

Assurance of Learning
Database Technology (COMP6799001)
Normalization and Entity Relationship Diagrams



Disusun oleh
Kelompok 1 - LM01
Bryan Tanjung - 2602117720
Elvan Nicola - 2602113256
Jakendra Lathaniel Sulistijo - 2602121106
Raissa Raffi Darmawan - 2602177146

Chapter 1: Description of The Case Used

Dataset tersebut merupakan kumpulan data yang dimiliki oleh OSHA (Occupational Safety and Health Administration), diambil dari link berikut:

<https://data.world/us-dol-gov/17697573-9de0-4886-aeab-78acf14fca3e>.

OSHA melakukan inspeksi terhadap keselamatan pekerjaan. Setiap data yang ada OSHA memberikan alasan atau dorongan mengapa melakukan inspeksi. Setiap kejadian kecelakaan atau pelanggaran standar OSHA diberikan sanksi dan denda yang diberikan. Dataset ini juga mencakup informasi investigasi kecelakaan. Ini mencakup deskripsi teks dari kecelakaan, dan spesifik tentang cedera dan kematian yang terjadi sebagai akibat dari kecelakaan tersebut.

Chapter 2: Problem Identification

A. Mengidentifikasi dan Mengatasi Bahaya di Tempat Kerja

Database yang kita buat dapat digunakan untuk mengidentifikasi bahaya yang sering terjadi di tempat kerja dan pelanggaran keamanan yang ada pada lapangan, kemudian dapat kita gunakan juga untuk mengembangkan strategi yang ditargetkan untuk meningkatkan keselamatan pekerja. Misalnya, jika database menunjukkan bahwa industri tertentu memiliki tingkat pelanggaran yang tinggi terkait dengan perlindungan pekerja akan cedera atas jatuh dari tempat tinggi, kita dapat mengembangkan program pelatihan atau regulasi keamanan yang ditargetkan untuk menurunkan kemungkinan terjadinya insiden tersebut.

B. Mengedukasikan Pekerja dan Pemberi kerja

Database dapat digunakan untuk mengembangkan materi pendidikan untuk pekerja dan pemberi kerja tentang keselamatan di tempat kerja. Misalnya, pemberi kerja dapat menggunakan database untuk mengembangkan fakta tentang bahaya umum dan cara mencegahnya.

C. Mengatur Volume Data yang Besar

Kita memerlukan aplikasi database karena data yang kita miliki terlalu banyak dan kompleks. Untuk mengatur dan memastikan data yang kita miliki dapat digunakan dengan efisien, kita memerlukan database agar penambahan, pengurangan, maupun pencarian data dapat dilakukan dengan lancar dan tanpa adanya kerusakan data.

Chapter 3: The sample data description and list of anomalies may occur.

Suatu anomali merupakan masalah-masalah yang ada pada database, biasanya terjadi apabila suatu tabel data memiliki redundansi data atau data yang berulang-ulang kali muncul dalam salah satu kolom database. Bila dilihat dari dataset yang kita pilih, tabel tersebut memiliki beberapa anomali yang disebabkan oleh redundansi data seperti pada kolom **project_cost**, **project_type**, dan **event_keyword**. Hal ini dapat menyebabkan beberapa masalah yang berkaitan dengan anomali-anomali yang ada pada database. Berikut merupakan anomali-anomali yang ada pada dataset tersebut:

- **Insertion Anomaly**

Anomali ini terjadi ketika kita akan memasukkan data baru pada suatu kolom pada dataset, namun pada saat kita memasukkan data tersebut kita diharuskan juga untuk memasukkan data-data pada baris yang sama karena data tersebut di boleh dikosongkan (NOT NULL). Hal ini menjadi sulit untuk dilakukan karena kita belum mengetahui data apa yang akan dimasukkan pada kolom-kolom tersebut.

Sebagai contoh kita akan memasukkan data baru pada dataset yang kita pilih, ketika kita memasukkan data baru pada kolom **project_type**, tetapi kita belum mengetahui data apa yang akan dimasukkan pada kolom **event_date**, **event_time**, **event_desc**, **event_keyword**, dan lain-lain. Sehingga kita masukkan data NULL sementara, maka hal ini merupakan salah satu anomali Insertion yang ada pada dataset tersebut.

event_date	event_time	event_desc	event_keyword	project_cost	project_type
16-AUG		Employ	EXCAVATI	B	A

-07		ee killed when struck by manhol e ring after cave-in	ON,WORK RULES,CO NSTRUCTI ON,TRENC H,STRUCK BY,CAVE-I N		
30-JUL- 07		Forklift Overtur n Fractur es Employ ee's Head	FRACTURE ,ROLL-OVE R,ROPS,IN DUSTRIAL TRUCK,SE AT BELT,FACE		A
NULL	NULL	NULL	NULL		Z (Inserted Data)

- **Updation Anomaly**

Anomali ini terjadi ketika kita ingin mengubah suatu data pada kolom tertentu yang saling berhubungan, seperti pada kolom ID dengan kolom lainnya. Namun, data lainnya dengan ID yang sama tidak akan ikut berubah dan hanya akan mempengaruhi data yang pertama kali diubah. Hal ini mengharuskan kita untuk mengubah setiap data yang kita telah kita ubah dengan ID yang sama. Jika tidak dilakukan maka akan terjadi ketidak konsistenan pada dataset.

