**Soluções** **Ciclo de Vida de Dados** Em cada etapa, o GCP oferece múltiplos serviços para gerenciar seus dados.

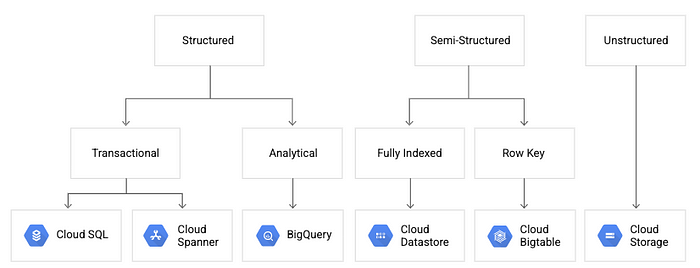
1. **Ingestão**: a primeira etapa é trazer os dados brutos, como dados de streaming de dispositivos, em lotes de serviços on-premises e de aplicações móveis, ou de usuários por meio de APIs que enviam os dados para armazenamento (Stackdriver Logging, Pub/Sub, CloudSQL, Datastore, Bigtable, Spanner).
2. **Armazenamento**: depois que os dados são recuperados, é necessário armazená-los em um formato durável e de fácil acesso.
3. **Processar e Analisar**: os dados são transformados de uma forma bruta em informações acionáveis.
4. **Explorar e Visualizar**: converte os resultados das análises em um formato que é fácil para desenhar insights e compartilhar com colegas e parceiros.

**Ingestão** Existem várias abordagens para coletar dados brutos, baseadas no tamanho dos dados, fonte e latência.

* **Aplicação**: dados de aplicações em eventos, logs de usuários etc., tipicamente coletados por uma API para enviar os dados para armazenamento (Stackdriver Logging, Pub/Sub, CloudSQL, Datastore, Bigtable, Spanner).
* **Streaming**: os dados consistem de um fluxo contínuo de pequenas mensagens assíncronas. Usos comuns incluem telemetria ou coleta de dados de dispositivos geograficamente distribuídos (IoT) e eventos de usuários analisados (Pub/Sub).
* **Batch**: grandes quantidades de dados são armazenadas em um conjunto de arquivos que são transferidos para armazenamento em lote. Casos de uso comuns incluem workloads científicos, backups, migração (Storage, Transfer Service, Appliance).

**Soluções - Parte II** **Armazenamento**

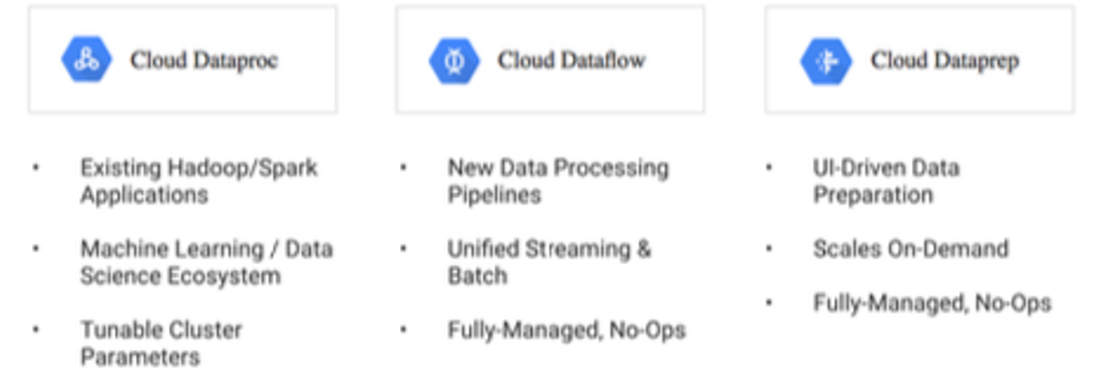
* **Cloud Storage**: armazenamento de objeto durável e altamente disponível para dados estruturados e não estruturados.
* **Cloud SQL**: banco de dados RDBMS totalmente gerenciado que oferece suporte tanto para MySQL quanto para PostgreSQL com réplicas internas, suporte multi-região, suportando workloads financeiros, transações RDBMS, credenciais de usuário (trabalhadores).
* **BigTable**: banco de dados NoSQL de alta performance projetado para workloads de escala petabyte. Adequado para grandes cargas de trabalho, como tecnologias publicitárias, jogos ou infraestrutura de IoT (não suporta SQL, transações ou consultas OLTP de baixa latência).
* **Cloud Spanner**: banco de dados relacional gerenciado escalável e mission-critical para aplicações OLTP, horizontalmente escalável, segurança robusta, esquemas ACID, transações globais e consultas SQL (suporta serviços financeiros e análises quantitativas de grandes quantidades de dados para consultas e processamento analítico).



**Exploração e Visualização** Exploração e visualização dos dados são a exploração e o processamento de dados brutos em análise.

* **Cloud Datalab:** ferramenta web interativa que você pode usar para explorar, analisar e visualizar dados construída em cima de notebooks Jupyter. Roda em uma VM e é automaticamente salva em discos persistentes e pode ser armazenada no GC Source Repo (repositório git).
* **Data Studio:** construtor de relatórios drag-and-drop que você pode usar para visualizar dados em relatórios e dashboards que podem ser compartilhados com outros, alimentados por dados ao vivo, que podem ser compartilhados e atualizados facilmente. Fontes de dados podem ser arquivos de dados, Google Sheets, Cloud SQL e BigQuery. Suporta consultas e o cache de pré-busca: a consulta lembra consultas anteriores e se os dados não forem encontrados, vai para o cache de pré-busca, que prevê dados que podem ser solicitados (desative se a fonte de dados mudar frequentemente/usar a fonte de dados incorrer em custos).

**Parte superior do formulário**



**Processar e Analisar** Para obter valor e insights dos dados, você precisa transformá-los e analisá-los. Isso requer um fluxo de trabalho que pode transformar diretamente os dados brutos para prepará-los para análise downstream, bem como analisar os dados transformados e converter os resultados processados em informações acionáveis.

* Normalizado: dados processados são limpos, processados e transformados em sistemas gerenciáveis, e analisados em sistemas capazes de armazenar dados processados.
* Análise: dados processados são armazenados em sistemas que permitem análise focada em descoberta e exploração.
* Compreensão: baseado em análises resultantes, os dados são usados para treinar e automatizar modelos de machine learning.

**Processamento**

* **Cloud Dataproc**: migra seus deployments Hadoop ou Spark existentes para um cluster gerenciado que simplifica a criação e gestão de clusters, facilitando a configuração e gestão do seu cluster, possui monitoramento embutido e alertas de utilização de recursos e pode ser desligado quando não estiver em uso.
* **Cloud Dataflow:** Projetado para simplificar tanto processos de streaming quanto de lote, foca em arquitetura aberta, transações em lote, agregação e transformações de dados.
* **Cloud Dataprep:** Serviço para explorar visualmente, limpar e preparar dados para análise, transformando dados em formatos como CSV, JSON ou tabelas relacionais.

**Analisando e Consultando**

* **BigQuery**: Consultas usando SQL, todos os dados criptografados, análise inteligente de dispositivos, métricas de negócios, TI - analíticos.
* **ML Específico para Tasks**: Visão, Fala, Linguagem Natural, Tradução, Inteligência de Vídeo
* **ML Engine:** plataforma gerenciada que você pode usar para executar modelos de aprendizado de máquina personalizados em escala