

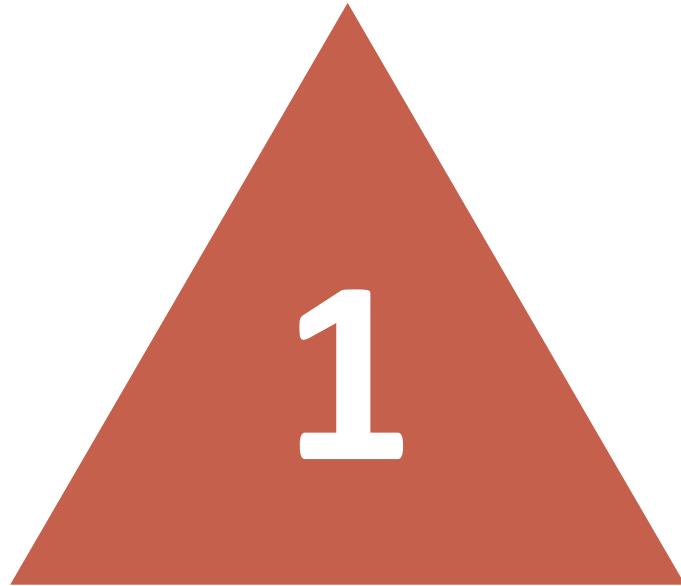
AGILE in DATA SCIENCE

Ngoc-Khanh Trinh @ KiotViet



1. Introduction to Data Science
2. Agile in Data science





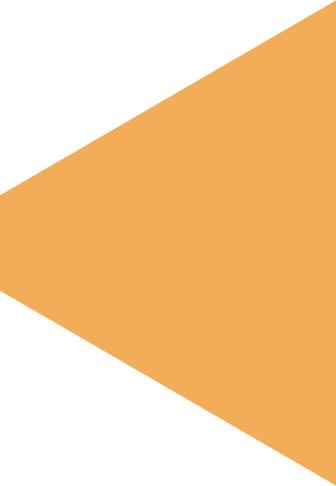
Introduction to

DATA SCIENCE

“

“Data Science is a set of disciplines and practices that extracts meaningful insights from data in a scientific manner”





Data Science

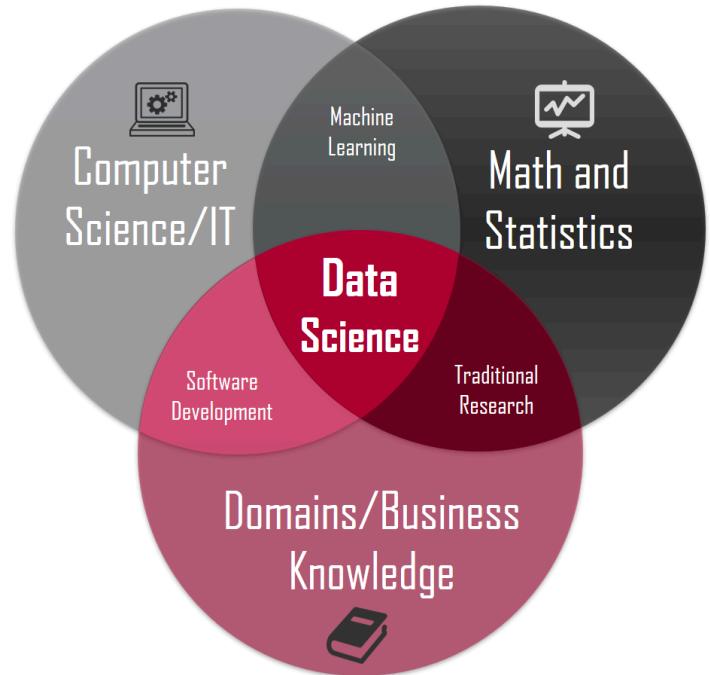
can help
organizations

- ▶ Understand their situations/environments
- ▶ Analyze existing issues
- ▶ Reveal new knowledge and hidden opportunities

