

Normalization

- Normalization is a technique which is been used to remove the redundancy (or we can say the duplicates) from a table.
- If we are not able to remove the duplicacy completely from a table, so it gives us the feasibility to reduce that duplicacy to an extent.

* There are basically two types of duplicacy in a table →

- (1) Row level
- (2) Column level

So, let's start with row level duplicacy, in which we will take data for a student →

M T W T F S S							M T W T F S S							M T W T F S S						
M T F S S		M T W T F S S		M T W T F S S		M T W T F S S		M T W T F S S		M T W T F S S		M T W T F S S		M T W T F S S		M T W T F S S		M T W T F S S		
12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	•	•	•	•
1																				

Primary Key

~~Ex →~~

	SID	Sname	Age
10	1	Rutvi	22
11	2	Ram	20
12	3	Ravi	19
1	1	Rutvi	22

So, in the above table we can see that the data of row 1 and row 4 have the exact same values.
i.e., ~~duplicacy~~.

So, in order to remove this row duplicacy we have a concept of "Primary key", in which we set any one attribute of our table as primary key.

So, here we can make SID as primary key.

2022

January • Tuesday

Day 004-361

04

and Properties of Primary Key are that it should be (Unique + Not Null)

↓
Values won't be repeated + (neither we can leave this empty)

So, this is how it will take care of the row level duplicacy.

* Now, let's talk about Column level duplicacy.

Ex → [sid, Sname, Cid, Cname, fid, Fname, Salary] Columns of our table.

Where, sid → student id

Sname → Student name

Cid → Course id

Cname → Course name

fid → Faculty id

Fname → Faculty name

Salary → Salary

S	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	FEB
																				2022

SID	Sname	Grd	Gname	Fid	Fname	Salary
1	Rutvi	C ₁	DBMS	F ₁	John	30000
2	Ram	C ₂	JAVA	F ₂	Bob	30000
3	Ravi	C ₁	DBMS	F ₁	John	30000
4	Monal	C ₁	DBMS	F ₁	John	30000
5	:	:	:	:	:	:
6	:	:	:	:	:	:
7	:	:	:	:	:	:

- As, it is quite evident from the table above that there is no row duplication, since we have already removed it, by setting 'SID' as our Primary key.
- But, we can see that there are columns which are quite similar to each other.
- Now, as we know that when the data is been added to a table it is # in thousands, lakhs, and what

DEC	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S
2021	•	•	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

2022

January • Thursday

Day 006-359

06

if there are columns with similar data, how does it even create a problem?

So, there are 3 types of Problems which can arise in case of column duplicacy.

- (a) Insertion Anomaly
- (b) Deletion Anomaly
- (c) Updation Anomaly

(These three are the operations used in database)

* [Anomaly is basically a type of problem, but a problem which arises at some special occasions.]

So, let's begin with insertion anomaly →

Suppose, I want to enter the data of a new student in my database, it will done easily.

Eg → SID	Sname
1	Varun
5	Varun

S	T	W	F	S	S	M	T	W	F	S	S	M	T	W	F	S	S
16	17	18	19	20		21	22	23	24	25	26	27	28	•	•	•	•

FEB
2022

Now, suppose there was a new course introduced in the university whose

Sid	Sname	CourseID	Cname
1	;	!	!
D ↓ 12		C10	Computer

- So, in here this data can't be directly inserted into the table since, we haven't talked about Sid and Sname.

When there is no specific SID for this where this will be inserted, and since SID is a Primary key we can't keep it null and it must be unique as well.

- So, in case of deletion anomaly, simply we are asked to delete the data for a student whose roll no (SID) is 1.

so, by simply writing a query for deletion we can do that →

DEC	M	T	W	T	F	S	S	M	T	W	F	S	S	M	T	W	F	S
2021	•	•	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

2022

(Table name)

January • Saturday

Day 008-357

08

"Delete from Student where SID = 1 ;"
 ↓
 (Column name)

So, this will completely remove the whole row where SID = 1.

So, now in case of updation anomaly suppose, I want to update Sname from Monal to Monal Thakkar where SID is 4.

So the query for this would be →

"Update Student Set Sname = "Monal Thakkar" Where SID = 4 . "

This would work fine for the updation.

Now, let's take another scenario for updation.

Sunday 09

Suppose we want to update the salary of the faculty whose FID = F₁.

We want to change it from 30,000 to 40,000.

M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	FEB
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	2022

10

Monday • January

Day 010-355

2022

So, now when try to update the salary all the values of FID, whose value is F1 will be updated.

10

11

12

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

2022

January - Tuesday,

Day, 011-354

inserted into table IInd without
affecting the data of other tables.

9 Similarly, if we delete a student's record
10 the other records like for
11 course and faculty still remains the
unaffected.

12 And same happens at the time of
1 update since FID is the Primary
1 key in Table IIIrd, no entries would
2 be repeated and salary can be
easily changed.