

Pengantar Google Cloud Platform Dan Database

**Sesi 1-Bootcamp Data Afialyst SQL & Python
using Google Platform**



Google Cloud Platform

Introduction

Google Cloud merupakan salah satu provider layanan Cloud Computing di dunia yang mengandalkan infrastruktur dari Google.

Google Cloud menawarkan layanan untuk komputasi, penyimpanan, jaringan, big data, machine learning, dan IoT, serta manajemen cloud, dan keamanan.

Dalam konteks Big Data Analytics, dalam GCP akan dibahas 3 tools berikut:

- BigQuery, yaitu Analytical Database untuk mengakses data dengan query.
- Google Colab, yaitu notebook yang digunakan untuk data explanatory dan machine learning.
- Looker, yaitu tools visualisasi

Perbandingan	On Premise	Cloud
Deployment	Dalam lingkungan on premise, sumber daya di-deploy secara in-house dan berada di dalam infrastruktur IT perusahaan. Sebuah perusahaan bertanggung jawab untuk memelihara seluruh proses dan isu yang akan muncul.	Meski ada beberapa jenis dari cloud computing yakni ada public cloud, private cloud, dan hybrid cloud, seluruh sumber daya ditaruh di lingkungan dari penyedia cloud. Namun, perusahaan memiliki akses sumber daya tersebut dan dapat menggunakan sebanyak yang perusahaan butuhkan.
Biaya	Untuk perusahaan yang men-deploy software pada on premise, mereka bertanggung jawab atas ongkos server, konsumsi power, dan ruangan.	Perusahaan yang memilih cloud hanya perlu membayar sumber daya yang mereka gunakan tanpa harus membayar maintenance. Harga akan naik dan turun tergantung dari penggunaan.
Kendali	Di lingkungan on premise, perusahaan memiliki hak penuh atas data mereka, dan punya kendali penuh atas apa yang terjadi, baik itu hal buruk atau baik. Perusahaan yang memiliki sensitivitas tinggi atas keamanan pribadi, biasanya akan ragu untuk menggunakan cloud	Dalam lingkungan cloud computing, pertanyaan akan kepemilikan data masih menjadi pertanyaan mendasar. Data dan kunci enkripsi ada di vendor cloud, sehingga apabila hal yang tidak terduga terjadi, Anda mungkin tidak bisa mengakses data-data tersebut.
Compliance	Banyak perusahaan yang kini beroperasi di bawah kendali regulasi tertentu seperti perusahaan yang bergerak di telehealth, e-learning, dan lain-lain. Untuk perusahaan-perusahaan yang berada dalam regulasi seperti itu, adalah suatu keharusan untuk tetap patuh dan tahu data mereka kapan pun.	perusahaan yang memilih model cloud computing harus memastikan bahwa penyedia cloud mereka sudah sesuai dan patuh dengan regulasi yang ada di industri. Data-data yang sensitif harus diamankan, sekaligus konsumen, partner, dan karyawan harus memastikan data-data mereka sudah aman.

BigQuery



BigQuery adalah layanan pengolahan data berbasis cloud yang disediakan oleh Google Cloud Platform. BigQuery dirancang untuk menganalisis data dalam skala besar dengan cepat dan efisien. Ini adalah solusi yang tangguh untuk kebutuhan analisis data, pemodelan, dan visualisasi.

BigQuery menggunakan arsitektur terdistribusi yang memanfaatkan infrastruktur komputasi dan penyimpanan Google untuk memproses dan menganalisis data yang sangat besar. Ini mendukung pemrosesan paralel yang kuat dan dapat menangani kumpulan data berukuran terabyte hingga petabyte dengan mudah.



Google Looker adalah platform analitik dan visualisasi data yang diperoleh oleh Google pada tahun 2020. Looker memungkinkan pengguna untuk menjelajahi, menganalisis, dan memvisualisasikan data dari berbagai sumber dengan cara yang intuitif dan interaktif.

