SODA

Helping the Enterprise Build Good Data Products

Presenters

**Maarten
Masschelein**

Founder & CEO

**Mathisse De
Strooper**

Lead Product Manager

Agenda

- The Rise of DataOps
- Data Observability
- Best-practices for building Data Products
- The Soda Cloud
- Soda in Action
- Q&A

The Rise of DataOps

What is DataOps?

A process framework that helps data teams deliver high quality, reliable data insights with high velocity.

Inspired from DevOps and Agile Principles, and applies them to Data Analytics
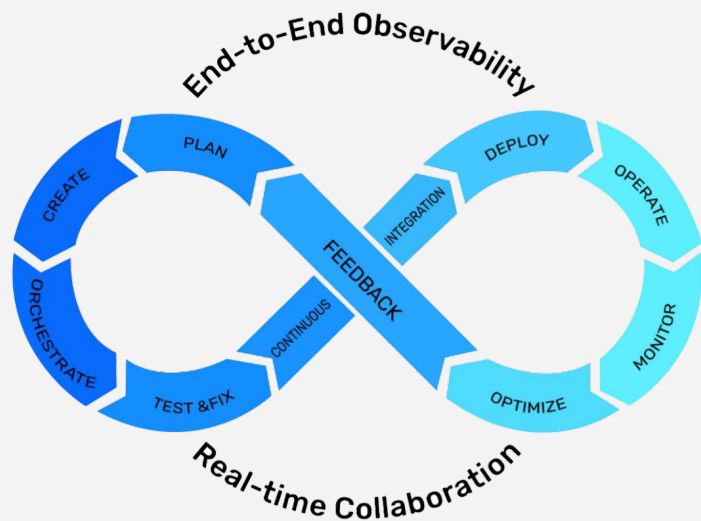
The DataOps framework

Image Source: DataOps Unleashed

Data Observability

Observability and Monitoring: the Evolution of the IT Stack

Infra & App Management

DevOps: Key is to monitor, observe, triage, diagnose, and resolve infrastructure and performance issues.

Data & Model Management

DataOps (& MLOps)**:** Key is to monitor, observe, triage, diagnose, and resolve (data) issues.

Data Product

Software that uses data to drive automation and business outcomes.

Silent data issue

Data problem that rendered the data product useless for an amount of time without the owner knowing.

Data teams want to get ahead of silent data issues.

Auto Monitor

Automatically track key data characteristics, and raise incidents across all your datasets.

Test & Validate

Test the data data flows into your data pipelines to ensure we keep running on good data.

Track Fitness

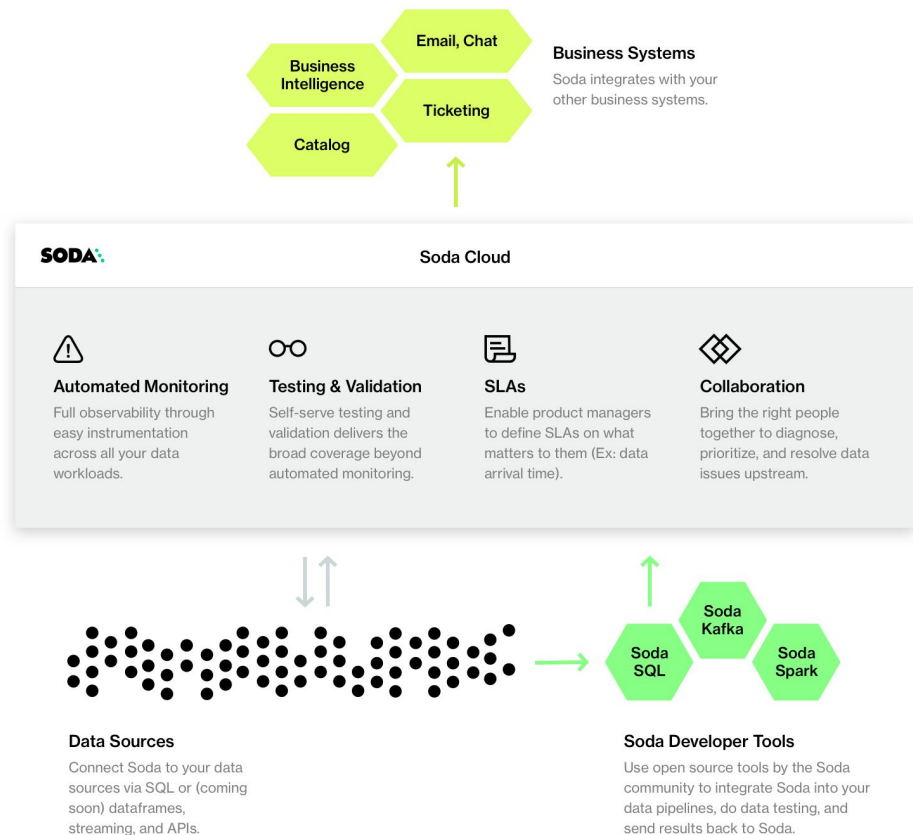Align data owners with data product teams to make sure data remains fit for purpose.

Collaborate

Break down data silos by collaboratively diagnosing, prioritizing, and resolving data issues.

Introducing Soda Cloud

Soda brings everyone closer to the data.

The Soda cloud is an integrated environment that allows you to easily connect to sources and destinations (e.g. catalogs, BI tools, …).

Email, Chat

Business Intelligence

Ticketing

Catalog

Business Systems
Soda integrates with your other business systems.

