

Terminology

Topic 11b – Part 2

1. What We'll Cover

This topic will introduce...

- What is data science.
- Key concepts – the scientific method.
- Useful terminology.

} Part 2

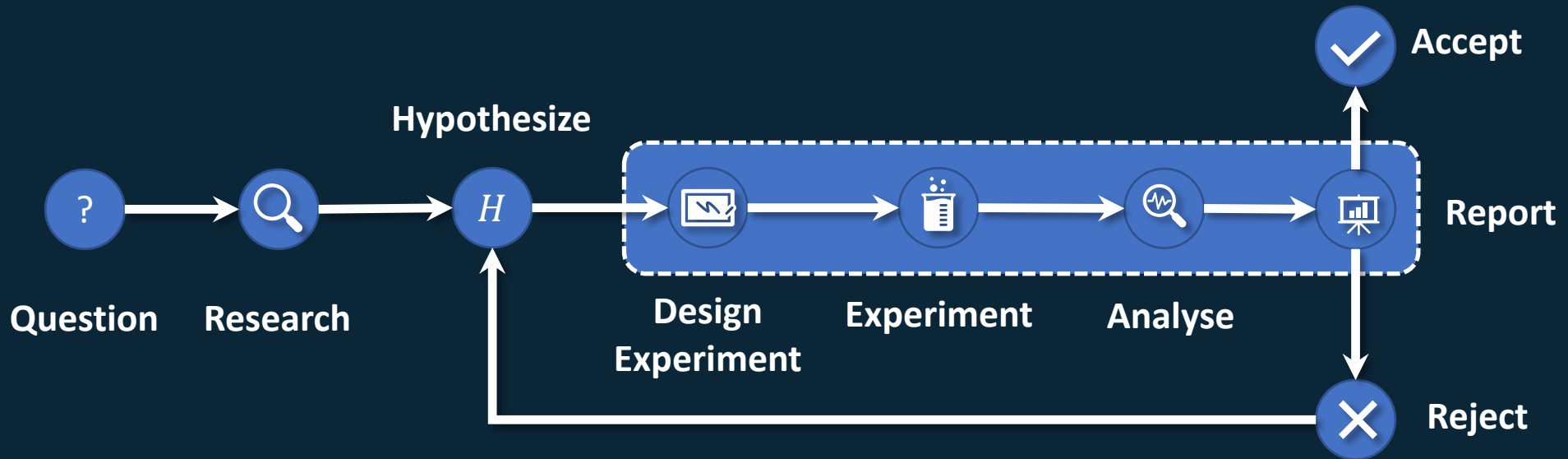
- Important tools - Statistics.
- Data collection & Experiment Design.
- Probability basics.
- Data distributions.
- Hypothesis testing.

The aim: to help you understand what it means to be a data scientist and to get you familiar with data science tools.

A word cloud featuring various terms related to data science and machine learning. The words are arranged in a circular pattern, with some appearing more prominently than others. The terms include: order, part, future, neural, past, train, like, values, allows, fit, clustering, Predictive, split, database, true, number, predict, sets, text, SQL, layers, model, uses, example, try, etc, help, value, test, many, training, layer, idea, Using, business, social, groups, contain, Granger, labelled, steps, manage, networks, best, Models, often, learning, deep, used, data, analytics, set, large, false, inputs, better, variety, Feature, point, advanced, mining, fed, go, label, big, different, rate, ETL, and 20. The words are in various colors and sizes, creating a dense and colorful visual representation of the field.

A sepia-toned historical map of the Americas, featuring various transportation modes: a steam locomotive in the north, a hot air balloon in the upper right, a three-masted sailing ship in the lower left, and an early motor car in the lower right. The map includes labels for 'AMERICA', 'NORTH AMERICA', 'SOUTH AMERICA', and 'ATLANTIC OCEAN'.

3. Data Science Process



4. Application Areas

Analytics

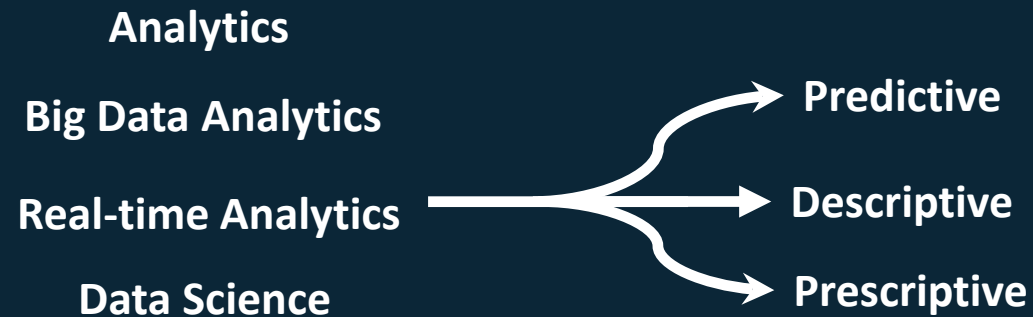
Big Data Analytics

Real-time Analytics

Data Science

“Data science is a technical discipline concerned with the extraction of new knowledge from data, via application of the scientific method, in conjunction with the tools of mathematics and statistics.”

5. Application Areas



Business Intelligence / Analytics

Healthcare Analytics

Fraud Analytics

...