Looker memadukan data dari berbagai sumber, termasuk data warehouse seperti Google BigQuery, serta sumber data lain seperti basis data SQL, alat cloud, layanan web, dan lainnya. Ini memungkinkan pengguna untuk memiliki pandangan holistik atas data mereka tanpa perlu mentransfer atau menggabungkan data secara manual.

Dengan menggunakan Looker, pengguna dapat membuat tampilan data yang kaya dan interaktif, membuat laporan dan dashboard, serta melakukan analisis data yang mendalam. Platform ini menyediakan antarmuka yang intuitif dan mendukung bahasa SQL untuk menjalankan query yang kompleks. Selain itu, Looker juga memiliki fitur-fitur seperti penjadwalan tugas, pemberitahuan, dan kolaborasi yang memudahkan pengguna untuk berbagi dan berkolaborasi dalam menganalisis data.

Google Colab



Google Colaboratory

Google Colab, singkatan dari Google Colaboratory, adalah platform pengembangan berbasis cloud yang disediakan oleh Google. Notebook Colab memungkinkan pengguna untuk menulis dan menjalankan kode Python, serta melakukan eksperimen, analisis data, dan pelatihan model machine learning. Beberapa fitur utama dari Google Colab meliputi:

- Lingkungan Berbasis Cloud:** Colab berjalan di cloud, sehingga pengguna tidak perlu menginstal atau mengkonfigurasi perangkat lunak di komputer mereka. Semua komputasi dan penyimpanan dilakukan di server Google.
- Jupyter Notebook:** Colab menggunakan format Jupyter Notebook yang memungkinkan pengguna untuk menulis dan menjalankan kode Python dalam sel-sel yang dapat dijalankan secara interaktif. Ini memungkinkan eksperimen dan iterasi cepat pada kode.
- Akses ke GPU dan TPU:** Colab menyediakan akses ke unit pemrosesan grafis (GPU) dan unit pemrosesan tensor (TPU) dari Google secara gratis. Ini sangat bermanfaat untuk melatih model machine learning yang membutuhkan kecepatan pemrosesan yang tinggi.
- Integrasi dengan Layanan Google:** Colab terintegrasi dengan berbagai layanan Google seperti Google Drive, Google Sheets, dan Google Cloud Storage. Ini memudahkan pengguna untuk mengimpor dan mengekspor data serta berbagi kode dengan mudah.
- Berbagi dan Kolaborasi:** Colab memungkinkan pengguna untuk berbagi notebook mereka dengan orang lain dan berkolaborasi. Hal ini memudahkan tim untuk bekerja bersama dalam pengembangan proyek atau analisis data.

Database

Basis data yang juga dikenal sebagai database, terdiri dari kata basis data data. Data merupakan catatan atas kumpulan fakta yang mewakili suatu objek. Data masih bersifat mentah dan tidak mewakili konteks. Sedangkan basis atau base diartikan sebagai markas, tempat berkumpul dari suatu objek atau representasi objek.

Dalam implementasinya, untuk memudahkan dalam mengakses data, data disusun dalam suatu struktur logis yang menjelaskan bahwa:

1. Kumpulan tabel menyusun basis data
2. Tabel tersusun atas sejumlah **record**
3. Sebuah **record** mengandung sejumlah **field**
4. Sebuah **field** disimpan dalam bentuk kumpulan **bit**.

Berikut merupakan pengertian dari istilah di atas:

1. **Field** menyatakan data terkecil yang memiliki makna. Istilah lain untuk field yaitu elemen data, kolom item, dan atribut. Contoh field yaitu nim, nama, prodi, no_hp, alamat.
2. **Record** menyatakan sekumpulan dari sejumlah elemen data yang saling terkait. sebagai contoh: nim, nama, prodi dan alamat dari mahasiswa menyusun sebuah record. Istilah lain yang menyatakan record yaitu tupel dan baris.

