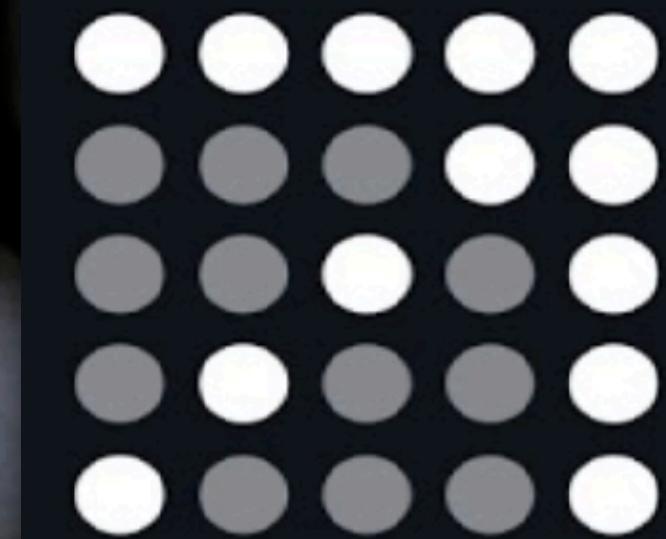




NEW  
PRO  
LAB



ALIGNED  
RESEARCH  
GROUP

# Кто такой Data Engineer?

Краткий курс юного трубопрокладчика

Николай Марков (@enchantner), Aligned Research Group, 2020

# Обо мне

- Меня зовут Николай Марков, и мой ник **enchantner** практически везде
- Я работаю Principal Architect'ом в компании Aligned Research Group
- Пишу на Python уже примерно 8 лет, еще использую Golang, C/C++, Scala
- Копался с сетями и облачными платформами (AWS, GCP, Azure, OpenStack)
- Пишу статьи, читаю лекции, организую мероприятия (заходите на PyData митапы и Data-завтраки!), записываем подкасты вместе с Moscow Python
- С удовольствием потреплюсь о современных технологиях и низкоуровневой магии



# Три пути



Программист

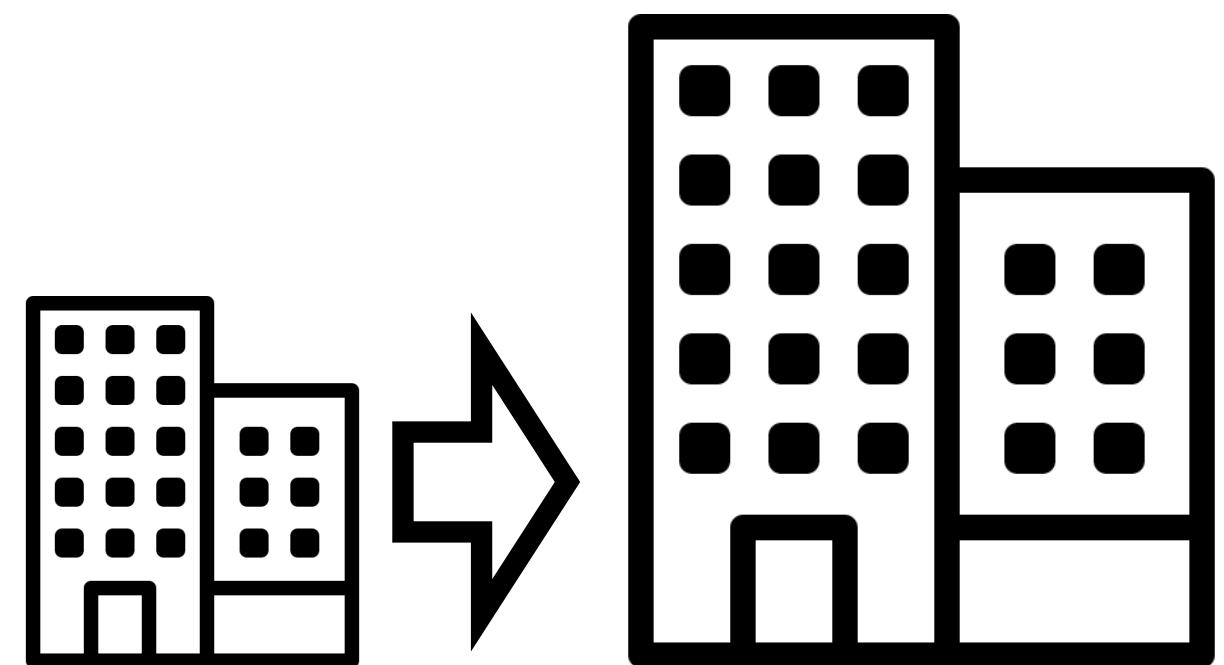


Ученый

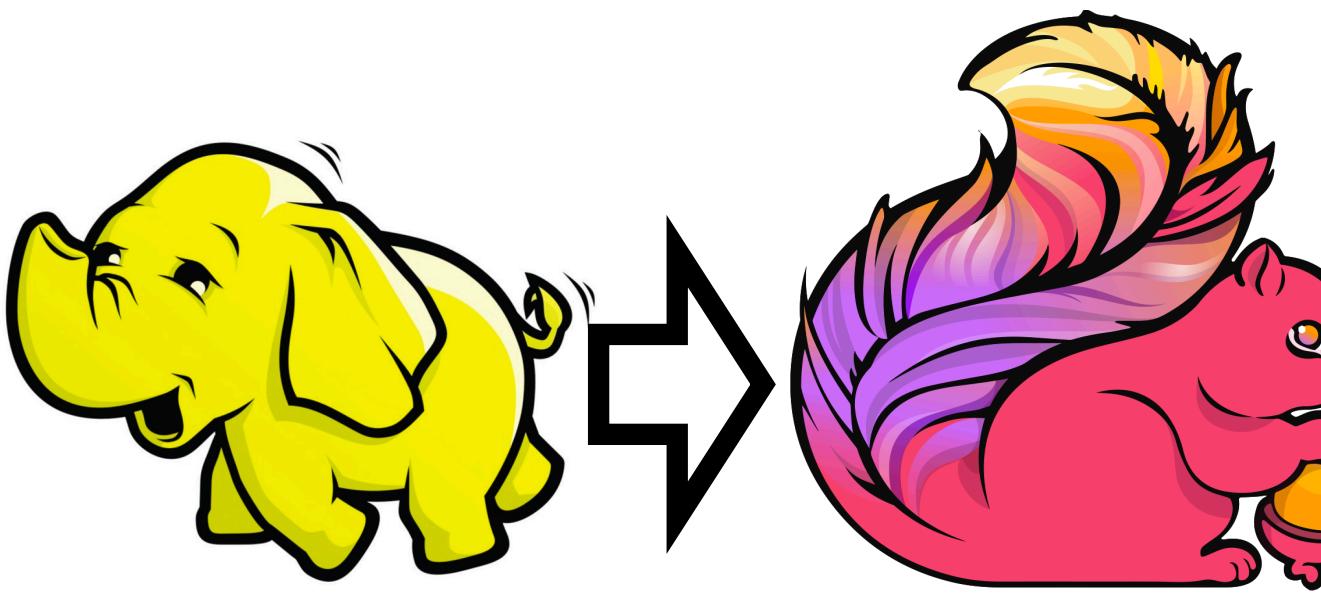


Из другой сферы

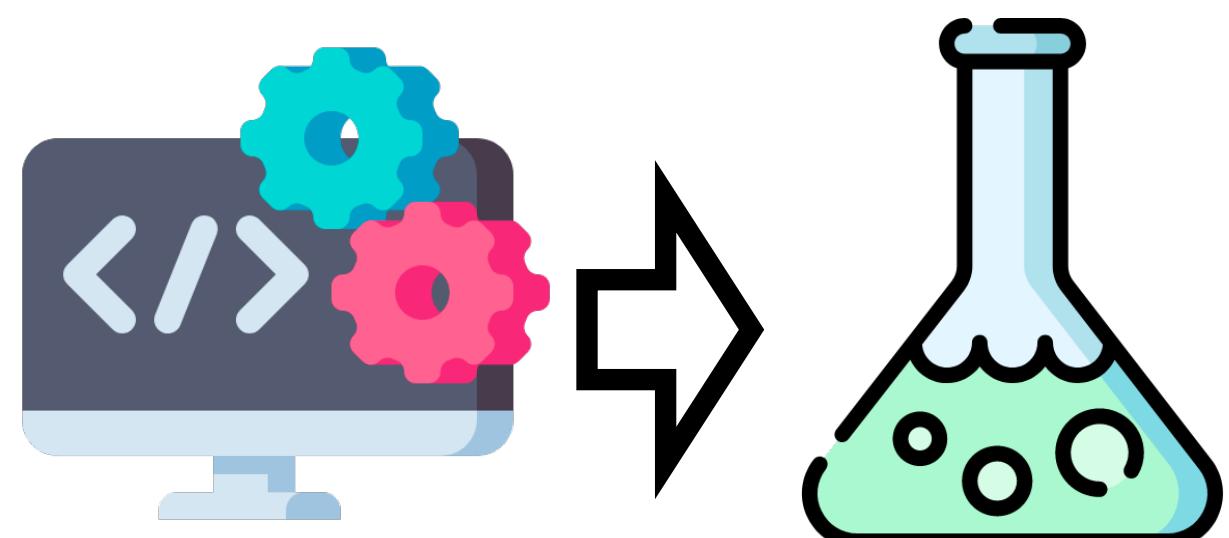
# Зависит от компании



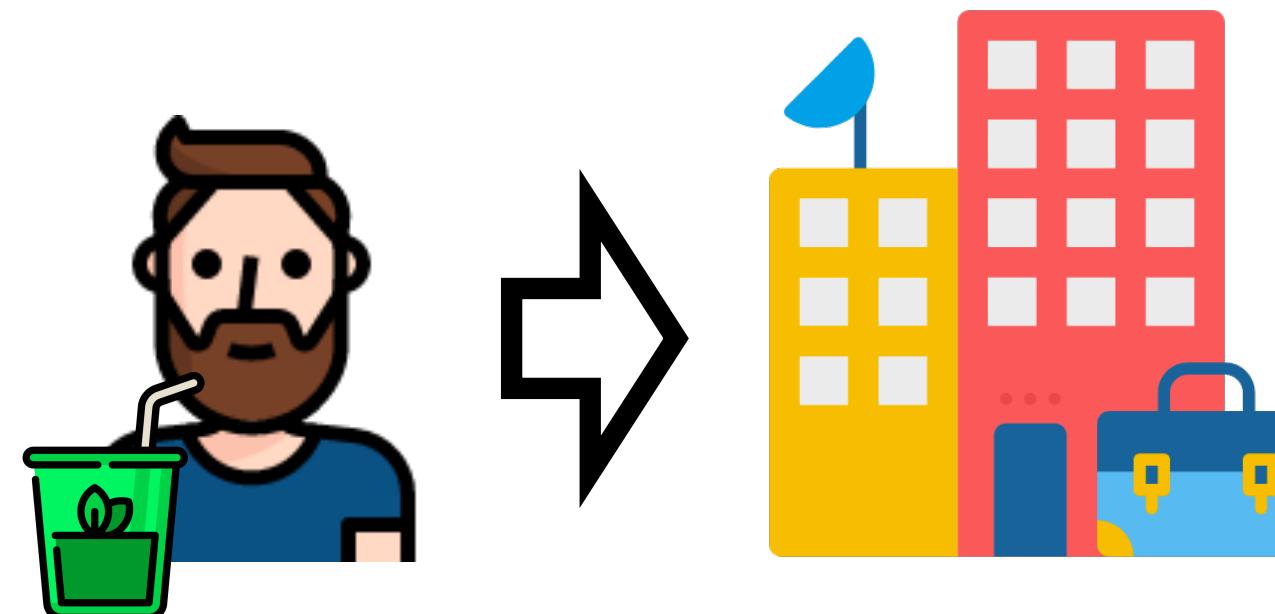
Маленькая или большая?



Legacy или свежак?



Продуктовая или R&D?

