# Functional Dependencies & Normalization for Relational DBs

Chapter 6

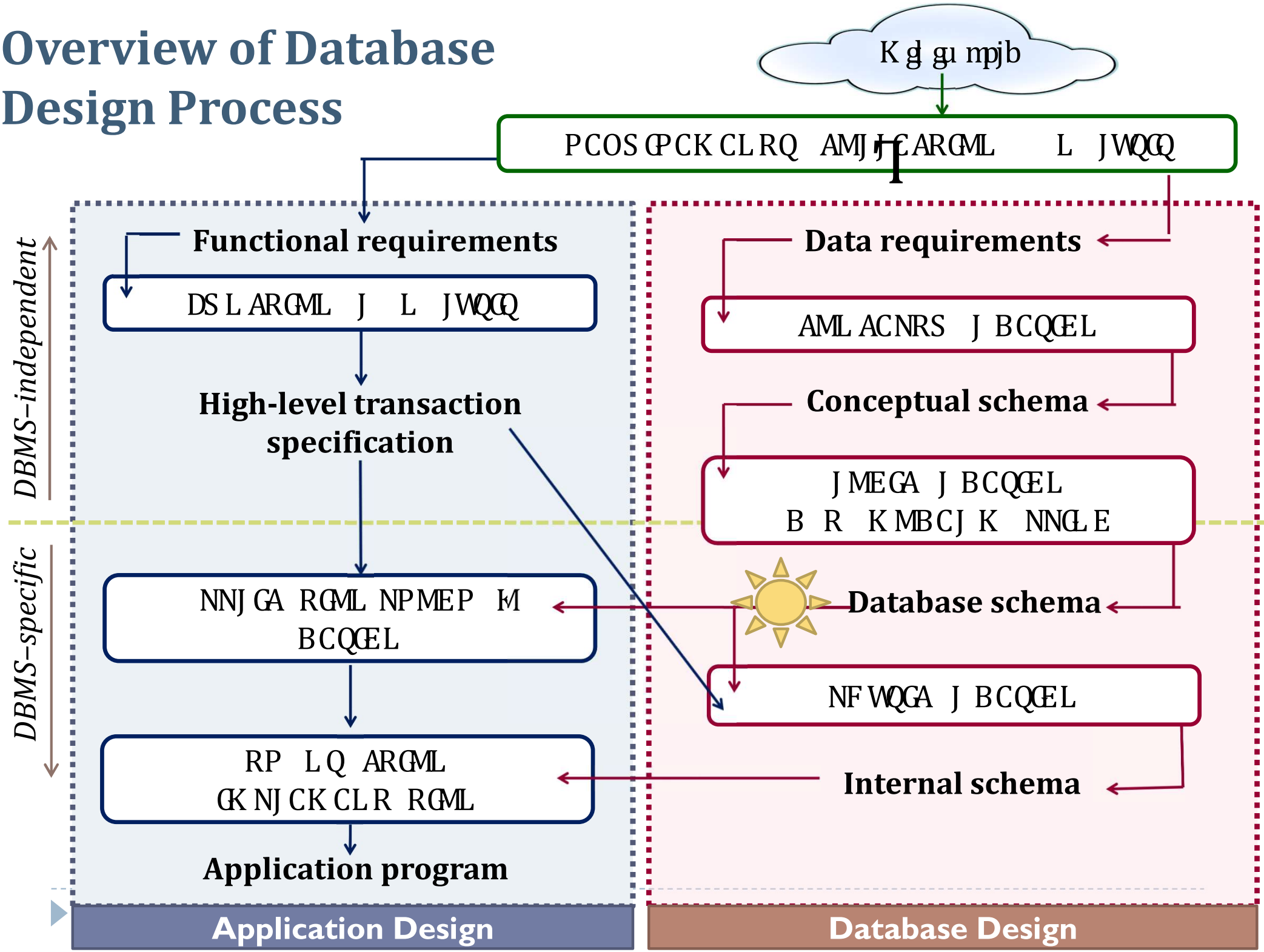
# Contents



# Contents



# Overview of Database Design Process



# Introduction

- C af pcj rgml qafck aml qggprq md lsk cp md rrpg srcq lb rfc pcj rgml jb r qc qafck aml qggprq md lsk cp mdpcj rgml qafck q
- rrpg srcq pc epms ncb rmdmpk pcj rgml qafck
- L ccb qmk c dmpk jk c qs pc mdu f wml c epms ngl e md rrpg srcq gl rm pcj rgml qafck k w c crrcp rf l l mrf cp

# Introduction

- Emmbl cqq k c qs pcq
  - K i gl e qs pc rf r rf c qck l rgaq mdrf c rrpg s rcq gq ajc p gl rf c qaf ck
  - Pcbs l b l r gl dmpk rgml gl rs njcq l b S nb rc l mk jgcq k mbgdga rgml bcjcrgml gl qc prgml
  - Pcbs agl e rf c L S J J t js cq gl rs njcq
  - Bgq jjmu gl e rf c nmqqg gjgrwmdecl cp rgl e qns pgms q rs njcq

# Introduction

- **Redundant information** gl rsnjcq rf c rrpg src t jscq ncpr gl gl e rm n prgasj p bcn prk cl r BLSK CP BL KC BKEPQQL pc pcnc rcb dmpct cpwck njmwcc u f mu mpi q dmprf r bcn prk cl r

EMP_DEPT

| Ename | Ssn | Bdate | Address | Dnumber | Dname | Dmgr_ssn |
|---|---|---|---|---|---|---|
| Smith, John B. | 123456789 | 1965-01-09 | 731 Fondren, Houston, TX | 5 | Research | 333445555 |
| Wong, Franklin T. | 333445555 | 1955-12-08 | 638 Voss, Houston, TX | 5 | Research | 333445555 |
| Zelaya, Alicia J. | 999887777 | 1968-07-19 | 3321 Castle, Spring, TX | 4 | Administration | 987654321 |
| Wallace, Jennifer S. | 987654321 | 1941-06-20 | 291 Berry, Bellaire, TX | 4 | Administration | 987654321 |
| Narayan, Ramesh K. | 666884444 | 1962-09-15 | 975 FireOak, Humble, TX | 5 | Research | 333445555 |
| English, Joyce A. | 453453453 | 1972-07-31 | 5631 Rice, Houston, TX | 5 | Research | 333445555 |
| Jabbar, Ahmad V. | 987987987 | 1969-03-29 | 980 Dallas, Houston, TX | 4 | Administration | 987654321 |
| Borg, James E. | 888665555 | 1937-11-10 | 450 Stone, Houston, TX | 1 | Headquarters | 888665555 |

# Introduction

- **Update anomalies** k mbgdga rgml bcjcrgml gl qcprgml
  - K mbgdga rgml
    - qrf c k l ecpmd bcnr af l ecqu c f tc rmsnb rc k l w t jscq aampbgl e rmck njmwccqu mpi gl e dmprf r bcnr
    - C qwrmk i c rf c B **inconsistent**

