

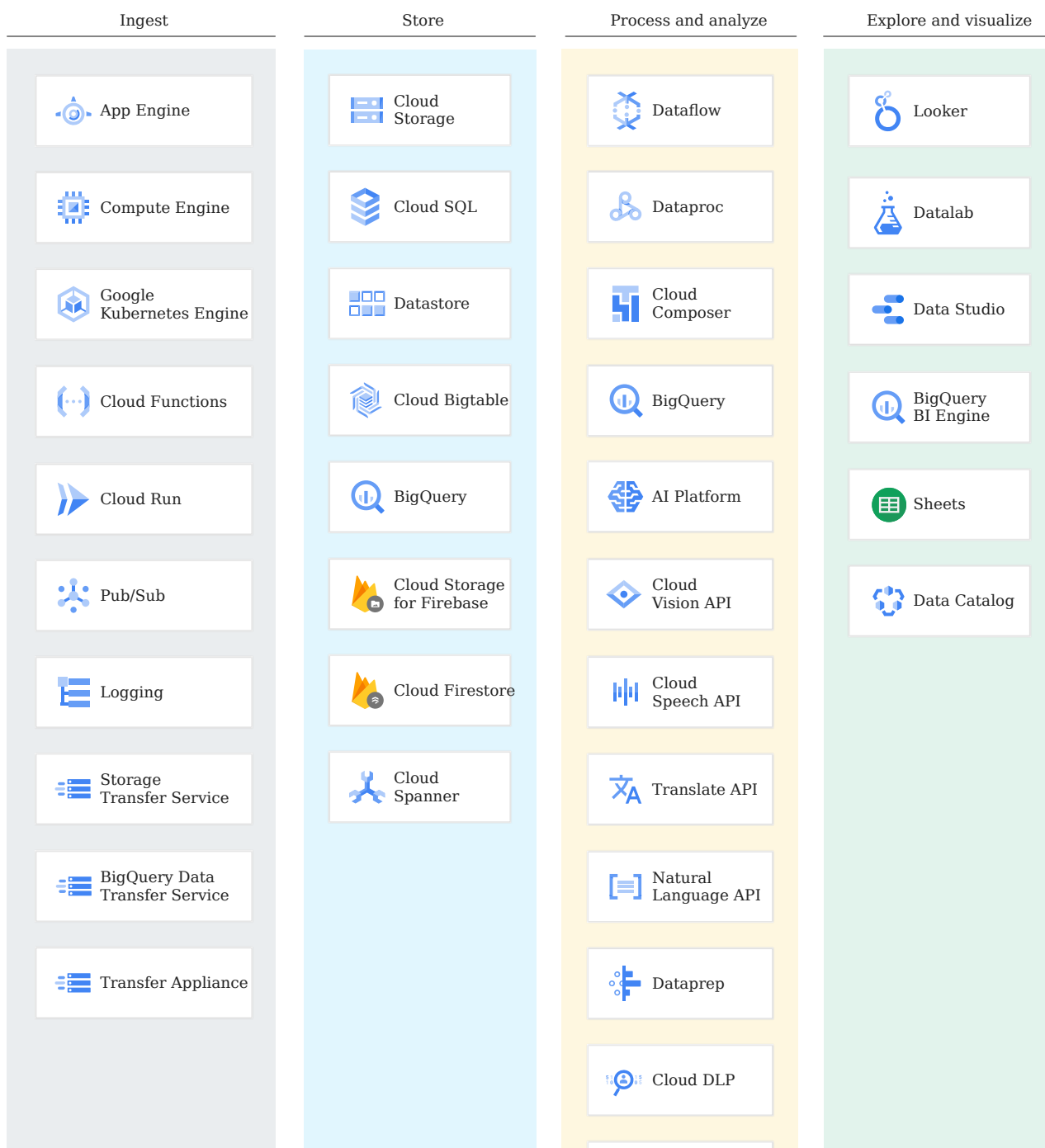
Ciclo de vida dos dados

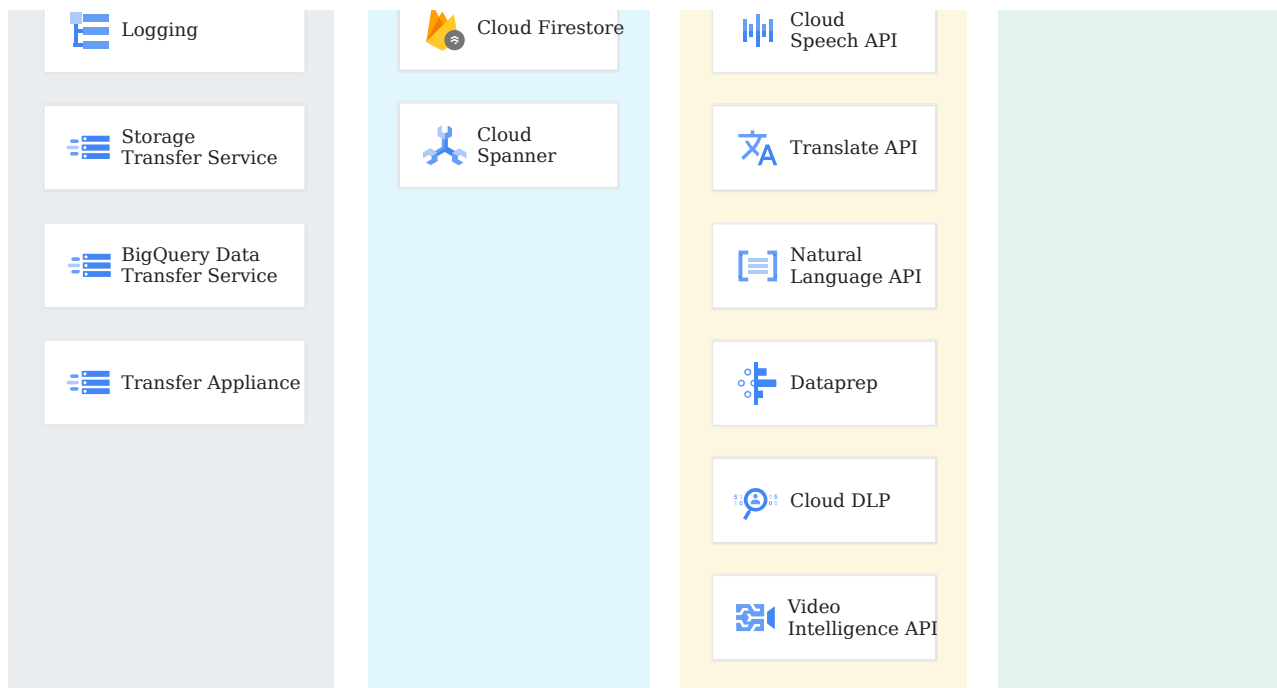
Neste artigo, descrevemos os serviços do Google Cloud que podem ser usados para gerenciar dados em todo o ciclo de vida, desde a aquisição inicial até a visualização final. Você aprenderá sobre os recursos e a funcionalidade de cada serviço para decidir melhor sobre os serviços que são mais adequados à sua carga de trabalho.

Há quatro etapas no ciclo de vida dos dados:

- **Ingerir:** a primeira etapa é extrair dados brutos, como de streaming de dispositivos, dados em lote no local, registros de aplicativos ou eventos e análises de usuários de aplicativos para dispositivos móveis.