Database System

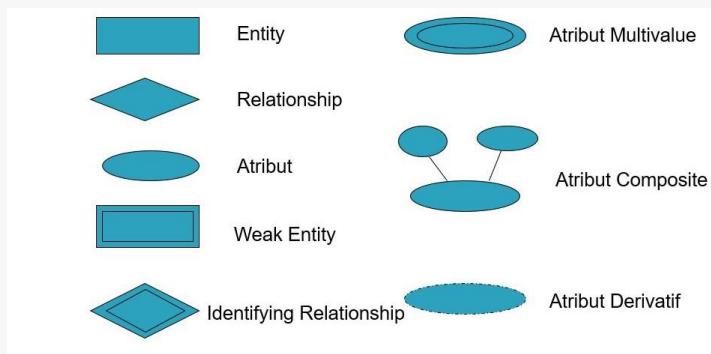
Sistem Basis Data atau **DBMS (Database Management System)** adalah software atau perangkat lunak pengolah Basis Data. Salah satu tugas penting dari DBMS adalah menerjemahkan perintah-perintah yang diberikan oleh user terkait pengelolaan database. Perintah-perintah tersebut disebut dengan bahasa SQL (Structure Query Language).

Tujuan dibuatnya DBMS adalah sebagai berikut :

1. Agar basis data dapat digunakan secara bersama
2. Agar proses akses data lebih mudah dan cepat
3. Untuk menghemat ruang penyimpanan data
4. Membantu menjaga keamanan data
5. Mencegah dan menghilangkan duplikasi dan inkonsistensi data
6. Menangani data dalam jumlah yang besar

Entity Relationship Diagram (ERD)

ERD (Entity Relationship Diagram) adalah model atau rancangan untuk membuat database, supaya lebih mudah dalam menggambarkan data yang memiliki hubungan atau relasi dalam bentuk sebuah desain. Dengan adanya ER diagram, maka sistem database yang terbentuk dapat digambarkan dengan lebih terstruktur dan terlihat rapi. ERD komponen utamanya terdiri dari Entitas, Atribut dan Relasi.



Jenis-Jenis Entitas

Entitas Lemah (Weak Entity)

Berisi entitas yang kemunculannya tergantung pada eksistensinya dalam sebuah relasi. misal entitas orangtua, hobby tergantung pada entitas Mahasiswa. Entitas tanggungan tergantung pada entitas pegawai.

Entitas kuat (Strong Entity)

Entitas yang berdiri sendiri, keberadaannya tidak tergantung dengan entitas lain. contoh: mahasiswa, pegawai.

Jenis-Jenis Atribut

- **Atribut Key**

Atribut yang digunakan untuk membedakan data secara unik. misal nik, nim, no_peserta.

- **Atribut Simple**

Atribut bernilai tunggal, misal nama, harga, status.

- **Atribut Multivalue**

Atribut memiliki sekelompok nilai yang banyak, misal gelar, hobby.

- **Atribut Composite**

Atribut yang masih dapat diuraikan lagi, misal alamat, korban, waktu.