EMP_DEPT

| Ename | Ssn | Bdate | Address | Dnumber | Dname | Dmgr_ssn |
|---|---|---|---|---|---|---|
| Smith, John B. | 123456789 | 1965-01-09 | 731 Fondren, Houston, TX | 5 | Research | 333445555 |
| Wong, Franklin T. | 333445555 | 1955-12-08 | 638 Voss, Houston, TX | 5 | Research | 333445555 |
| Zelaya, Alicia J. | 999887777 | 1968-07-19 | 3321 Castle, Spring, TX | 4 | Administration | 987654321 |
| Wallace, Jennifer S. | 987654321 | 1941-06-20 | 291 Berry, Bellaire, TX | 4 | Administration | 987654321 |
| Narayan, Ramesh K. | 666884444 | 1962-09-15 | 975 FireOak, Humble, TX | 5 | Research | 333445555 |
| English, Joyce A. | 453453453 | 1972-07-31 | 5631 Rice, Houston, TX | 5 | Research | 333445555 |
| Jabbar, Ahmad V. | 987987987 | 1969-03-29 | 980 Dallas, Houston, TX | 4 | Administration | 987654321 |
| Borg, James E. | 888665555 | 1937-11-10 | 450 Stone, Houston, TX | 1 | Headquarters | 888665555 |

# Introduction

- Bcjcrgml gdl mpe Hk cq C jc tcq u c bcjcrc f gqrs njc l b jmqc rf c cvggrgl e mdbcnr rf c l k c mdbcnr l b u f mgqrf c k l ecp mdbcnr

EMP_DEPT

| Ename | Ssn | Bdate | Address | Dnumber | Dname | Dmgr_ssn |
|---|---|---|---|---|---|---|
| Smith, John B. | 123456789 | 1965-01-09 | 731 Fondren, Houston, TX | 5 | Research | 333445555 |
| Wong, Franklin T. | 333445555 | 1955-12-08 | 638 Voss, Houston, TX | 5 | Research | 333445555 |
| Zelaya, Alicia J. | 999887777 | 1968-07-19 | 3321 Castle, Spring, TX | 4 | Administration | 987654321 |
| Wallace, Jennifer S. | 987654321 | 1941-06-20 | 291 Berry, Bellaire, TX | 4 | Administration | 987654321 |
| Narayan, Ramesh K. | 666884444 | 1962-09-15 | 975 FireOak, Humble, TX | 5 | Research | 333445555 |
| English, Joyce A. | 453453453 | 1972-07-31 | 5631 Rice, Houston, TX | 5 | Research | 333445555 |
| Jabbar, Ahmad V. | 987987987 | 1969-03-29 | 980 Dallas, Houston, TX | 4 | Administration | 987654321 |
| Borg, James E. | 888665555 | 1937-11-10 | 450 Stone, Houston, TX | 1 | Headquarters | 888665555 |

# Introduction

- Cl qcprgml
  - F mu a l u c apc rc bcn prk cl r cdmpc l wck njmwccq pc qqgel cb rmgr

EMP_DEPT

| Ename | Ssn | Bdate | Address | Dnumber | Dname | Dmgr_ssn |
|---|---|---|---|---|---|---|
| Smith, John B. | 123456789 | 1965-01-09 | 731 Fondren, Houston, TX | 5 | Research | 333445555 |
| Wong, Franklin T. | 333445555 | 1955-12-08 | 638 Voss, Houston, TX | 5 | Research | 333445555 |
| Zelaya, Alicia J. | 999887777 | 1968-07-19 | 3321 Castle, Spring, TX | 4 | Administration | 987654321 |
| Wallace, Jennifer S. | 987654321 | 1941-06-20 | 291 Berry, Bellaire, TX | 4 | Administration | 987654321 |
| Narayan, Ramesh K. | 666884444 | 1962-09-15 | 975 FireOak, Humble, TX | 5 | Research | 333445555 |
| English, Joyce A. | 453453453 | 1972-07-31 | 5631 Rice, Houston, TX | 5 | Research | 333445555 |
| Jabbar, Ahmad V. | 987987987 | 1969-03-29 | 980 Dallas, Houston, TX | 4 | Administration | 987654321 |
| Borg, James E. | 888665555 | 1937-11-10 | 450 Stone, Houston, TX | 1 | Headquarters | 888665555 |

# Introduction

- Pcbsagl e rf c LSJJ t jscqgl rsnjcq
  - Ck njmwccql mr qqgel cb rm l wbcnr u qrc rf c qrmp ec qn ac
  - Mrf cp bgdbgasjrgcq eepce rgml mncp rgml q c e AMSLR QSK l b hmgl q

# Introduction

- Bgq jjmu gl e rf c nmqqg gjgrwmdecl cp rgl e qnspgmsqrsnjcq

CK N NPMH QQL NLsk cp Fmspq CL k c NL k c NJma rgml

CK N JMAQ CL k c NJma rgml
CK N NPMH QQL NLsk cp Fmspq NL k c NJma rgml

- Ecl cp rgml mdgl t jgb l b qnspgmsqb r bspgl e HMCLQ NJma rgml gqrf c rrpg src rf r pcj rcq CK N JMAQ l b CK N NPMH l b NJma rgml gql cgrf cp npgk pwi cwl mp dmpcgel i cwgl cgrf cp CK N JMAQmp CK N NPMH

# Introduction

- Bgq jjmu gl e rf c nmqqg gigrwmdecl cp rgl e qns pgms q rs njcq

**EMP_LOCS**

| Ename | Plocation |
|---|---|
| Smith, John B. | Bellaire |
| Smith, John B. | Sugarland |
| Narayan, Ramesh K. | Houston |
| English, Joyce A. | Bellaire |
| English, Joyce A. | Sugarland |
| Wong, Franklin T. | Sugarland |
| Wong, Franklin T. | Houston |
| Wong, Franklin T. | Stafford |
| Zelaya, Alicia J. | Stafford |
| Jabbar, Ahmad V. | Stafford |
| Wallace, Jennifer S. | Stafford |
| Wallace, Jennifer S. | Houston |
| Borg, James E. | Houston |

**EMP_PROJ1**

| Ssn | Pnumber | Hours | Pname | Plocation |
|---|---|---|---|---|
| 123456789 | 1 | 32.5 | ProductX | Bellaire |
| 123456789 | 2 | 7.5 | ProductY | Sugarland |
| 666884444 | 3 | 40.0 | ProductZ | Houston |
| 453453453 | 1 | 20.0 | ProductX | Bellaire |
| 453453453 | 2 | 20.0 | ProductY | Sugarland |
| 333445555 | 2 | 10.0 | ProductY | Sugarland |
| 333445555 | 3 | 10.0 | ProductZ | Houston |
| 333445555 | 10 | 10.0 | Computerization | Stafford |
| 333445555 | 20 | 10.0 | Reorganization | Houston |
| 999887777 | 30 | 30.0 | Newbenefits | Stafford |
| 999887777 | 10 | 10.0 | Computerization | Stafford |
| 987987987 | 10 | 35.0 | Computerization | Stafford |
| 987987987 | 30 | 5.0 | Newbenefits | Stafford |
| 987654321 | 30 | 20.0 | Newbenefits | Stafford |
| 987654321 | 20 | 15.0 | Reorganization | Houston |
| 888665555 | 20 | NULL | Reorganization | Houston |

