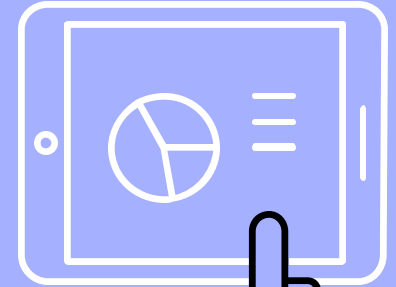
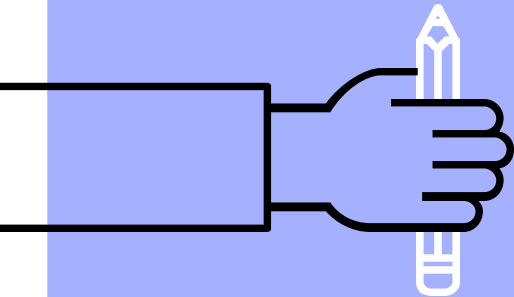
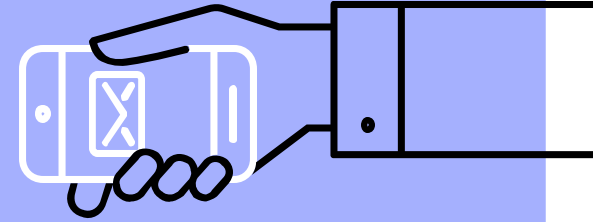
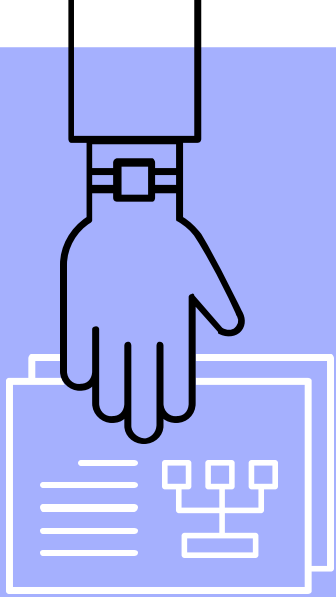


NEAR REAL-TIME
DATA
WAREHOUSING –
THE FINAL
FRONTIER?



Elffar Analytics

Joel Acha

Oracle Technologies since 1997



[@ElffarAnalytics](https://twitter.com/ElffarAnalytics)