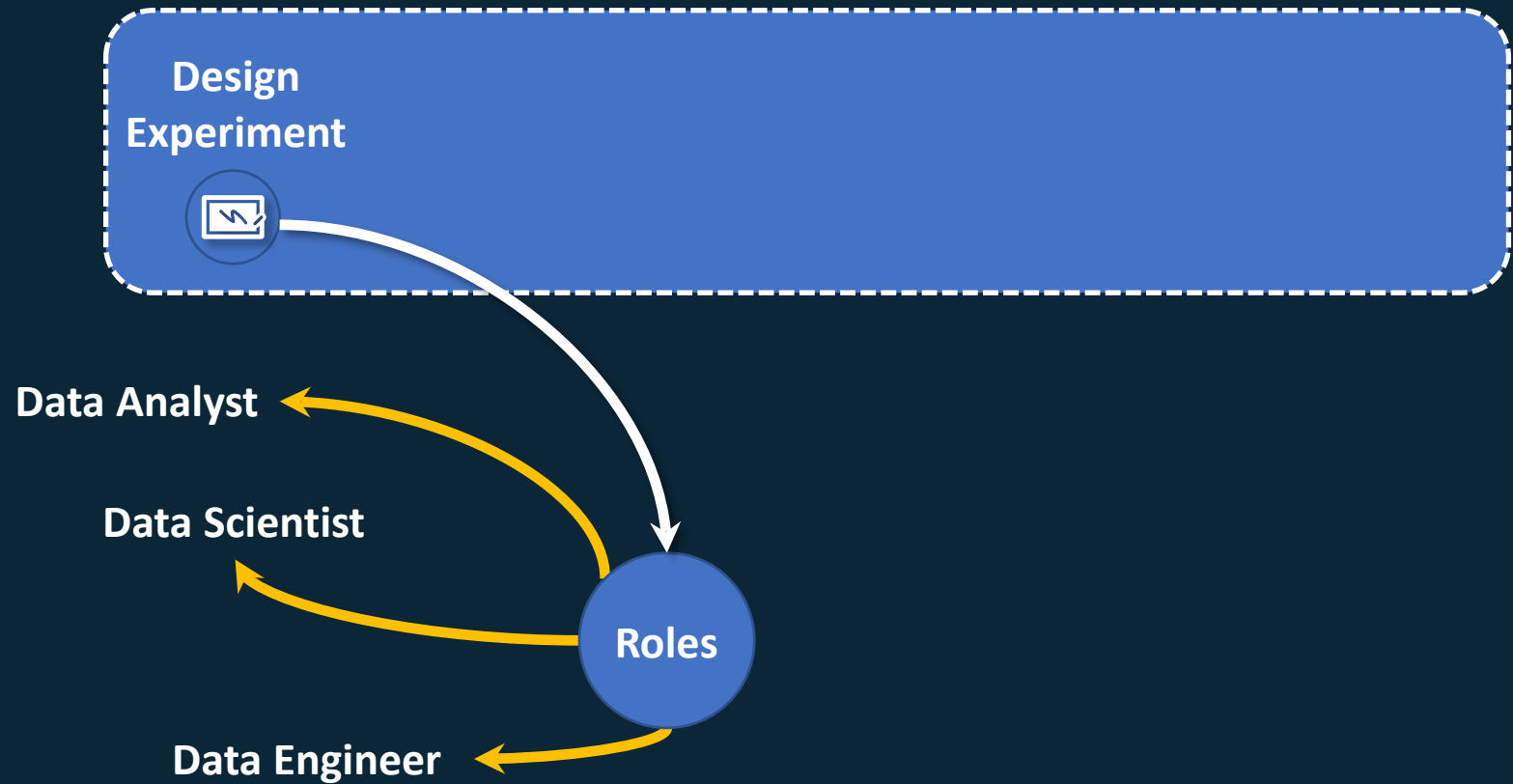
“Data science is a technical discipline concerned with the extraction of new knowledge from data, via application of the scientific method, in conjunction with the tools of mathematics and statistics.”