# Introduction

- Bgq jjmu gl e rf c nmqqg gigrwmdecl cp rgl e qns pgms q rs njcq

| | Ssn | Pnumber | Hours | Pname | Plocation | Ename |
|---|---|---|---|---|---|---|
| | 123456789 | 1 | 32.5 | ProductX | Bellaire | Smith, John B. |
| * | 123456789 | 1 | 32.5 | ProductX | Bellaire | English, Joyce A. |
| | 123456789 | 2 | 7.5 | ProductY | Sugarland | Smith, John B. |
| * | 123456789 | 2 | 7.5 | ProductY | Sugarland | English, Joyce A. |
| * | 123456789 | 2 | 7.5 | ProductY | Sugarland | Wong, Franklin T. |
| | 666884444 | 3 | 40.0 | ProductZ | Houston | Narayan, Ramesh K. |
| * | 666884444 | 3 | 40.0 | ProductZ | Houston | Wong, Franklin T. |
| * | 453453453 | 1 | 20.0 | ProductX | Bellaire | Smith, John B. |
| | 453453453 | 1 | 20.0 | ProductX | | |

# Introduction

- Emmbl cqq k c qs pcq
  - Pcbs l b l r gl dmpk rgml gl rs njcq
  - S nb rc l mk jgcq k mbggba rgml bcjcrgml gl qcprgml
  - Pcbs agl e rf c LS J J t js cq gl rs njcq
  - Bgq jjmu gl e rf c nmqqg gjgrwmdecl cp rgl e qns pgms q rs njcq

☞ **Normalization**

# Introduction

- L mpk jgx rgml f cjnq B bcqgel cpq bcrcpk gl c rf c cqr pcj rgml qaf ck q
  - dmpk jdp k cu mpi dmp l jwxgl e pcj rgml qaf ck q qcb ml rf cgp i cwq l b ml rf c ds l argml j bcncl bcl agcq k ml e rf cgp rrpg s rcq
  - qcpgcq mdl mpk jdmpk rcqrq rf r a l c a ppgcb ms r ml gl bgt gbs j pcj rgml qaf ck q qmrf r rf c pcj rgml j b r qc a l c l mpk jgxcb rm l wbcqgpcb bcepcc
- Gr gq qcb ml rf c aml acnr mdl mpk jdmpk LD LD LD ALD LD LD
- Gr gq npmacqq u f gaf cl qs pcq rf r rf c b r gq qrps ars pcb gl qs af u wrf r rrpg s rcq pc epms ncb u gf rf c NI rrpg s rcq rf r b ml mr bgpc arjwbcncl b ml NI k w c cvrp arcb rmdmpk l cu pcj rgml

# Introduction

- Rf cpc pc ru mgk nmpr l r npmmcprgcq mdbcamk nmqgrgml q

  l ml bbgrgqc mp jmqqjcqql cqq mdrf c ampcqnml bgl e hmgl
  npcqcpt rgml mdrf c dsl argml j bcncl bcl agcq

- L mrc rf r npmmcprw gq cvrpck cjwgk nmpr l r l b a l l mr
  c q apggbacb Npmmcprw gq jcqq qrpgl ecl r l b k w c
  q apggbacb

# Contents



# Functional Dependencies (FDs)

- **Definition of FDs**
- Bgpcar gl bgpcar n prg j bcncl bcl agcq
- Cl dc pcl ac Ps jcq dmp DBq
- Cos gt jcl ac mdQcrq mdDBq
- K gl gk j Qcrq mdDBq

# Functional Dependencies (FDs)

- Dslargml jbcncl bclagcq DBq pc sqcb rmqncagdw dmpk jk c qspcq mdrfc emmblcqq mdpcj rgml j bcqgelq
- DBq lbicwq pc sqcb rmbcdglc lmpk jdmpk qdmp pcj rgml q
- DBq pc aml qrp gl rqrf r pc bcpgtcb dpmk rfc k c lgle lbgl rcppcj rgml qf gnq mdrfc b r rrpg srcq
- qcr md rrpg srcqV **<u>functionally determines</u>** qcr md rrpg srcqWgdrfc t jsc mdV bcrcpk gl cq slgpsc t jsc dmpW

**X → Y**

# Functional Dependencies (FDs)

- V → Wf mjbqgdu f cl ct cp ru mrs njcq f tc rf c q k c t jsc dmp
  V rf cwk s qr f tc rf c q k c t jsc dmp W
- Dmp l wru mrs njcq r l b r gl l wpcj rgml gl qr l ac p P
  Gdr V r V rf cl r W r W
- V → Wgl P qncagdgcq aml qrp gl r ml jj pcj rgml gl qr l acq p P
- Cv k njcq
  - qmag j qcas pgrwl s k cp bcrcpk gl cq ck njmwcc l k c
    QQL → CL k c
  - npmhcar l s k cp bcrcpk gl cq npmhcar l k c l b jma rgml
    NL s k cp → NL k c NJ ma rgml
  - ck njmwcc qql l b npmhcar l s k cp bcrcpk gl cq rf c f ms pq ncp
    u cci rf r rf c ck njmwcc u mpi q ml rf c npmhcar
    QQL NL s k cp → F ms pq

# Functional Dependencies (FDs)

- GdI gq icwmdP rfcl I dslargml jjwbcrcpkglcq jj rrpg srcqgl P qgl acuclctcpf tcrumbggrglarrsnjcq ugrf r I r I

# Functional Dependencies (FDs)

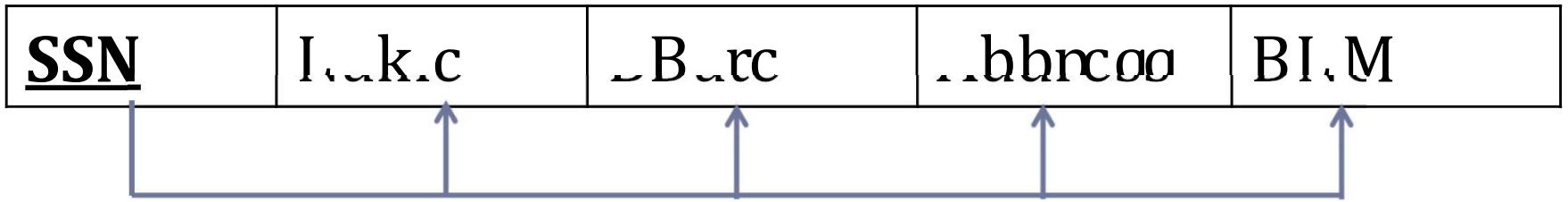
- Bcdgl grgml mdDBq
- **Direct, indirect, partial dependencies**
- Gldcpclac Psjcqdmp DBq
- Cosgt jclac mdQcrq mdDBq
- Kglgk jQcrq mdDBq