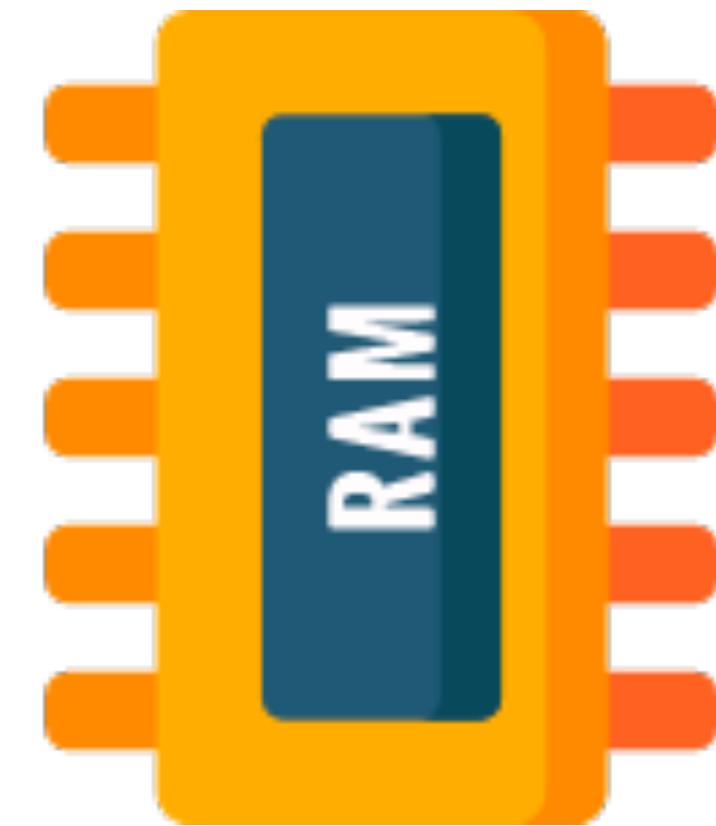


Смузи в коворкинге или  
кровавый энтерпрайз?

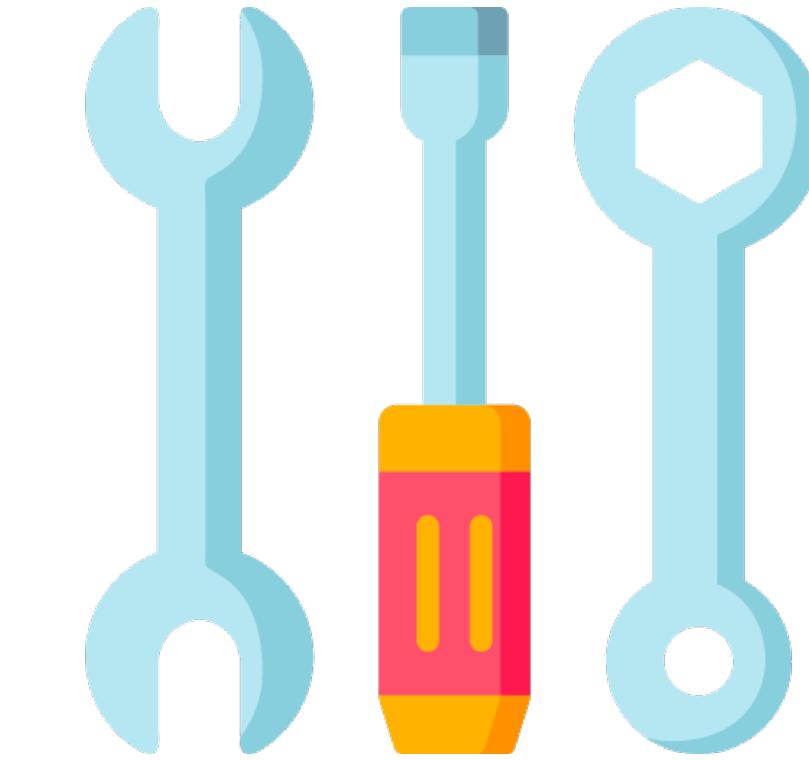
# Надооп (Java/maintenance)



**Задача:** как эффективнее преобразовывать XML в стектрейсы?



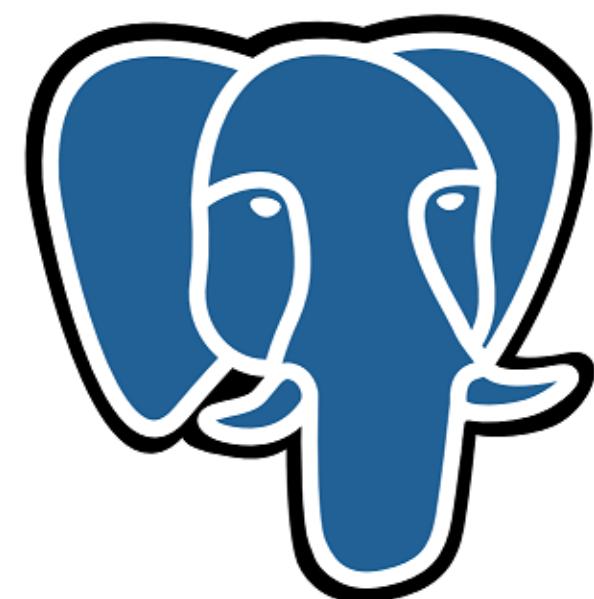
**Задача:** как эффективнее утилизировать имеющиеся ресурсы?



**Задача:** как управлять железом и деплоить новые ноды?

# ETL-разработчик

[https://en.wikipedia.org/wiki/Data\\_janitor](https://en.wikipedia.org/wiki/Data_janitor)

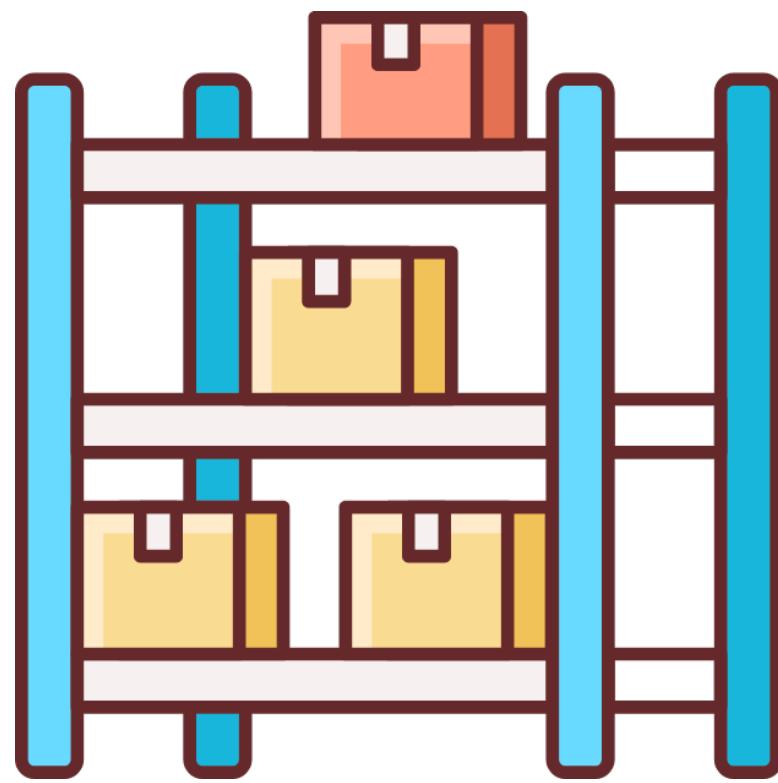


PostgreSQL

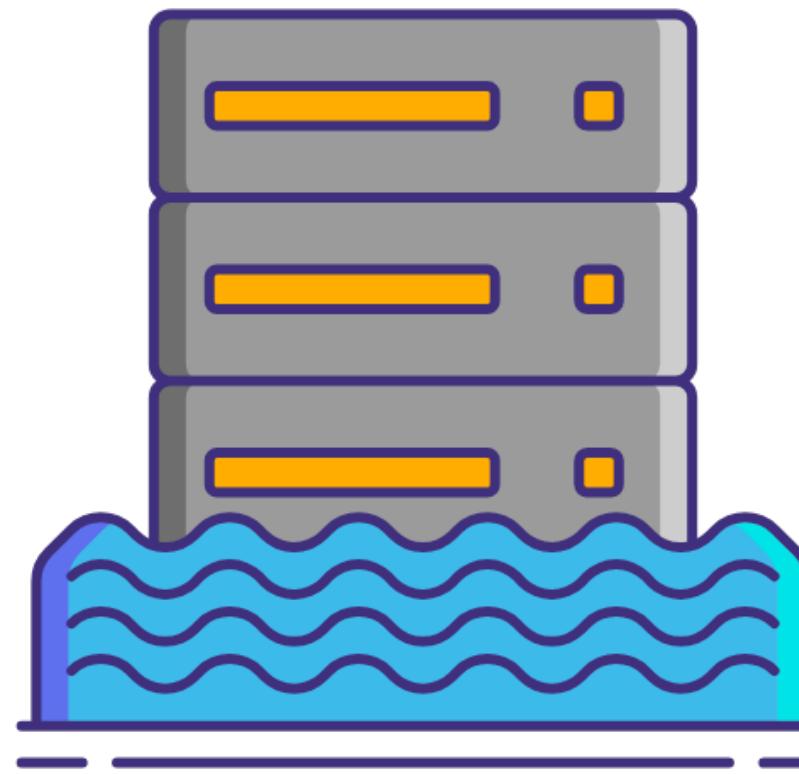


Stack

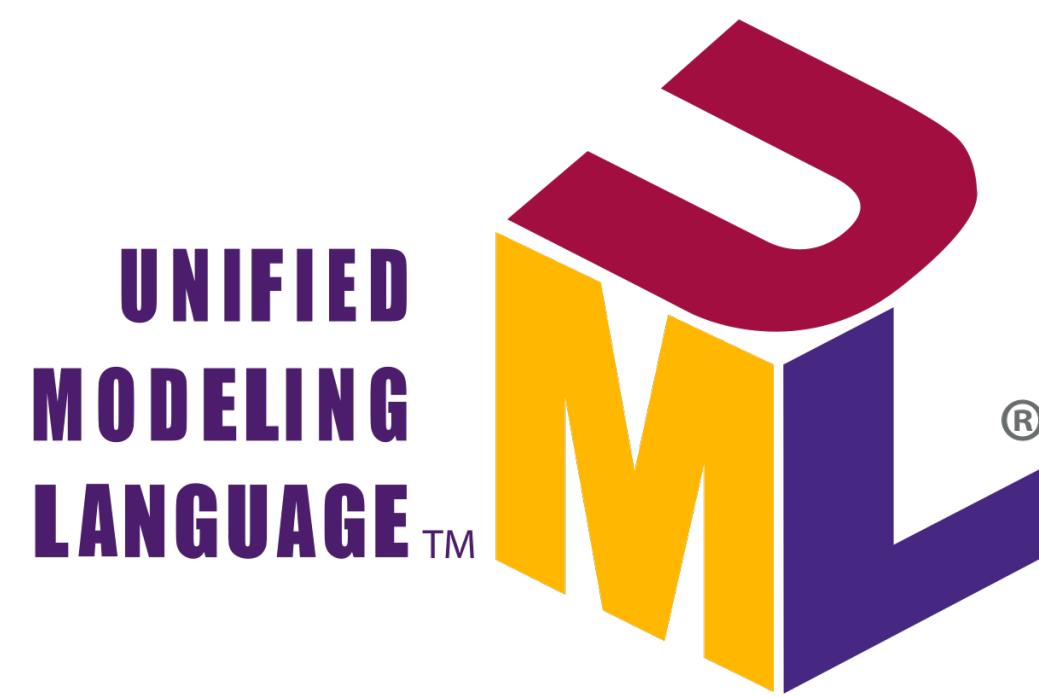
# Архитектор хранилищ



Data Warehouse



Data Lake



Icons by [www.flaticon.com](http://www.flaticon.com)