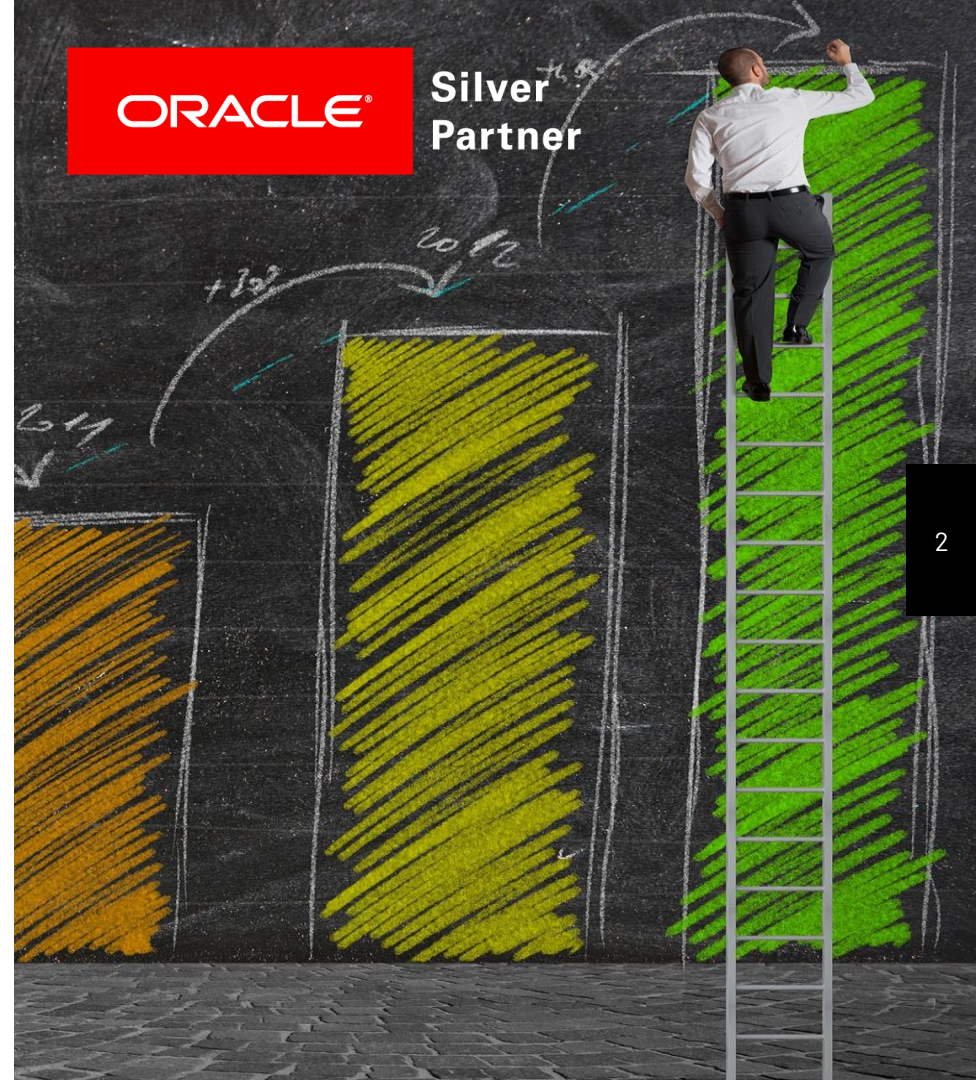
elffar.co.uk

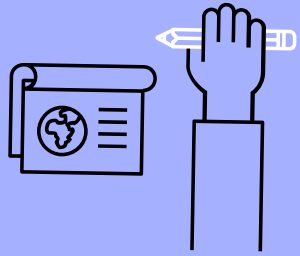


info@elffar.co.uk



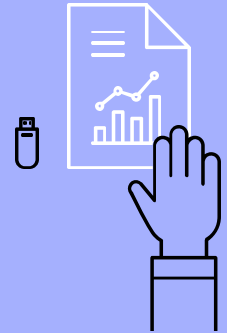
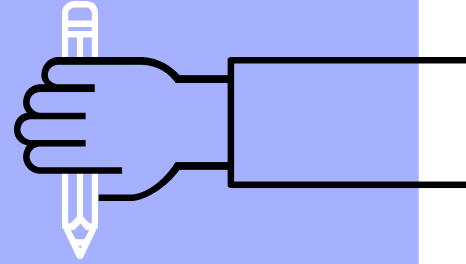
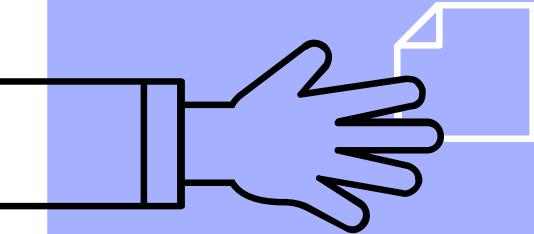
[#obihackers IRC channel](#)





1. Data Warehouse

Let's start with a simple
definition





“

"A data warehouse is a copy of transaction data specifically structured for query and analysis."