6. Application Areas

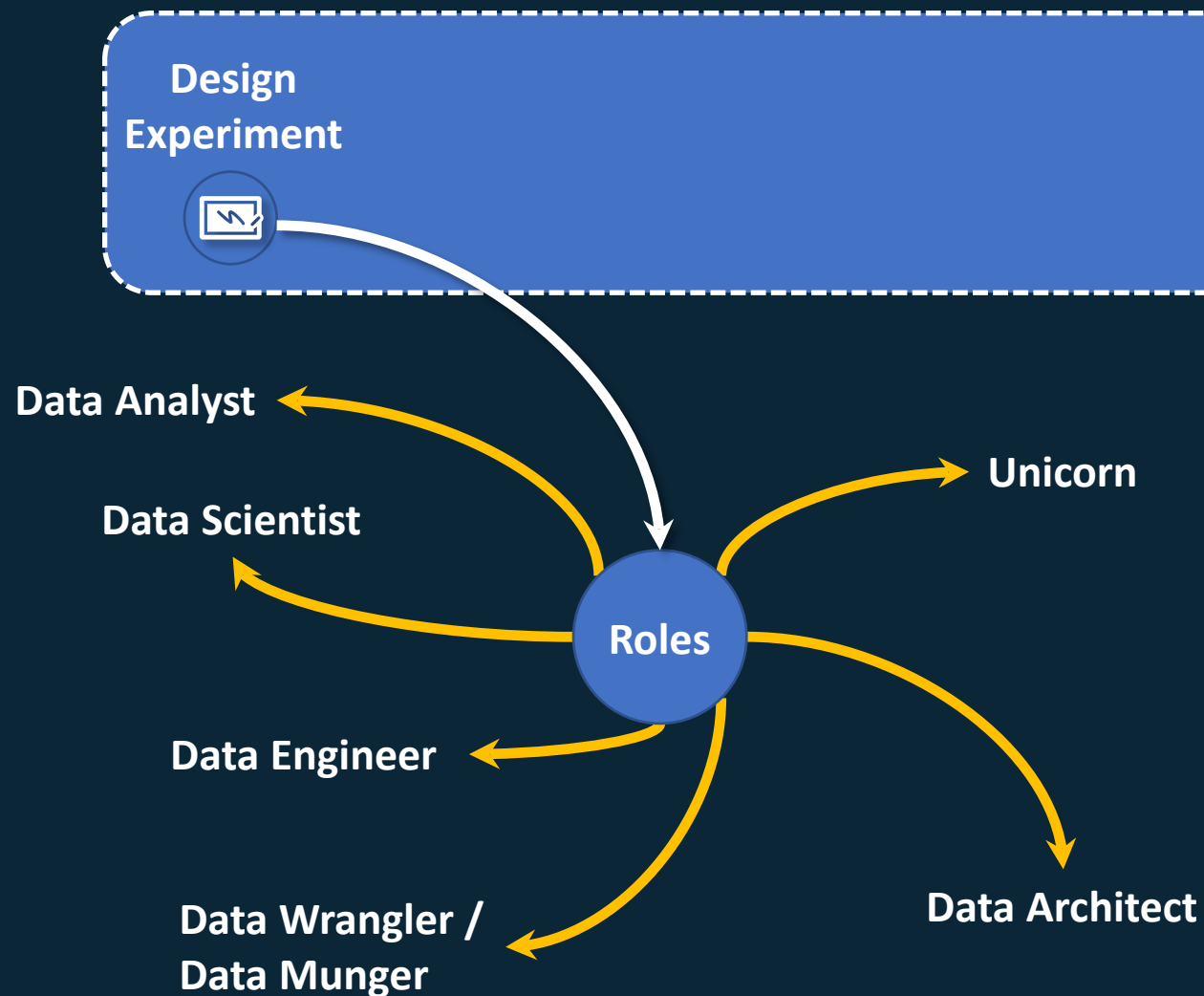


Credit: 365 Data Science

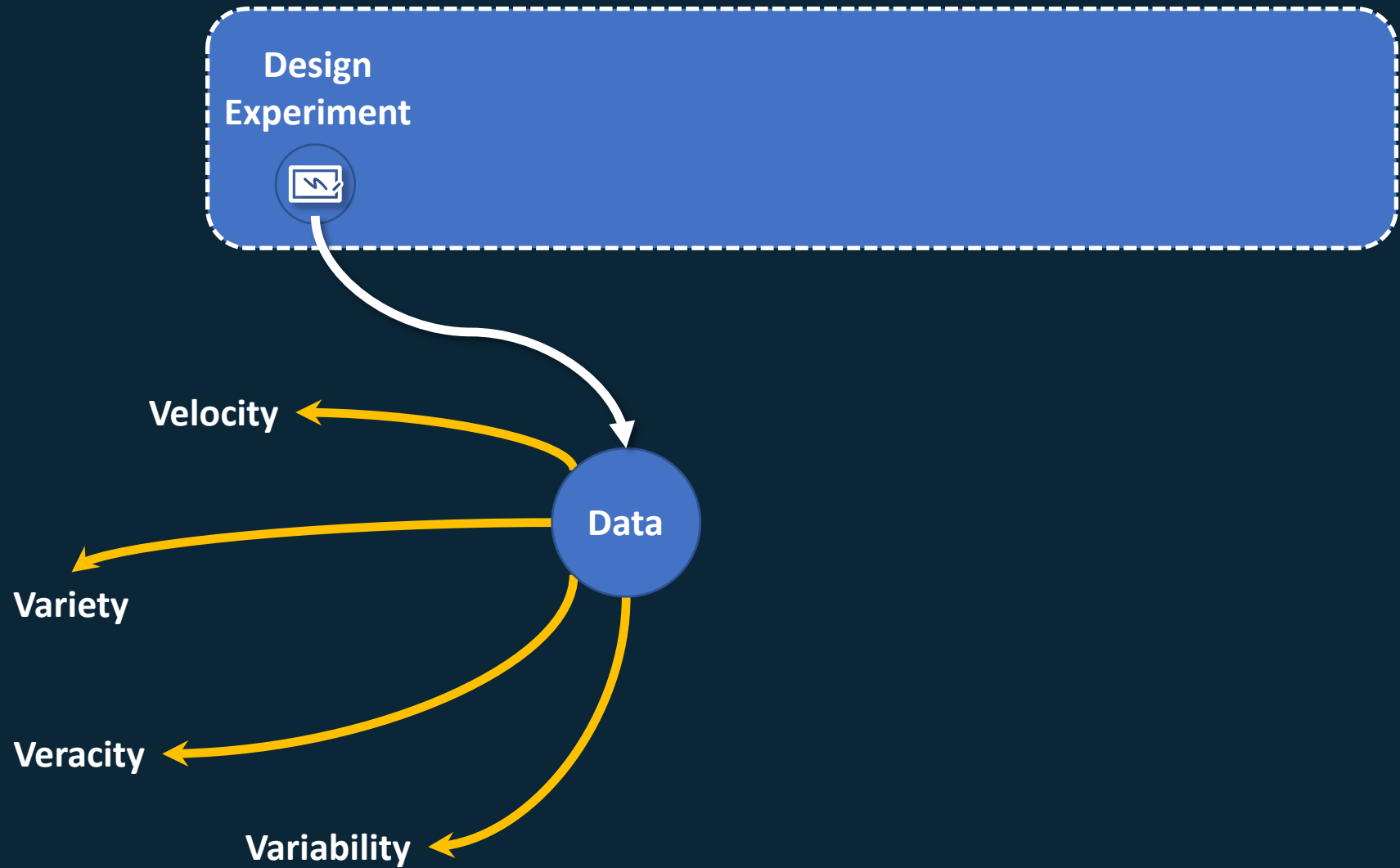