SODA   Soda Cloud

Automated Monitoring
Full observability through easy instrumentation across all your data workloads.

Testing & Validation
Self-serve testing and validation delivers the broad coverage beyond automated monitoring.

SLAs
Enable product managers to define SLAs on what matters to them (Ex: data arrival time).

Collaboration
Bring the right people together to diagnose, prioritize, and resolve data issues upstream.

Soda Kafka

Soda SQL

Soda Spark

Data Sources
Connect Soda to your data sources via SQL or (coming soon) dataframes, streaming, and APIs.

Soda Developer Tools
Use open source tools by the Soda community to integrate Soda into your data pipelines, do data testing, and send results back to Soda.

Uncover data issues before your customers do.

Soda APP 10:37 AM

Issues with your data have triggered 3 critical alerts and 2 warnings.
Access your Soda Cloud account to view your data monitoring results.

View Results

customer_users 10:36 AM

⊗ Missing opt-in values in Canada: 100% (0% - 1.34% expected)

⊗ Row count of customer_users: 958,005 (958,403 - 1,753,020 expected)

⊗ Inactive users over 3 years: 138 (> 50 critical)

⚠ Users receiving more than 1 promotion per day: 340 (> 100 warn)

⚠ Schema changes to customer_users: 1 ("Region" column added)

⊘ New categorical values in user_country: 2 ("IN" and "BD" added)

⊘ Data arrival time for customer_users: 10:28 AM (by 10:30 AM expected)

⊘ Hubspot email opt_out count: 22,430 (11,327 - 31,843 expected)

Soda <team@soda.io>
to me

10:53 AM (0 minutes ago)

SODA.

Maarten,

Issues with your data have triggered 3 critical alerts and 2 warnings.
Access your Soda Cloud account to view your data monitoring results.

View Results

Dataset123 10:53 AM

⊗ Missing opt-in values in Canada: 100% (0% - 1.34% expected)

⊗ Sales transactions row count: 43,842 (958,403 - 1,753,020 expected)

⊗ Inactive customers over 3 years: 148 (> 5 critical)

⚠ Users receiving more than 1 promotion per day: 340 (> 0 warn)

⚠ Schema changes in product dataset: 1 ("Region" column added)

⊘ New categorical values in user_country: 2 ("IN" and "BD" added)

⊘ Active users dataset arrival time: 9:55 AM (by 10:00 AM expected)

⊘ Hubspot email opt_out count: 22,430 (11,327 - 31,843 expected)

Soda in Action

Summary

Soda, The Data Observability & Collaboration Platform

Soda brings everyone closer to the data.

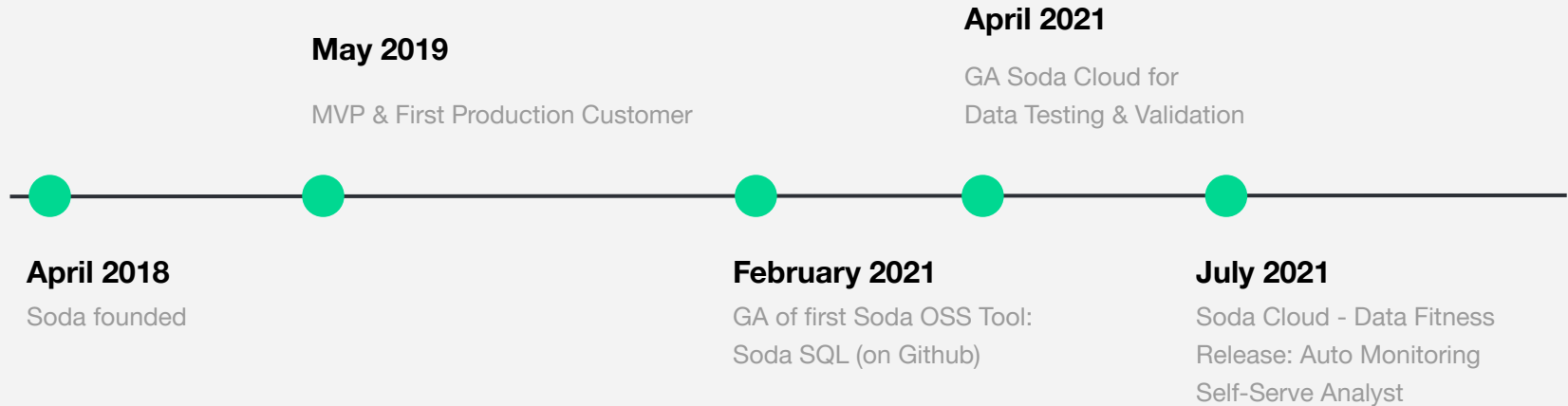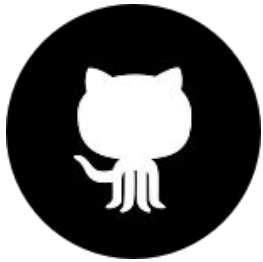**Cloud and open-source developers tools.**

It works, and we can show you how!

What's Coming

May 2019

MVP & First Production Customer

April 2021

GA Soda Cloud for
Data Testing & Validation

April 2018

Soda founded

February 2021

GA of first Soda OSS Tool:
Soda SQL (on Github)

July 2021

Soda Cloud - Data Fitness
Release: Auto Monitoring
Self-Serve Analyst

Join The Soda Community

Github

Repo:
https://github.com/sodadata/soda-sql

Free Trial Version

Available on Soda.io

Slack

soda-community.slack.com

Q&A