Contohnya, Kita asumsikan bahwa terjadi kesalahan data pada salah satu kolom event_desc dengan report_id 1054112, maka kita dapat melakukan perubahan pada data tersebut. Namun, setelah kita ubah, **event_desc** dengan **report_id** 1054112 lainnya akan memiliki event_desc yang belum berubah sehingga terjadi ketidakkonsistenan data.

summary _nr	report _id	event_d ate	event_ti me	event_des c	event_keywo rd
20163704 8	105411 2	16-AUG -07 04.45.00 .000000 PM		Employee slip by a banana peel (UPDATE D)	EXCAVATIO N,WORK RULES,CON STRUCTION ,TRENCH,ST RUCK BY,CAVE-IN
20163707 1	105411 2	20-AUG -07 09.30.00 .000000 AM		Employee Fractures Legs in Fall	FRACTURE, UNGUARDE D FL OPENING,C ONSTRUCTI ON,FALL,FA LL PROTECTIO N,FLOOR OPENING,E LEVATOR SHAFT,LEG
20163708 9	105411 2	25-AUG -07 04.00.00		Employee Is Killed, Another Is	AIR PRESSURE, ABRASION,

		.000000 PM		Injured by Generator	CONTUSIO N,FRACTUR E,MAINTEN ANCE,COLL APSE,LOCK OUT
--	--	---------------	--	-------------------------	---

- **Deletion Anomaly**

Anomali ini terjadi apabila kita ingin menghapus salah satu kolom data pada tabel tersebut. Namun, data yang kita hapus tersebut akan mempengaruhi data-data lain yang ada pada baris yang sama. Hal ini menyebabkan semua data pada baris tersebut akan ikut terhapus. Sehingga, kita akan kehilangan informasi-informasi penting yang seharusnya masih dapat digunakan nantinya.

Contohnya, misalkan kita ingin menghapus kolom dengan **summary_nr** 201637048, maka yang terjadi adalah kita juga akan menghapus seluruh data yang ada pada baris tersebut.

summary _nr	report_id	event_da te	event_ti me	event_de sc	event_keyw ord
20163704 8	1054112	16-AUG- 07	04.45.PM	Employee killed when struck by manhole ring after cave-in	EXCAVATI ON,WORK RULES,CO NSTRUCTI ON,TRENC H,STRUCK BY,CAVE-I N
20163707	1054112	20-AUG-		Employee	FRACTUR

1		07 09.30.00. 000000 AM		Fractures Legs in Fall	E,UNGUA RDED FL OPENING, CONSTRU CTION,FAL L,FALL PROTECTI ON,FLOOR OPENING, ELEVATOR SHAFT,LE G
---	--	---------------------------------	--	------------------------------	---

summary _nr	report_id	event_da te	event_ti me	event_de sc	event_keyw ord
DELET ED	*DELET ED*	*DELET ED*	*DELET ED*	*DELET ED*	*DELETED *
20163707 1	1054112	20-AUG- 07	09.30 AM	Employee Fractures Legs in Fall	FRACTUR E,UNGUA RDED FL OPENING, CONSTRU CTION,FAL L,FALL PROTECTI ON,FLOOR OPENING, ELEVATOR

					SHAFT,LE G
--	--	--	--	--	---------------

Berikut merupakan functional dependency untuk dataset kami:

- ProjectID -> ProjectCost, ProjectType, sic_list, ProjectState
- ReportID -> loadDate
- BuildingID -> BuildStories, NonbuildHT
- SummaryNo -> EventKeyword
- BuildingID -> BuildStories, NonbuildHT
- SummaryNo -> ReportID, EventID, EventTime, EventDesc, Fatality

Chapter 4: Implementation of Normalization (Ken)

A. Un-Normal Form

Secara singkat, dataset berbentuk Un-Normal Form adalah bentuk database yang belum sama sekali memenuhi syarat dan aturan tingkatan-tingkatan normalisasi yang ada.

Dataset yang kami pilih masih belum bersifat single-valued, partial dependency masih ada, dan transitive dependency masih ada.