Ralph Kimball



Data Integration Methods

Traditional ETL

Batch Based
High Latency

CDC Replication

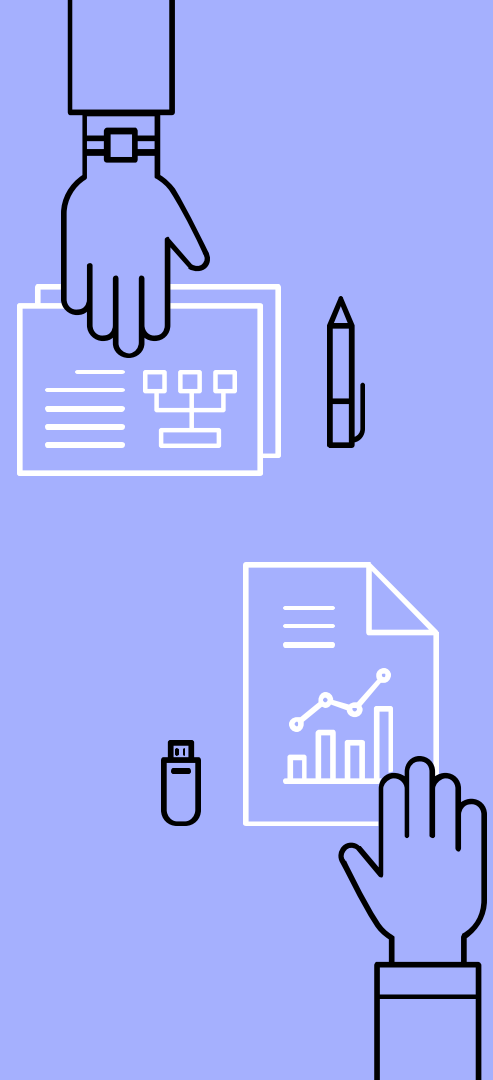
"Real Time"
Low Latency

Real Time Streaming

Low Latency
In-line in-memory transformation

Data Movement

- ▶ Batching
- ▶ Micro Batching
- ▶ Streaming



Main Characteristics

Batching

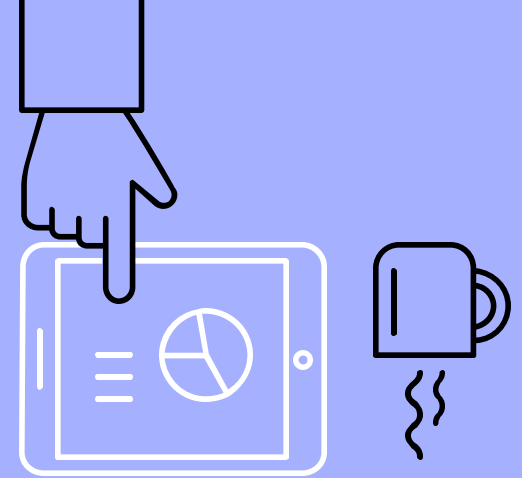
- ▶ New data elements grouped into a batch
- ▶ Based on a time-based batch interval

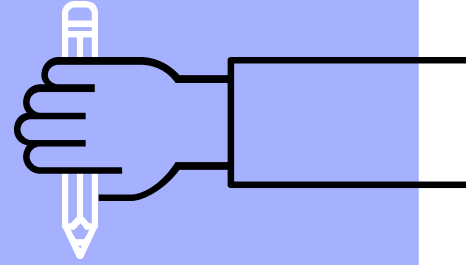
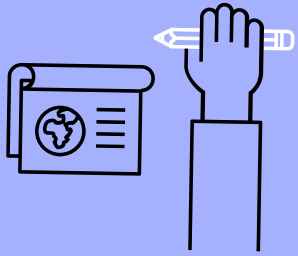
Micro Batching

- ▶ New data elements more frequently grouped into a batch
- ▶ Real-time analytics not essential

Streaming

- ▶ Event driven architecture
- ▶ Low latency is critical

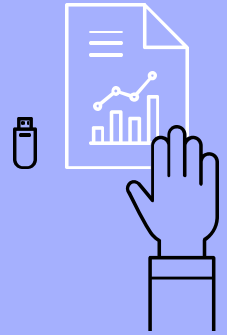
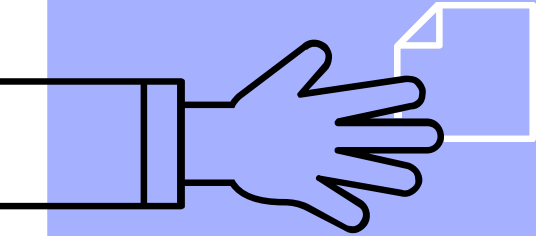




2.

Big Data

Yet another simple
definition...



“

**“larger, more
complex data sets,
especially from new
data sources.”**

[source: oracle.com](https://www.oracle.com)



BIG

DATA



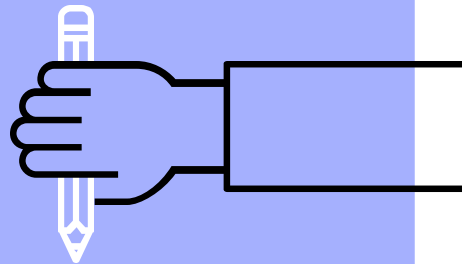
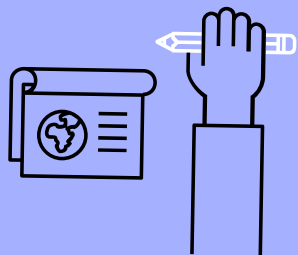
3 V's:

1. Velocity
2. Volume
3. Variety



The final nail in the coffin?