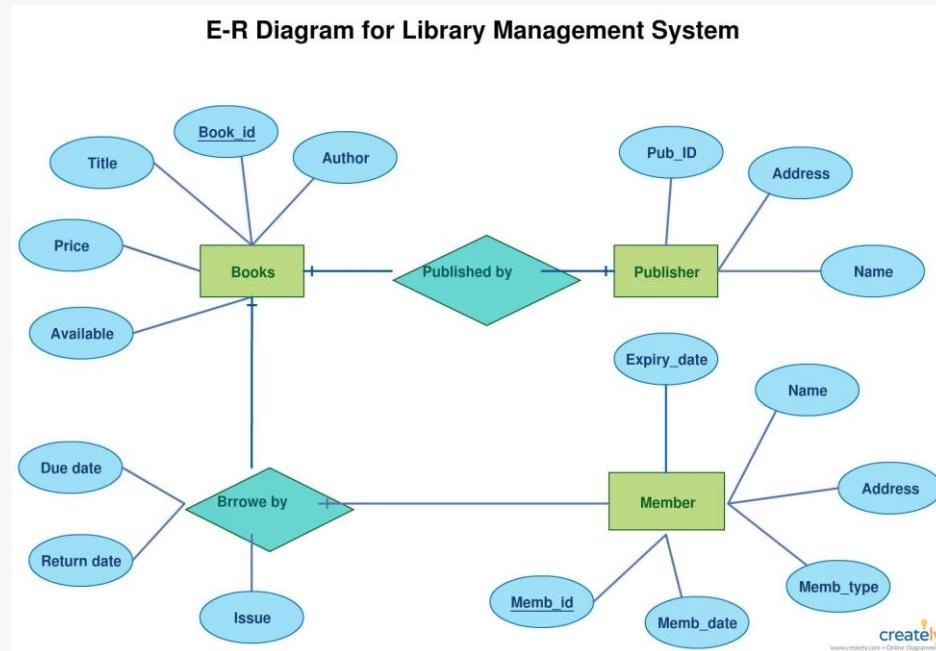
- **Atribut Derivatif**

Atribut yang dihasilkan dari turunan atribut lain, misal umur dari tanggal lahir.

Tahapan Umum Pembuatan ERD

- Tahap 1 – Menentukan entitas yang akan terlibat atau menentukan tabel.
- Tahap 2 – Menentukan atribut-atribut key dari masing-masing himpunan entitas.
- Tahap 3 – Menetapkan seluruh himpunan relasi di antara himpunan entitas yang ada beserta foreign key-nya dan kardinalitas relasi.
- Tahap 4 – Membuat model Entity Relationship Diagram.

Contoh Diagram ERD



Normalisasi Database

Normalisasi merupakan sebuah teknik dalam logical desain sebuah basis data yang mengelompokkan atribut dari suatu relasi sehingga membentuk struktur relasi yang baik (tanpa redundansi).

Normalisasi adalah proses pembentukan struktur basis data sehingga sebagian besar ambiguity bisa dihilangkan.

Tujuan Normalisasi Database:

- Untuk menghilang kerangkapan data
- Untuk mengurangi kompleksitas
- Untuk mempermudah pemodifikasiyan data

Proses Normalisasi

Data diuraikan dalam bentuk tabel, selanjutnya dianalisis berdasarkan persyaratan tertentu ke beberapa tingkat.

Apabila tabel yang diuji belum memenuhi persyaratan tertentu, maka tabel tersebut perlu dipecah menjadi beberapa tabel yang lebih sederhana sampai memenuhi bentuk yang optimal.

Contoh Normalisasi

Untuk melakukan normalisasi database kita harus mengidentifikasi data seperti apa yang akan disimpan. Sebagai contoh terdapat data dari struk penjualan. Contohnya adalah data struk penjualan karena data ini yang paling umum di sekitar kita.

kode_faktur	tanggal	kode_barang	nama_barang	harga	qty
KD_001	13/07/2020	BRG_001	Indomie Goreng	2500	5
		BRG_002	Indomie Goreng Jumbo	3000	8
		BRG_003	Mie Sedaap Goreng	2500	9
		BRG_004	Mie Sedaap Soto	2300	3
KD_002	13/07/2020	BRG_005	Intermie Goreng	1500	6
		BRG_006	Intermie Soto	1500	6
KD_003	14/07/2020	BRG_007	Pop Mie Ayam	4500	3

Contoh Normalisasi 1NF

Suatu tabel dikatakan 1NF jika dan hanya jika setiap atribut dari data tersebut hanya memiliki nilai tunggal dalam satu baris.

1NF

kode_faktur	tanggal	kode_barang	nama_barang	harga	qty
KD_001	13/07/2020	BRG_001	Indomie Goreng	2500	5
KD_001	13/07/2020	BRG_002	Indomie Goreng Jumbo	3000	8
KD_001	13/07/2020	BRG_003	Mie Sedaap Goreng	2500	9
KD_001	13/07/2020	BRG_004	Mie Sedaap Soto	2300	3
KD_002	13/07/2020	BRG_005	Intermie Goreng	1500	6
KD_002	13/07/2020	BRG_006	Intermie Soto	1500	6
KD_003	14/07/2020	BRG_007	Pop Mie Ayam	4500	3

Contoh Normalisasi 2NF

Syarat 2NF adalah tidak diperkenankan adanya partial “functional dependency” kepada primary key dalam sebuah tabel.