# Functional Dependencies (FDs)

- **Direct and fully dependency** jj rrpg srcqgl P k s qr c ds jjwds l argml jjwbcncl bcl r ml rf c npgk pw i cw mp rf c NI gq bcrcpk gl l r md jj rrpg srcqgl P

  QQL → L k c B rc bbpcqq BL M

CK NJ MMCC



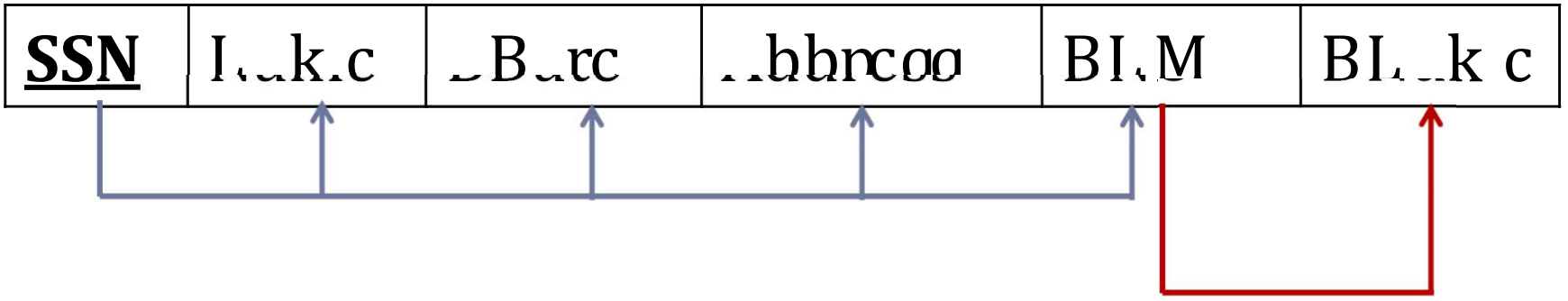

# Functional Dependencies (FDs)

- **Indirect dependency** T jsc md l rrpg src gql mr bcrcpk gl cb bgpcarjw wrf c npgk pwi cw

  BLM → BL k c

- QQL → Bl k c gq *transitive dependency* rf pms ef BLM

CK N BCNR

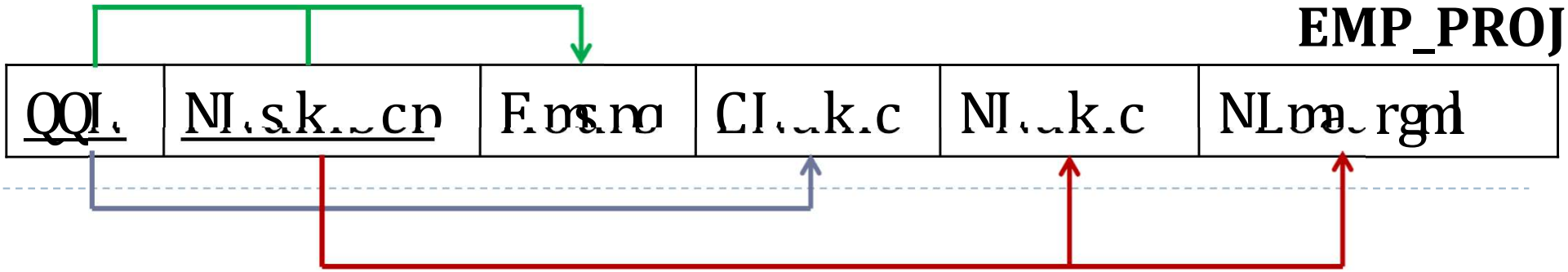


# Functional Dependencies (FDs)

- **Partial dependency**
  - Amk nmqgrc bcrcpk gl lr k mpcrf l ml c t jsc gqpcosgpcb rmbcrcpk gl c rf c t jsc md l mrf cp rrpg s rc rf c amk gl rgml mdt jscqgqa jjcb amk nmqgrc bcrcpk gl lr

    QQL NLsk cp gl CKN NPMH QQL NLsk cp → Fmspq
  - N prg j bcncl bcl aw gdrf c t jsc md l rrpg s rc bmcql mr bcncl b ml l cl rgpc amk nmqgrc bcrcpk gl lr s r ml jw n pr mdgr rf c pcj rgml qf gn gqi l mu l qrf c n prg j bcncl bcl aw

    QQL → CL k c  Nl s k cp → NL k c NJ ma rgml



# Functional Dependencies (FDs)

- Bcdgl grgml mdDB
- Bgpcar gl bgpcar n prg j bcncl bcl agcq
- **Inference Rules for FDs**
- Cosgt jcl ac mdQcrq mdDBq
- Kgl gk j Qcrq mdDBq

# Functional Dependencies (FDs)

- Egtcl qcrmdDBqDucalgldcp bbgrgml jDBqrf r fmjbufclctcprfcDBqgl Dfmjb
- **Armstrong's inference rules:**
  - **IR1. (Reflexive)** GdW⊆V rfcl V → W
  - **IR2. (Augmentation)** GdV → Wrfcl VX → WX

    Lmr rgml VXqr lbqdmpV∪X
  - **IR3. (Transitive)** GdV → W lbW→ X rfcl V → X

# Functional Dependencies (FDs)

- Qmk c bbgrgml j gl dcpcl ac ps jcq rf r pc s qcds j
  - **Decomposition:** GdV WX rf cl V W rf b V X
  - **Union:** GdV W rf b V X rf cl V WX
  - **Psuedotransitivity:** GdV W rf b U W X rf cl U V X
- Rf c j qr rf pcc gl dcpcl ac ps jcq qu cjj q l wm rf cp gl dcpcl ac ps jcq a l c bcbs acb dpmk CP CP l b CP amk njcrcl cqq npmmcprw
- **Armstrong's inference rules:**
  - **IR1. (Reflexive)** GdW⊆ V rf cl V → W
  - **IR2. (Augmentation)** GdV → W rf cl VX → WX
    L mr rgml VX qr l bq dmp V ∪ X
  - **IR3. (Transitive)** GdV → W l b W→ X rf cl V → X