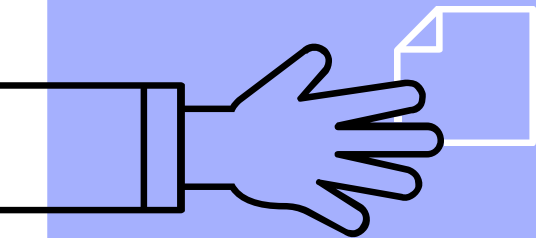
WHAT IS BIG DATA ANYWAY?



3.

Data Lake

Last definition - I promise...



“

"A centralized repository that allows you to store all your structured and unstructured data at any scale."

[source: amazon.com](https://aws.amazon.com/redshift/)



Are Apples & Pears the same?



≠

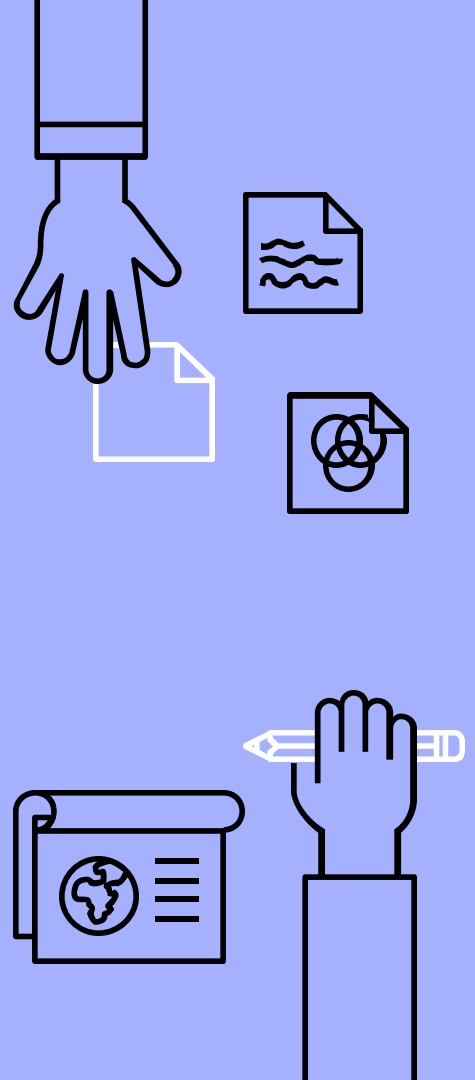


Big Data - Data Lake

- ▶ Schema on Read
- ▶ Data stored in raw form
- ▶ "Freeform"

Data Warehouse

- ▶ Schema on Write
- ▶ Structured Data
- ▶ Query limitations
- ▶ Value of data clear from the outset