2NF

Tabel Barang

kode_barang	nama_barang	harga
BRG_001	Indomie Goreng	2500
BRG_002	Indomie Goreng Jumbo	3000
BRG_003	Mie Sedaap Goreng	2500
BRG_004	Mie Sedaap Soto	2300
BRG_005	Intermie Goreng	1500
BRG_006	Intermie Soto	1500
BRG_007	Pop Mie Ayam	4500

Tabel Transaksi

kode_faktur	tanggal	kode_barang	qty
KD_001	13/07/2020	BRG_001	5
KD_001	13/07/2020	BRG_002	8
KD_001	13/07/2020	BRG_003	9
KD_001	13/07/2020	BRG_004	3
KD_002	13/07/2020	BRG_005	6
KD_002	13/07/2020	BRG_006	6
KD_003	14/07/2020	BRG_007	3

Contoh Normalisasi 3NF

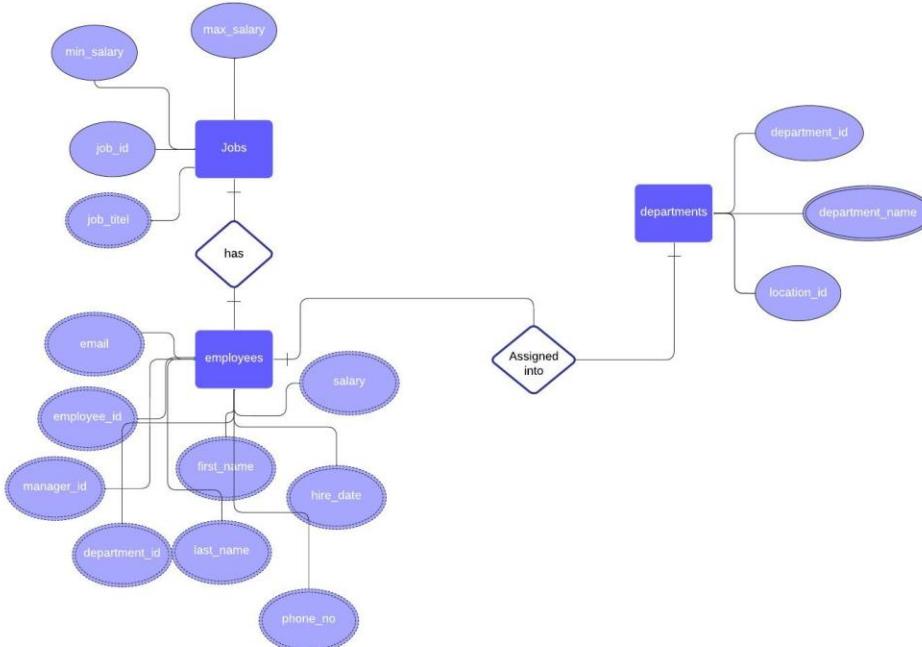
Pada 3NF tidak diperkenankan adanya partial “transitive dependency” dalam sebuah tabel.

3NF			
Tabel Barang		Tabel Transaksi	
kode_barang	nama_barang	harga	kode_faktur
BRG_001	Indomie Goreng	2500	KD_001
BRG_002	Indomie Goreng Jumbo	3000	KD_002
BRG_003	Mie Sedaap Goreng	2500	KD_003
BRG_004	Mie Sedaap Soto	2300	
BRG_005	Intermie Goreng	1500	
BRG_006	Intermie Soto	1500	
BRG_007	Pop Mie Ayam	4500	

Tabel Detail Barang			
kode_faktur	kode_barang	qty	harga
KD_001	BRG_001	5	2500
KD_001	BRG_002	8	2500
KD_001	BRG_003	9	2500
KD_001	BRG_004	3	2500
KD_002	BRG_005	6	3000
KD_002	BRG_006	6	3000
KD_003	BRG_007	3	2500

Quiz

Buatlah ER sederhana untuk use case HR Database pada tabel Employee, Job, dan Department.



Pengenalan Misi

Misi dapat diakses di: <https://academy.dqlab.id/>