- **Armazenar:** depois de recuperados, os dados precisam ser armazenados em um formato durável e facilmente acessível.
- **Processar e analisar:** nessa etapa, os dados são transformados de brutos em informações acionáveis.
- **Explorar e visualizar:** a etapa final é a de conversão dos resultados da análise em um formato que facilite a extração de informações e o compartilhamento com os colegas.

Em cada etapa, o Google Cloud fornece vários serviços para gerenciar seus dados. Isso significa que é possível selecionar um conjunto de serviços adaptados aos dados e ao fluxo de trabalho.



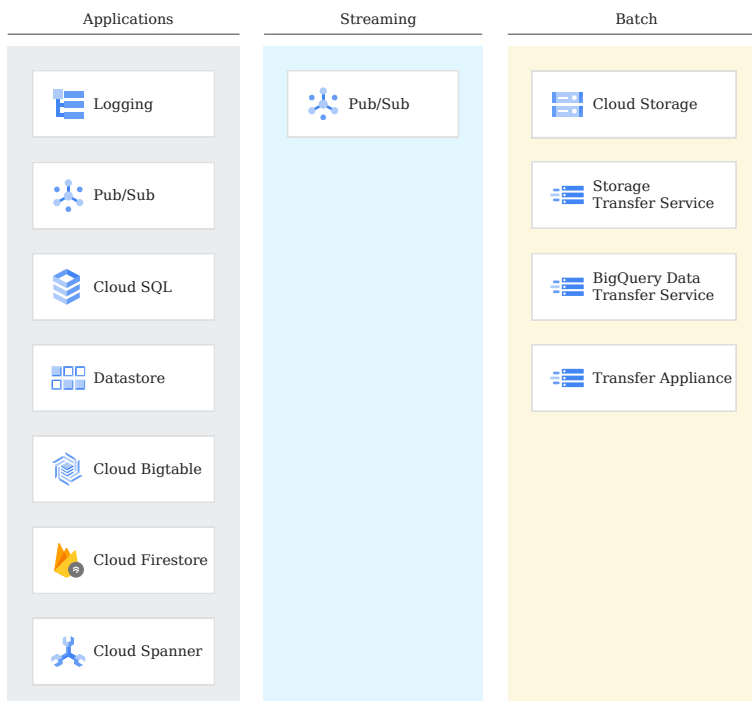


Ingestão

Há diversas abordagens para a coleta de dados brutos com base no tamanho, fonte e latência deles.

