



# Data Architectures

Batch & Real time processing





# Overview

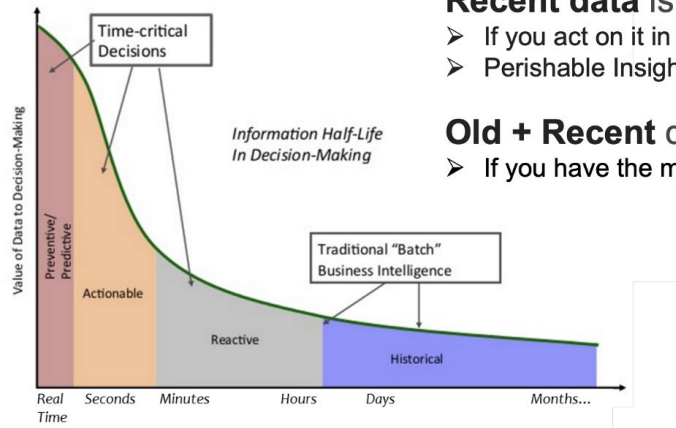
- Batch processing
- Real time processing
- Near-real time processing





# Value of Data

## The Diminishing Value of Data



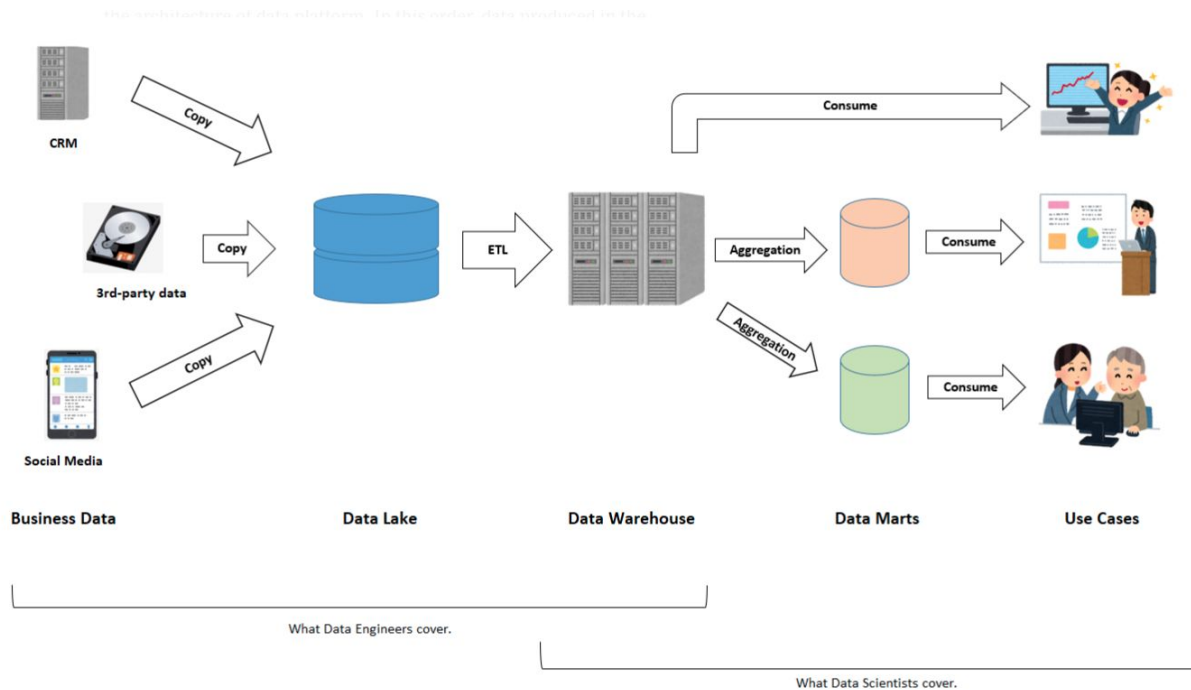
### Recent data is highly valuable

- If you act on it in time
- Perishable Insights (M. Gualtieri, Forrester)