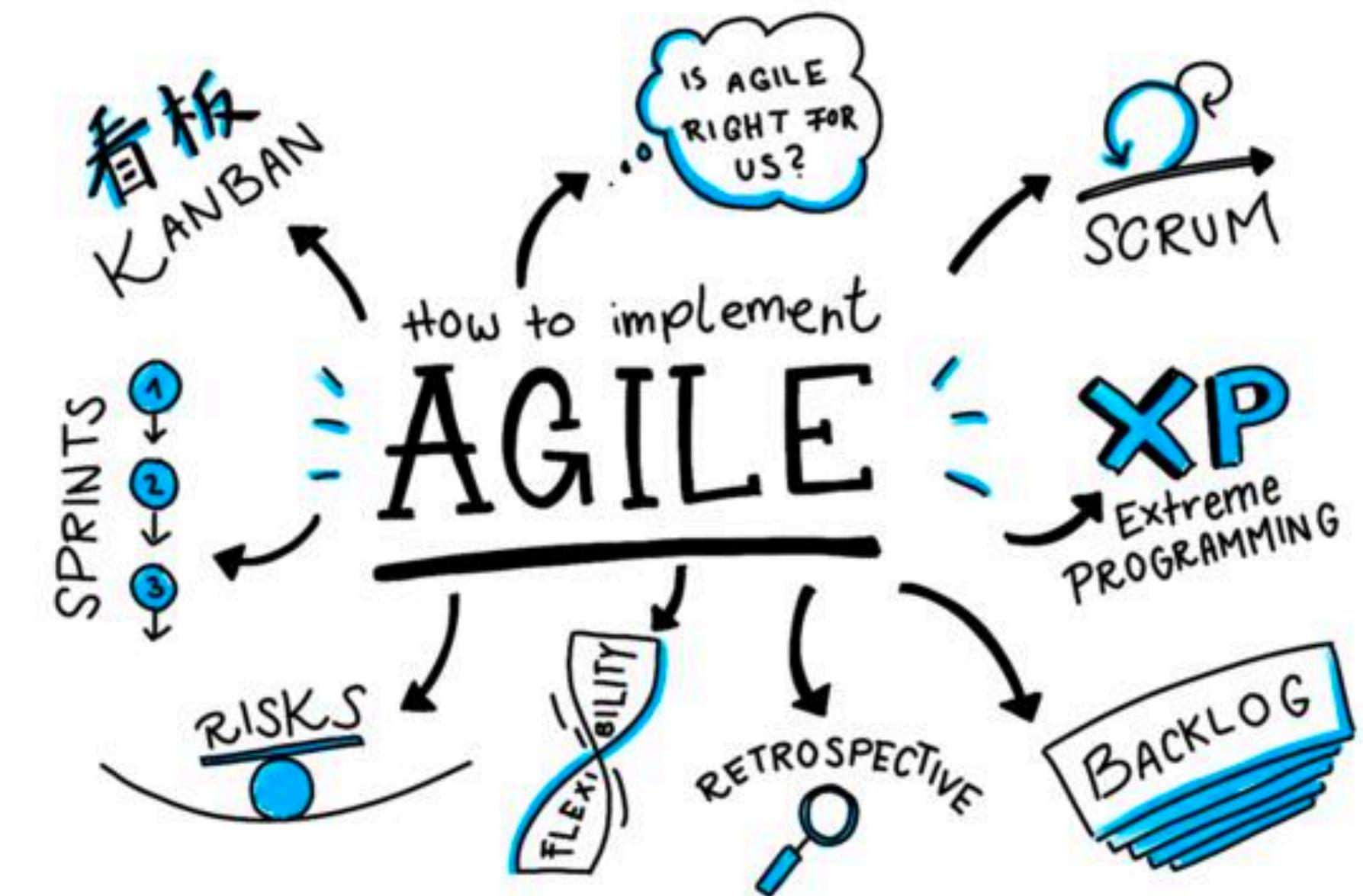
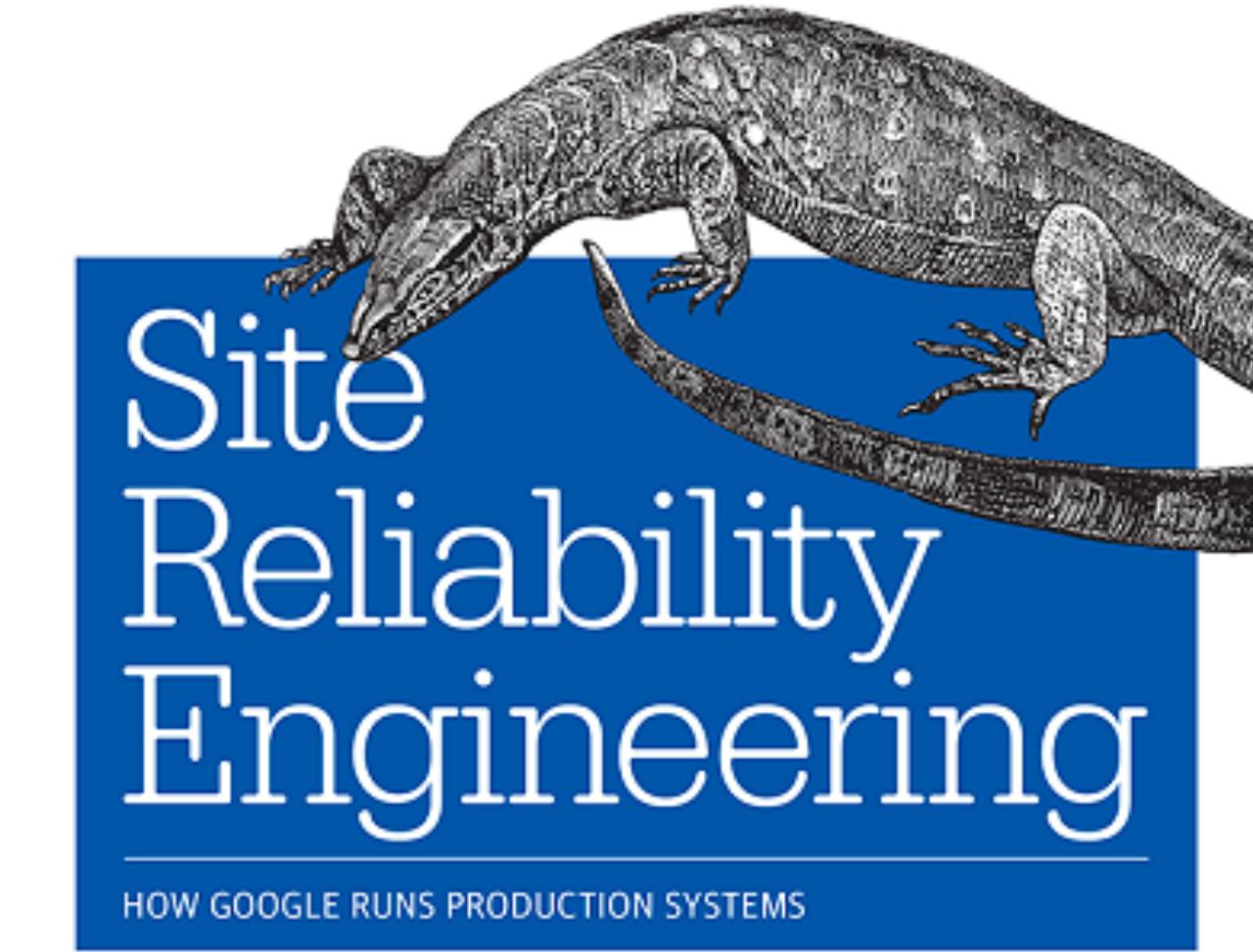
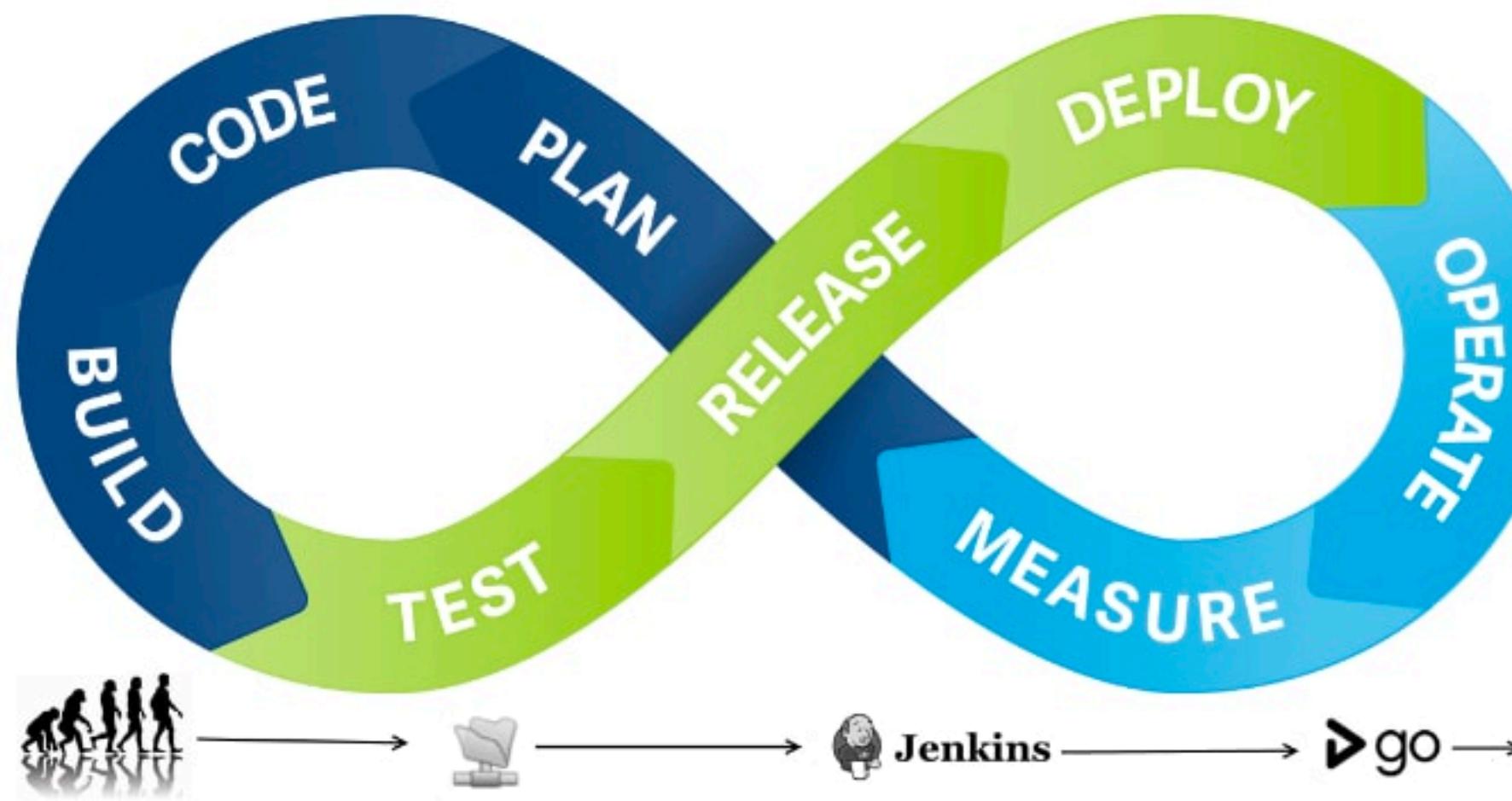


Dungeons & Dragons

# DevOps



# Jenkins



Pic source: <https://proglib.io/p/primenenie-principov-agile-na-praktike-2019-09-21>