*“Explore the best way to provide value to
the business using data”*

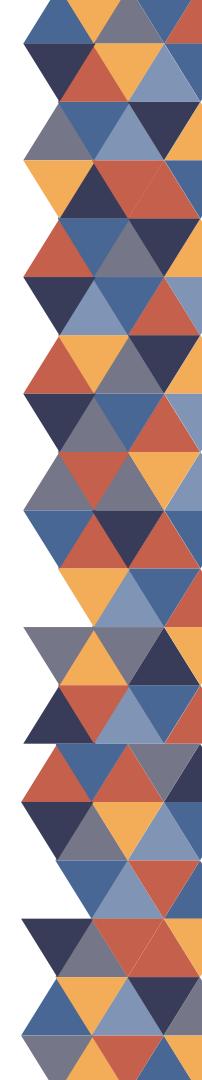
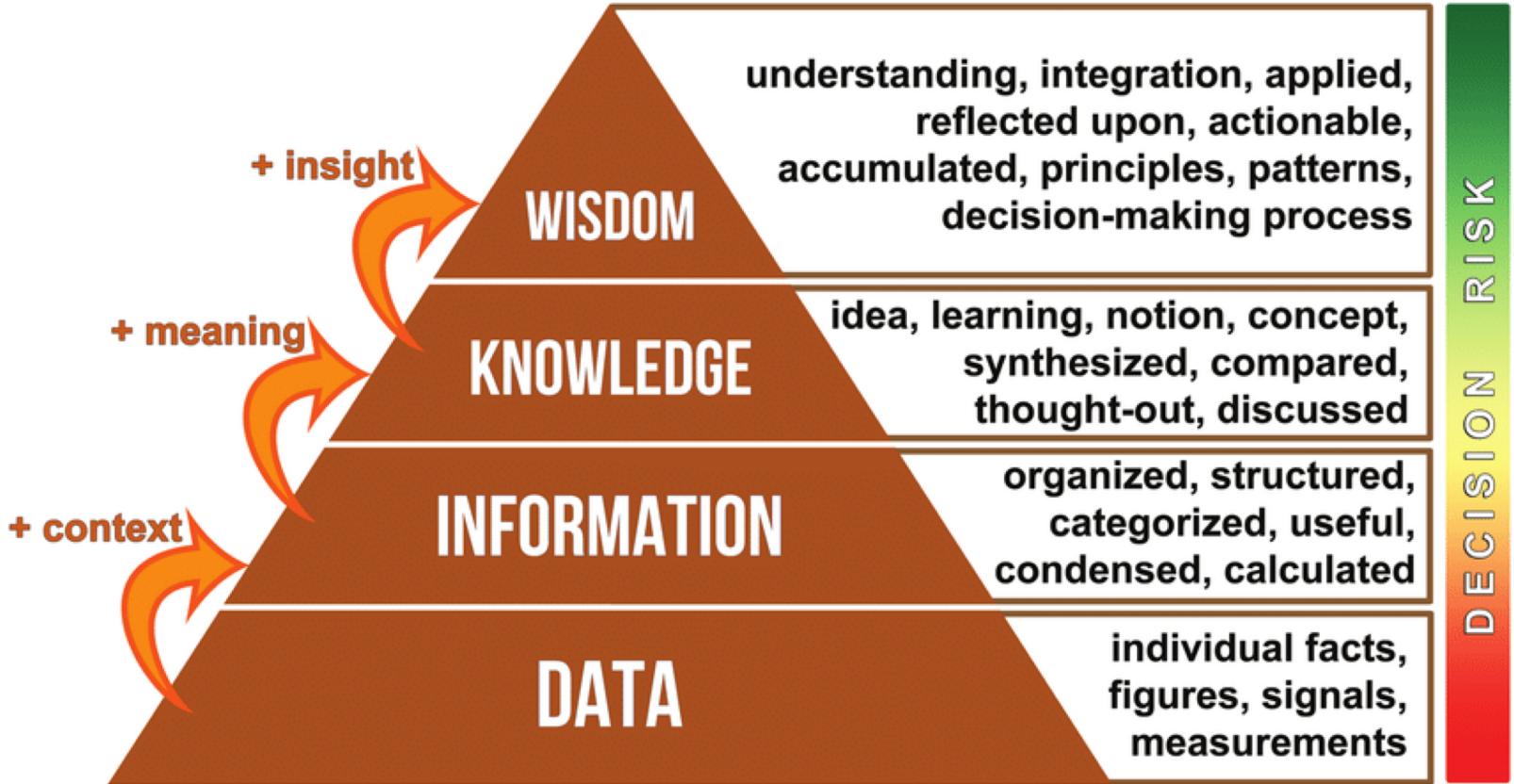
DATA SCIENCE

An inter-disciplinary field of study that combines domain expertise, mathematics, scientific methods and computing skills to leverage data for better decision-making