# Functional Dependencies (FDs)

- Ajmqspc md qcr Dmd DBqggqrfc qcr **F+** md jj DBqrf r a l cgl dcppcb dpmk D
- Ajmqspc md qcr md rrpg srcqVugrf pcqncarrmDggqrfc qcr **X+** md jj rrpg srcqrf r pc dslargml jjwbcrcpkgl cb wV
- **X+** a l c a jasj rcb wpcnc rcbjw nnjwgl e CP CP CP sqgl e rfc DBqgl D

# Exercise

- Aml qgbcp pcj rgml *R A, B, C, D, E* u grf rf c dmjjmu gl e bcncl bcl agcq D

  → A

  AB → C

  BC →

- Dgl b B
- Dgl b D

# Functional Dependencies (FDs)

- Bcdgl grgml mdDB
- Bgpcar gl bgpcar n prg j bcncl bcl agcq
- Cl dcpcl ac Ps jcq dmp DBq
- **Equivalence of Sets of FDs**
- K gl gk j Qcrq mdDBq

# Functional Dependencies (FDs)

- Ru mqcrq mdDBq D l b E pc **equivalent** gd **$F^+$** **$G^+$**
- Bcdgl grgml
  - D **covers** E gd **$G^+ \subseteq F^+$**
  - D l b E pc **equivalent** gd D amt cpq E l b E amt cpq D
- Rf cpc gq l jempgrf k dmp af cai gl e cos gt jcl ac mdqcrq mdDBq

# Functional Dependencies (FDs)

- qcr mdDBqggk gl gk jgdgrq rgqdgcqrf c dmjjmu gl e aml bgrgml q
  - Ctcpwbcncl bcl awgl Df q qgl ejc rrpg src dmpgrq PFQ
  - U c a l l mr pck mtc l wbcncl bcl awdpmk D l b f t c qcr mdbcncl bcl agcqrf r gq cosgt jcl r rmD
  - U c a l l mr pcnj ac l wbcncl bcl awV→ gl Du gf bcncl bcl awW→ u f cpc Wnpmmcp qs qcr mdV⌐W qs qcr mdV⌐l b qrgjj f t c qcr mdbcncl bcl agcqrf r gq cosgt jcl r rmD

# Functional Dependencies (FDs)

- Ctcpwqcrmd DBqf q l cosgt jclrkglgk jqcr
- Rfcpca l cqctcp jcosgt jclrkglgk jqcrq
- Rfcpcgql mqgk njc jempgrfk dmpamk nsrgle kglgk j qcrmd DBqrf rgqcosgt jclrrm qcrDmd DBq
- Rmqwlrfcqgxc qcrmdpcj rgmlquc qqskcrf ruc qr prugrf qcrmdbcnclbclagcqrf rgq **minimal set**

# Contents



# Normalization

- **Normalization** Rf c npmacqq md **decomposing** s l q rgqd armpw b pcj rgml q w pc i gl e s n rf cgp rrpg s rcq gl rmqk jjcp pcj rgml q
- L mpk j dmpk S qgl e i cwq l b DBq md pcj rgml rm acprgdwu f crf cp pcj rgml qaf ck gq gl n prgas j p l mpk j dmpk
- L mpk jgx rgml gq a ppgc b ms r gl np arga c qmrf r rf c pcqs jrgl e bcqgel q pc mdf gef os jgrw l b k ccr rf c bcqgp jc npmn c prgcq
- Rf c b r qc bcqgel cpq **need not** l mpk jgxc rmrf c f gef cqr nmqqg jc l mpk j dmpk L D AL D mp L D

# Normalization

- Ru ml cu aml acnrq
  - **Prime attribute** k s qr c k ck cp mdqmk c a l bgb rc i cw
  - **Nonprime attribute** gq l mr npgk c rrpg s rc gr gq l mr k ck cp md l wa l bgb rc i cw

# Normalization

- L D l b bcncl bcl awnpm jck q
- L D qmjt cq n prg j bcncl bcl aw
- L D qmjt cq gl bgpcar bcncl bcl aw
- AL D u cjj l mpkr T jgxcb pcj rgml q

# Normalization

- **First normal form (1NF):** rf cpc gq ml jwml c t js c r rf c gl rc pqc argml mdc af pmu l b amjs k l md pcj rgml l mqcr t js cb rrpg s rcq gl L D

  → Bgq jjmu q amk nmqgrc rrpg s rcq k s jrgt js cb rrpg s rcq l b **nested relations**

- Rm c n pr mdrf c dmpk j bcdgl ggml md pcj rgml gl rf c qga dj r pcj rgml j k mbcj

# 1NF

(a)

**DEPARTMENT**

| Dname | Dnumber | Dmgr_ssn | Dlocations |
|---|---|---|---|

**(b)**

**DEPARTMENT**

| Dname | Dnumber | Dmgr_ssn | Dlocations |
|---|---|---|---|
| Research | 5 | 333445555 | {Bellaire, Sugarland, Houston} |
| Administration | 4 | 987654321 | {Stafford} |
| Headquarters | 1 | 888665555 | {Houston} |

**(c)**

**DEPARTMENT**

| Dname | Dnumber | Dmgr_ssn | Dlocation |
|---|---|---|---|
| Research | 5 | 333445555 | Bellaire |
| Research | 5 | 333445555 | Sugarland |
| Research | 5 | 333445555 | Houston |
| Administration | 4 | 987654321 | Stafford |
| Headquarters | 1 | 888665555 | Houston |

# 1NF

**EMP_PROJ**

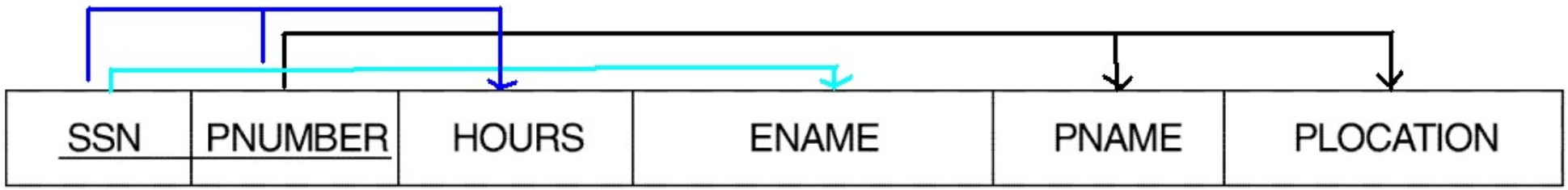
| SSN | ENAME | PROJS | |
|---|---|---|---|
| | | PNUMBER | HOURS |

**EMP_PROJ**

| Ssn | Ename | Pnumber | Hours |
|---|---|---|---|
| 123456789 | Smith, John B. | 1 | 32.5 |
| | | 2 | 7.5 |
| 666884444 | Narayan, Ramesh K. | 3 | 40.0 |
| 453453453 | English, Joyce A. | 1 | 20.0 |
| | | 2 | 20.0 |

