

Data Discovery Self-Service Analytics

sonra
Unleash the Value of your Data

About Uli

15+ years in big & small data

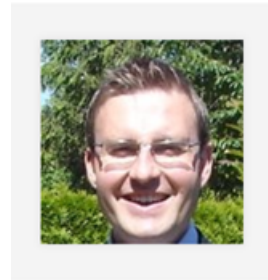
Oracle ACE

Speaker, Writer

Books

Co-founder Sonra

HUG/OUG Ireland



Uli Bethke

Co-founder Sonra. Organizer Hadoop User Group Ireland. Oracle ACE. Blogger, Writer, Speaker.

Ireland | Computer Software

Current Sonra, OUG Ireland Business Intelligence SIG, Hadoop User Group Ireland

Education Freie Universität Berlin



• **Twitter:** @ubethke

• **LinkedIn:** ie.linkedin.com/in/ulibethke



• **Email:** uli.bethke@sonra.io

About Sonra

Big Data Consulting (Spark, Hadoop)

Data Hub / Data Warehouse

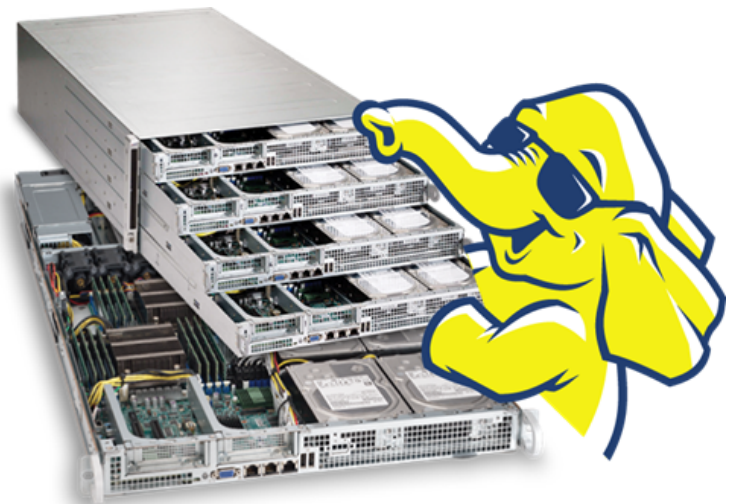