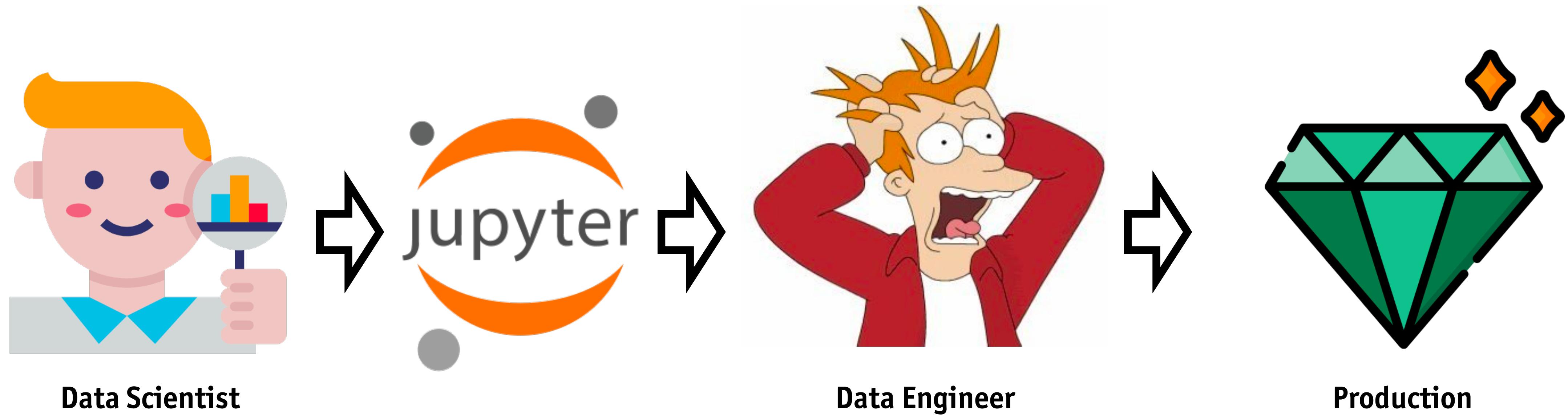
Мое определение



**Data Engineer - это человек, который  
умеет правильно использовать  
компьютеры**

**Мое определение**

# Сопровождение DS



**Каждый делает свою работу**

Обязательно посмотрите: <https://www.youtube.com/watch?v=7jiPeIFXb6U>



# Практики разработки



# Практики разработки

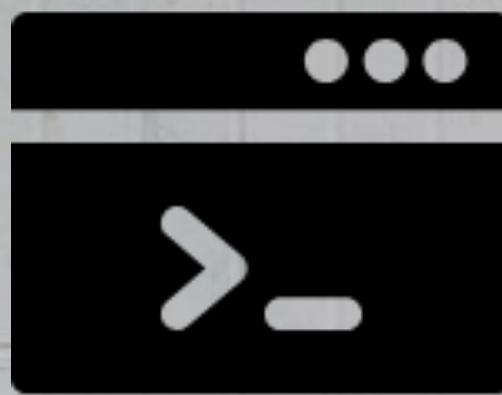


Храните деньги в  
сберегательной  
кассе код в системе  
контроля версий

# Практики разработки



Храните деньги в  
сберегательной  
кассе код в системе  
контроля версий

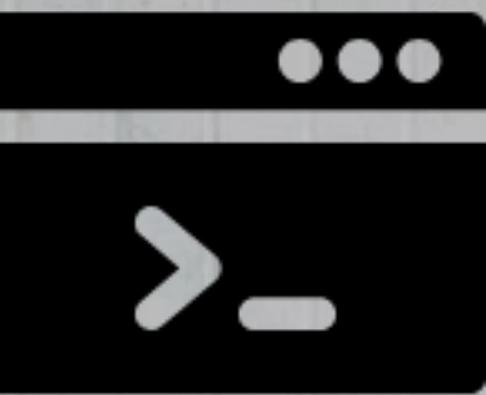


SSH - не инструмент  
развертывания