7. Roles



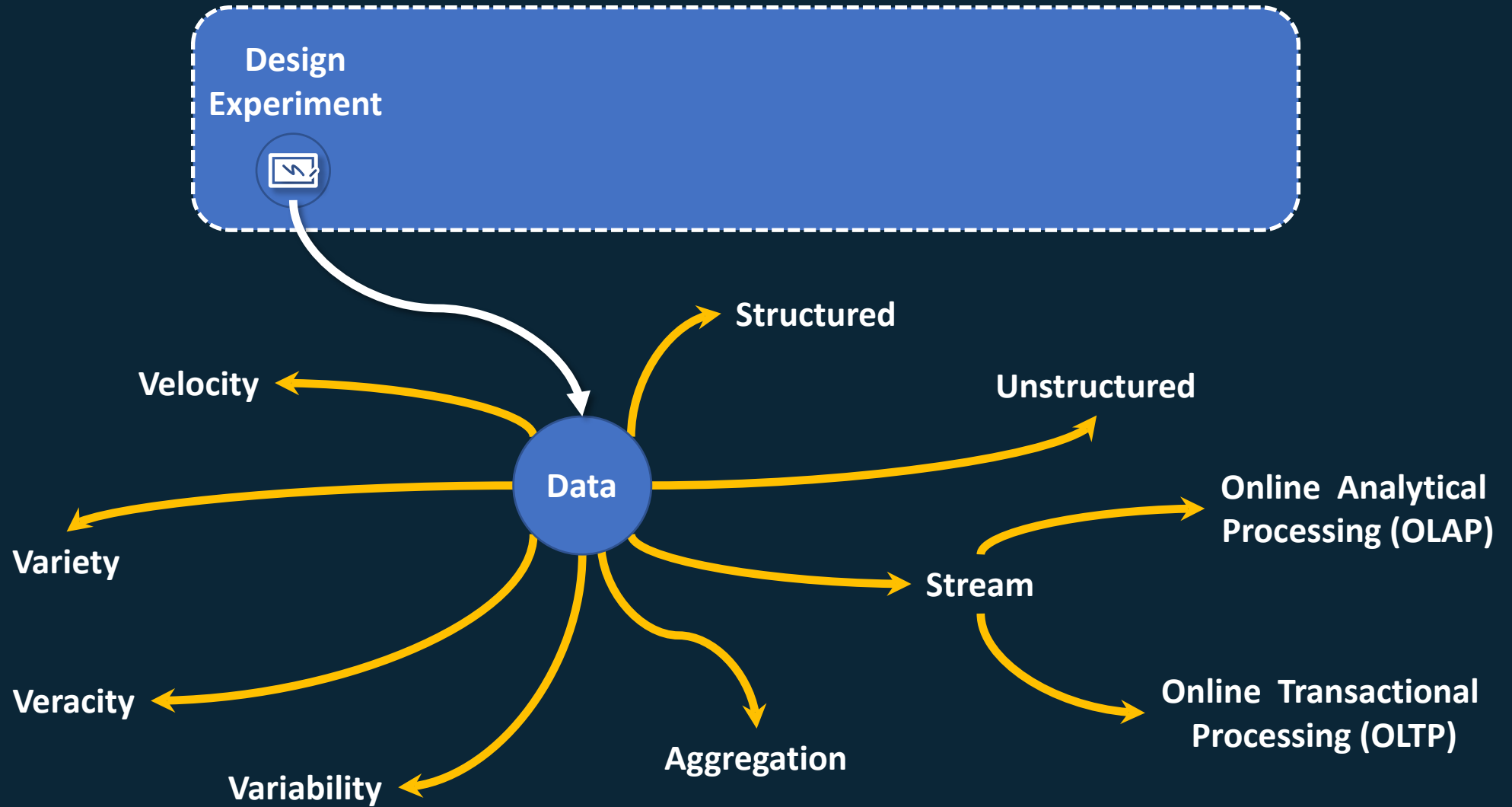
8. Roles



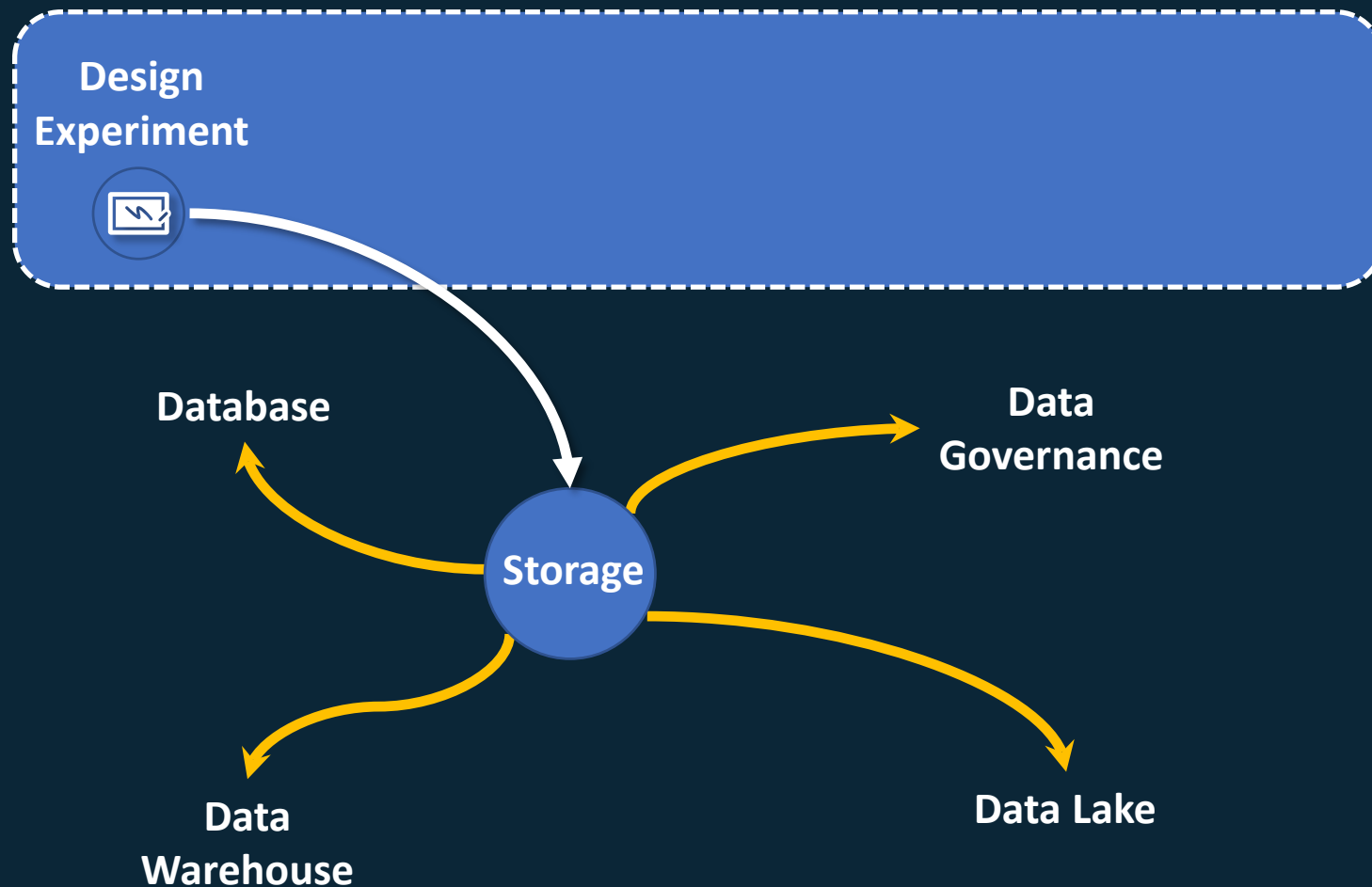
9. Data



