

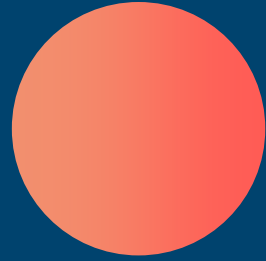


# How to evaluate your Data Platform's maturity

**Lucie Bailly**

Data Engineer @ Criteo

**NABD 2019**





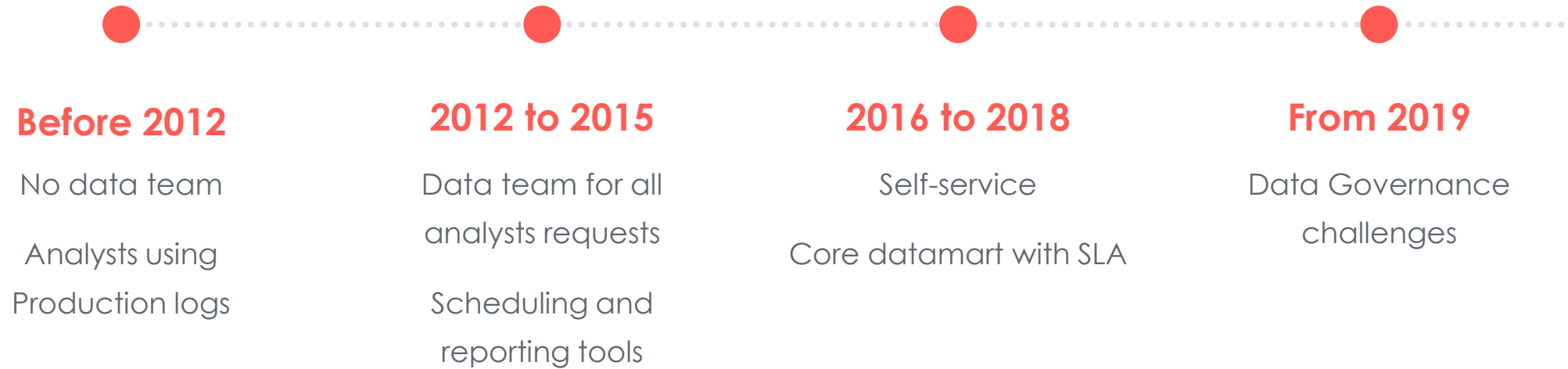
**On a scale on 1 to 5, how would you score your data platform (as a user or a provider)?**



- I joined Criteo in **2012**
- Lead of the **Data Models** team, in charge of the creating **data for Analytics** and Reporting tools, and sharing **Data Governance** best practices

# Criteo from an Analytics data perspective

A bit of history



# Criteo from an Analytics data perspective



**200Tb**

**Data** ingested  
daily



**200**

**Core Datamart**  
tables



**300+**

Analysts



**1400+**

**Self-service** jobs

# Why

1

Core Datamart was  
mature

2

Data Governance  
initiative

# How

What kind of service, process, tool... should I expect as a data producer or data consumer?

How far are we from that?

Is it applicable to other teams?

What is a « perfect » dataset?

Is it really what we want to reach?



**1 to 5 Scale**







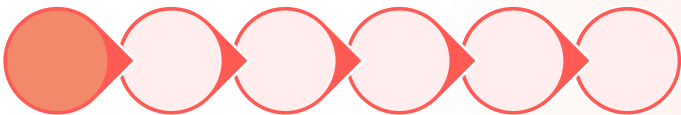
**6 Services**

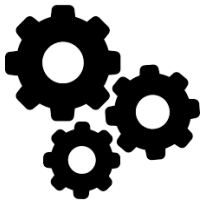


## Platform

Keep the platform  
healthy and  
rationalize resource  
usage on the cluster

- Right amount of **resource**
- Usage **measured** and **rationalized**
- **Impact** estimated
- **Optimized** jobs

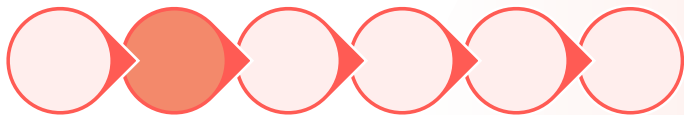




# Operations

Validate and measure impacts of any operations on data. Detect and fix incidents

- Non-regression **tests**
- **Backfills**

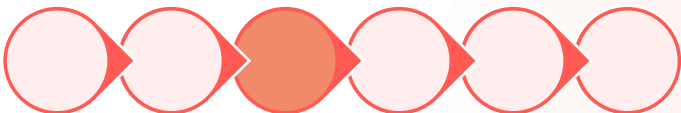




## Documentation & Discovery

Provide a clear and  
comprehensive  
overview of all datasets

- **Up-to-date** documentation
- Standardize **template**
- Central **tool**
- **Metadata**

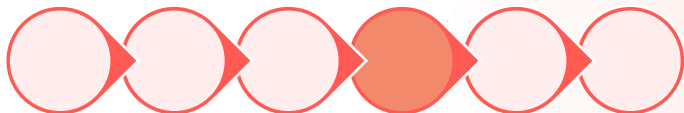


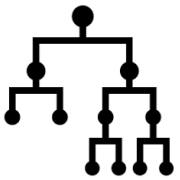


# Monitoring

Declare and monitor SLA. Provide Data contract for each dataset

- **SLI/SLO/SLA**
- **Data Contract**
- Incidents processes

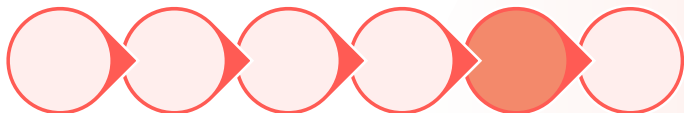




## Lineage & Ownership

Provide a full view of data lineage across the platform, from extraction to business usage

- Table, field and partition **lineage**
- **Technical owners**
- Defined **responsibilities**
- Leverage other services

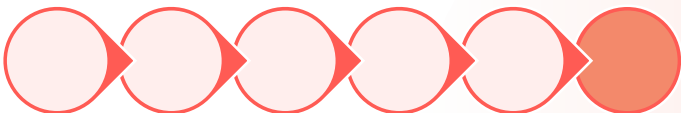




## Business value

Optimize platform and processes to fit business values and needs

- **Functional owners**
- **Data lifecycle**
- **Critical** data
- Privacy and legal constraints





# What

Define the goals

1

Coherent  
datasets groups

2

Score on the 6  
criteria

3

Maximum score  
to reach

4

Goal for the next  
months

5

Tasks to reach the  
next level

# Core Datamart

	1- Initial	2- Managed	3- Tool based	4-Automated	5- Optimised
Platform		✓			
Operations			✓		
Documentation and Discovery			✓		
Monitoring		✓			
Lineage and Ownership		✓			
Business value		✓			

**Total: 16 / 27**

**Common scale**



**Alignment of data  
governance strategy**



**Tools to share best  
practices and vision**

# Takeaways

**1**

**6 services  
1 to 5 scale**

**2**

**Perfection isn't  
the goal**

**3**

**Build your own  
matrix, set your  
goals!**

# Thank you!

Email: [l.bailly@criteo.com](mailto:l.bailly@criteo.com)

Twitter: [@lucie\\_bailly](https://twitter.com/lucie_bailly)



<http://tiny.cc/qveegz>

criteo<sup>l</sup>.