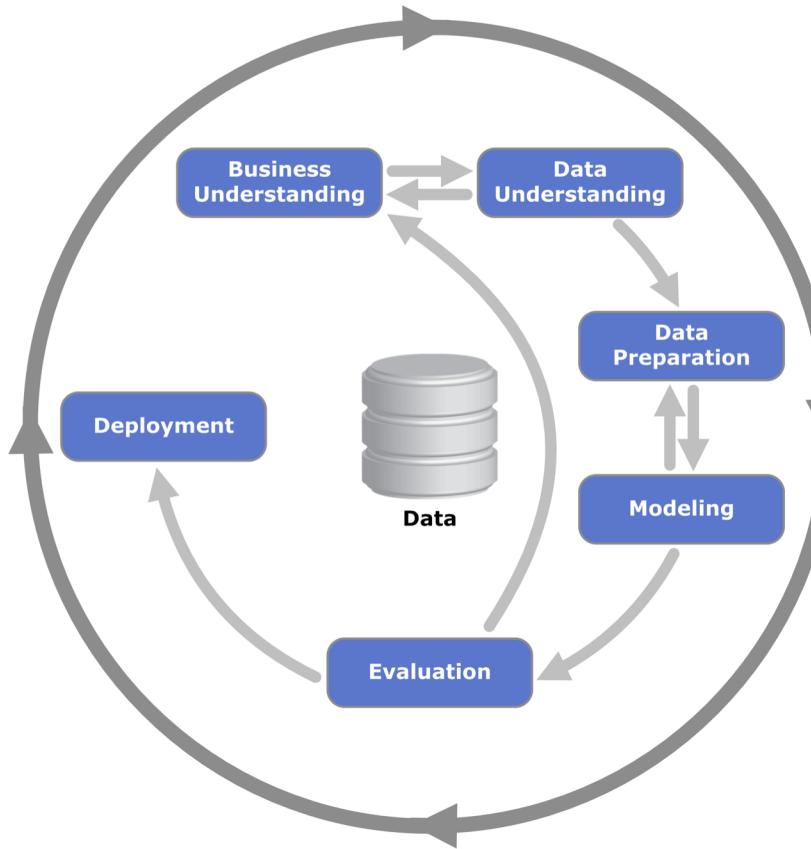




The ultimate objective of data science is to find potentially useful conclusions that can be acted upon by the users of the analysis



Data Science Process (CRISP-DM)



The **CRoss Industry Standard Process for Data Mining (CRISP-DM)** is a process model with six phases that naturally describes the data science life cycle.



Business understanding

What does the business need?



Data understanding

What data do we have/need?
Is it clean?



Data preparation

How do we organize the data
for modeling?



Modeling

What modeling techniques
should we apply?



Evaluation

Which model best meets the
business objectives?



Deployment

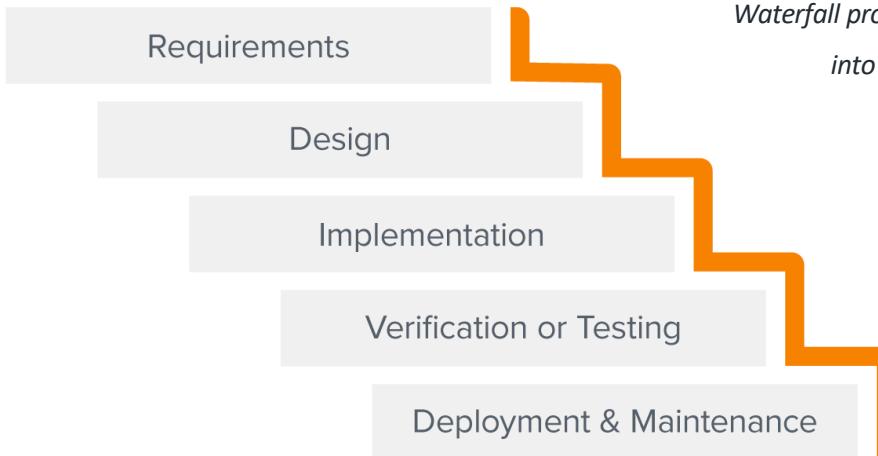
How do stakeholders access the
results?