Analytics / Data Discovery / BI

Real-time / Batch

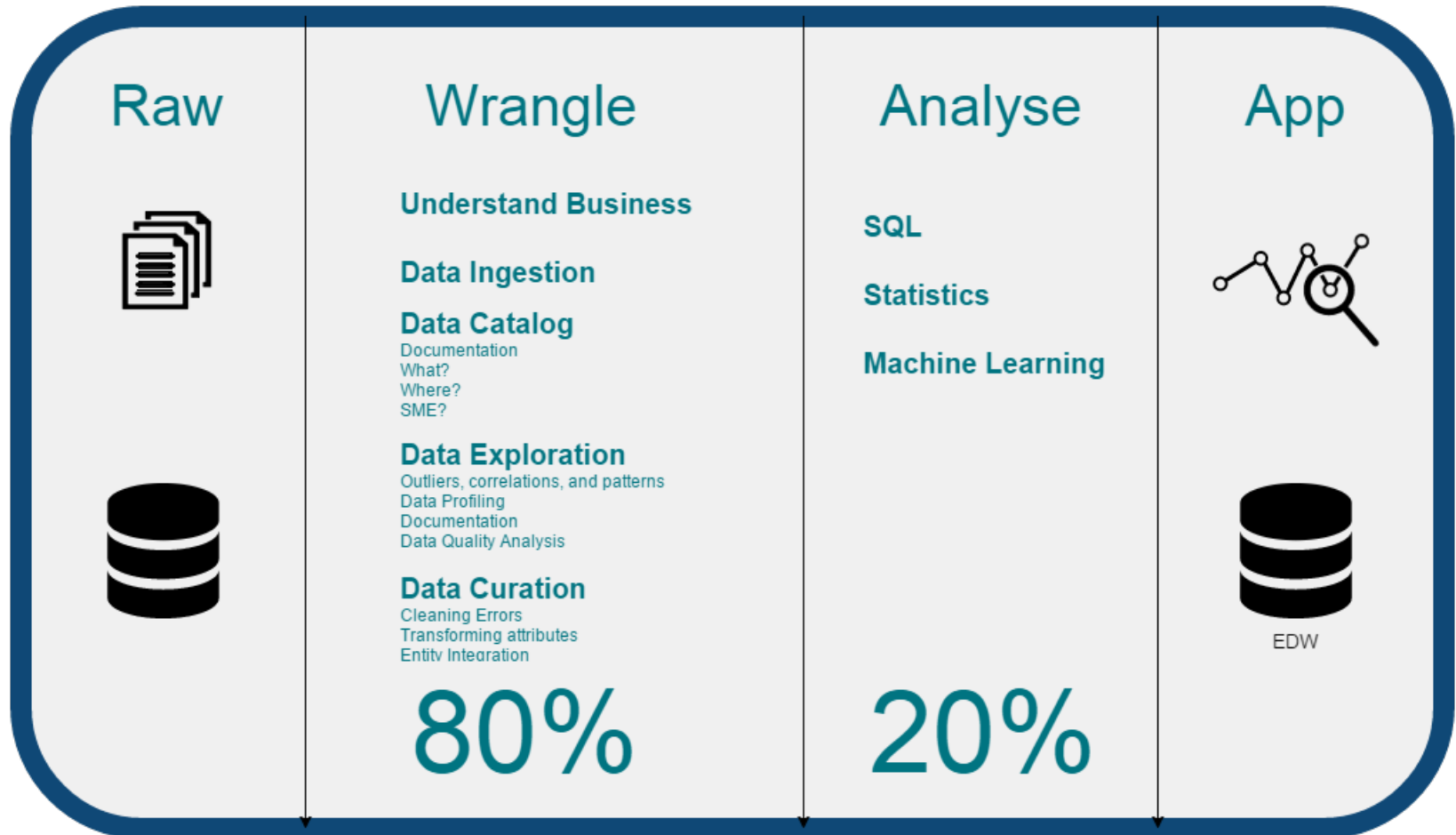
Bare Metal Hosting

Member CeADAR

Predict Conference



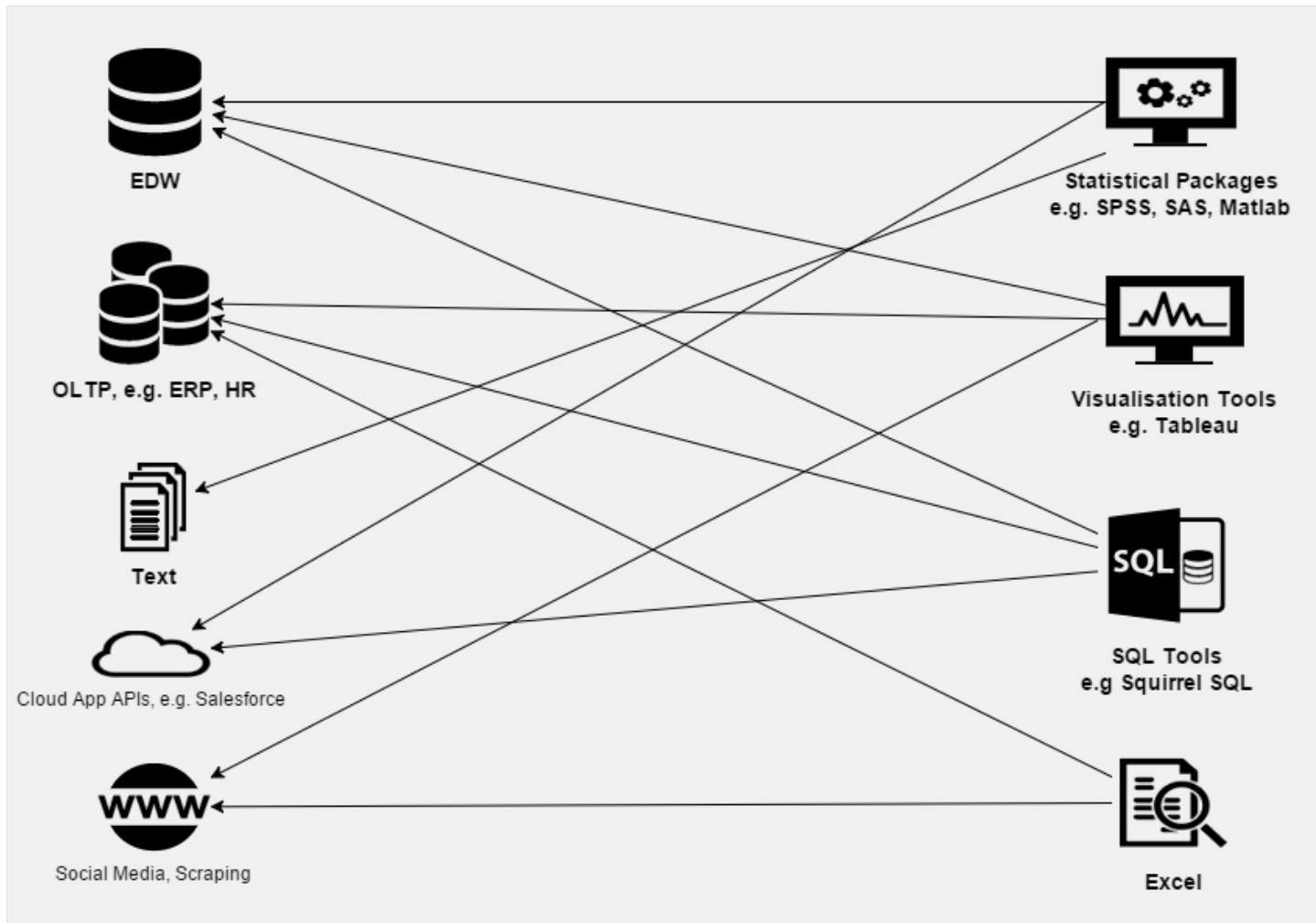
Data Discovery Process



Data Discovery & BI

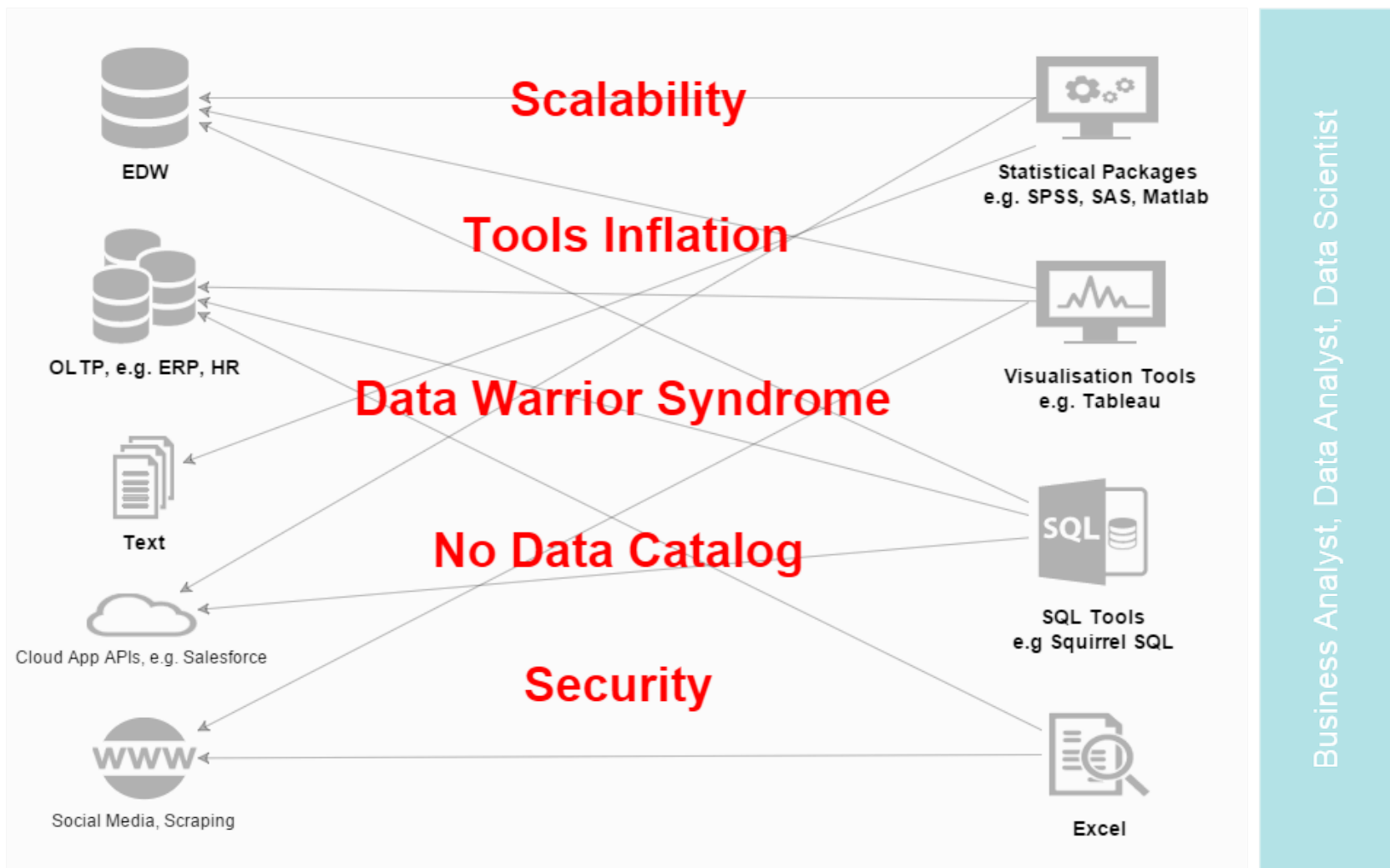
	Business Intelligence	Data Discovery
Data Format	Cleansed, Processed, Integrated	Raw, Unknown
Data Types	Structured	Any
Data Profiling	N/A	Exploratory Data Analysis
Data Scope	EDW	Long tail of data
Time to value	Long time to value	Reduced time to value
Data Transformations	Formal: ETL, IT	Ad-hoc: Iterative, Self-Service
Tool	ETL & BI Tool	Data Discovery Tool
Self-Service	Confined to EDW aka ad-hoc queries	All data

