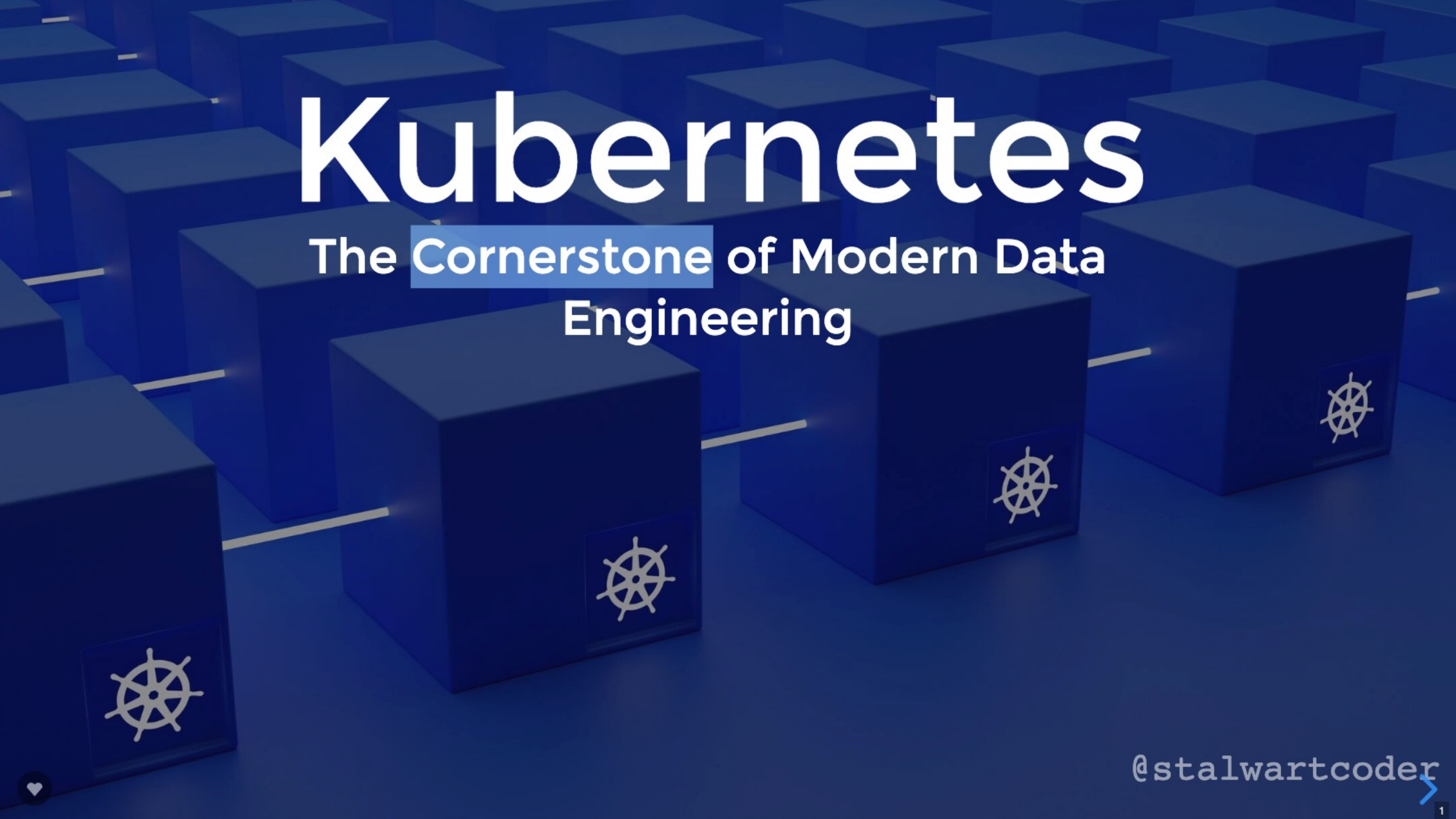

Kubernetes
The Cornerstone of Modern Data Engineering
@stalwartcoder

# What this talk is, and is not!?

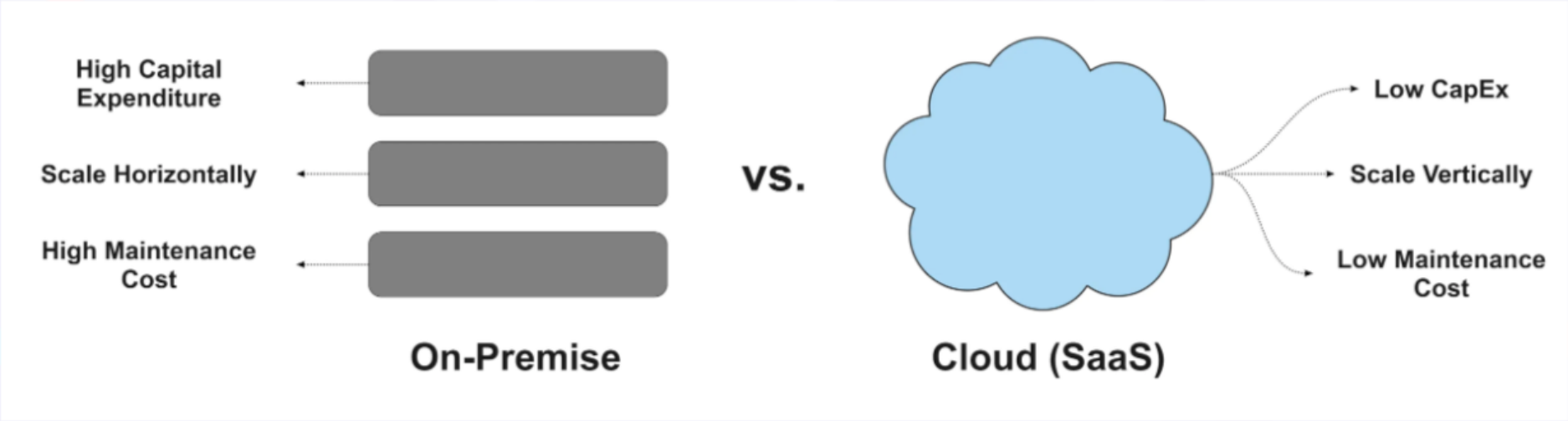
✅ Trip around bird's eye view of how Kubernetes being used in data engineering