# 1NF

**EMP_PROJ**



| SSN | PNUMBER | HOURS | ENAME | PNAME | PLOCATION |
|---|---|---|---|---|---|
| 123456789 | 1 | 32.5 | Smith,John B. | ProductX | Bellaire |
| 123456789 | 2 | 7.5 | Smith,John B. | ProductY | Sugarland |
| 666884444 | 3 | 40.0 | Narayan,Ramesh K. | ProductZ | Houston |
| 453453453 | 1 | 20.0 | English,Joyce A. | ProductX | Bellaire |
| 453453453 | 2 | 20.0 | English,Joyce A. | ProductY | Sugarland |
| 333445555 | 2 | 10.0 | Wong,Franklin T. | ProductY | Sugarland |
| 333445555 | 3 | 10.0 | Wong,Franklin T. | ProductZ | Houston |
| 333445555 | 10 | 10.0 | Wong,Franklin T. | Computerization | Stafford |
| 333445555 | 20 | 10.0 | Wong,Franklin T. | Reorganization | Houston |
| 999887777 | 30 | 30.0 | Zelaya,Alicia J. | Newbenefits | Stafford |
| 999887777 | 10 | 10.0 | Zelaya,Alicia J. | Computerization | Stafford |
| 987987987 | 10 | 35.0 | Jabbar,Ahmad V. | Computerization | Stafford |
| 987987987 | 30 | 5.0 | Jabbar,Ahmad V. | Newbenefits | Stafford |

# 1NF

CKNNPMH QQL NLsk cp Fmspq CL kc NL kc NJma rgml

QQL NLsk cp → Fmspq

QQL → CL kc

NLsk cp → NL kc NJma rgml

# 1NF

CKN NPMH QQL NLsk cp Fmspq

CKN QQL CL kc

NPMH NLsk cp NL kc NJma rgml

QQL NLSK CP → FMSPQ

QQL → CL KC

NLSK CP → NL KC NJMA RGML

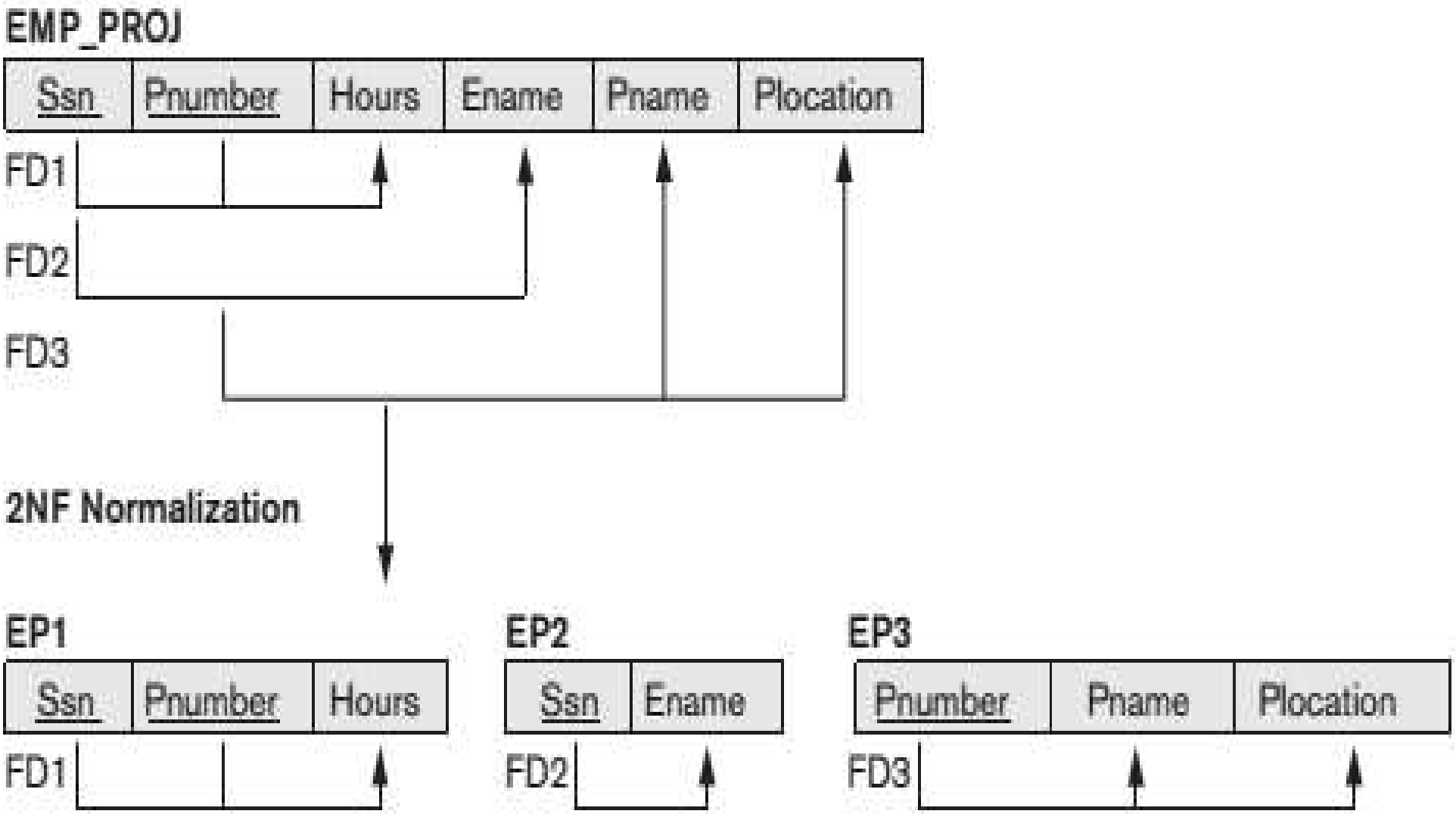
# Normalization

- L D l b bcncl bcl awnpm jck q
- **2NF – solves partial dependency**
- L D qmjt cq gl bgpcar bcncl bcl aw
- AL D u cjj l mpkr T jgxcb pcj rgml q

# Normalization

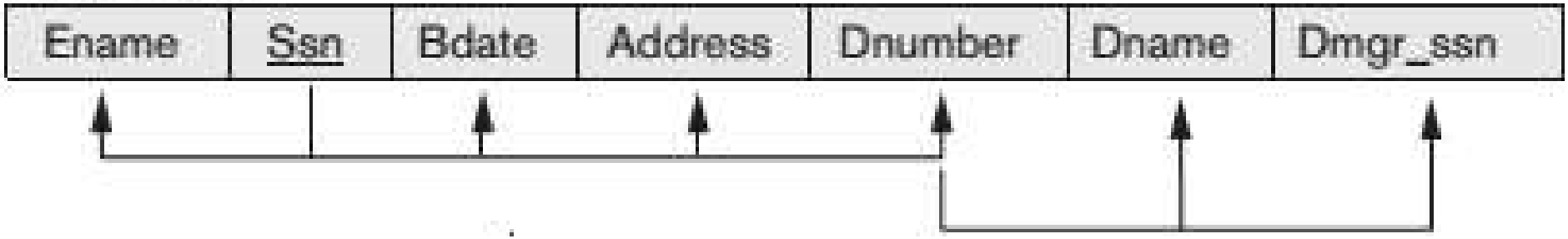
- Qcaml b l mpk j dmpk LD jj rrpg s rcq k s qr c ds jjw ds l argml jjwbcncl bcl r ml rf c npgk pwi cw
- LDqmjtcq n prg j bcncl bcl awnpm jck gl LD
- K crf mb gbcl rgdwnpgk pwi cwq l b epms n rrpg s rcq rf r pcj rc rmrf c i cwrme crf cp rmdmpk qcn p rc l cu pcj rgml q