# Практики разработки



Храните деньги в сберегательной кассе код в системе контроля версий



SSH - не инструмент развертывания

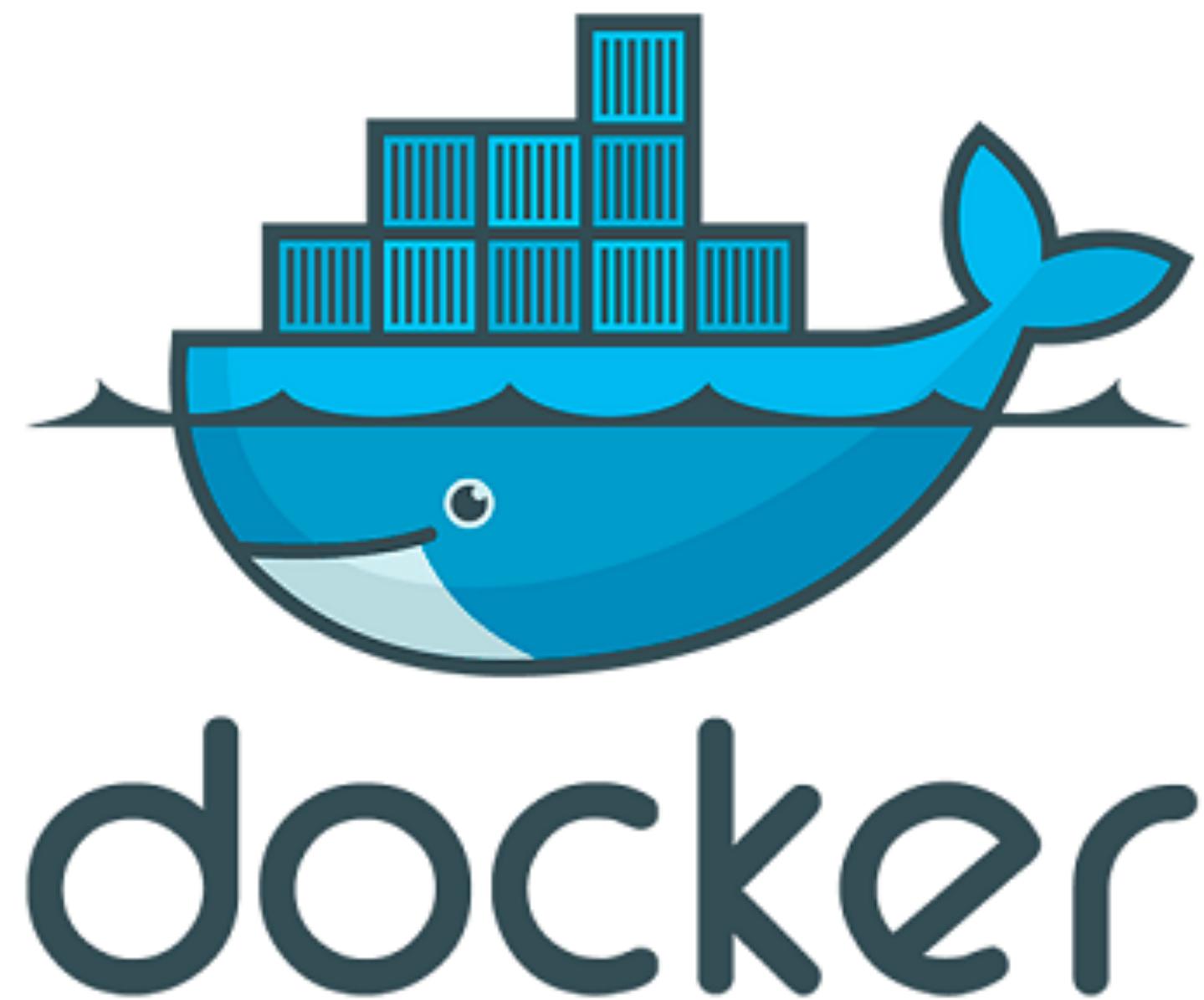


Трэш-сервера

# Практики разработки

# Практики развертывания

# Практики развертывания



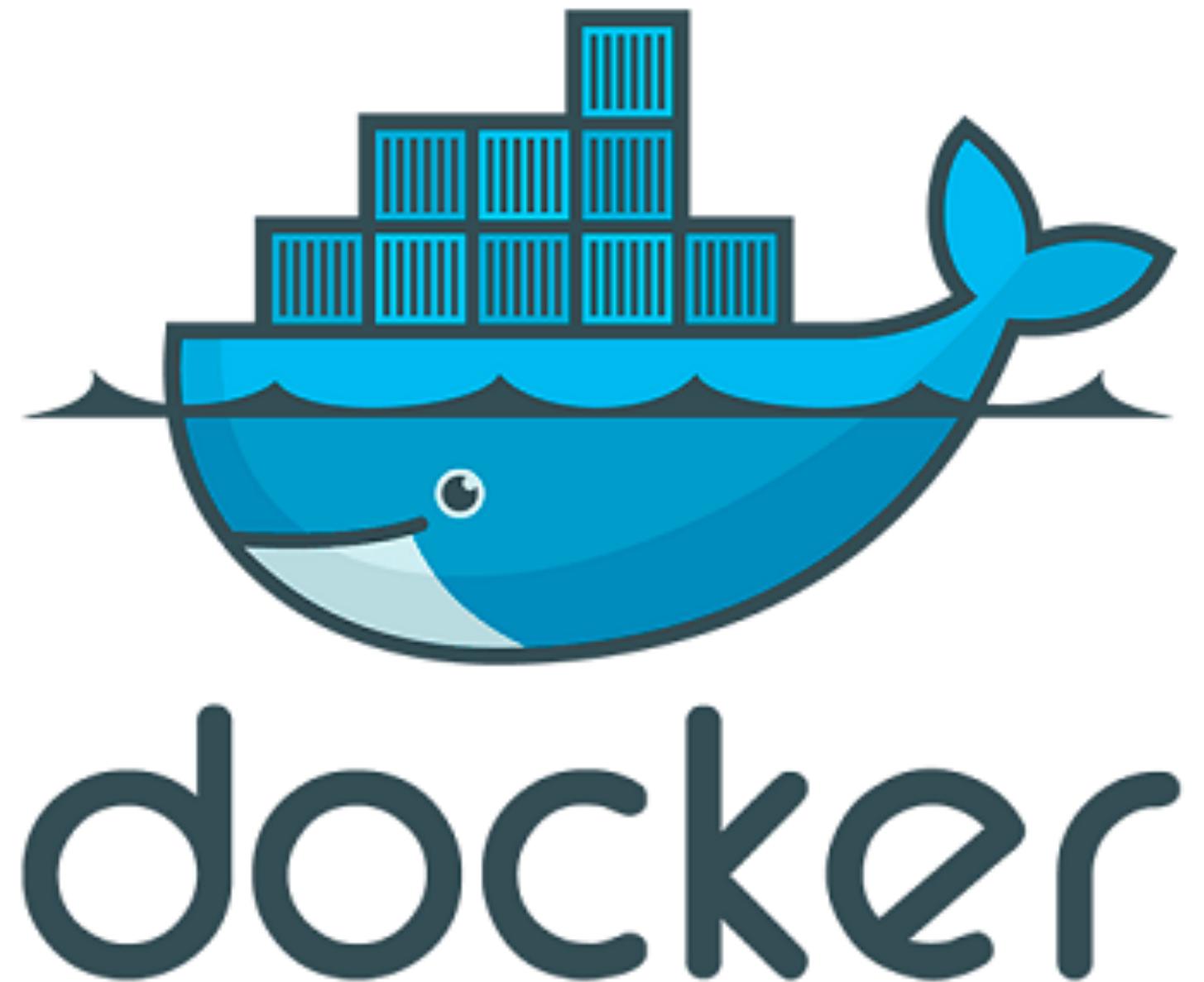
# Практики развертывания

Приватные репозитории  
для пакетов

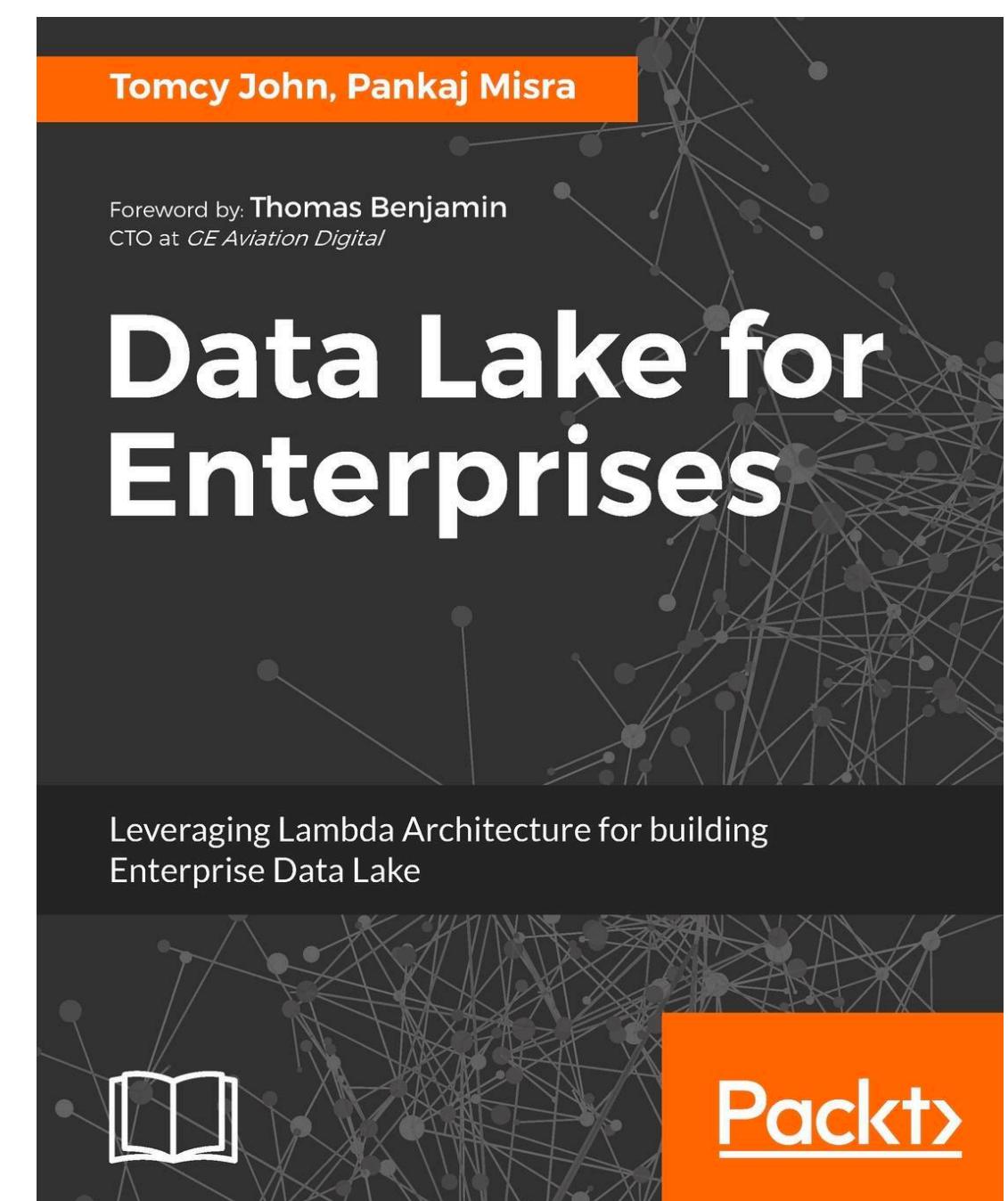
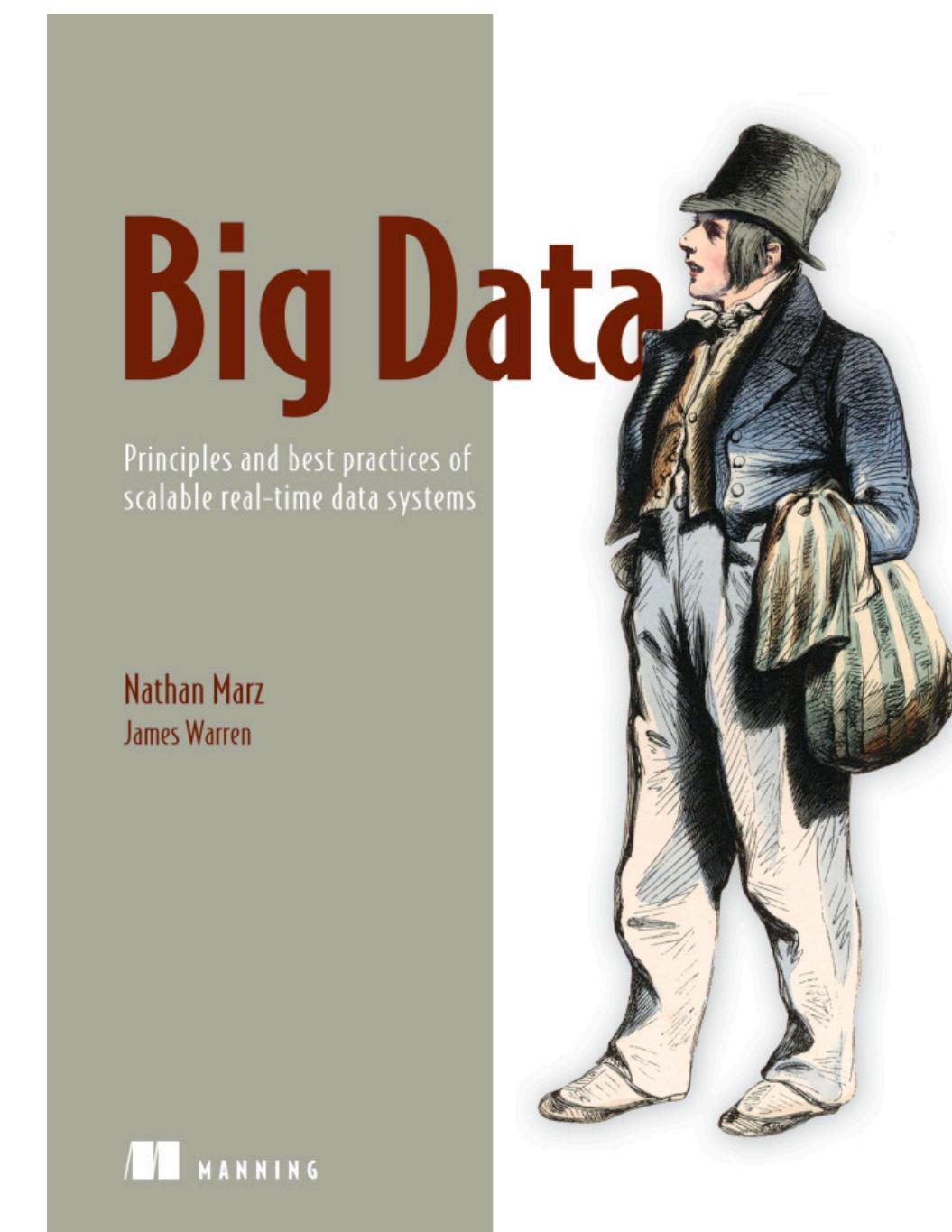
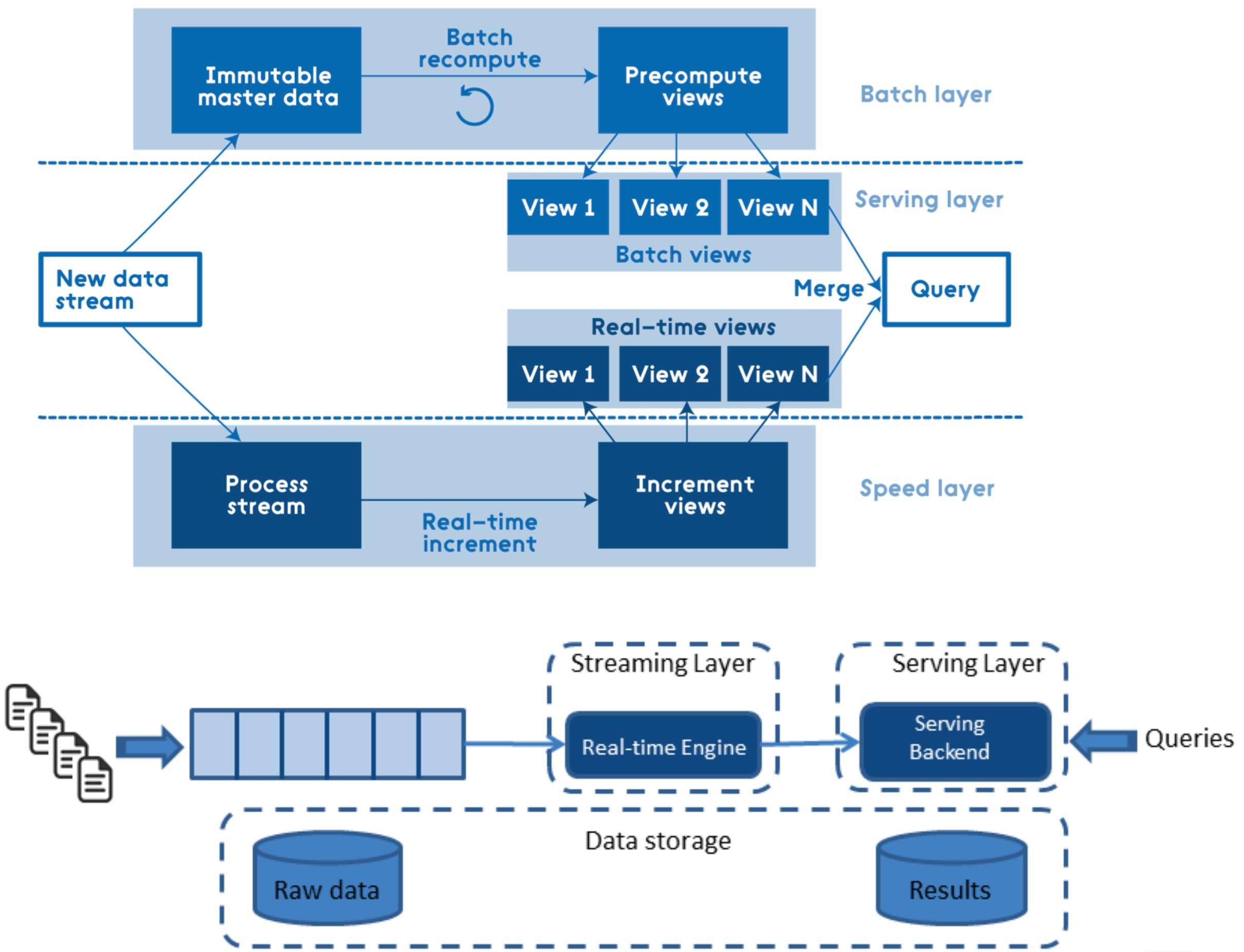
Базовая практика с  
публичными облаками

Документация по  
форматам данных

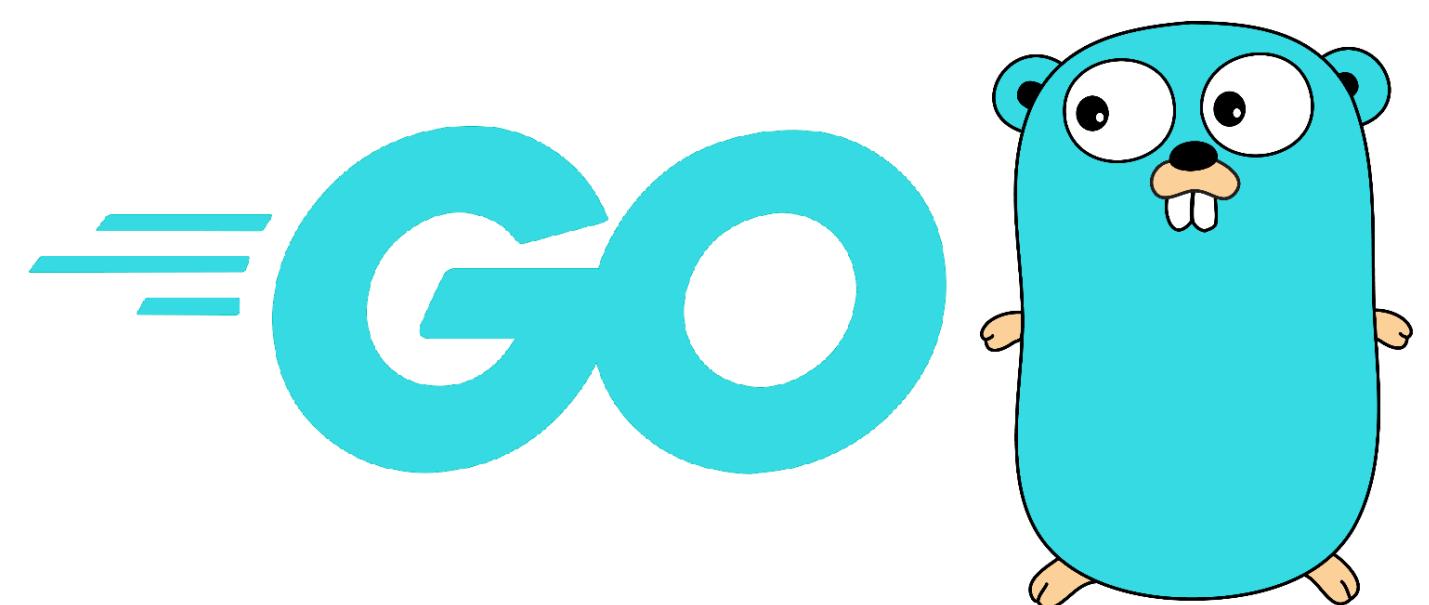
Понимание принципов  
построения API (REST, RPC)