✅ Intro of data engineering and modern data stack

✅ Quick tour of data eng. tools around Kubernetes

🙅‍♂️ Not a tutorial

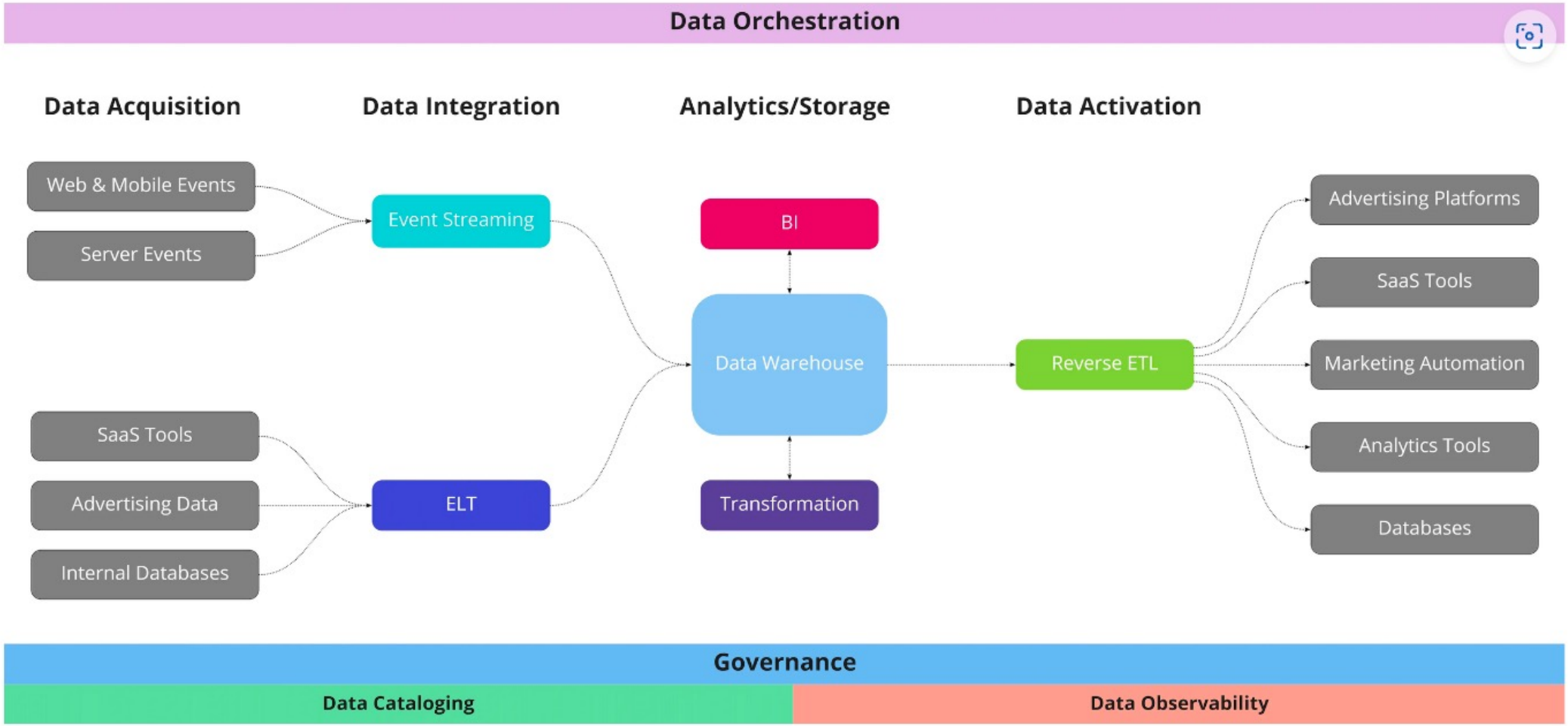
# Modern Data Stack




Legacy Data Stack vs. Modern Data Stack

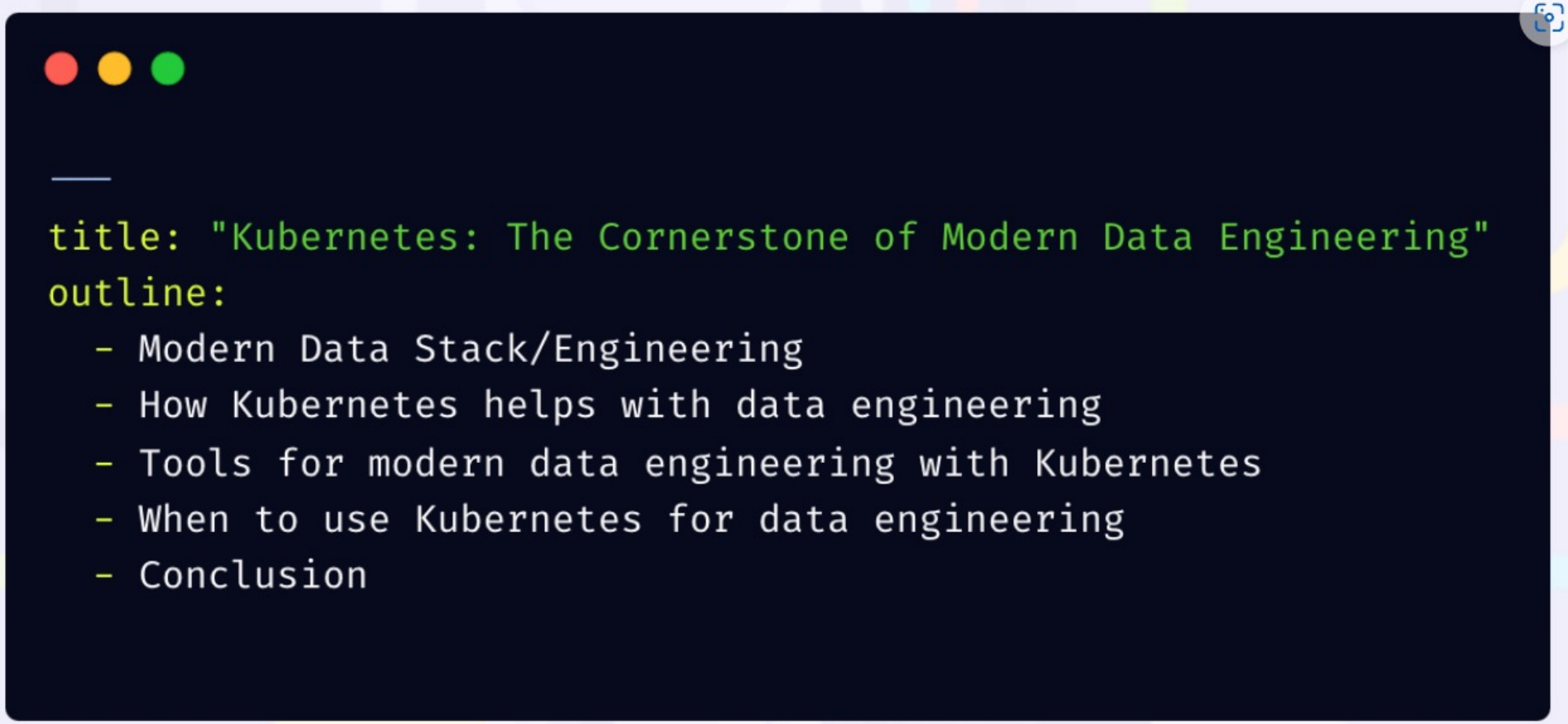
# Modern Data Stack

**Core differentiator:** the difference between on-premises hardware and cloud-native tools.




Modern Data Stack Components

# agenda_workflow.yaml



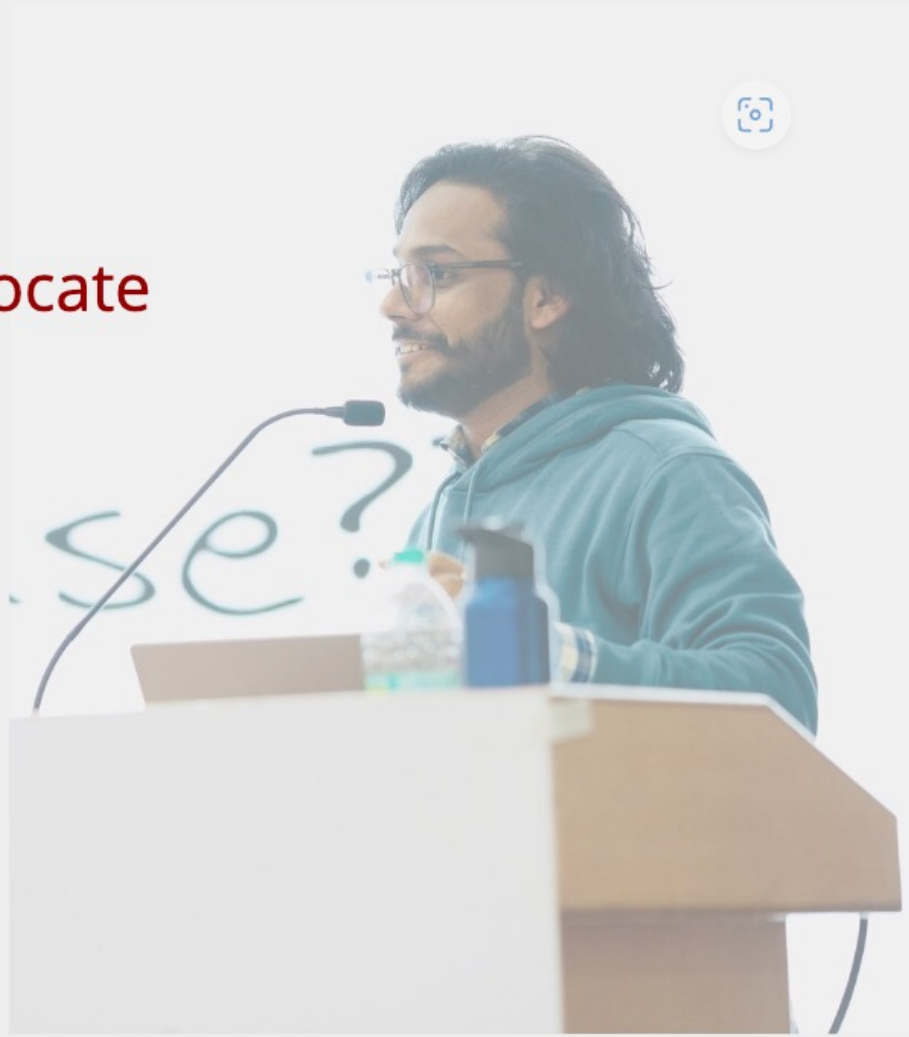
```yaml
---
title: "Kubernetes: The Cornerstone of Modern Data Engineering"
outline:
  - Modern Data Stack/Engineering
  - How Kubernetes helps with data engineering
  - Tools for modern data engineering with Kubernetes
  - When to use Kubernetes for data engineering
  - Conclusion
```

# Abhishek

- 🥑 Engineer & Aspiring Dev Advocate
- 🐍 Pythonista & Byte of 🦀
- 👨‍👩‍👧‍👦 Community first person 💛
- 👨‍💻 **"1x.engineer"**

***connect with me:*** @stalwartcoder

@stalwartcoder



# Why

- Dynamically scaling the computing scaling capacity relative to current needs by adding or removing servers to/from the cluster.

- Containerized data workloads.

- Orchestrating containers
- Declarative definition
- Communicating between teams
- Seamless Scale as data grows
- Iterating faster

# An Example

https://bit.ly/k8s-spark-data

# Some Tools ⚒

- Apache Airflow
- Argo Workflows
- Prefect
- KubeFlow Pipeline
- Dask on k8s

# When to use k8s

- Run scalable and reliable data pipelines
- Automate the management of ML models
- Track experiments and data lineage

# kubectl conclusion

**Kubernetes is like a data engineer's GPS –
it helps you navigate through the cloud-native terrain and keeps you from getting 'lost in containers' !**


**Access Slide:** https://slides.come/stalwartcoder