Legacy Data Discovery Architecture

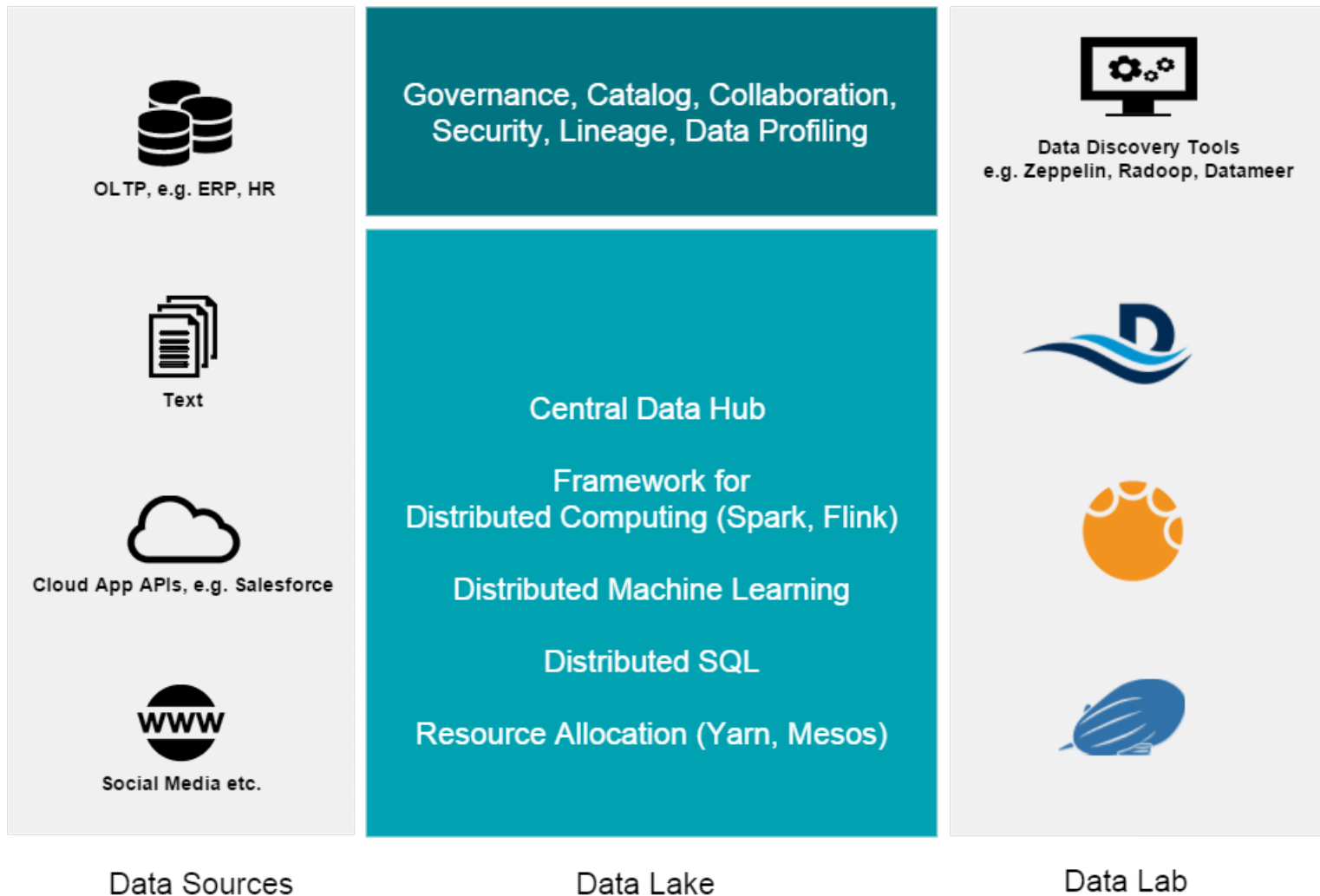


Business Analyst, Data Analyst, Data Scientist

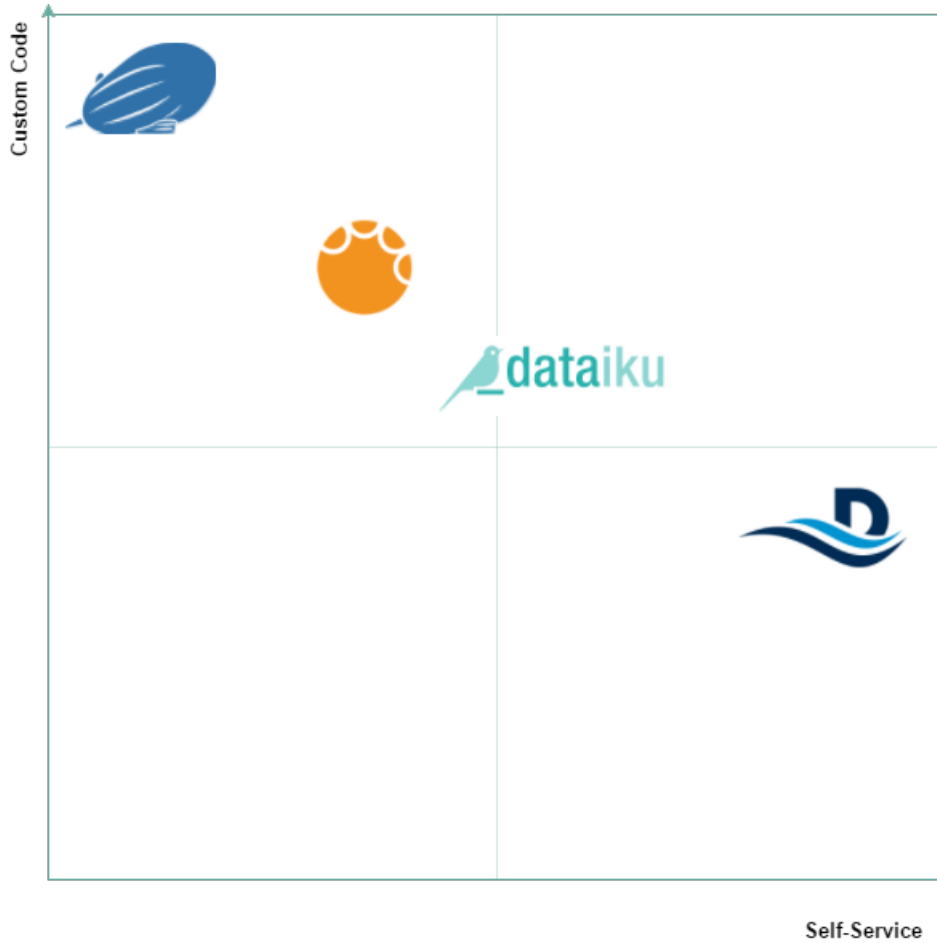
Issues Legacy Architecture



Modern Data Discovery Platform



Data Discovery Tools I



Zeppelin Notebook - Interpreter

```
SQL
select city, count(*) as num_members from online_member_journey where country = 0 group by city having count(*) > 10 order by num_members desc
```