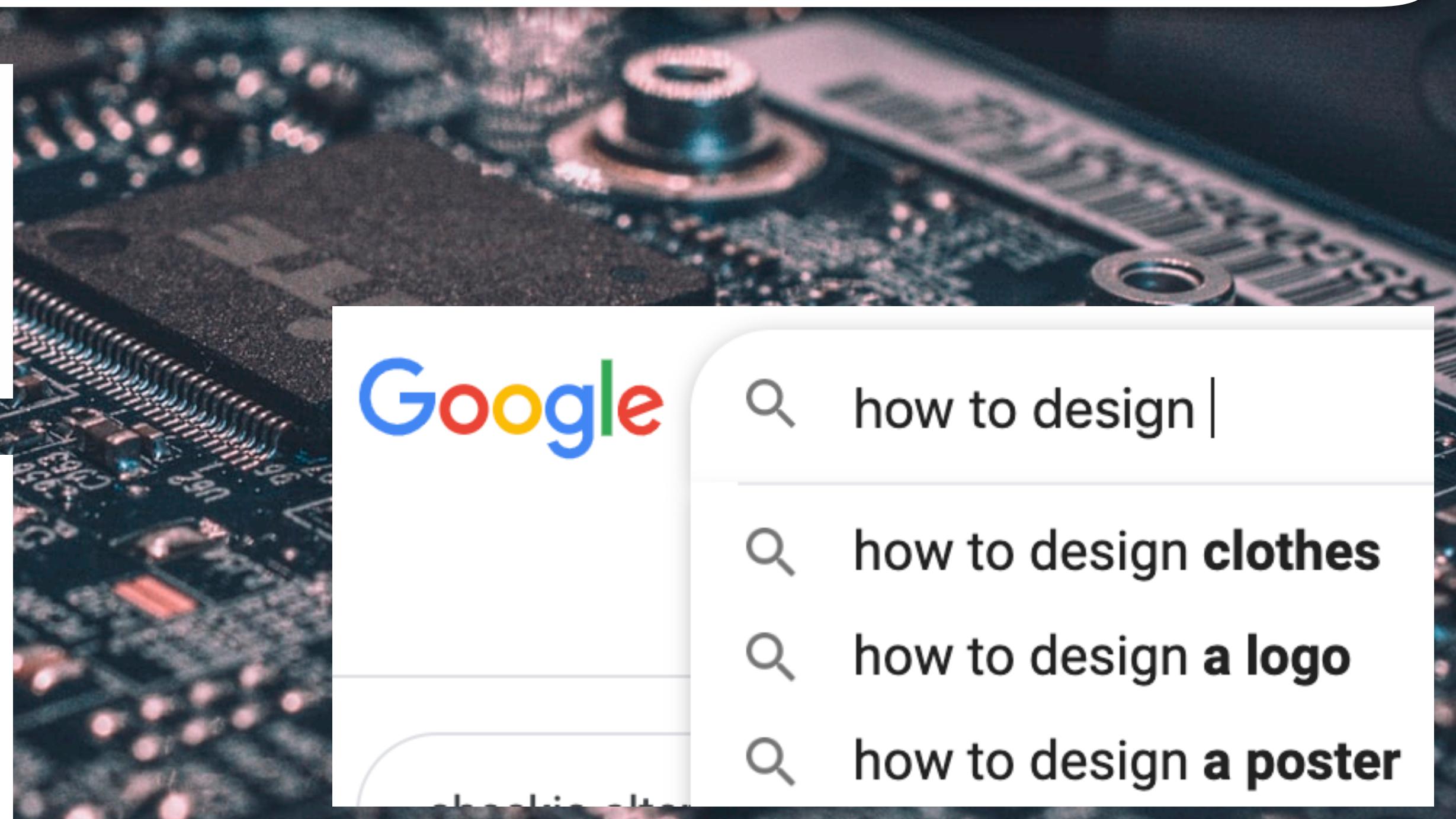
# Lambda и Kappa



# Языки программирования



# Задачки + System design



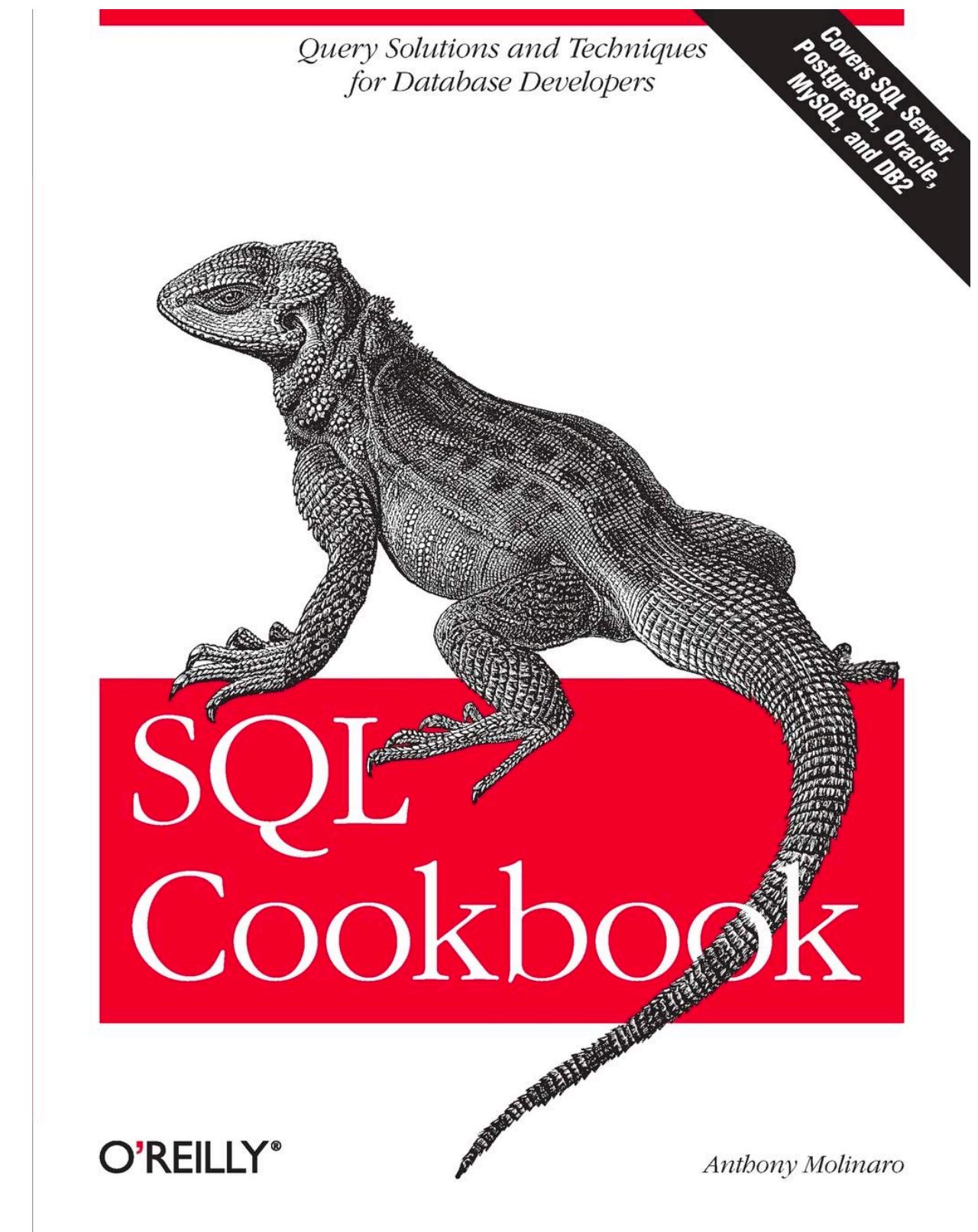
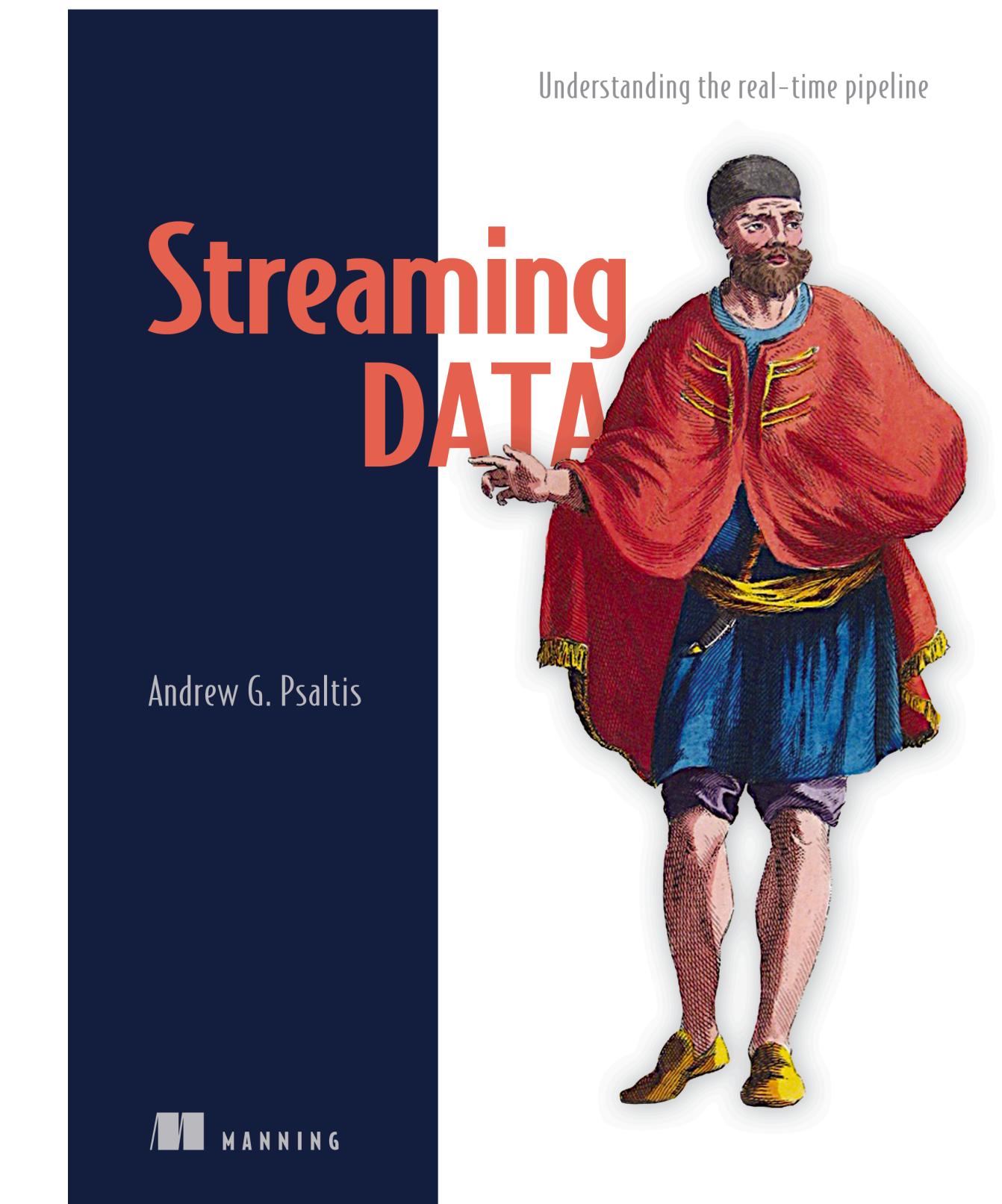
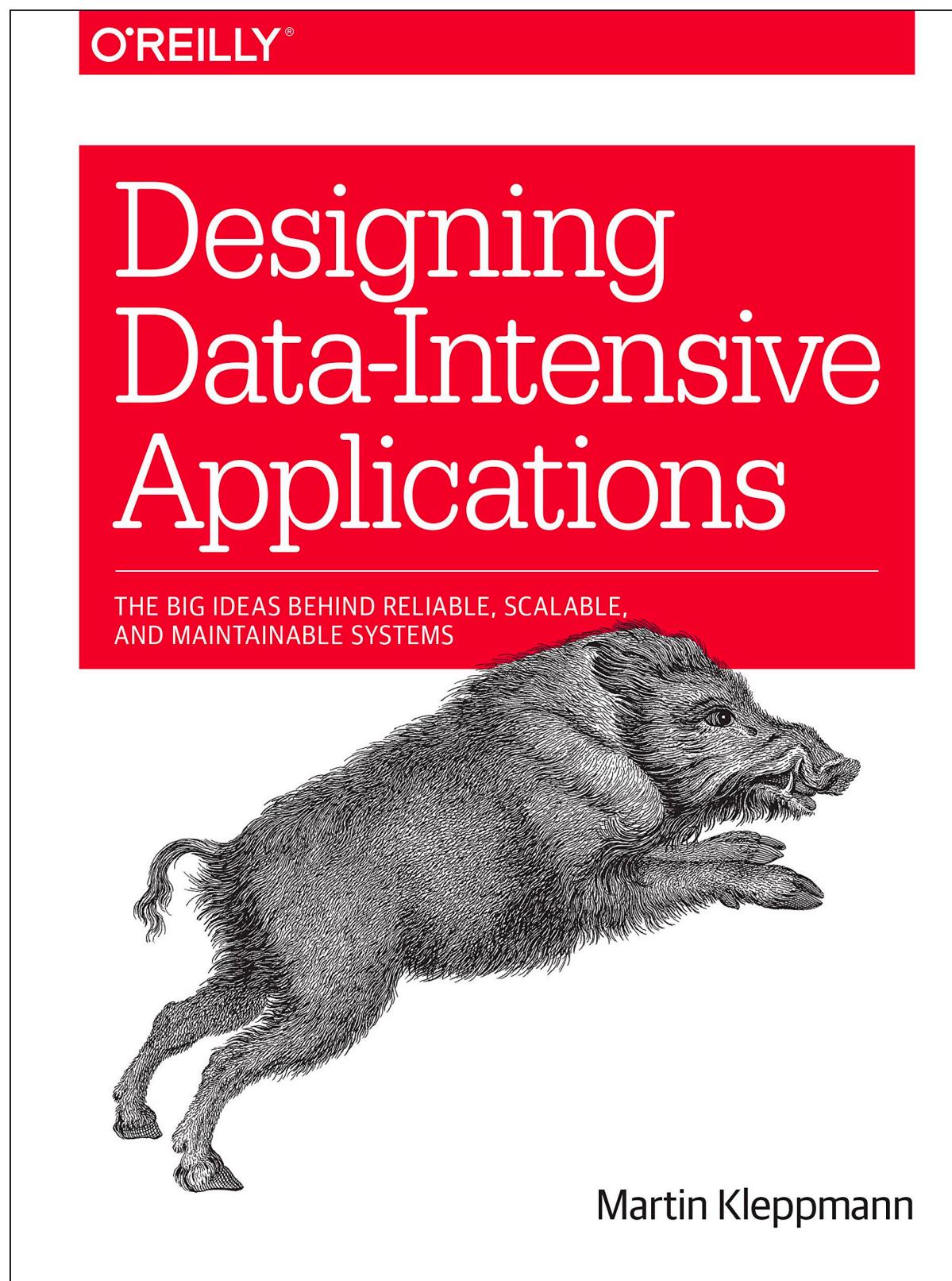
# Облачная сертификация

