



BEST PRACTICES FOR BUILDING STREAMING DATA ARCHITECTURES

RICARDO FERREIRA
BIG DATA CONFERENCE EU 2020



@RIFERREI



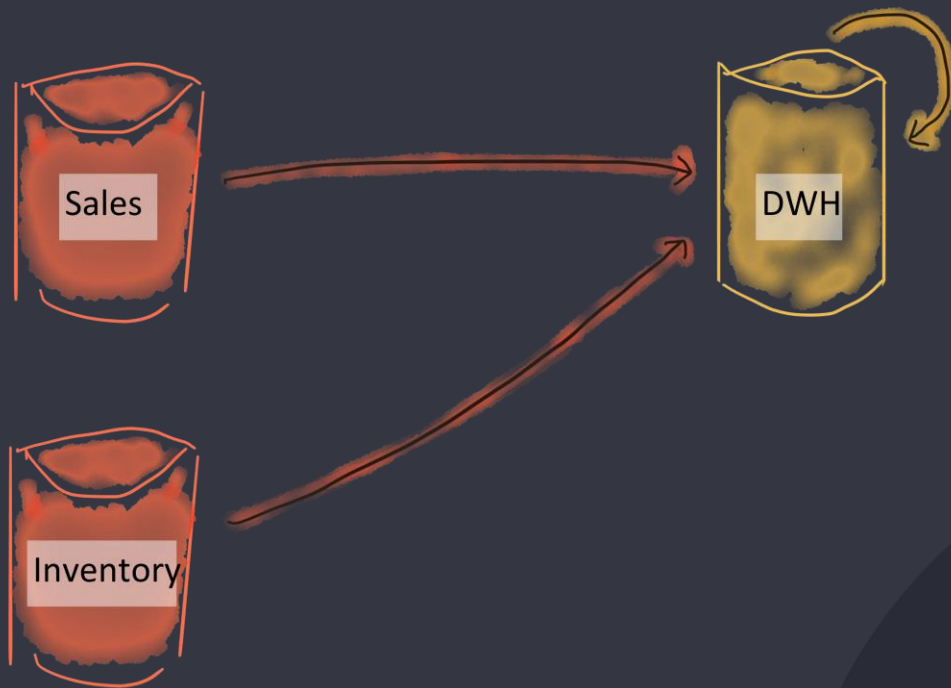
RICARDO FERREIRA

DEVELOPER ADVOCATE 🥑

- ❑ ELASTIC COMMUNITY TEAM
- ❑ BEFORE JOINING ELASTIC:
CONFLUENT, ORACLE, RED HAT
- ❑ STREAMING DATA, BIG DATA,
ANALYTICS, DATABASES, CLOUD
- ❑ RIFERREI@ELASTIC.CO
- ❑ RIFERREI@RIFERREI.COM
- ❑ [HTTPS://RIFERREI.COM](https://riferrei.com)

ONCE UPON
A TIME...

ANALYTICS WAS ALL ABOUT SQL DATABASES



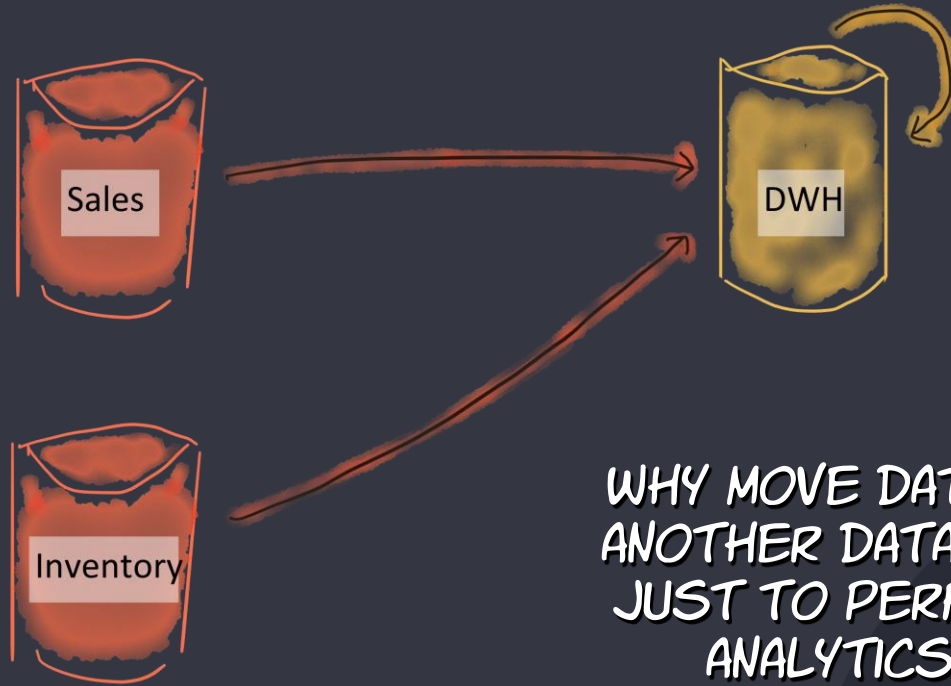
BUT SQL
DATABASES HAVE
LIMITATIONS



WHAT DO
YOU MEAN
BY...

