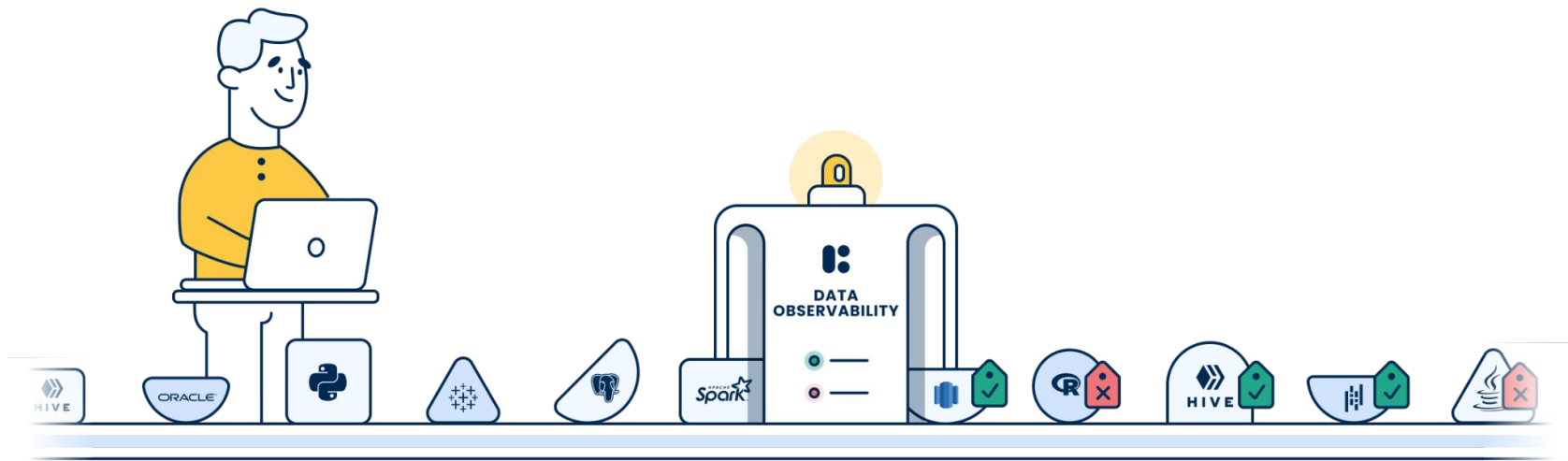




Manage your Data as a Product

Best practice of Data Management / Data Governance



My 20 years of scars with data

First 10 years

- Software Engineer in Geospatial – Map/Coverage/Catalog (Java, C++, YUI )
- (Satellite) Images and Vector Data Miner (Java/Python/R/Fortran/Scala)

Next 5 years

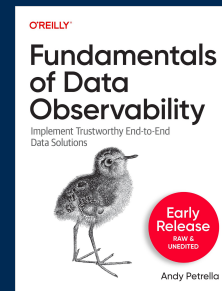
- Spark evangelist, teaching, and consulting in Big Data/AI in the Silicon Valley
- Creator of Spark-Notebook (pre-jupyter): open-source (3100+ ) and community-drive (20K+)

Last 5 years

- Brainstorming on how to bring quality and monitoring DevOps best practices to data (aka DODD)
- Founded [Kensu](#): easing data teams to embrace best practices and create trust in deliveries

Meanwhile, “serial author”

- “[What is Data Governance](#)”, O’Reilly, 2020
- “[What is Data Observability](#)”, O’Reilly, 2021
- “[Fundamentals of Data Observability](#)”, O’Reilly, 2023





How do you think about data?

- Data table
- Data file
- Data stream
- Dashboard
- Customers / ... (domain data)
- Representation of the real world
- Other?

How do I think about data?





Similar to “Web Service”

We think about them as:

- online store,
- movie streaming,
- document management system,
- email,
- calendar,
- ...

Not: API, Streaming hose, recommendation system, WebUI, ...



Data as a Product: Consider users as customers

Customers are looking for products to achieve their goals:

- They need to find them (discovery)
- They evaluate them (access)
- They need to trust them (reputation)



Data as a Product: Consider producers as product teams

Product owners/managers want to address existing/new customers:

- They need to find uncovered needs
- They need to develop and test quickly the adequacy
- They need to know what are the expectations in the long run



Domains

Enterprises are generally:

- Organized in domains
- Implementing strategies cross-domains

All domains, such as Finance, are relying on data:

- For their own needs – e.g., controlling the cash flow
- To participate to the company objectives – e.g., new product line



Domains – Data Product 🖱️ Data as a Product

Domains are creating and managing products.

They feel responsible for them...

They are supporting their own objectives directly.

Because all products, nowadays, are *data products*:

- Find them – look at API, dashboards, Portals, ... (*).
- Uncover the data they rely on.
- Turn them into Data as Products (**).
- Implement strategies by connecting them at higher level (projects)

(*) a catalog of data products would be great



(**) a data catalog becomes a catalog of products (each being a data)



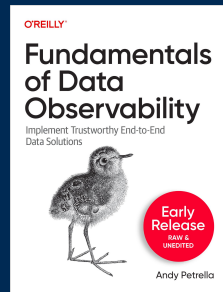
Thank YOU!

Try it by yourself: <https://sandbox.kensuapp.com>

Powered by  KENSU

- Connect with Google in 10 seconds 😊.
-  of use.
-  Get started with examples in Python, Spark, DBT, SQL, ...

#GoGrabIt



<https://kensu.io>

Ping me: [@noootsab](#) – [LinkedIn](#) – andy.petrella@kensu.io