# 2NF


EMP_PROJ
Ssn
Pnumber
Hours
Ename
Pname
Plocation
FD1
FD2
FD3
2NF Normalization
EP1
Ssn
Pnumber
Hours
FD1
EP2
Ssn
Ename
FD2
EP3
Pnumber
Pname
Plocation
FD3

# 2NF

EMP_DEPT




EMP_DEPT

➢**Problem with 2NF**

| Ename | Ssn | Bdate | Address | Dnumber | Dname | Dmgr_ssn |
|---|---|---|---|---|---|---|
| Smith, John B. | 123456789 | 1965-01-09 | 731 Fondren, Houston, TX | 5 | Research | 333445555 |
| Wong, Franklin T. | 333445555 | 1955-12-08 | 638 Voss, Houston, TX | 5 | Research | 333445555 |
| Zelaya, Alicia J. | 999887777 | 1968-07-19 | 3321 Castle, Spring, TX | 4 | Administration | 987654321 |
| Wallace, Jennifer S. | 987654321 | 1941-06-20 | 291 Berry, Bellaire, TX | 4 | Administration | 987654321 |
| Narayan, Ramesh K. | 666884444 | 1962-09-15 | 975 FireOak, Humble, TX | 5 | Research | 333445555 |
| English, Joyce A. | 453453453 | 1972-07-31 | 5631 Rice, Houston, TX | 5 | Research | 333445555 |
| Jabbar, Ahmad V. | 987987987 | 1969-03-29 | 980 Dallas, Houston, TX | 4 | Administration | 987654321 |
| Borg, James E. | 888665555 | 1937-11-10 | 450 Stone, Houston, TX | 1 | Headquarters | 888665555 |

# Normalization

- L D l b bcncl bcl awnpm jck q
- L D qmjt cq n prg j bcncl bcl aw
- **3NF – solves indirect dependency**
- AL D u cjj l mpk r jgxc b pcj rgml q

# Normalization

- pcj rgml qafck P gqlg **third normal form 3NF** gdgr gqlg LD *and* l ml ml npgk rc rrpg s rc gl P gq rp l qgrgt cjwbcncl bcl r ml rf c npgk pwi cw

**NOTE:**

Cl V → W l b W→ X u grf V qrf c npgk pwi cwu c aml qgbcp rf gq npm jck ml jwgdWgql mr a l bgb rc i cwU f cl Wgq a l bgb rc i cwrf cpc gql mnpm jck u grf rf c rp l qgrgt c bcncl bcl aw

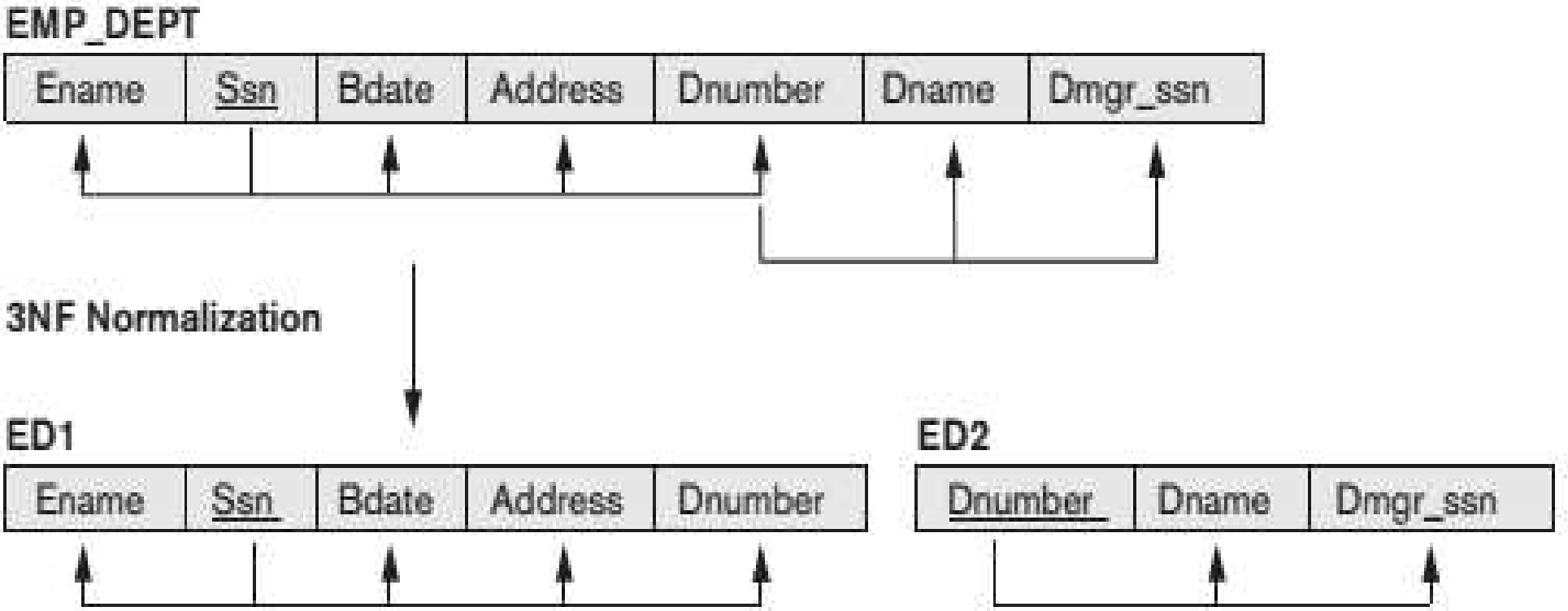
Ce Aml qgbcpCKN QQL Ck n Q j pw

Fcpc QQL → Ck n →Q j pw l b Ck n gq a l bgb rc i cw

# Normalization

- LDqmjtcqgl bgpcar rp lqggtc bcncl bclagcqnpm jck gl LD lb LD
- Kcrfmb gbcl rgdw jj rp lqggtc bcncl bclagcq lbc af rp lqggtc bcncl bclawu gjdmpk lcu pcj rgml ugrf lml npgkrc rrpg srcqn prgagn rgl e gl rfc rp lqggtc bcncl bclaw lbrfc rrpg src u fgaf bcrcpk gl cqmrfcpq qrfc rrpg srcqdmprfc lcu pcj rgml

