the challenges faced while creating a data pipeline is :

1. data coming from various sources IoT sensors, log files etc.
2. data at scale, high velocity of data coming in.
3. real-time batch or stream ingestion.
4. data may be low quality.

google pub/sub(publisher/subscriber): Pub/Sub is a distributed messaging service that can receive messages from a variety of different streams,

upstream data systems like gaming events, IoT devices, applications streams, and more. It ensures at least once delivery of messages and passes them to subscribing applications and no provisioning is required.

Where there's a ton of messages or none at all, Pub/Sub will scale to meet that demand. Also, the APIs are open, and the service is global by default and offers end-to-end encryption for those messages.

pub/sub is the first point of contact of data with the pipeline, it then passes the data to google dataflow to transform/ process the data and then the data is fed to analytics or storage.

**Apache Beam** is a portable data processing programming model.

It's open-source, and it can be run in a highly distributed fashion.

**Cloud Dataflow.**

* It's designed to be easy to deploy your pipeline and have it just run.
* It's serverless and fully managed.
* Same code can both batch and streaming at scale.
* It's built on Apache Beam's open-source programming model.
* It can autoscale, auto-heal, and re-balance to millions of queries per second.

Data Studio:

it is a data visualization tool, which can be used to make live and dynamic charts and bars which represent data. the data can be from various sources.