



Common Issues



Data analysts face query, infrastructure, and storage challenges

"My queries are taking way **too long** to run and is stalling my analysis."

"We're a data department, not an **infrastructure** department. Maintaining and upgrading our own servers is unsustainable."

"We can only **afford to store a subset** of the data our business generates"

"I have no easy way to **combine and query** all the data I've collected"

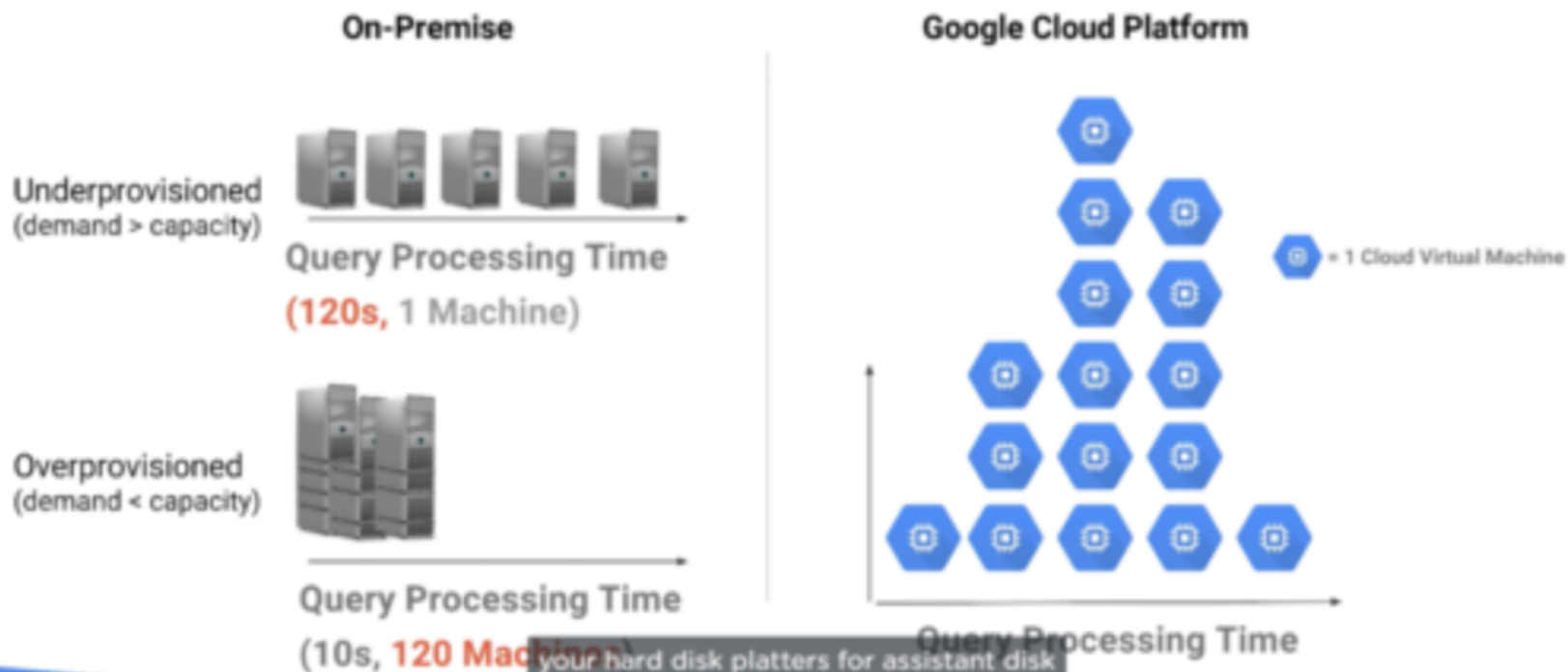
"My on premise clusters **aren't scaling** with my analysis"

"We don't have a **central data** analytics warehouse or set of tools"

Google Cloud Platform Scalability



Google Cloud Platform enables on-demand scalability



On-Premise vs GCP: Resource Allocation



Separation of storage and computing power enables efficient resource allocation

On-Premise



Pay for ability to use processing power
even when no queries running

Google Cloud Platform



Pay for only the resources you
are using and no more



How does BigQuery handle
scaling and pricing?



BigQuery scales automatically
and you only pay for what you
use.



Data Analysis Steps and Challenges



Challenges in each task prevent data analysts from getting to scalable insights



Ingest

Get data in.



Challenges

Data Volume
Data Variety
Data Velocity



Transform

Prepare, clean, and transform data.



Challenges

Slow Exploration
Slow Processing
Unclear Logic



Store

Create, save, and store datasets.



Challenges

Storage Cost
Hard to Scale
Latency Issues



Analyze

Derive insights from data.



Slow Queries
Data Volume
Siloed Data



Visualize

Explore and present data insights.



Dataset Size
Tool Latency

Ingest (Get data in). Transform (Clean the data). Store (Save the data). Analyze (Derive insights from the data). Visualize (Explore and present data and findings).



BigQuery Terminology















Creating and Querying Datasets: BigQuery Terminology



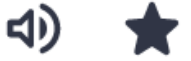
Data Analyst vs Data Scientist vs Data Engineer



Each data-related role uses a different suite of tools

Roles:	Data Analyst	Data Scientist	Data Engineer
What they do:	Derive data insights from queries and visualization.	Analyze data and model systems using statistics and machine learning.	Designs, builds, and maintains data processing systems.
Background:	Data analysis using SQL	Statistical analysis using SQL, R, Python	Computer Engineering
GCP Tools Used:	    	   	         

Data Analysts derive insights from queries and visualization. Data Scientists analyze and model systems using statistics and machine learning. Data Engineers design, build, and manage data processing systems.



Highlighting GCP Tools



End-to-end gaming analytics example highlighting GCP tools



Why would you not want to use
stylistic formatting inside of
BigQuery?



It can unintentionally affect the
data type.



Why should you avoid using *
when writing queries?



Selecting only the columns you need greatly increases query speed and helps with readability.