Analyse Data in Less Time with Google BigQuery

University of Michigan
Google Cloud Platform provides the foundation for your data analytics and machine learning workloads.
The State of Big Data
“Information is the oil of the 21st century, and analytics is the combustion engine.”

Peter Sondergaard
Gartner Research
Big Data growth is **explosive**, and accumulating **exponentially**

**Varied sources and types**
- Transactions, logs
- Real-time streaming data
- Structured and unstructured data
- Web, social, sensor
- Audio, video, documents

**Countless uses**
- Discover new trends
- Explore new business opportunities
- Increase innovation
- Fraud detection
- Prediction models
- Identify inefficiencies
Big Data **amplifies resource shortages**

- Big Data analytics is **constantly evolving** and demand continues to **build**
- Most development methods do not accommodate **data discovery** and **interactive** analytics
- Large scale on-premise computing resources are **capital** and set-up time **intensive**
- Capitalizing on Big Data insights requires dynamic system resizing and reconfiguration
Google Big Data services offer cost, performance and reliability advantages over on-premises services.

- **On premises**
  - **Exponential cost**: 2x processing power equals 5x the cost
  - **Limited protection** against software/hardware failure
  - Requires infrastructure support

- **Google Cloud**
  - **Linear cost**: 2x processing power equals 2x cost
  - **Fault tolerance designed in**: resilient when hardware or software fails
  - **Google-managed infrastructure**: included in price
Defining BigQuery
BigQuery is the key product in many big data workloads.
BigQuery is a fully-managed enterprise data warehouse

- Provides **near real-time interactive analysis** of massive datasets
- Reliable
  - Data replicated across multiple data centers
- Economical
  - Only pay for storage and processing used
BigQuery is secure

- **Secured** through Access Control Lists (ACLs) and Identity and Access Management (IAM)
  - Data is **encrypted** in **transport** and at **rest**
- Google Cloud **audit logs** track **admin activity** and **data access**
  - Logs are immutable - “who did what, where, and when?”
BigQuery is incredibly fast

- **Familiar** standard SQL query language
- **Intuitive UI**
- **Simple** to load data from multiple sources
  - Batch
  - Streaming
- **Public datasets** to explore and integrate
BigQuery is easy to use

```sql
SELECT
    language,
    SUM(views) as views
FROM
`bigquery-samples.wikipedia_benchmark.Wiki10B`
WHERE
    REGEXP_CONTAINS(title,"G.*o.*o.*g")
GROUP by language
ORDER by views DESC
```
Processes 10 billion rows in 10 seconds
Welcome to BigQuery!

Google BigQuery is a web service that lets you do interactive analysis of massive datasets—up to billions of rows. Scalable and easy to use, BigQuery lets developers and businesses tap into powerful data analytics on demand.

To get started, try one of the following options:

- Read our BigQuery Quickstart guide
- Run a query against our sample data by clicking “Compose Query”
- Create a new dataset and load some of your own data into a table using the menu on the left
- For more information on the UI, see the BigQuery Web UI guide
Big Query does not replace every data store

<table>
<thead>
<tr>
<th>Transactional RDBMS</th>
<th>Operational Data Store</th>
<th>On-premises solution or appliance</th>
</tr>
</thead>
<tbody>
<tr>
<td>BigQuery is not an OLTP system</td>
<td>BigQuery is not geared towards capturing live data and applying updates/deletes as they happen in the system of record</td>
<td>BigQuery is a self-contained, cloud-based solution</td>
</tr>
</tbody>
</table>
Google Big Query Demo
Resources

Product Documentation
https://cloud.google.com/bigquery/

Try it yourself in a Codelab!
https://goo.gl/Z2uv7F (setup)
https://goo.gl/3xl5Qt (Big Query)
Thanks