



Dataflow for Real-time Clickstream Analytics

What is Dataflow?
It's a serverless data processing service for both streaming & batch data.

How can I personalize my site's content by visitor if I can only analyze their behavior days after they visited?

Wouldn't it be great if we could deliver relevant content and offers in the moment?

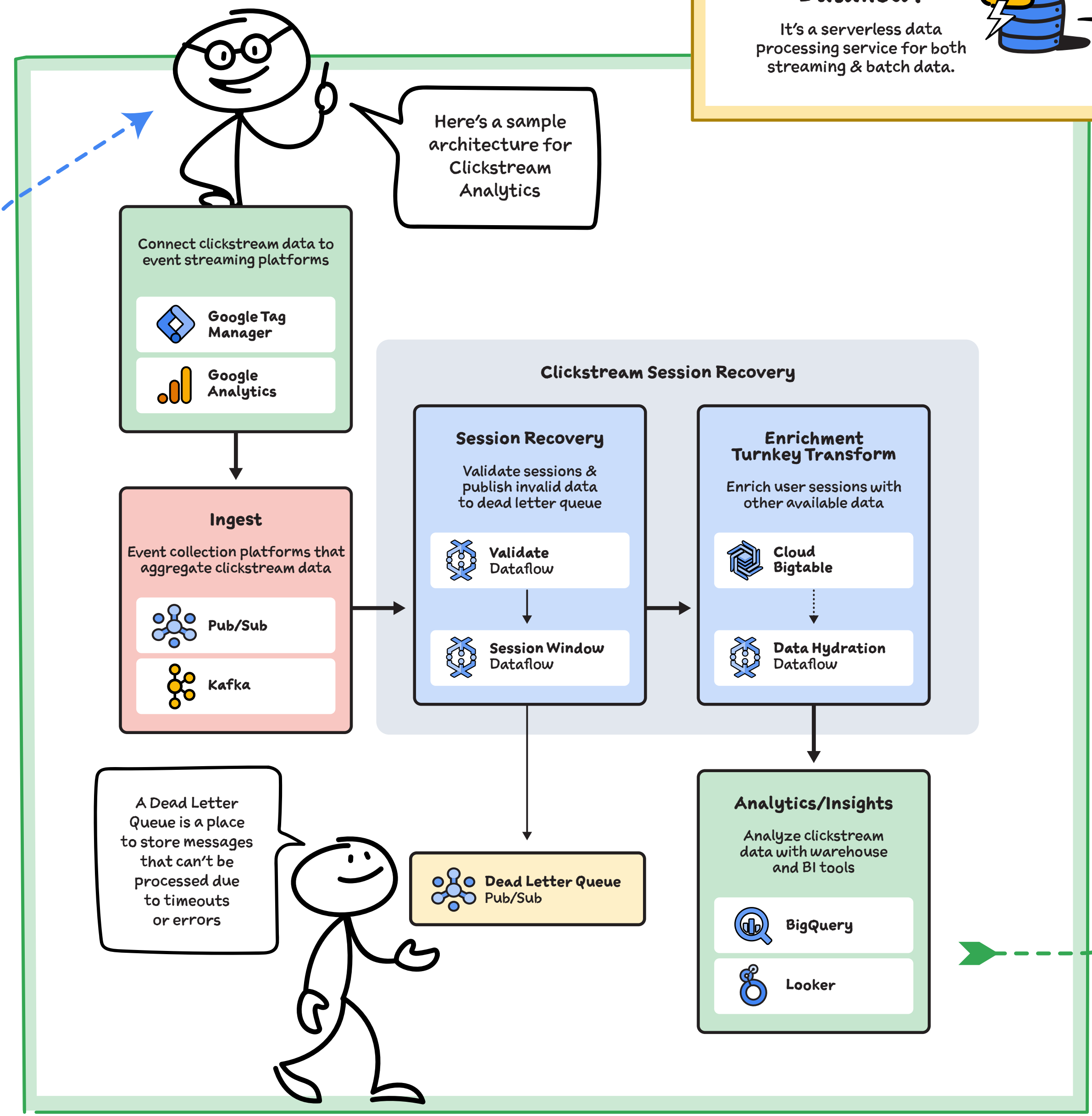
Dataflow enables real-time clickstream analysis, so we can personalize the user journey & boost engagement!

Here's how Dataflow can help your business...

- Real-time Personalization
- Real-time A/B Testing
- Decreased Churn & Bounce Rate
- Contextual Product Support
- Dynamic Pricing & Cart Upsell

Take a look at these technical benefits...

- Autoscaling that provisions infrastructure for your unpredictable traffic patterns
- Simplify ML pipelines with easy-to-use code blocks
- Flexible I/O framework that can adapt to your clickstream solution
- Standardize on your data model & business logic with Dataflow Templates
- High resiliency with snapshots & built-in checkpointing



Keep in mind...

- Use Apache Beam's built-in session windows to simplify sessionization, or use State & Timer APIs for more advanced sessionization
- Send messages that don't meet your session definition to a dead letter queue
- Weigh tradeoffs between completeness & latency when windowing your data
- Use Enrichment with fast-lookup databases like BigTable or Vertex AI Feature Store for feature engineering
- Using your session ID as a primary key for databases that support upserts avoids duplication & ensures consistency