CURRENT STATE

BI/DW CHALLENGES

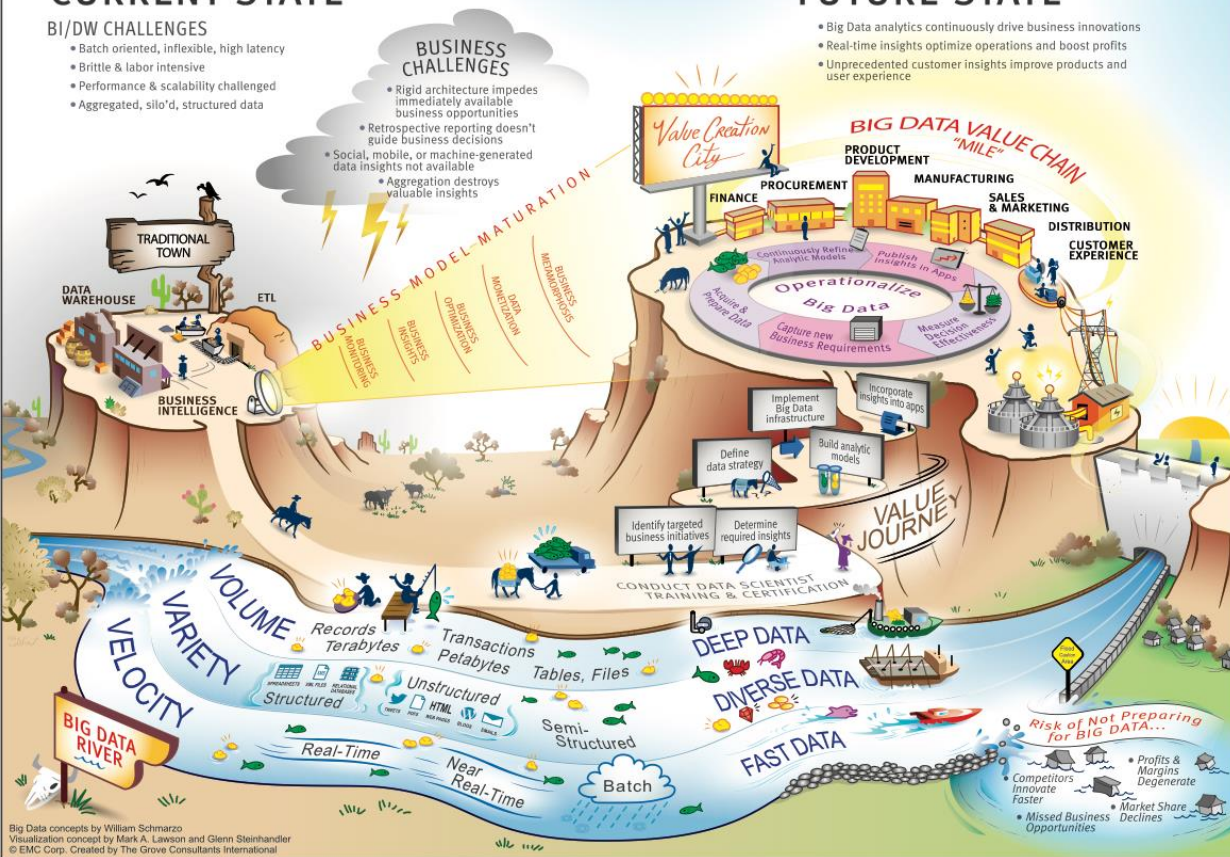
- Batch oriented, inflexible, high latency
- Brittle & labor intensive
- Performance & scalability challenged
- Aggregated, silo'd, structured data

BUSINESS CHALLENGES

- Rigid architecture impedes immediately available business opportunities
- Retrospective reporting doesn't guide business decisions
- Social, mobile, or machine-generated data insights not available
- Aggregation destroys valuable insights

FUTURE STATE

- Big Data analytics continuously drive business innovations
- Real-time insights optimize operations and boost profits
- Unprecedented customer insights improve products and user experience



Big Data concepts by William Schmarzo
 Visualization concept by Mark A. Lawson and Glenn Steinhardt
 © EMC Corp. Created by The Grove Consultants International