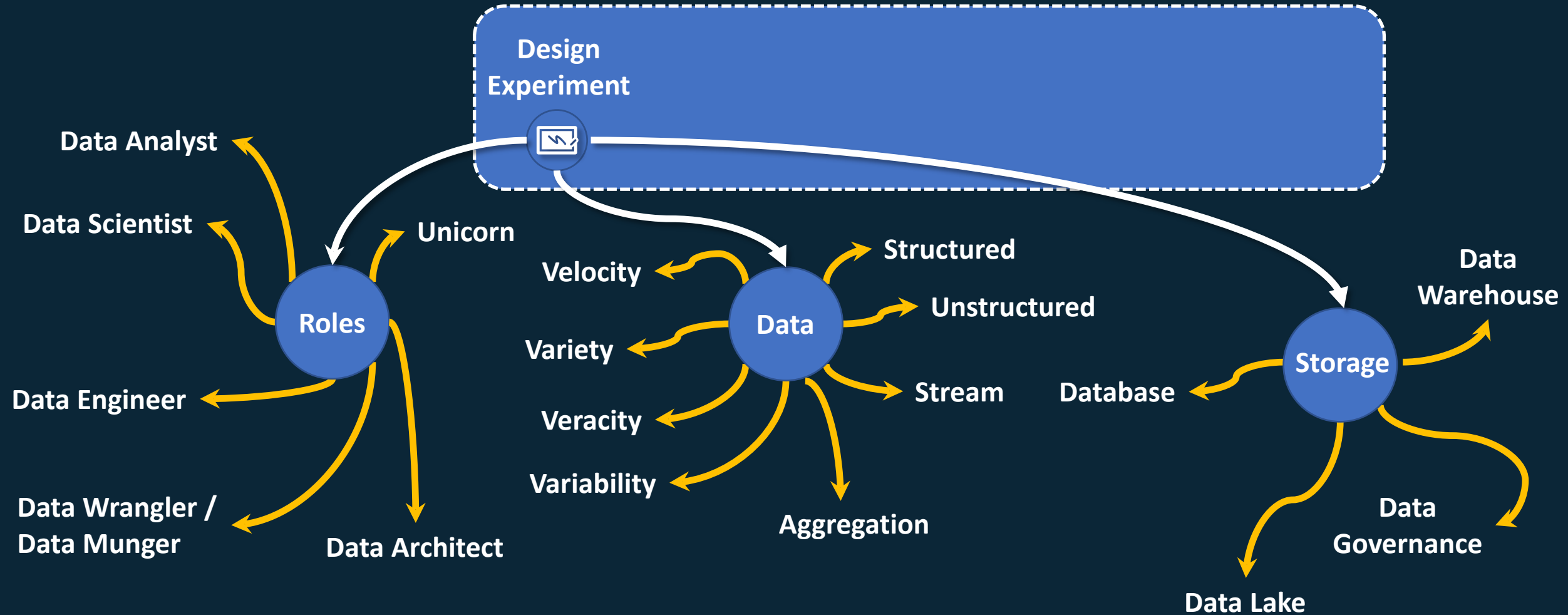
10. Data



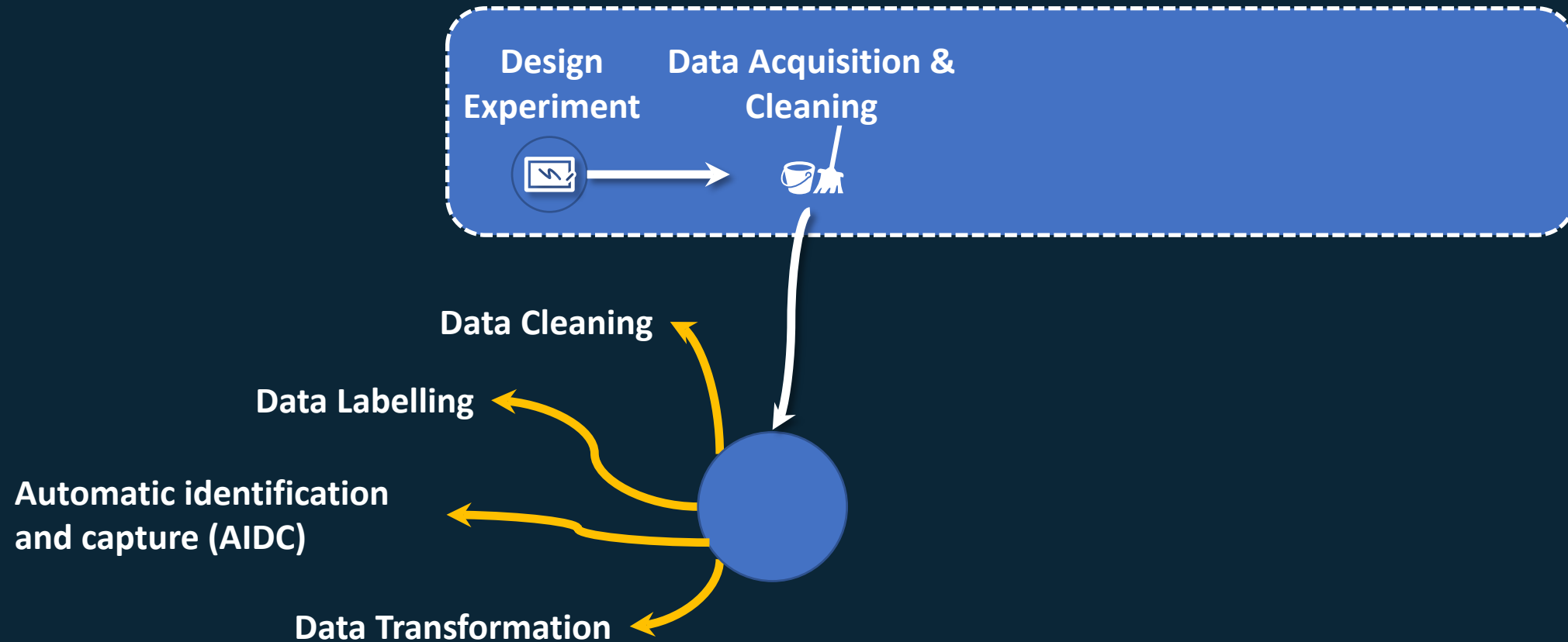
11. Storage



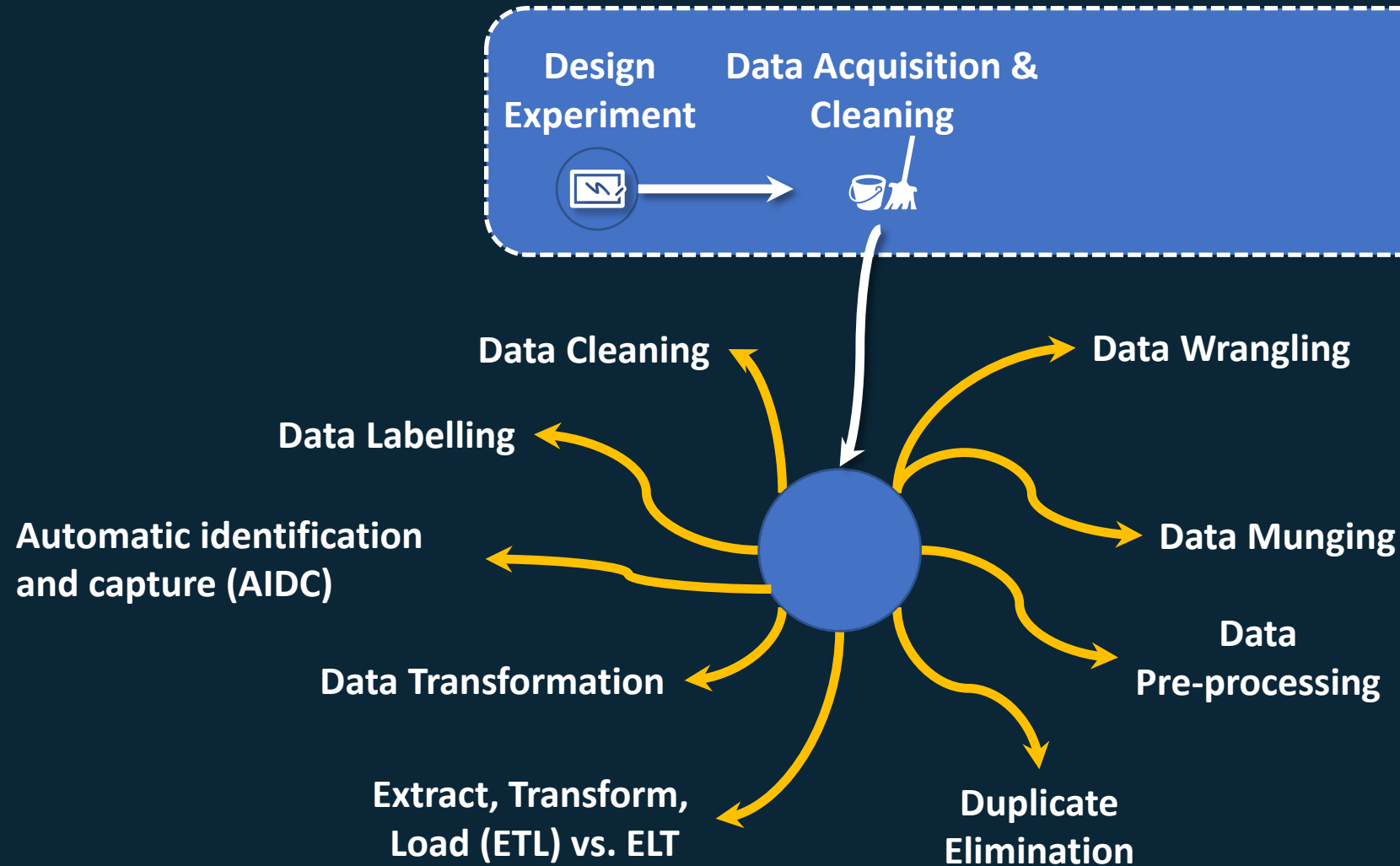