<https://www.udemy.com/course/aws-certified-solutions-architect-associate>

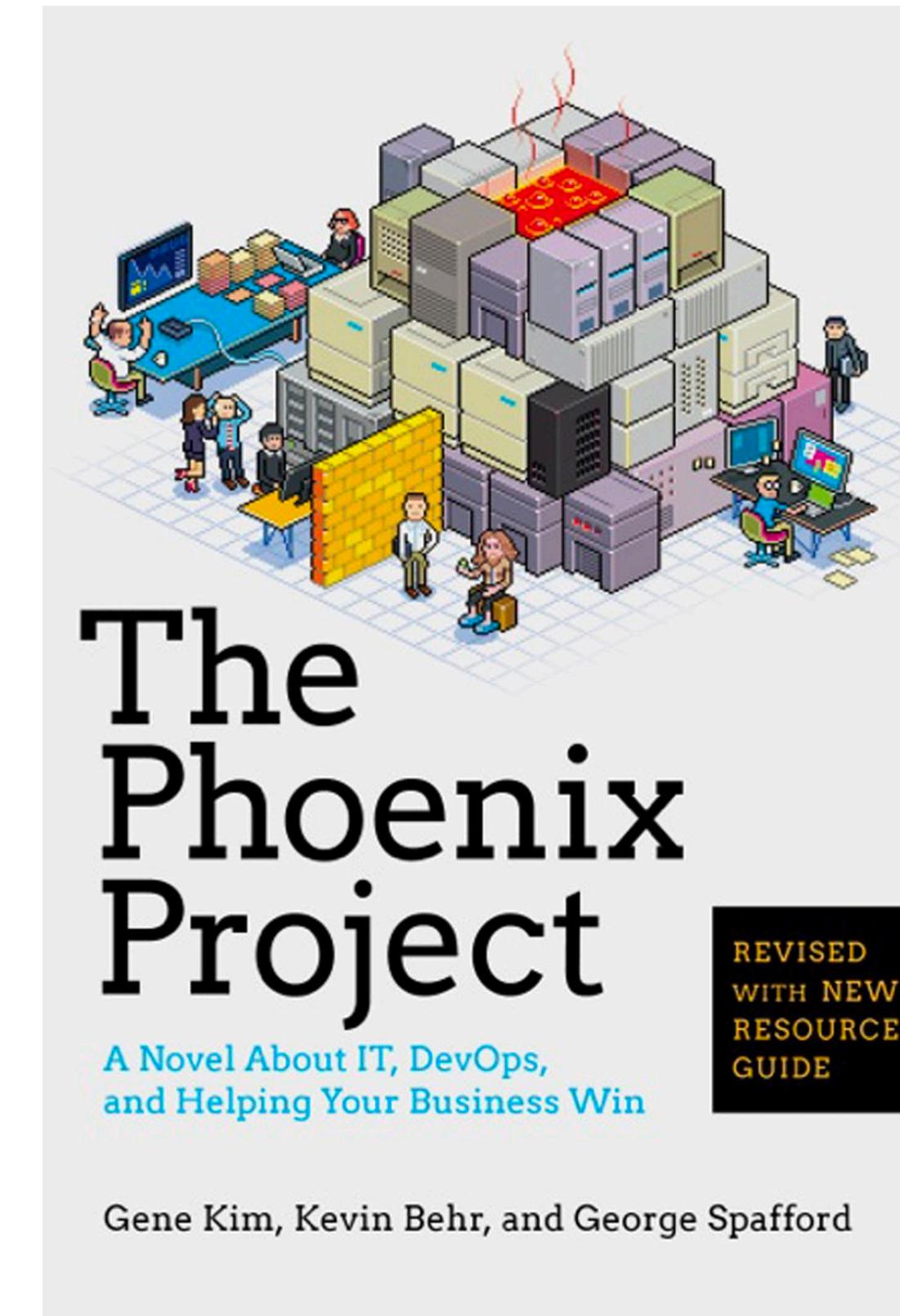
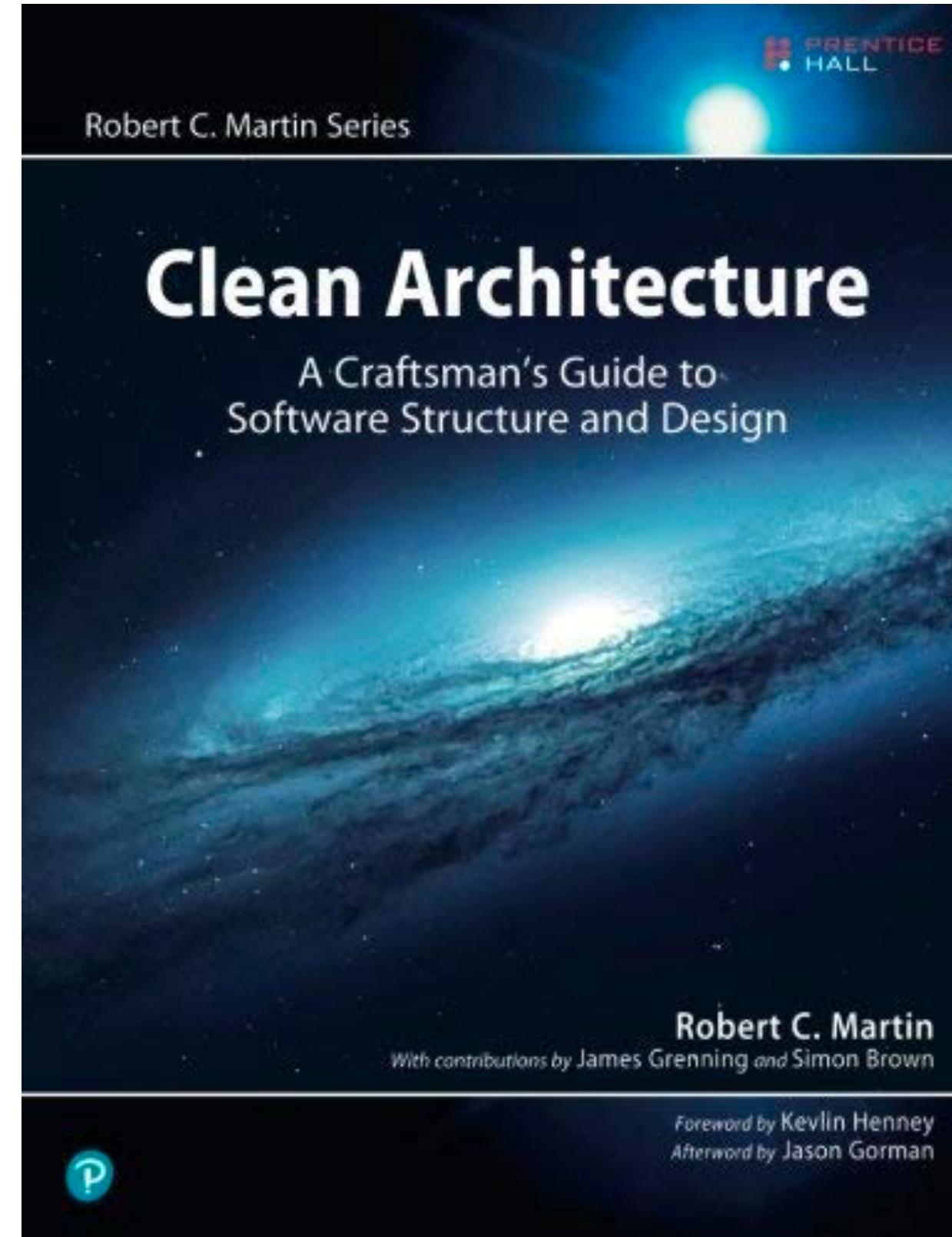
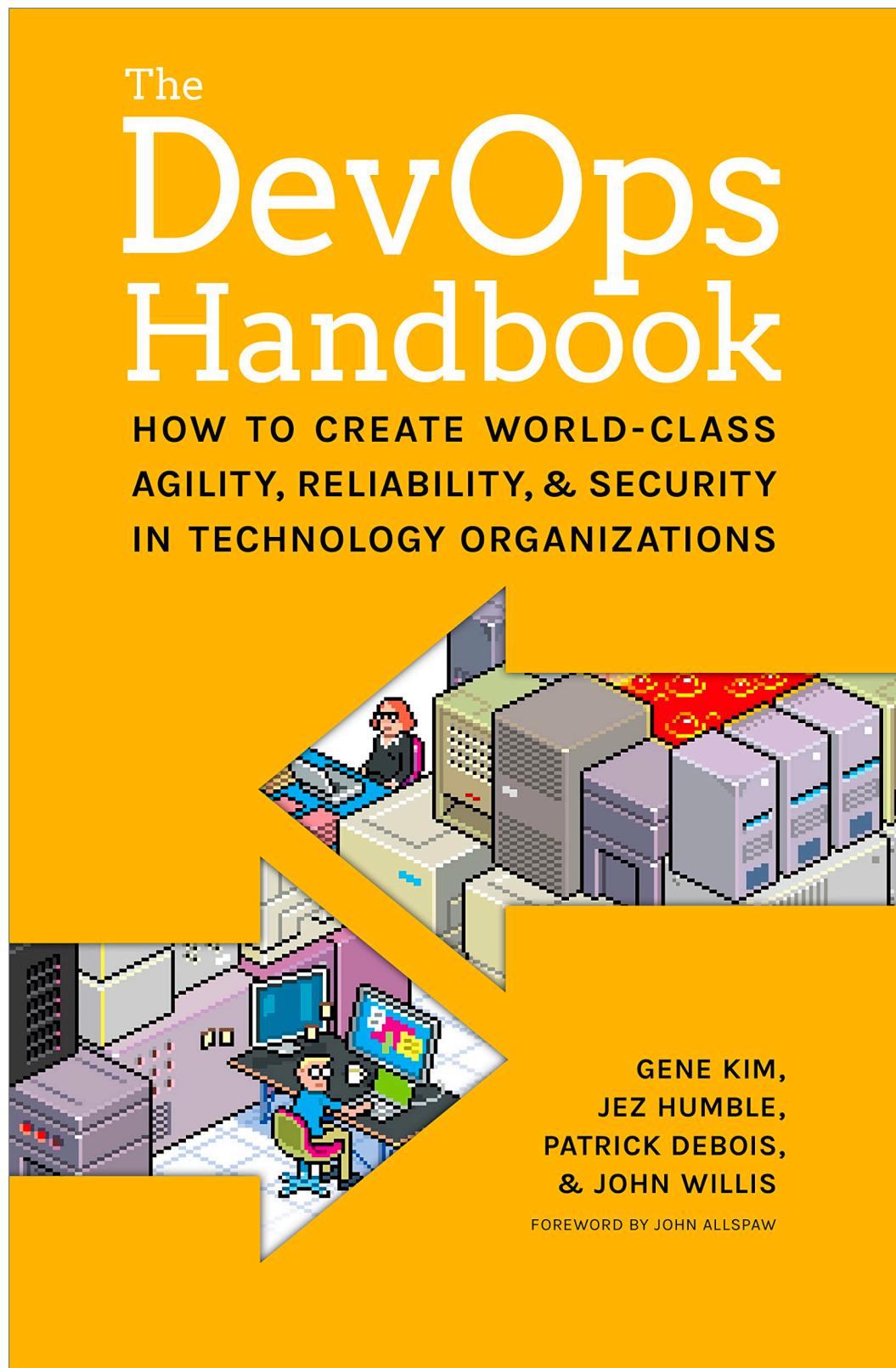
<https://www.udemy.com/course/learning-terraform-on-microsoft-azure/>