B. First Normal Form (1NF)

Dalam bentuk 1NF, semua nilai (value) yang terdapat dari setiap kolom attribute dalam suatu tabel harus bersifat single-valued (atomic). Dalam kata lain, setiap masukkan baris data hanya boleh memiliki satu value untuk setiap kolom yang ada. Apabila terdapat suatu many-to-many relationship, maka semua datanya harus terpisah menjadi entry baru.

Dalam kasus dataset kami, pada kolom Event_Keyword menampung data yang bersifat multi-valued. Sebagai contoh, pada suatu entry, suatu kecelakaan dapat memiliki banyak keyword, maka malah menjadi tergabung.

Oleh karena itu, setiap entry harus terpisah menjadi entry baru dengan membuat baris baru untuk setiap entry..

C. Second Normal Form (2NF)

Dalam normalisasi ke bentuk 2NF, semua partial dependency harus dihilangkan dengan melakukan dekomposisi data. Partial dependency adalah ketika dalam suatu tabel terdapat composite primary key, namun terdapat kolom yang hanya dependent terhadap sebagian dari composite primary key tersebut.

Dalam dataset kami hingga 1NF, kami dapat melihat bahwa terdapat 2 candidate key, yakni “Summary no” dan composite primary key yang terdiri dari “Event_desc”, “Event_time”, “Event_date”, dan “Event_Keyword”. Lalu, composite primary key yang ada dapat menjadi primary key yang lebih bermakna dan lebih pasti.

D. Third Normal Form (3NF)

Dalam normalisasi ke bentuk 3NF, semua transitive dependency harus dihilangkan. Caranya adalah dengan memisahkan data atau kolom yang dapat menimbulkan transitive dependency tersebut.

Dalam dataset yang telah kami dekomposisi hingga 2NF, kami mendapatkan bahwa terdapat banyak kolom data yang sebenarnya dependent secara transitive ke “Summary number”. Oleh karena itu, segala kolom yang dependent terhadap “Summary number” dapat dipisahkan ke dalam tabel baru dengan “Summary number” sebagai primary key.

E. Boyce-Codd Normal Form (BCNF)

Dalam BCNF, dependency yang perlu dihapus adalah dependency di mana suatu primary key atau bagian dari primary key malah menjadi dependent terhadap attribute yang bersifat non-key.

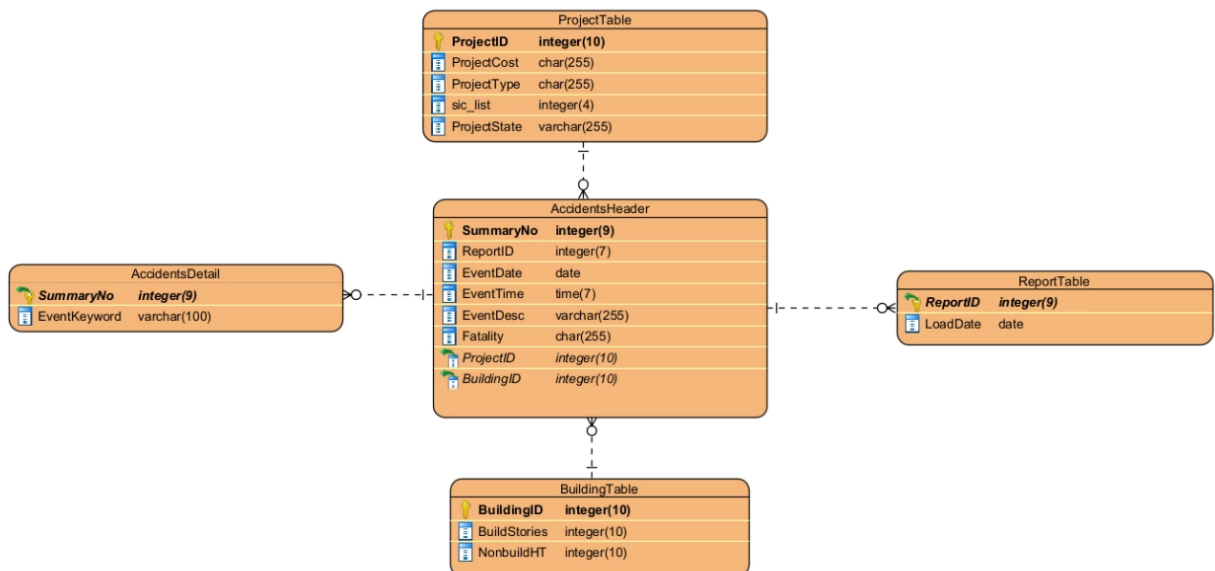
Dalam normalisasi dataset kami hingga 3NF, kebetulan tidak ada perubahan ketika memasuki BCNF karena modelnya sudah memenuhi syarat BCNF

Link sheets normalisasi hingga BCNF:

