



University of
South Australia

Which Big Data technology?

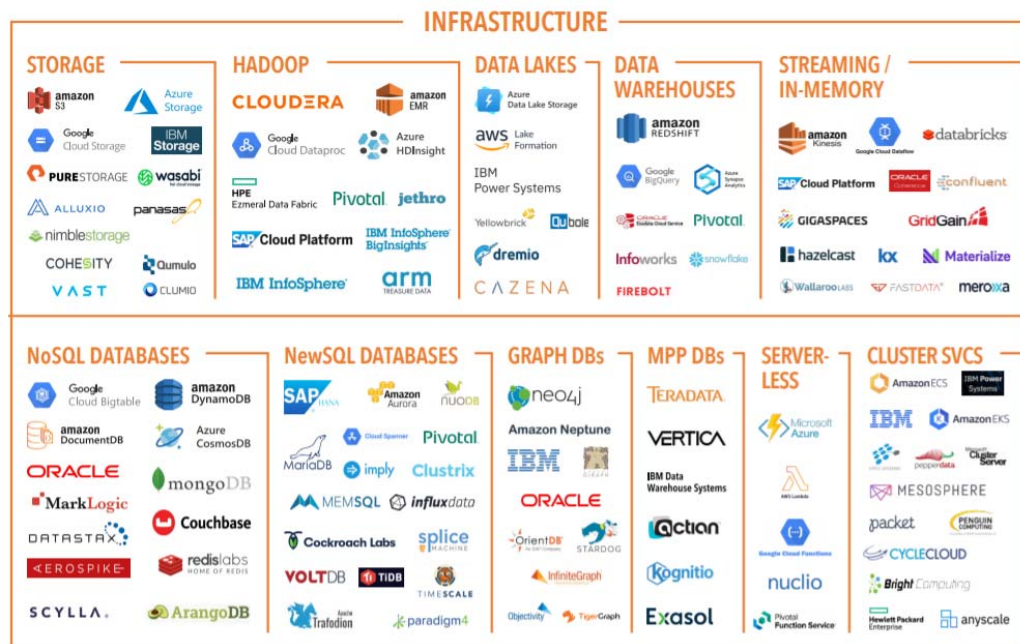
1

Keep in mind
your future role



2

Lots of
choice



<https://mattturck.com/data2020/>

3

Differing technology requirements

Capacity

Can we store it all?

Consistency

Do we need strong guarantees?

Data loss

Do we care if a write fails?

Data type

Structured or unstructured?

Speed

Read or write?
Real time or batch?

Persistence

How long do we need to keep the data?

Different sources of Big Data



Open data



'Private'
(company) data



Streaming data

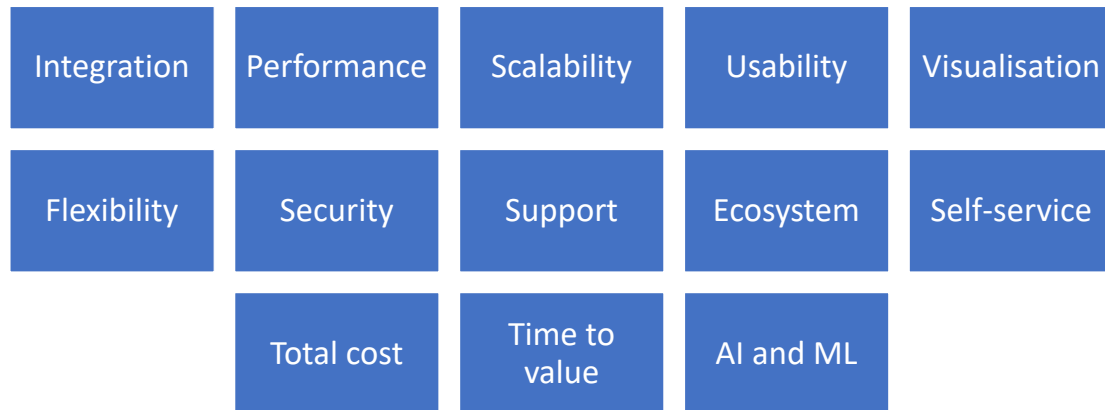
Time series e.g. financial
Internet of Things
Security
Broadcast networks
Web activity