LIMITATIONS?

YES... LIMITATIONS.



WHY MOVE DATA TO
ANOTHER DATABASE
JUST TO PERFORM
ANALYTICS?

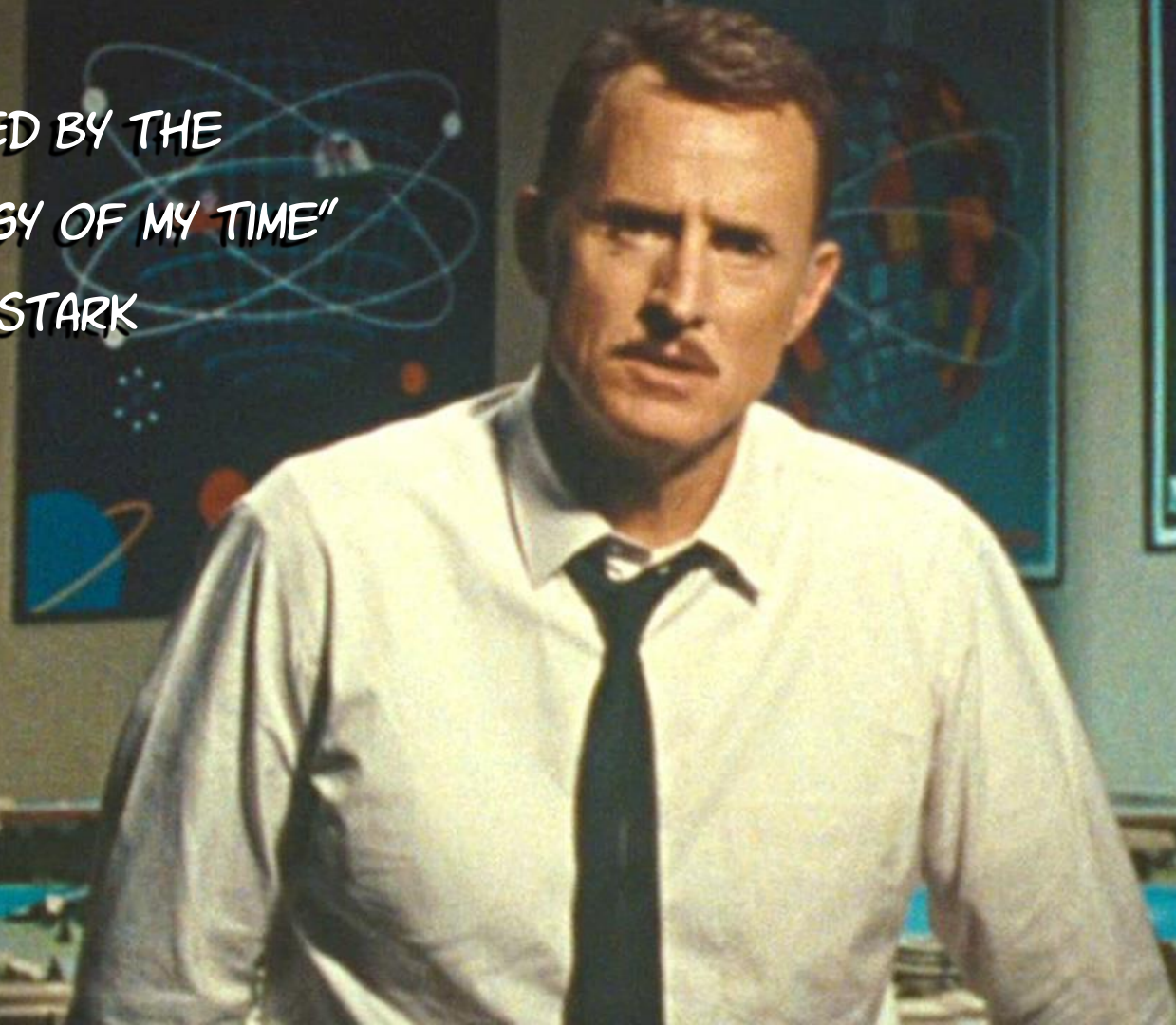


DATABASES WITH ONLY OLTP



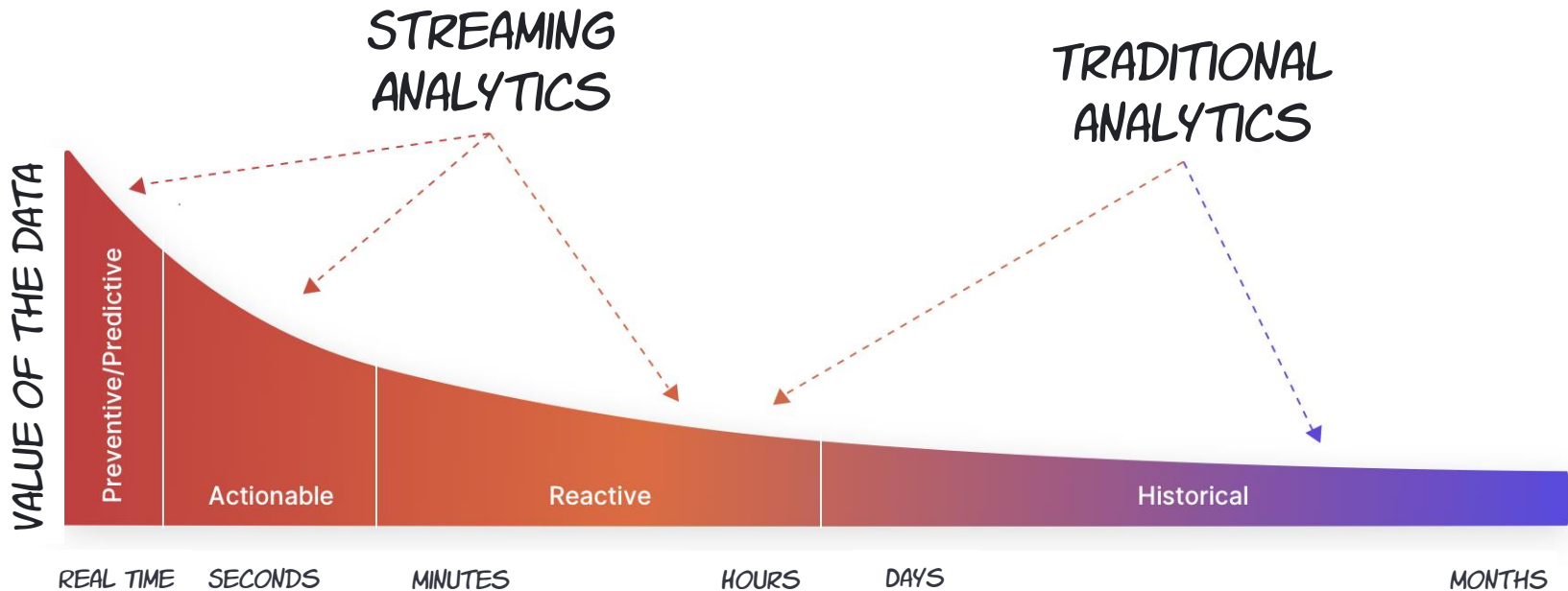
...WITH OLTP AND OLAP

**"I AM LIMITED BY THE
TECHNOLOGY OF MY TIME"
- HOWARD STARK**



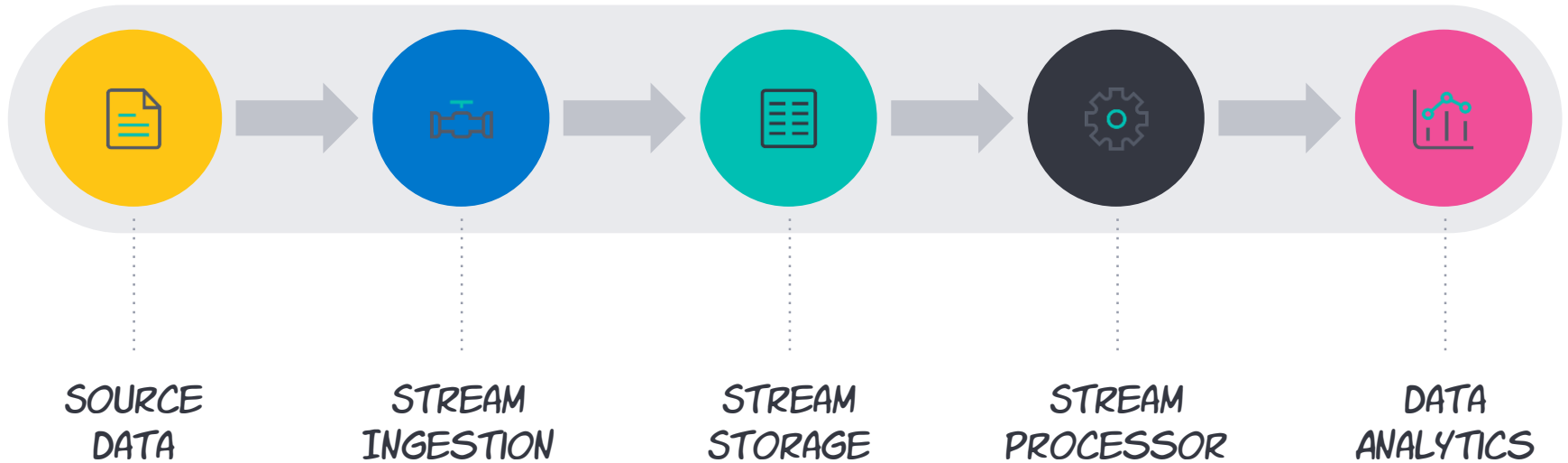
SAY HELLO
TO ACTIONABLE
INSIGHTS

THE RISE OF ACTIONABLE INSIGHTS









SOURCE: PERISHABLE INSIGHTS, FORRESTER

STREAMING DATA ARCHITECTURE



IMPLEMENTATION STACKS

STACK		STREAM STORAGE	STREAM PROCESSOR	DATA ANALYTICS
APACHE KAFKA		KAFKA CLUSTER	KAFKA STREAMS	ELASTICSEARCH AND KIBANA
APACHE PULSAR		PULSAR CLUSTER	APACHE FLINK	ELASTICSEARCH AND KIBANA
ELASTIC		ELASTICSEARCH CLUSTER	ELASTICSEARCH CLUSTER	ELASTICSEARCH AND KIBANA
AMAZON WEB SERVICES		KINESIS DATA STREAMS	KINESIS DATA ANALYTICS	AMAZON REDSHIFT
MICROSOFT AZURE		AZURE EVENT HUBS	AZURE STREAM ANALYTICS	AZURE SQL DATABASE