12. Macro View



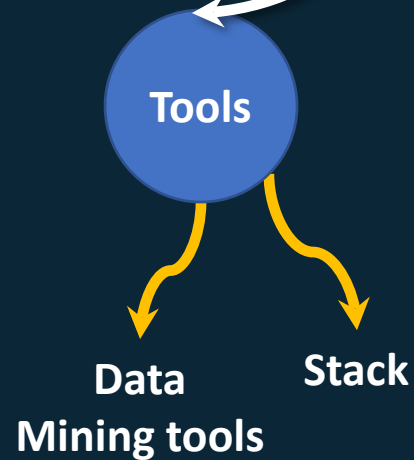
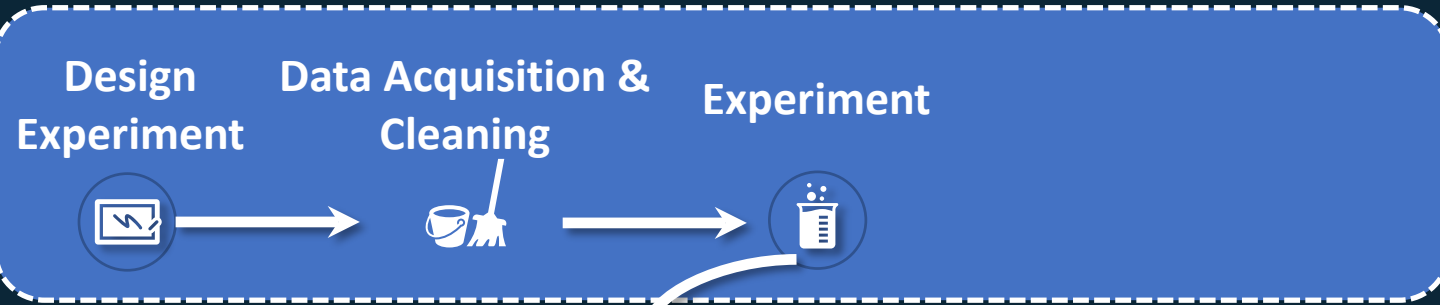
13. Acquiring and Preparing Data