Key decisions to narrow down options

- On-premise vs cloud-based
- Proprietary vs open source
- Batch vs streaming



Characteristics to look for



Keys to success - Gartner

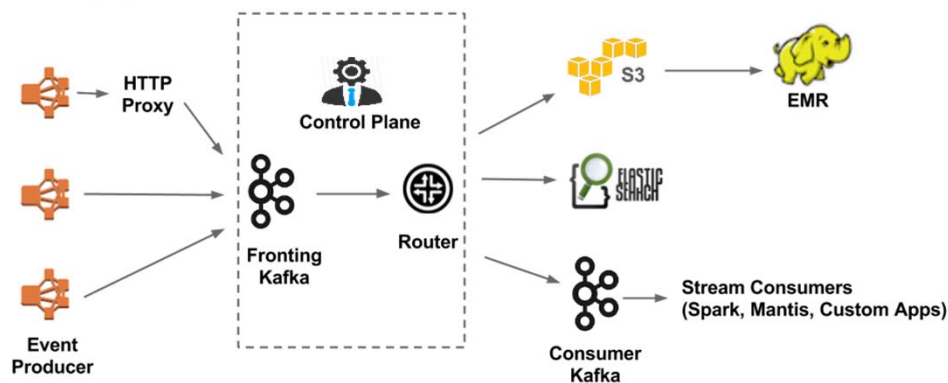
- Understand goals
- Start small
- Take holistic approach
- Work together
- Go viral



Architecture of 'giants'

500 billion events per day = 1.3 PB per day
8 million events per second at peak hours
Over 100 data engineers and analysts

NETFLIX

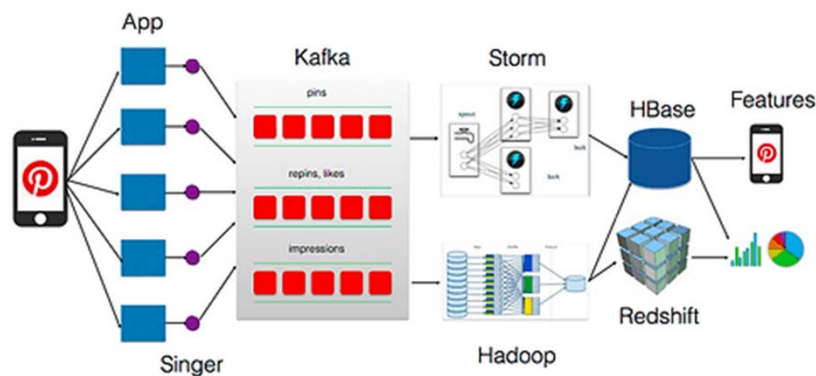


9

Architecture of 'giants' - Pinterest



10B+ pageviews per month
Data team of 250+ engineers



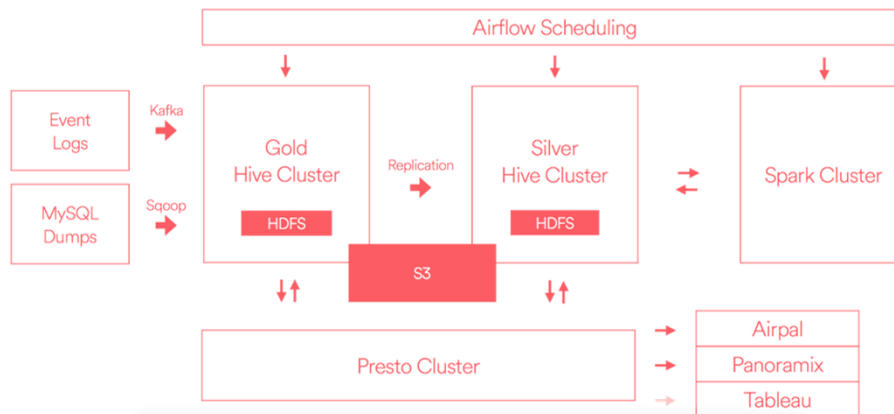
10

Architecture of 'giants' - Airbnb



AIRBNB DATA INFRA

Data team of 30+ engineers



11

WARNING

This material has been reproduced and communicated to you by or on behalf of the **University of South Australia** in accordance with section 113P of the *Copyright Act 1968 (Act)*.

The material in this communication may be subject to copyright under the Act. Any further reproduction or communication of this material by you may be the subject of copyright protection under the Act.

Do not remove this notice

12