AGILE in DATA SCIENCE



The Waterfall Method



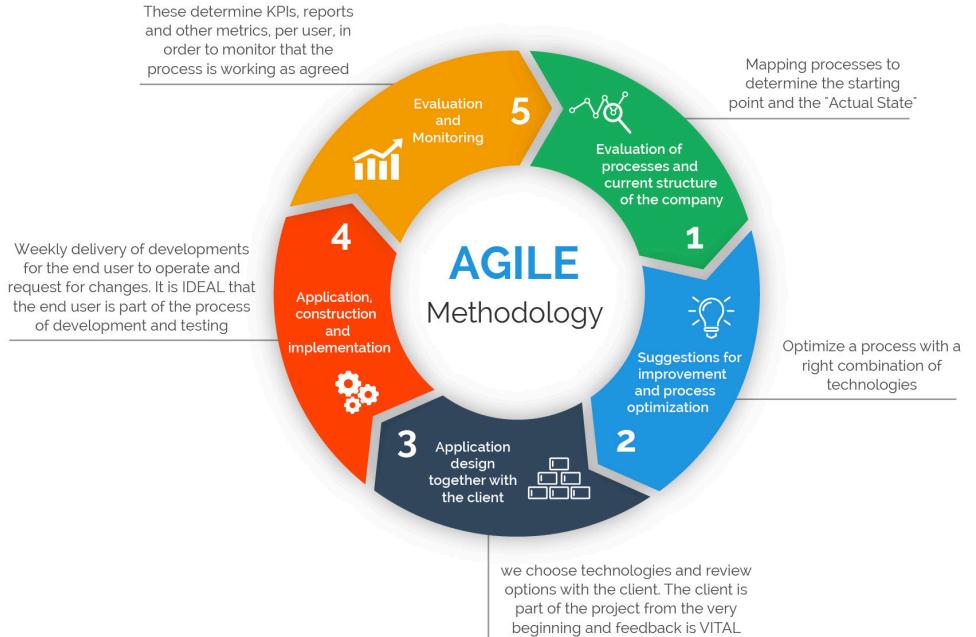
Waterfall project management maps out a project into distinct, sequential phases, with each new phase beginning only when the previous one has been completed.

“

“Waterfall is linear and sequential, and does not encourage changing up the process once it’s started”



Agile methodology is an **iterative** approach to delivering a project throughout its life cycle.