- **Aplicativo:** os dados de eventos de aplicativos, como arquivos de registro ou eventos de usuário, normalmente são coletados em um modelo de push. Nele, o aplicativo chama uma API para enviar os dados ao armazenamento.
- **Streaming:** os dados consistem em um stream contínuo de pequenas mensagens assíncronas.
- **Lote:** grandes quantidades de dados são armazenadas em um conjunto de arquivos que são transferidos em massa para o armazenamento.

No gráfico a seguir, mostramos como os serviços do Google Cloud são mapeados para cargas de trabalho de app, streaming e lote.



O modelo de transferência de dados escolhido depende da carga de trabalho. Cada modelo tem requisitos de infraestrutura diferentes.

Como ingerir dados do aplicativo

em tempo real. Em segundo app registrado nas mensagens transfere os dados para o armazenamento ou os processa imediatamente em situações como detecção de fraudes.

Cloud Logging: gerenciamento centralizado de registros

O Cloud Logging (<https://cloud.google.com/logging?hl=pt-br>) é um serviço centralizado de gerenciamento de registros que coleta dados de registros de apps executados no Google Cloud e em outras plataformas de nuvem pública e privada. Exporte os dados coletados pelo Cloud Logging usando ferramentas integradas que enviam os dados ao Cloud Storage (<https://cloud.google.com/storage?hl=pt-br>), Pub/Sub (<https://cloud.google.com/pubsub?hl=pt-br>) e BigQuery (<https://cloud.google.com/bigquery?hl=pt-br>).

Muitos serviços do Google Cloud gravam dados de registro automaticamente no Cloud Logging. Por exemplo, os apps executados no App Engine (<https://cloud.google.com/appengine?hl=pt-br>) registram automaticamente os detalhes de cada solicitação e resposta ao Cloud Logging. Também é possível gravar mensagens de registro personalizadas em `stdout` e `stderr`, que são coletadas e exibidas automaticamente pelo Cloud Logging no Visualizador de registros.

O Cloud Logging inclui um agente de geração de registros, com base no fluentd (<http://www.fluentd.org/>), que pode ser executado em instâncias de máquina virtual (VM, na sigla em inglês) hospedadas no Compute Engine (<https://cloud.google.com/compute/docs?hl=pt-br>) e também em clusters de contêiner gerenciados pelo GKE (<https://cloud.google.com/kubernetes-engine?hl=pt-br>). O agente transmite dados de registro de apps de terceiros comuns e do software do sistema para o Cloud Logging.

Como ingerir dados de streaming

Os dados de streaming são fornecidos de modo assíncrono, sem esperar resposta, e as mensagens individuais são pequenas. Em geral, os dados de streaming são usados para *telemetria*, coletando dados de dispositivos dispersos geograficamente. Os dados de streaming podem ser usados para acionar disparadores de eventos, realizar análise de sessão complexa e como entrada para tarefas de aprendizado de máquina.

Estes são dois usos comuns de streaming de dados.

- **Dados de telemetria:** os dispositivos de Internet das coisas (IoT, na sigla em inglês) são conectados na rede e coletam dados do ambiente próximo por meio de sensores. Cada dispositivo pode enviar apenas um pouco de dados por minuto. Porém, quando esses dados são multiplicados por

- **Cargas de trabalho científicas:** fazer o upload de dados genéticos, armazenados em arquivos de texto Variant Call Format (VCF), para o Cloud Storage (<https://cloud.google.com/storage?hl=pt-br>) e importá-los posteriormente no Genomics (<https://cloud.google.com/genomics?hl=pt-br>).
- **Migração para a nuvem:** mover dados armazenados em um banco de dados Oracle local para um totalmente gerenciado do Cloud SQL (<https://cloud.google.com/sql?hl=pt-br>) usando o Informatica.
- **Backup de dados:** replicar dados armazenados em um bucket da AWS para o Cloud Storage usando o Serviço de transferência do Cloud Storage (<https://cloud.google.com/storage/transfer?hl=pt-br>).

