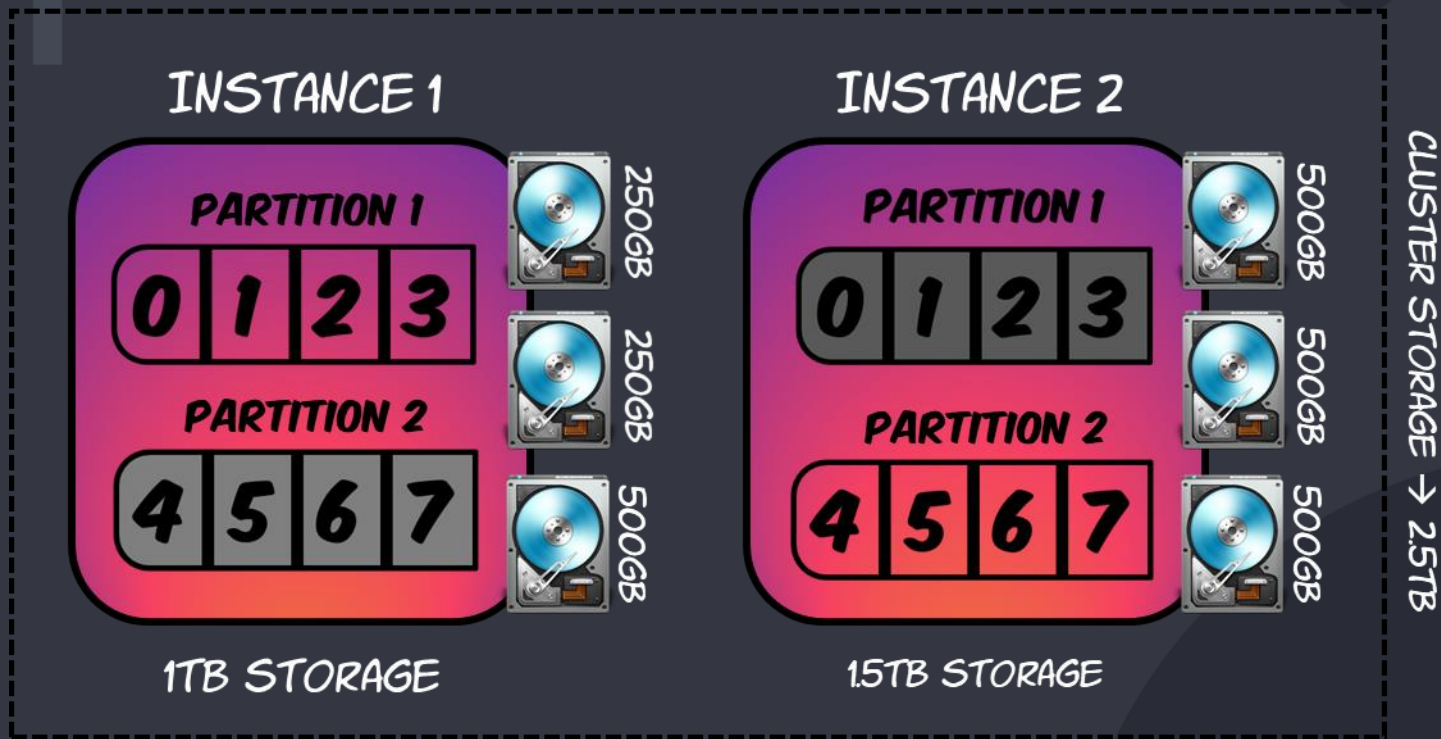
OUTSOURCE DATA STREAMS CREATION



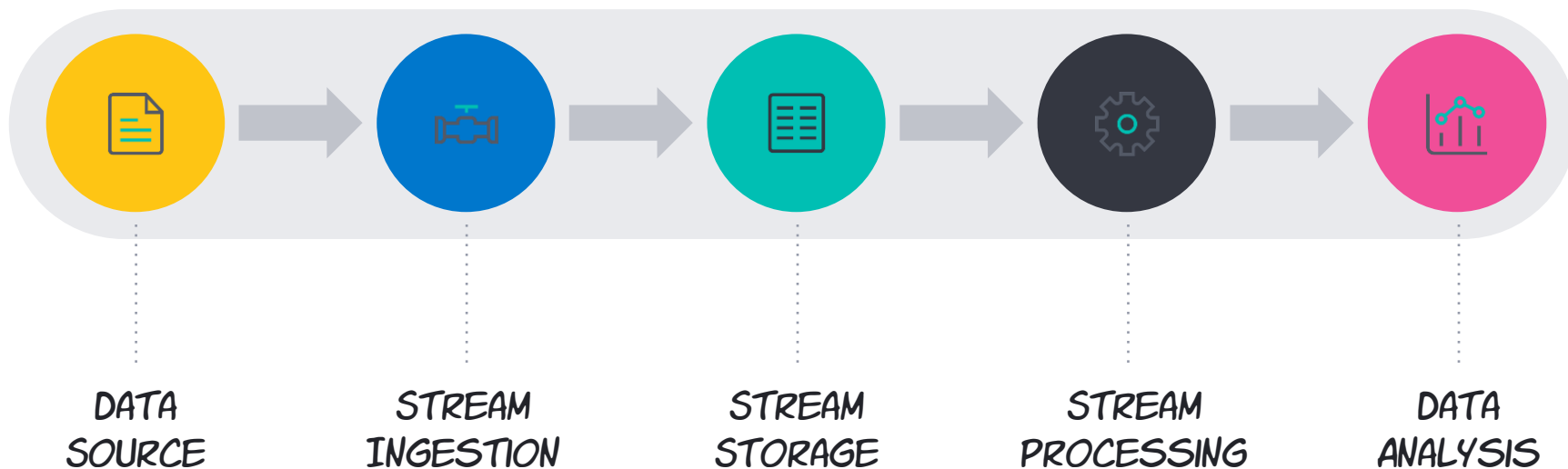
STREAM STORES PER PROJECT



SCALING STORAGE WITH COMPUTE









UNIDIRECTIONAL FLOW OF DATA



IMPLEMENTATION TECHNOLOGIES

IMPLEMENTATION TECHNOLOGIES

TECHNOLOGY		OPEN-SOURCE	ON-PREMISES	CLOUD
APACHE KAFKA		YES	YES	YES
APACHE PULSAR		YES	YES	YES
ELASTIC STACK		YES	YES	YES
AMAZON WEB SERVICES		NO	NO	YES
MICROSOFT AZURE		NO	NO	YES
GOOGLE CLOUD		NO	NO	YES



THANK YOU



@RIFERREI