<https://docs.google.com/spreadsheets/d/1SXRw4Mp043KuYN7nsh7v96oGkqaz2tKJBjsyCybK4Po/edit?usp=sharing>

F. Entity Relationships Diagram (ERD)

Berikut merupakan Entity Relationships Diagram dari database model yang sudah kami normalisasikan hingga BCNF:



Dibuat dengan website Visual Paradigm via <https://www.visual-paradigm.com/>

Chapter 5: Conclusions

Dari dataset yang kami dapat mengenai kejadian kecelakaan di tempat kerja yang dipungut oleh OSHA (Occupational Safety and Health Administration), kasus tersebut mencakup alasan inspeksi, kejadian kecelakaan, pelanggaran standar OSHA, sanksi dan denda yang diberikan. Dataset tersebut juga memiliki informasi investigasi kecelakaan, deskripsi teks

dari kecelakaan dan detail tentang cedera dan kematian akibat kecelakaan. Perlu juga diketahui dari dataset tersebut diperlukannya dibentuk menjadi suatu database.

Database memudahkan untuk mengorganisir suatu dataset dengan lebih mudah dan efektif. Topik dataset ini adalah mengenai kecelakaan dan kesehatan tempat kerja. Sebab itu juga diperlukan suatu penyimpanan data yang aman dan memastikan data tersebut tidak adanya redundansi dan inkonsistensi. Dataset tersebut ada lebih dari ribuan data yang ada, tentu saja kita tidak mungkin bisa mengecek satu-satu dengan mudah. Salah satu kelebihan utama dari suatu database adalah mengatur jumlah data yang besar, database dapat membantu kita untuk penambahan, pengurangan, maupun pencarian data dengan mudah.

Alasan kita memerlukan dataset tersebut menjadi suatu database karena data ini dapat menyelamatkan orang lain dan mencegah suatu kecelakaan yang bisa terjadi. Database tersebut dapat digunakan untuk mengidentifikasi bahaya pada tempat kerja dan pelanggaran keamanan. Misalnya jika suatu industri memiliki masalah kecelakaan yang tinggi pada kasus tertentu, dapat dilakukan regulasi keamanan yang ditargetkan terhadap kejadian kecelakaan tersebut. Pekerja dan perusahaan juga akan terbantu dengan database ini. Sebab informasi yang ada dapat digunakan sebagai bahan materi untuk edukasi.

Sebelum dataset tersebut dijadikan menjadi database, diperlukan adanya identifikasi anomali insert, delete dan update. Sebab anomali tersebut akan memberikan suatu permasalahan seperti redundansi data. Anomali insertion terjadi saat memasukkan data baru ke dataset, tetapi harus memasukkan data lain pada baris yang sama karena tidak boleh kosong. Anomali update terjadi saat mengubah data pada kolom yang saling berhubungan, seperti ID dengan kolom lainnya. Data dengan ID yang sama tidak ikut berubah dan hanya mempengaruhi data yang pertama kali diubah. Jika tidak dilakukan, maka akan terjadi ketidak konsistenan pada dataset. Delete anomaly terjadi saat ingin menghapus salah satu kolom data pada tabel, tetapi data yang dihapus akan mempengaruhi data lain pada baris yang sama.

Normalisasi adalah suatu proses untuk mengorganisir suatu dataset. Normalisasi akan membagi tabel yang lebih besar menjadi lebih kecil dan menghubungkan mereka menggunakan relationship. Proses tersebut adalah hal yang penting dalam database untuk mencegah hal-hal yang tidak diinginkan. Proses ini juga dapat membuat database menjadi lebih efisien dan meningkatkan integritas database. Dalam dataset tersebut dilakukan normalisasi sampai BCNF.

Suatu dataset akan disebut 1NF jika mengandung nilai atomik. Untuk dijadikan menjadi 2NF data sebelumnya harus menjadi 1NF terlebih dahulu dan semua atribut non-kunci sepenuhnya functional dependency pada kunci primer. Sama seperti sebelumnya bentuk data harus 2NF dan tidak ada transitive dependency yang ada. BCNF memastikan bahwa setiap penentu dari suatu relasi adalah kunci kandidat, yang berarti semua functional dependency non-trivial telah dieliminasi.

Jadi dari proses data yang telah dilakukan, dapat diketahui bahwa suatu database adalah hal yang penting bagi dataset. Suatu database akan lebih rapi dan mudah dilihat dengan dilakukannya normalisasi terlebih dahulu. Normalisasi akan menghilangkan masalah anomali insert, update, dan deletion. Hal tersebut penting bagi dataset yang kami pakai yaitu, data kecelakaan yang terjadi di tempat kerja. Dengan proses yang dilakukan, dataset ini diharapkan dapat membantu proses penggunaan dataset ini lebih baik dengan cepat.