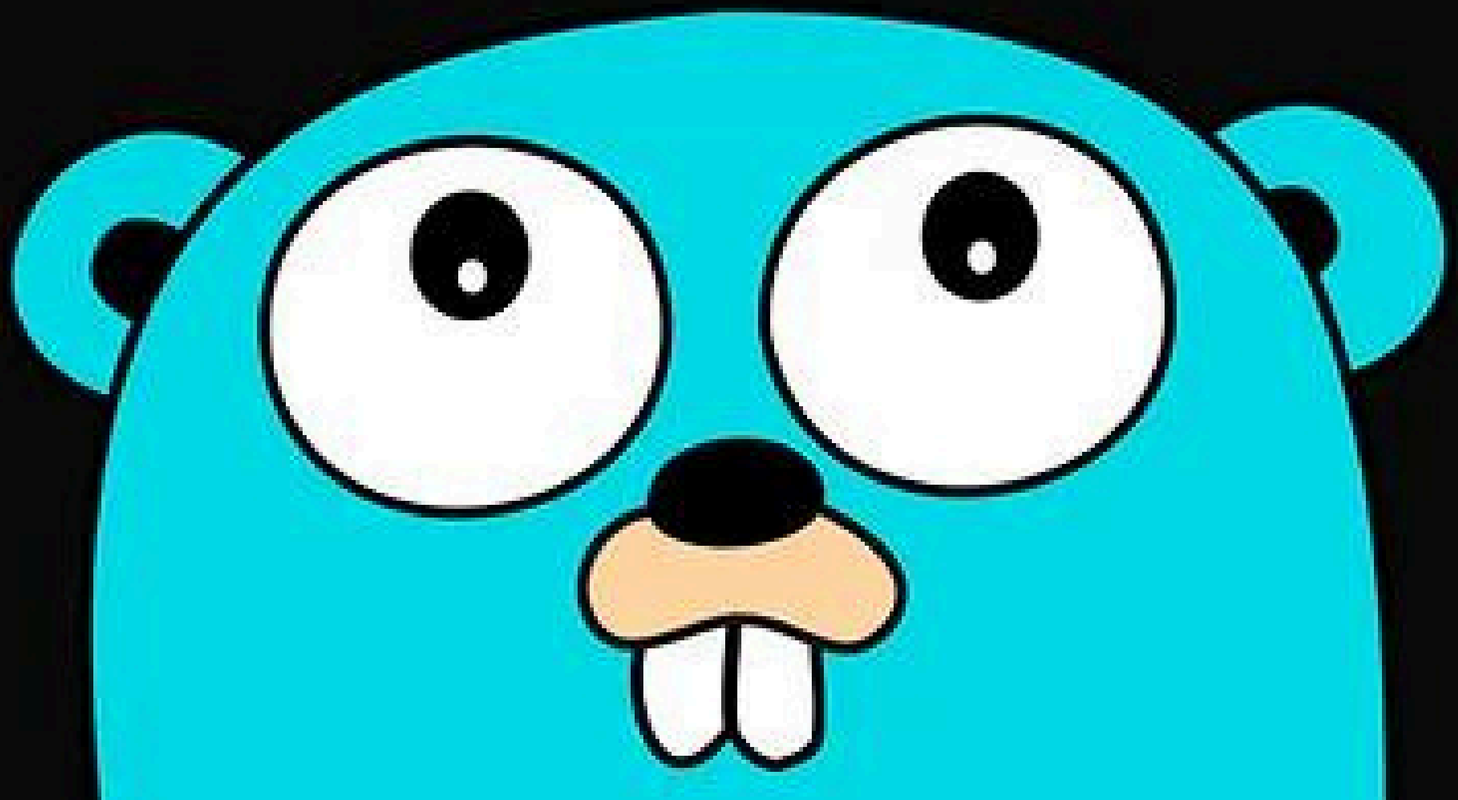
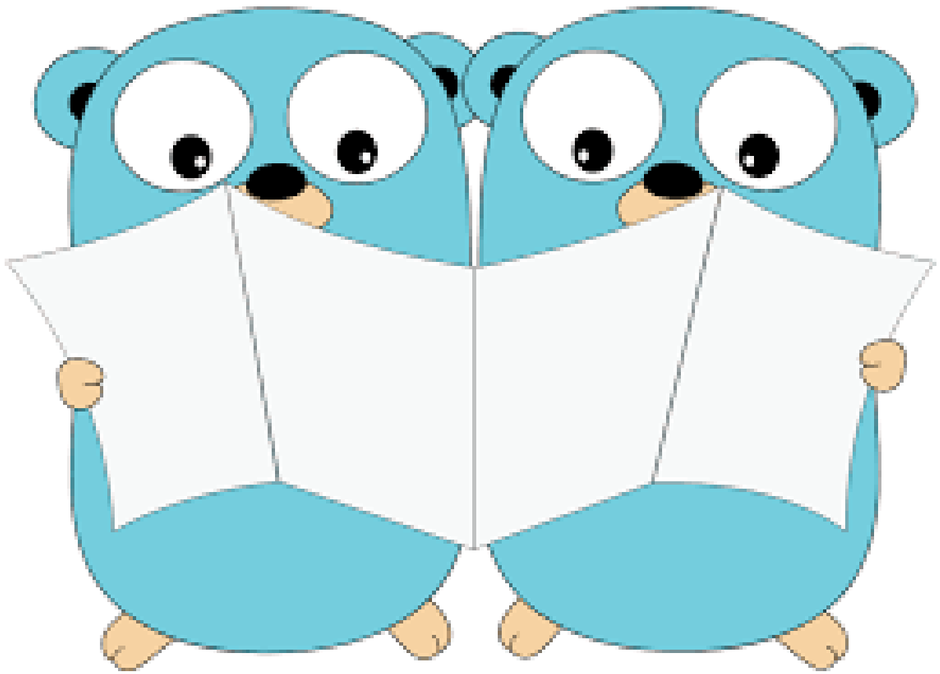
Real-Time Data Processing Pipeline