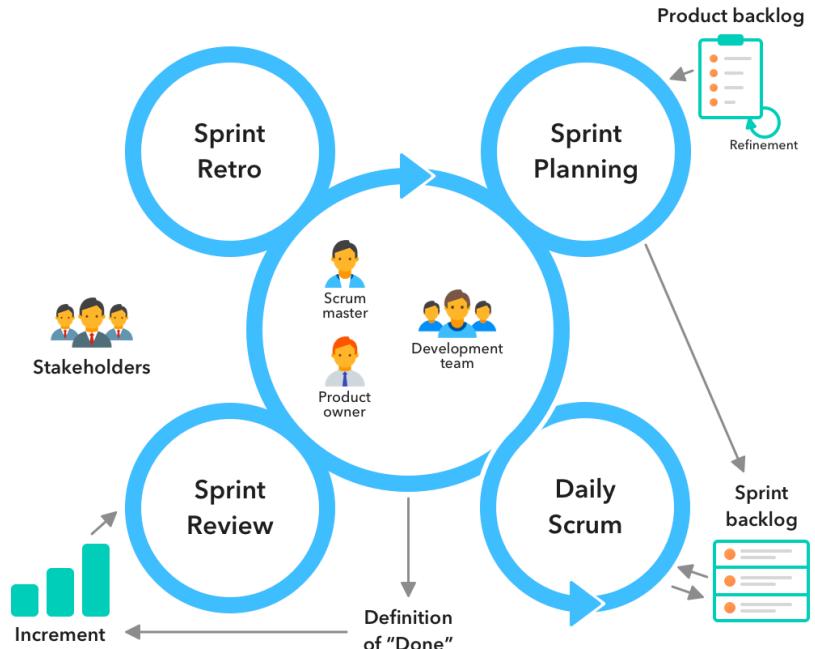


“

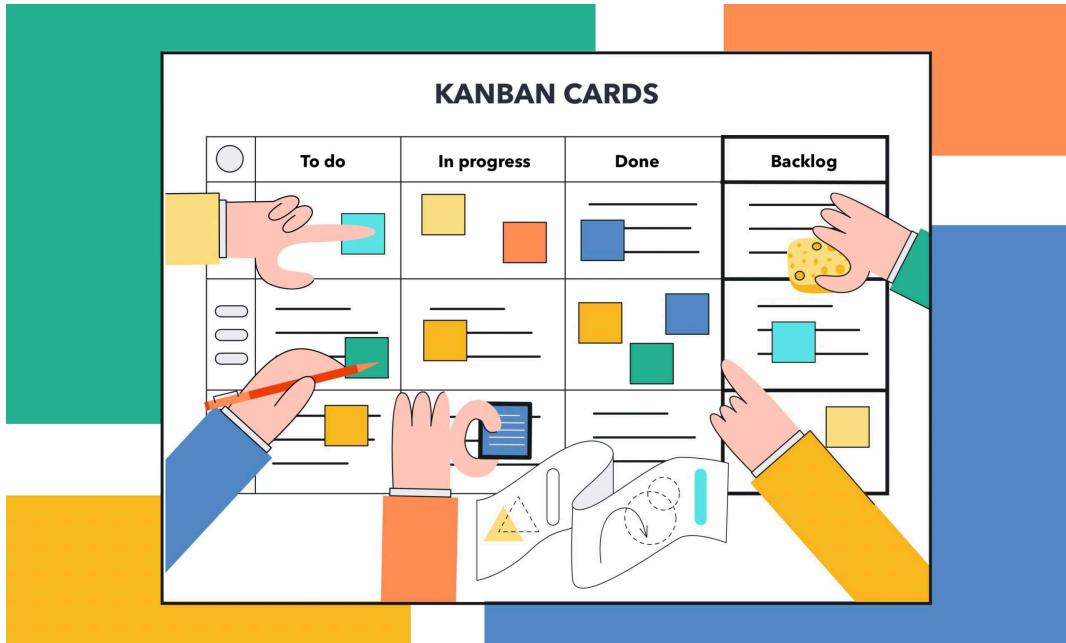
**“Agile’s iterative approach enables a project to move quickly,
as well as making it adaptive to change”**



Scrum is a project management framework that is used to efficiently produce quality work while adapting quickly to change (rather than viewing Scrum as methodology, think of it as a framework for managing a process).



Kanban is a workflow management method for defining, managing and improving services that deliver knowledge work. It aims to help you visualize your work, maximize efficiency, and improve continuously.



Is Agile for Data Science?

More Relevant Deliverables

Quicker Delivery of Customer Value

Real Feedback

Cut Losses Early

Improved Communication

Challenges of Applying Agile to Data Science?

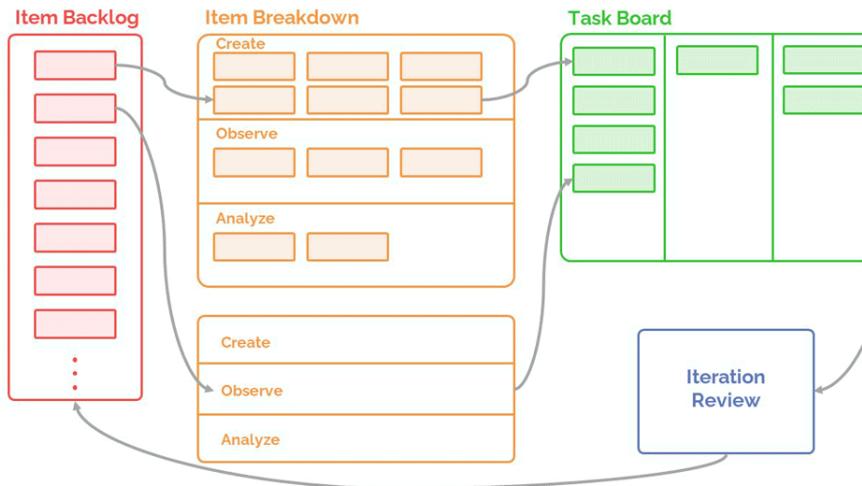
Less Straight-forward

Longer Time Horizons



Scrum-Kanban hybrid approach

- While Scrum requires all iterations (sprints) to be of equal length in time. In Scrum-Kanban, iterations vary in duration to allow a logical increment of work to be done in one iteration (rather than defining the amount of work that can be done in a specific unit of time).
- Scrum-Kanban employs Kanban principles (e.g., there is a Kanban board, DS teams need to limit WIP, and work items flow across the board).