GOOGLE CLOUD		GOOGLE PUB/SUB	GOOGLE DATAFLOW	GOOGLE BIGQUERY

BEST PRACTICES

DRAWING A LINE...

ABOUT KEEPING DATA ON STREAM
STORAGE OR IN THE DATA ANALYTICS

STREAM STORAGE OR DATA ANALYTICS?

- USE STREAM STORAGE FOR WRITE-ONCE-READ-MANY PATTERN
- THE TYPE OF DATA YOU ARE HANDLING IS TIME SENSITIVE?
- CHECK WHERE IS YOUR PROCESSING CENTER-OF-GRAVITY



IT IS ALL ABOUT STORAGE

IT IS ALL ABOUT STORAGE

- CAN YOU CONTROL STREAM STORAGE GROWTH COSTS?
- EVALUATE DATA SPILLAGE WITH THE STREAM PROCESSORS
- HOW LONG DATA MUST BE RETAINED? MONTHS? YEARS?

SCHEMAS ARE CONTRACTS FOR DEVELOPERS

On behalf of the Client (authorized signatory)



On behalf of the Developer (authorized signatory)



SCHEMAS ARE CONTRACTS FOR DEVS

- USING SCHEMA-FIRST APPROACH CAN AVOID THE HEADACHES
- USING SCHEMAS MEANS MORE AUTOMATION AVAILABLE TO USE
- DISTRIBUTED ARCHITECTURE: CHANGES HAPPEN EVERYWHERE!

TWO IS BETTER THAN ONE