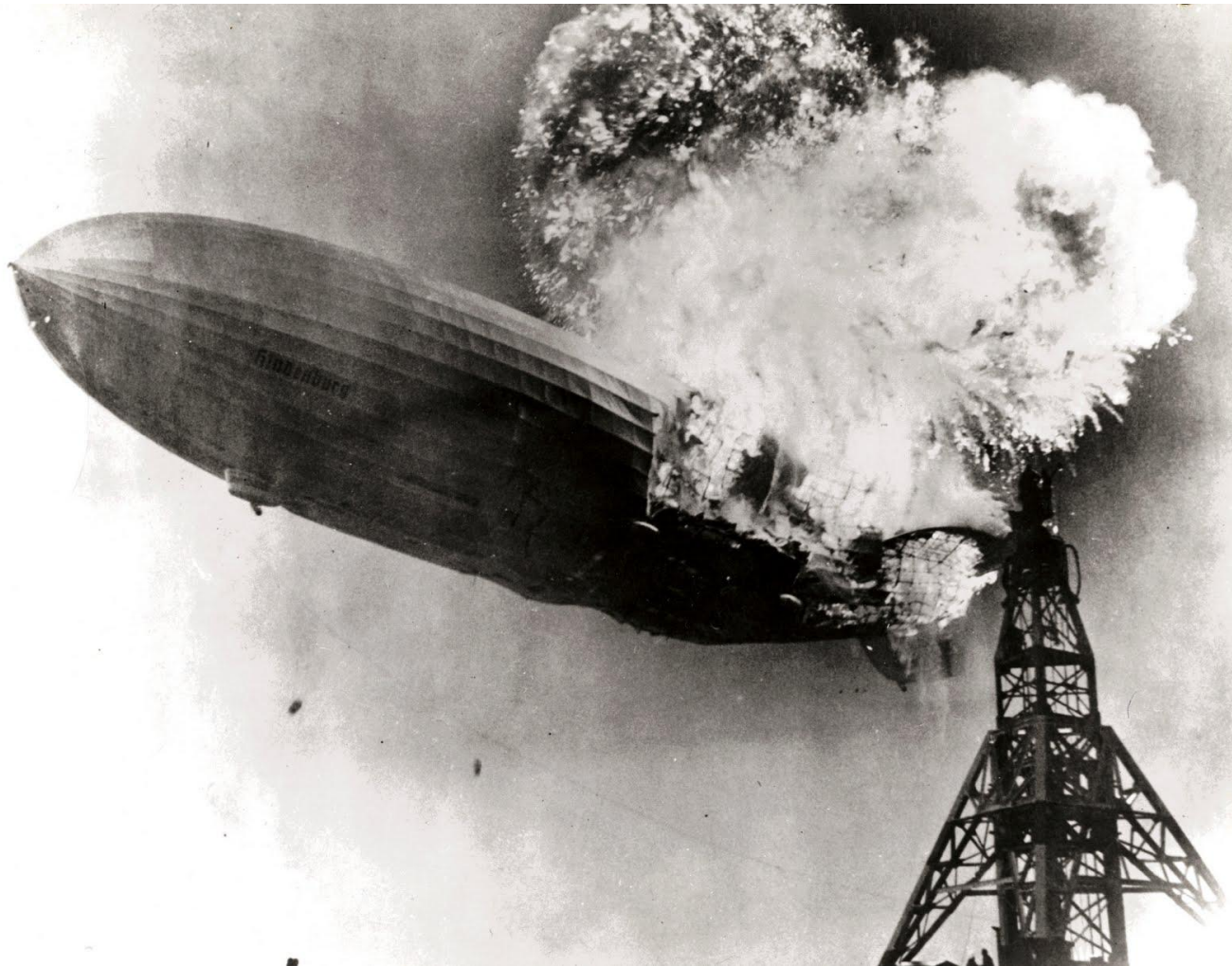
Retrieve out → Filter Exam... (exa, ori) → Aggregate (exa, ori) → Store (inp, out)

Sample	Time	Location	Country
2014-02-07 10:00:00	10:00:00	USA	USA
2014-02-07 10:00:00	10:00:00	USA	USA
2014-02-07 10:00:00	10:00:00	USA	USA
2014-02-07 10:00:00	10:00:00	USA	USA
2014-02-07 10:00:00	10:00:00	USA	USA
2014-02-07 10:00:00	10:00:00	USA	USA
2014-02-07 10:00:00	10:00:00	USA	USA
2014-02-07 10:00:00	10:00:00	USA	USA
2014-02-07 10:00:00	10:00:00	USA	USA
2014-02-07 10:00:00	10:00:00	USA	USA

Data Discovery Tools II

Feature	Zeppelin	Datameer
Self-service Transformations		GUI + SDK
Rich Connectivity	Code (framework)	GUI + SDK
Unstructured Data	Code	GUI
Machine Learning	Code	GUI + Zementis
Security		
Data Enrichment		
Collaboration		
Data Orchestration		
Data Profiling	Code (Spark)	GUI
Data Catalog		
Visualisations		
Distributed & Scalable	Spark, Flink, Ignite etc.	MR, Tez

Zeppelin Demo



p: +353 1 254 2897
t: @sonra_io

e: hello@sonra.io
w: www.sonra.io

sonra
Unleash the Value of your Data

Final Words

- Data Discovery Rocks
- Thanks to Michael
- Raffle Predict Conference
- Thanks to Bank of Ireland
@BOIGrandCanalSq
- Next meetup 6 October AIB
- Ferryman

sonra

WWW.SONRA.IO