Differences from Traditional Scrum

Functional Iterations

Uncertain Task Duration

Collective Analysis

Iteration-Independent Meetings



Tips for Agile-oriented Data Science team

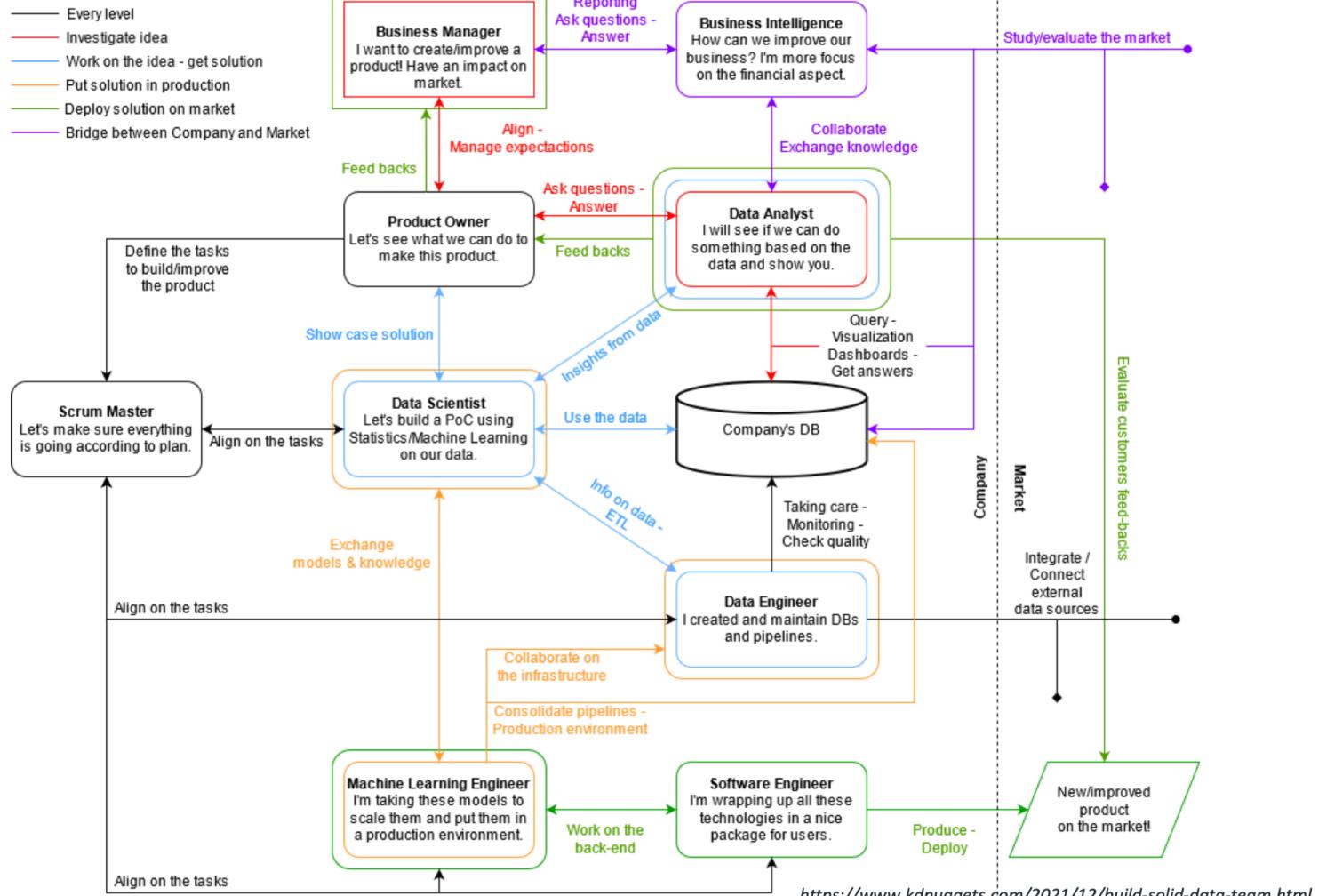
- Have Fully-Functional Teams
- Allow Teams to Self-Manage
- Start Simple and Iterate Quickly
- Measure Effectively
- Frequently Collaborate
- Have Flexible Plans



An Agile-oriented Data Science team

- **Product owner** decides which features and functionality to build, the order in which to build them, and what aspects of them to observe and analyze. The product owner owns the Item Backlog and is responsible for prioritizing its BItems, ensuring that each BItem is clearly defined, and that the upcoming work and priorities of the team are visible and transparent.
- **Scrum master** serves as the process master and acts as a coach, facilitator, impediment remover as well as helping everyone involved understand and embrace the values, principles, and practices to aid the organization obtain exceptional results from applying Agile in Data Science.





Thanks!

Q&A

References

- What is CRISP DM?
- Why using CRISP-DM will make you a better Data Scientist
- What is Data Science?