STREAM STORAGE AVAILABILITY

STREAM STORAGE AVAILABILITY

- UNDERSTAND HOW DURABILITY WORKS ON STREAM STORAGE
- DID YOU KNOW THAT YOU CAN SHOULD BUILD DR HERE TOO?
- MIND THE RED FLAGS OF THE CAP THEOREM HAUNTING YOU

BUILD VERSUS BUY



BUILD VERSUS BUY

- ASK YOURSELF: ARE YOU A DISTRIBUTED SYSTEMS SHOP?
- BUILD CONFIDENCE THROUGHOUT THE ENTIRE ARCHITECTURE
- HOW COSTLY IS TO FIND SOMEONE TO IMPLEMENT CODE?



UNDERSTAND THE TECHNOLOGY

UNDERSTAND THE TECHNOLOGY

- DON'T FOOL YOURSELF OR THE TEAM: IT IS NOT THAT SIMPLE!
- MOST PROJECTS FAIL BECAUSE OF UNKNOWN TECH DETAILS
- PUBLIC BENCHMARKS DON'T EXCUSE YOU TO BUILD YOUR OWN



TIME IS
RELATIVE

TIME IS RELATIVE

- NETWORK IS YOUR BIGGEST VILLAIN WHEN DEALING WITH TIME
- WHENEVER POSSIBLE SET THE TIME IN THE SOURCE DATA
- IF EVER IN DOUBT ABOUT TIME – ASK THE STREAM STORAGE



THANK YOU

RICARDO FERREIRA
BIG DATA CONFERENCE EU 2020



@RIFERREI