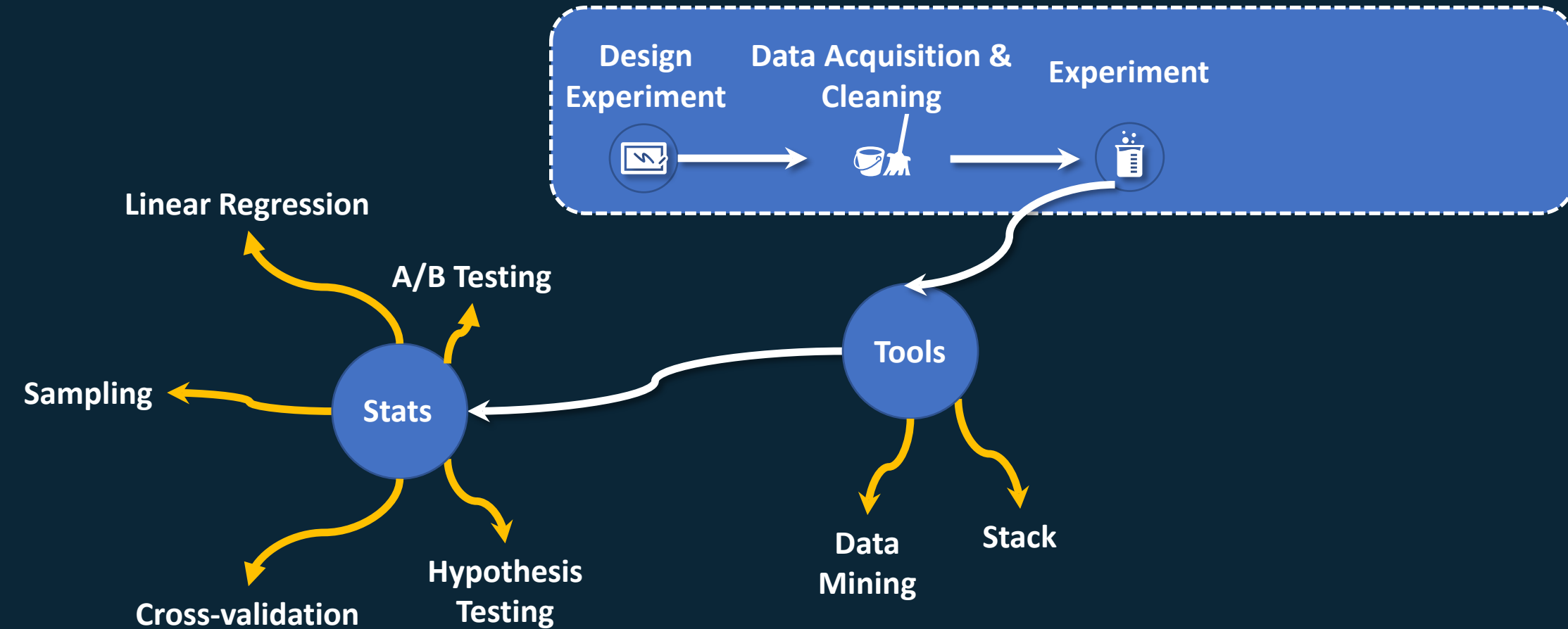
14. Acquiring and Preparing Data



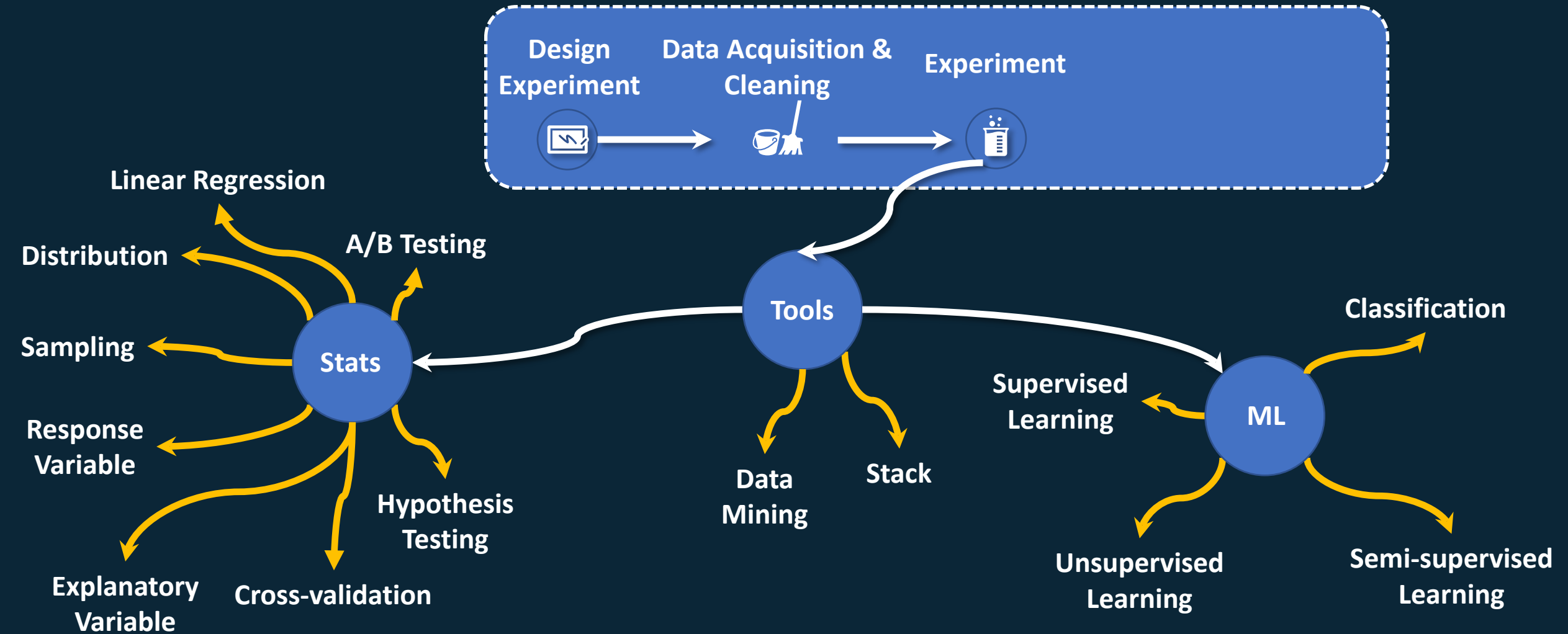
15. Statistics & Machine Learning



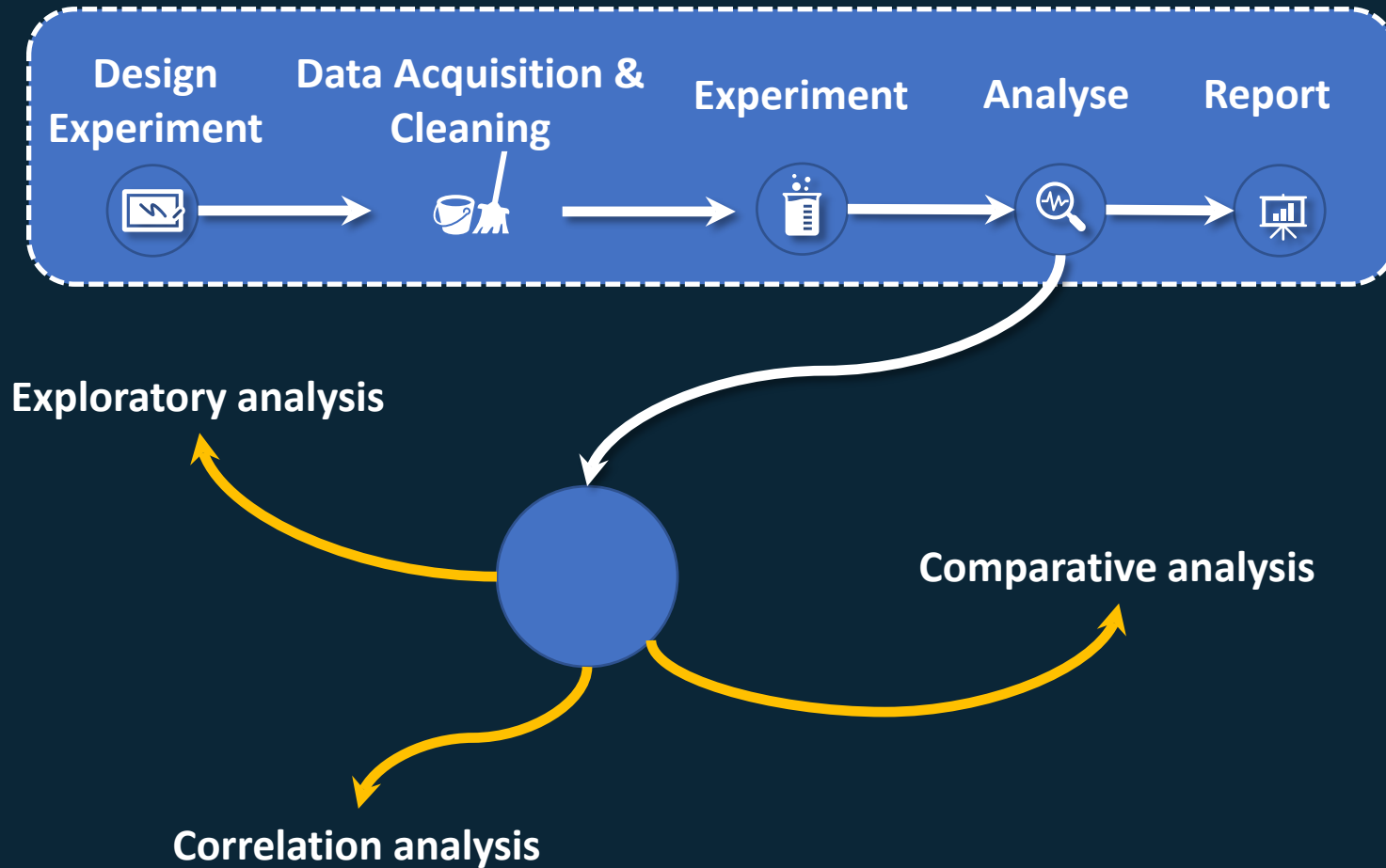
16. Statistics & Machine Learning



17. Statistics & Machine Learning



18. Analysis



20. Resources

Here are some resources you may find useful:

The data science “glossary” – fairly complete resource:

<http://www.datascienceglossary.org/>

Microsoft’s Machine Learning / Data Science Glossary:

<https://docs.microsoft.com/en-us/dotnet/machine-learning/resources/glossary>

Then there are some other useful online discussions, such as this:

<https://towardsdatascience.com/the-a-z-of-ai-and-machine-learning-comprehensive-glossary-fb6f0dd8230>