### Old + Recent data is more valuable

- If you have the means to combine them

# Generic Data Architecture

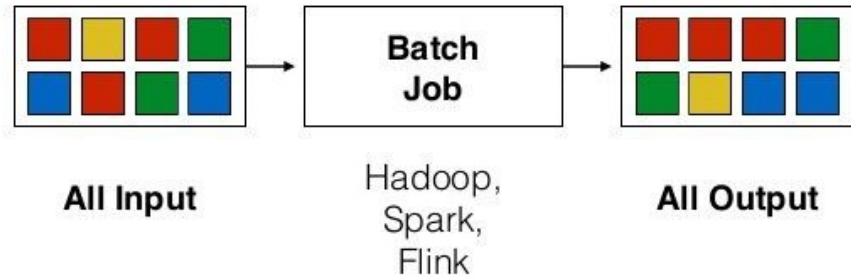




# Batch Processing

- Minutes to days time delay
- Process massive data
- Takes minutes to hours

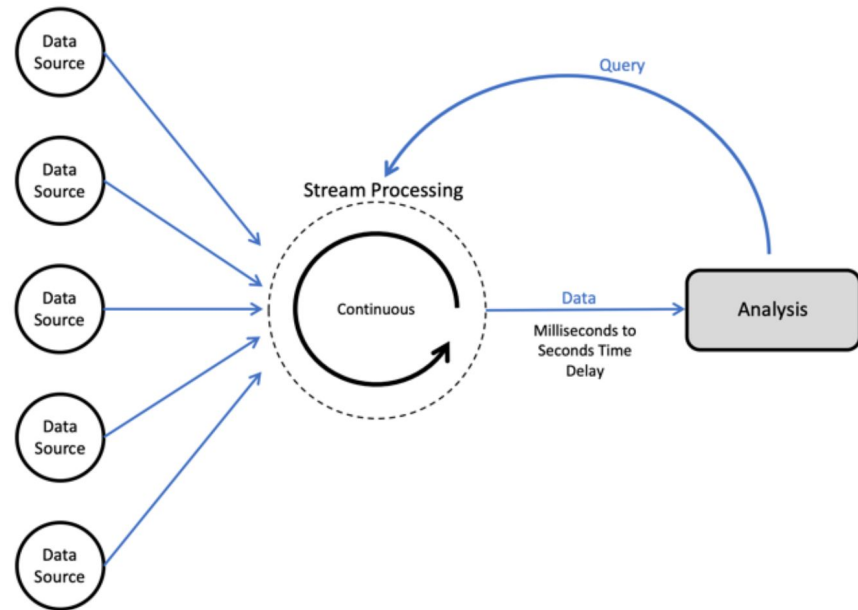
## Batch Processing





# Real time Processing

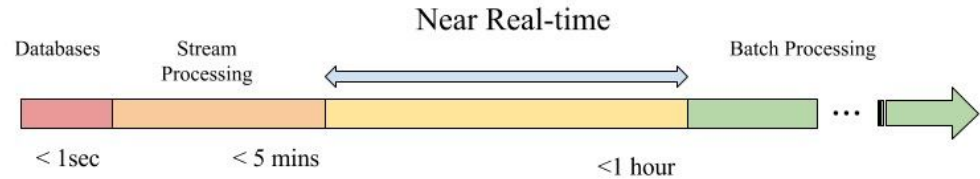
- Process within seconds
- Process small amount of data
- Takes seconds





# Near-Real time Processing

- Process within minutes
- Process small amount of data
- Takes seconds





# Popular Data Processing Architectures





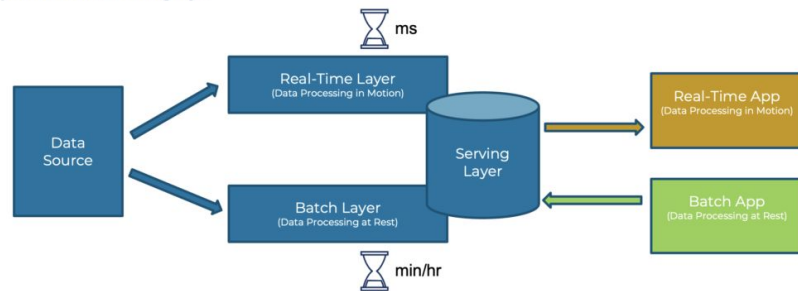


# Lambda Architecture

- Allow to implement both real time and batch
- Has multiple layers
- Real time layer provides real time processing
- Batch layer provides batch processing

## Lambda Architecture

Option 1: Unified serving layer



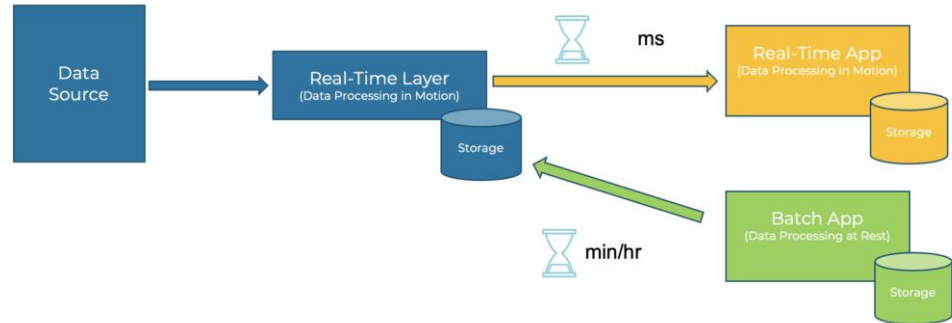


# Kappa Architecture

- Allow to implement real time processing
- Has multiple layers
- Real time layer provides real time processing
- If there is a need for batch processing, real time layer is used for batch processing with using storage database

## Kappa Architecture

One pipeline for real-time and batch consumers





# Next : Serverless