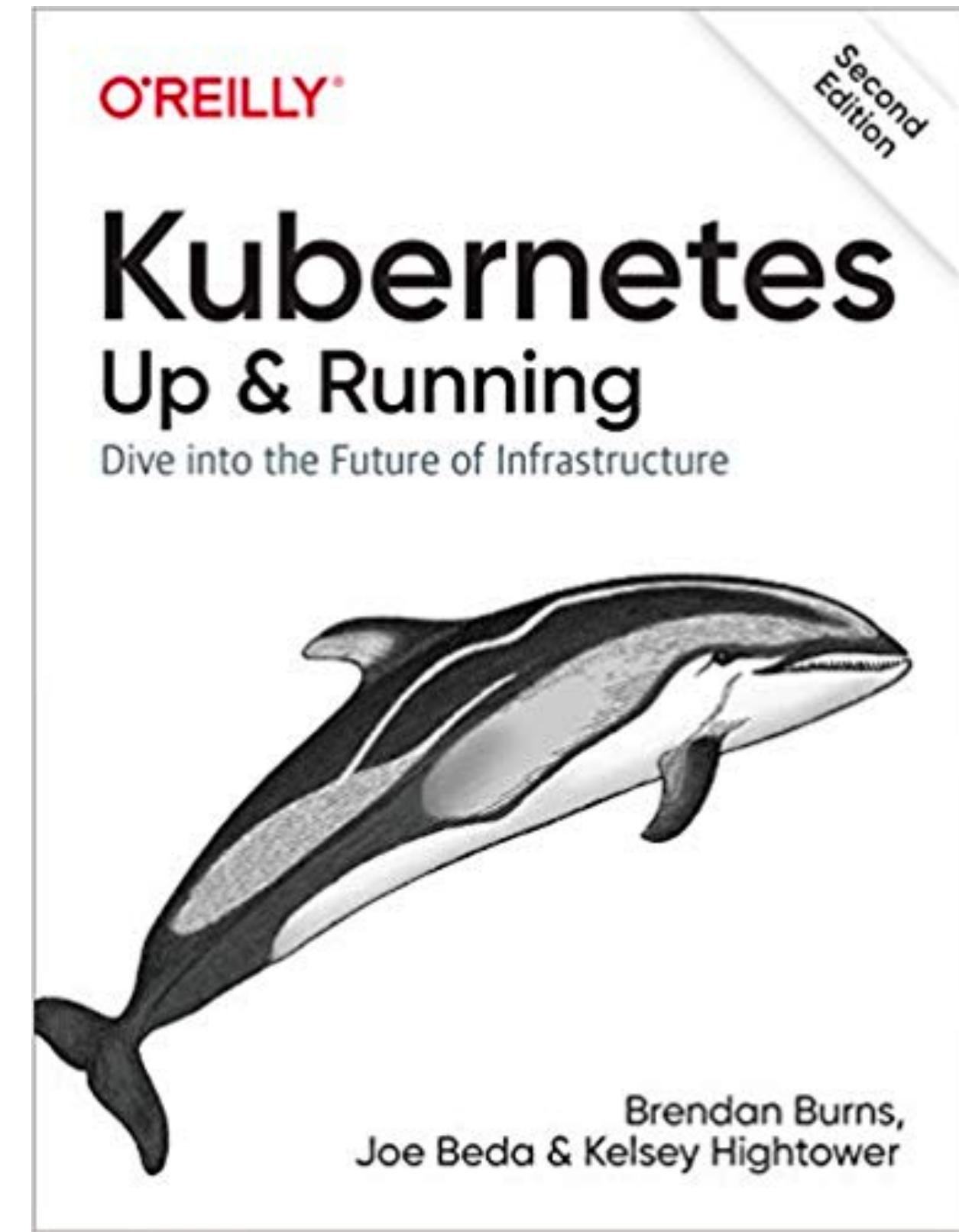
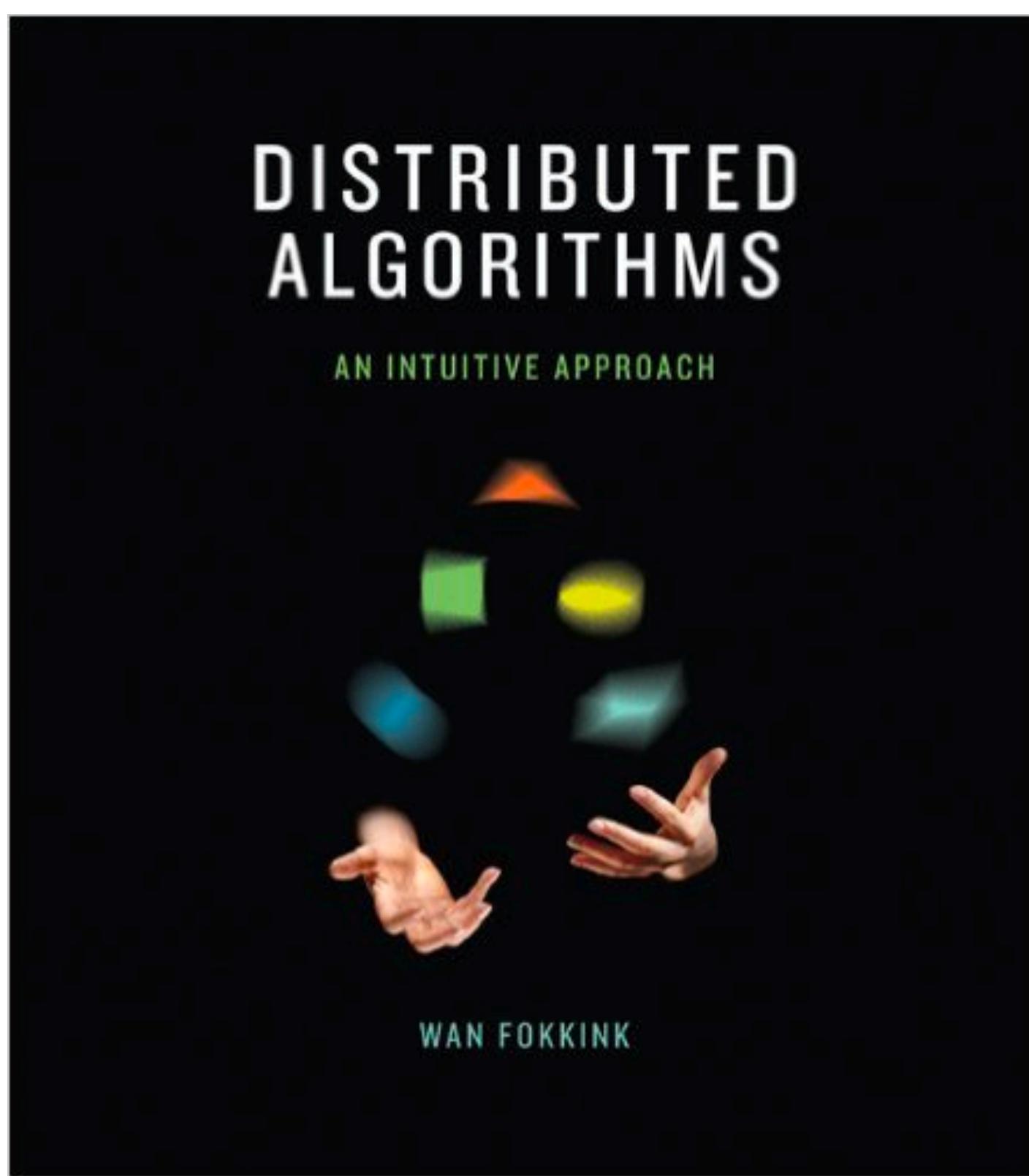
# Чо почитать



# Чо почитать



# Чо почитать



# Больше жести

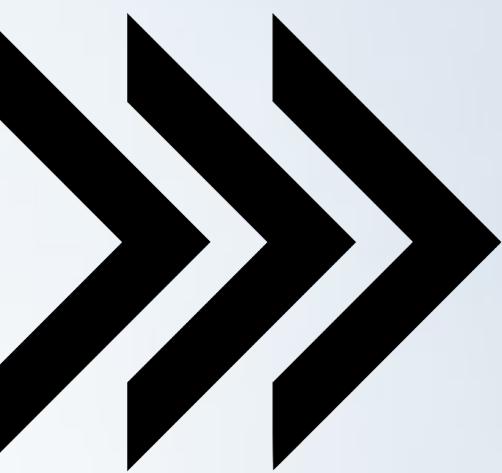


**samza**

Raft, Queue theory, low level

APACHE

**ARROW**



Performance/queue theory: <https://www.youtube.com/watch?v=bEYY3M0d-w8>

Distributed systems (MIT): [https://www.youtube.com/channel/UC\\_7WrbZTC0Du1o\\_kfUMq88g](https://www.youtube.com/channel/UC_7WrbZTC0Du1o_kfUMq88g)

Modern Hardware: <https://www.youtube.com/watch?v=0FgxAFdxYAQ>

Containers from scratch: <https://www.youtube.com/watch?v=8fi7uSYl0dc>

Julia Evans: <https://twitter.com/b0rk>

Cindy Sridharan: <https://medium.com/@copyconstruct>

# Чо погамать

Factorio: <https://store.steampowered.com/app/427520/Factorio/>

TIS-100: <https://store.steampowered.com/app/370360/TIS100/>

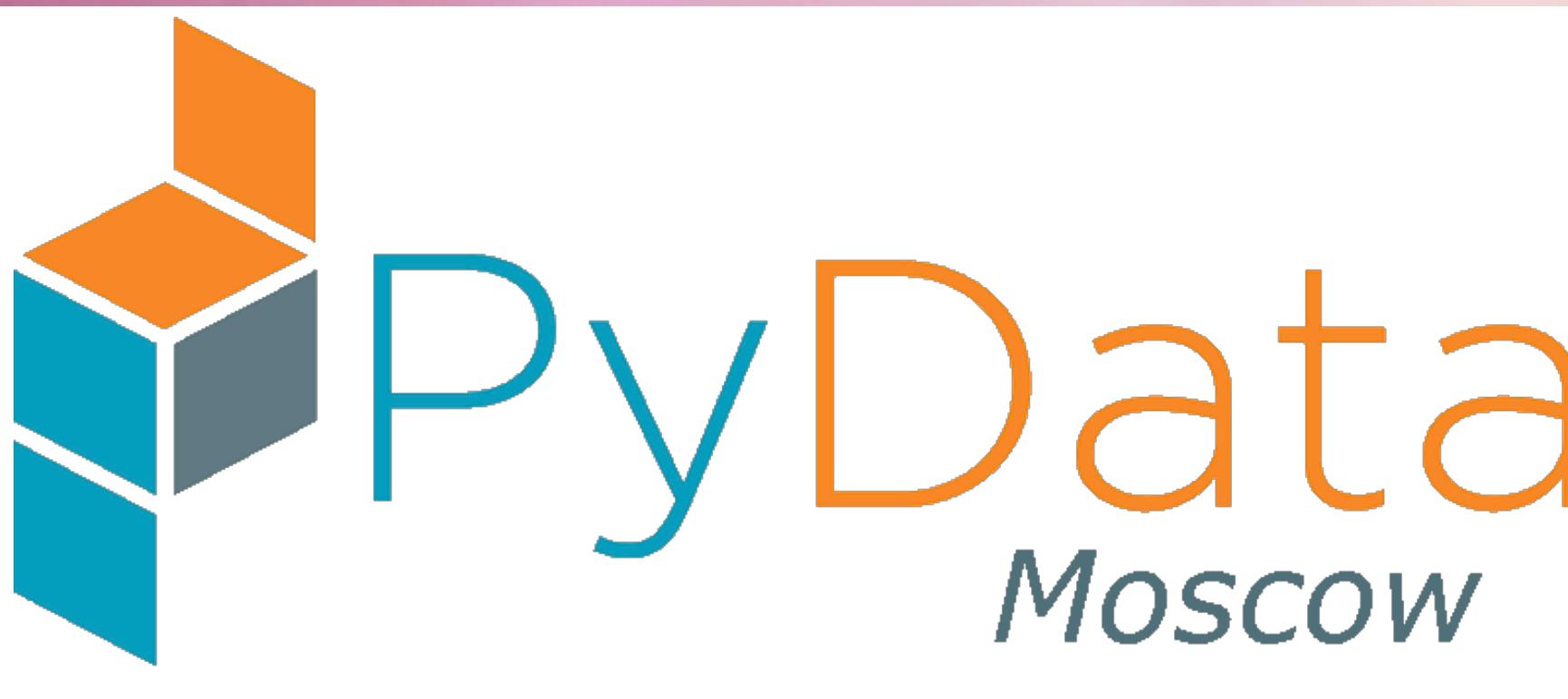
MHRD: <https://store.steampowered.com/app/576030/MHRD/>

Human Resource Machine:

[https://store.steampowered.com/app/375820/Human\\_Resource\\_Machine/](https://store.steampowered.com/app/375820/Human_Resource_Machine/)

Baba is You: [https://store.steampowered.com/app/736260/Baba\\_Is\\_You/](https://store.steampowered.com/app/736260/Baba_Is_You/)

# Где потусить



# Вопрос на засыпку

Какой навык является  
первым и одним из  
самых главных для Data  
Engineer'а?

# Вопросы?