Contents



What is Stream Processing

Stream processing is processing data continiously on arrival. not waiting for data to process in batch. Processes large volumes in near realtime.

used in industrie such as

1 Stock/ finance / crypto
2 IOT systems with thousands of sensors
3 fraud detection
4 Telemetry at scale

Why Golang for stream processing

Golang has

1 Concurrency, channels and mutex that can run millions of processing functions at the same time as compared that is hundreds of times better compared based languages like java, python.

2 It is highly scalable

3 Very low memory footrint

4 Easy to understand and develop

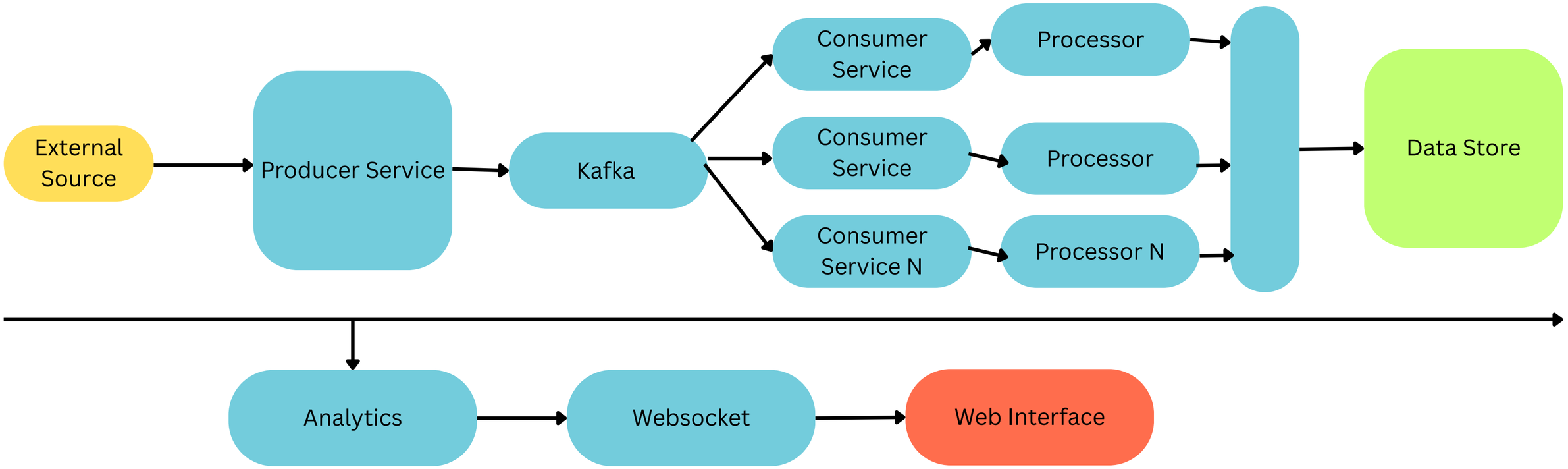
5 very high performance

Tools

1. Golang = for service development
2. Kafka as a service or self setup = highly writable
3. docker or kubernetes depending on requirement
4. cloud infra such as aws, gcp
5. Terraform = optional for infra setup
6. datastore service such as s3 , redshift , GCP Bigquery, Azure synapse analytics

Architecture

Producer →kafka-> consumer → processor → destination



Architecture 1

Producer →kafka-> consumer → processor → destination

1 Go Producer gets data from sources sich as stocksrpc, open telemetry or iot/ mobile clients.
2 Producer sends data to kafka as it is highly writable.
3 Kafka sends this data to consumer go service it can be single or group or consumers
4 Processor stage takes place when data from consumer is processed as per business logic like filtering, cleaning, pattern detection, transformation etc.
5 data after processing is sent to desitnation

Architecture Analytics

Analytics->websocket->webinterface

For data analysis the data logs /metadata are sent to a dedicated analytics service that collects logs and sends them to a websocket service backed . It can be both in same analytics service.

This websocket is connected to a frontend to show realtime graphs and analytics , or can be consumed as per the requirements.

Github Link

Github link :-

https://github.com/subpxl/streamprocessing

Github.com/subpxl

Shubham Panchal