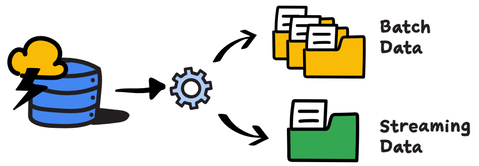
Real-Time ML & Generative AI with Dataflow

What is Dataflow?

It's a serverless data processing service for both streaming & batch data.



Our customers want instant responses from our AI models, but our pipelines are too slow!

And we really need real-time predictions, not hours-old results!

I've got it! **We'll use Dataflow!** Here's how it can help ...

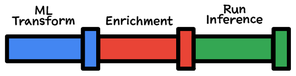
First let's look at increasing business impact...

- Realtime Predictions
- Instant Personalization
- Intelligent Agents
- Enhanced Decision Agility
- Improved Customer Support
- Predictive Maintenance

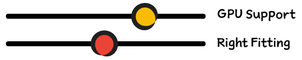
Hey, take a look at this! Here's an **architecture diagram** that shows just how Dataflow works.

Next, let's look at some **technical benefits...**

Simplify ML pipelines with low code turnkey processing.

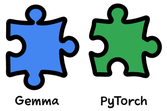


Optimize resource utilization; minimize expenses.

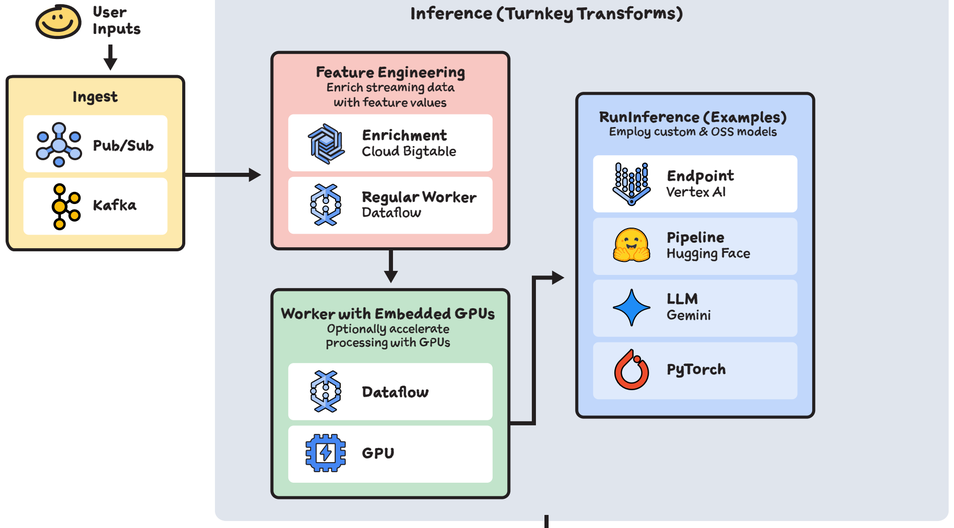


Develop streaming AI pipelines with interactive notebooks.

Open-source compatibility for popular industry standard models.



Write advanced business logic with the Apache Beam SDK.



Finally, keep in mind...

Use message queues like Pub/Sub or Kafka as your source & sink to decouple predictions from the main application loop.

Reduce operational overhead & code complexity with turnkey transformations.

Use Enrichment with fast-lookup databases like BigTable or Vertex AI Feature Store for feature engineering.

Use RunInference to generate predictions from models hosted in Vertex AI.

Combine GPUs and Dataflow right fitting to optimize efficiency and minimize costs.