Complimentary
Technologies

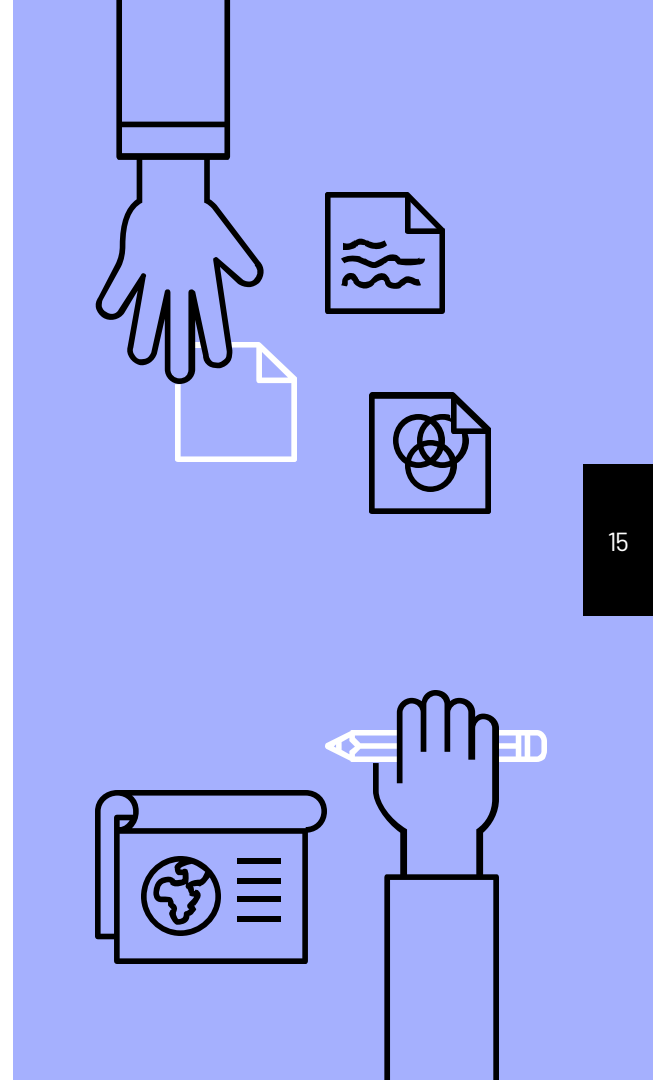
Use Cases

Data Warehouse

- ▶ Highly curated data
- ▶ Structured standard reporting

Data Lake

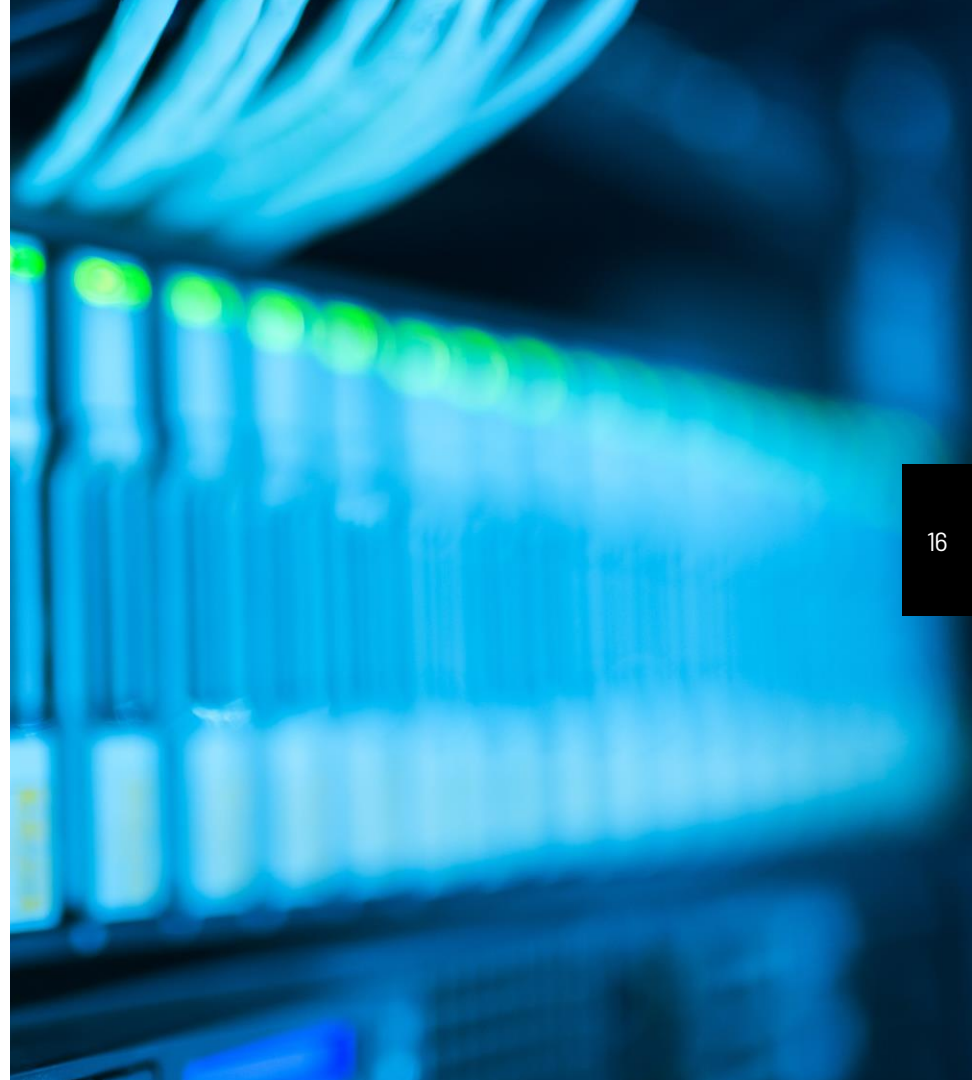
- ▶ Data Scientist access to raw data
- ▶ Flexible, reactive business model



THE BEST OF BOTH WORLDS?

Streaming

- ▶ Next generation Data Integration
- ▶ In-flight
- ▶ Real-time
- ▶ Can load into data lakes and data warehouses



THANKS!

Any questions?



[@ElffarAnalytics](https://twitter.com/ElffarAnalytics)



elffar.co.uk



info@elffar.co.uk



[#obihackers IRC channel](#)

