Choosing Technologies Across the Data Engineering Lifecycle

Architecture versus tools

A lot of people confuse architecture and tools. Architecture is strategic; tools are tactical. Architecture is the what, why, and when. Tools are used to make the architecture a reality; tools are the how.

Architecture first, technology second.

Considerations for choosing data technologies across the data engineering lifecycle

Team size and capabilities: use as many managed and SaaS tools as possible, and dedicate your limited bandwidth to solving the complex problems that directly add value to the business. Take an inventory of your team's skills.

Speed to market: work in a tight feedback loop of launching, learning, iterating, and making improvements. Deliver value early and often.

Interoperability: you'll need to ensure that it interacts and operates with other technologies. Interoperability describes how various technologies or systems connect, exchange information, and interact.

Cost optimization and business value: consider three main lenses: total cost of ownership, opportunity cost, and FinOps.

Today versus the future: immutable versus transitory technologies

Location (cloud, on prem, hybrid cloud, multicloud)

Build versus buy

Monolith versus modular

Serverless versus servers

Optimization, performance, and the benchmark wars

The undercurrents of the data engineering lifecycle