# 3NF


EMP_DEPT
Ename
Ssn
Bdate
Address
Dnumber
Dname
Dmgr_ssn
3NF Normalization
ED1
Ename
Ssn
Bdate
Address
Dnumber
ED2
Dnumber
Dname
Dmgr_ssn

# Exercise

Aml qgbcprf c s l gtcpq j pcj rgml P *A, B, C, D, E, F, G, H, I, J*
l b rf c qcr mddslargml j bcncl bcl agcq

1. $A, B \rightarrow C$
2. $A \rightarrow D, E$
3. $B \rightarrow F$
4. $F \rightarrow G\ H$
5. $D \rightarrow I\ J$

U f r gq rf c i cwdmp *R?* Bcamk nmqc *R* gl rm L D rf cl L D
pcj rgml q

# Normalization

- **1NF and dependency problems**
- **2NF – solves partial dependency**
- **3NF – solves indirect dependency**
- AL D u cjj l mpk¬Tjgxcb pcj rgml q

# SUMMARY OF NORMAL FORMS based on Primary Keys

Summary of Normal Forms Based on Primary Keys and Corresponding Normalization

| Normal Form | Test | Remedy (Normalization) |
|---|---|---|
| First (1NF) | Relation should have no multivalued attributes or nested relations. | Form new relations for each multivalued attribute or nested relation. |
| Second (2NF) | For relations where primary key contains multiple attributes, no nonkey attribute should be functionally dependent on a part of the primary key. | Decompose and set up a new relation for each partial key with its dependent attribute(s). Make sure to keep a relation with the original primary key and any attributes that are fully functionally dependent on it. |
| Third (3NF) | Relation should not have a nonkey attribute functionally determined by another nonkey attribute (or by a set of nonkey attributes). That is, there should be no transitive dependency of a nonkey attribute on the primary key. | Decompose and set up a relation that includes the nonkey attribute(s) that functionally determine(s) other nonkey attribute(s). |

# General Normal Form Definitions

- Rf c mt c bcdgl grgml q aml qgbcp rf c npgk pwi cwml jw
- Rf c dmjjmu gl e k mpc ecl cp j bcdgl grgml q r i c gl rm aamsl r pcj rgml q u grf k s jrgnjc a l bgb rc i cwq

# General Normal Form Definitions

- pcj rgml qafck Pgqgl **second normal form (2NF)** gl ctcpwl ml npgk c rrpg src gl Pgql mrn prg jjwbcncl bcl r ml l wi cwmdP
- pcj rgml qafck Pgqgl **third normal form 3NF** gl ufcl ctcp DBV-> fmjbqgl P rfcl cgrfcp

  Vgq qsncpi cwmdP mp

  gq npgk c rrpg src mdP

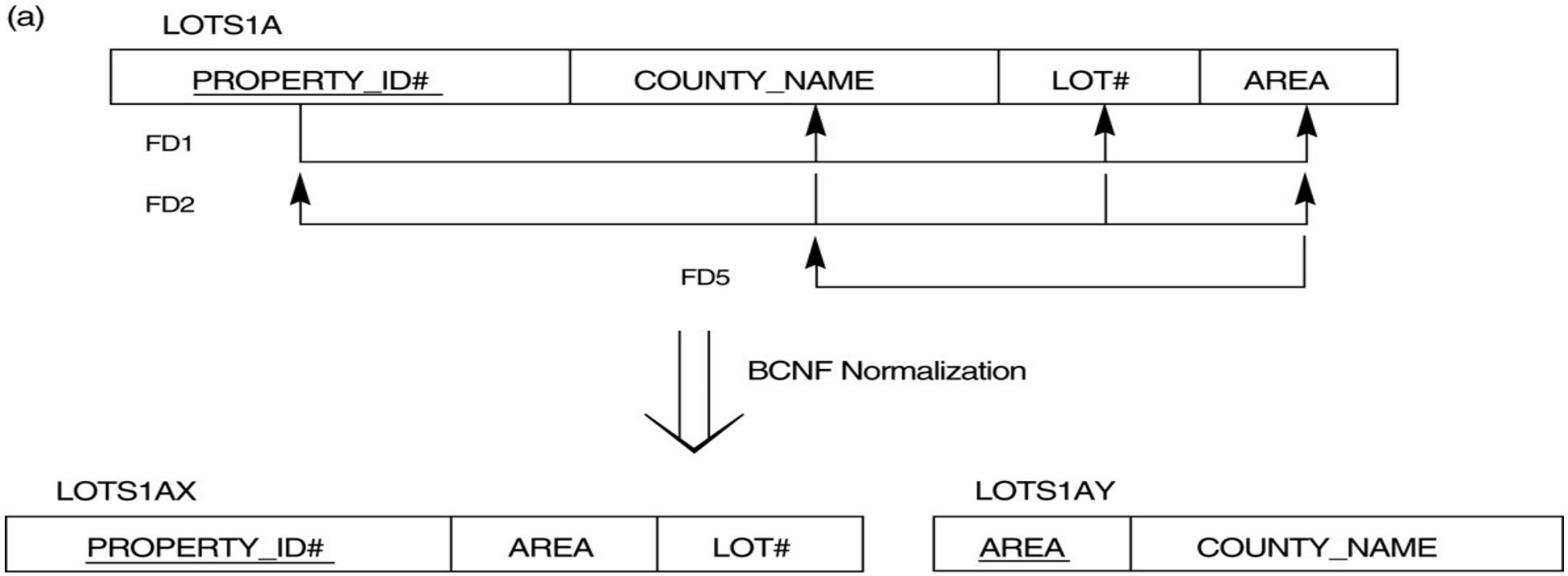
# Normalization

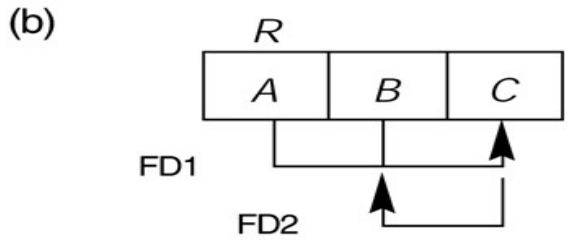
- L D l b bcncl bcl awnpm jck q
- L D qmjt cq n prg j bcncl bcl aw
- L D qmjt cq gl bgpcar bcncl bcl aw
- **BCNF – well-normalized relations**

# Normalization

- pcj rgml qafck P gqlg **Boyce-Codd Normal Form BCNF** gdu f cl ct cp l DB V -> f mjbqlg P rf cl V gq qs ncpi cwmdP

# BCNF


(a)
LOTS1A
PROPERTY_ID#
COUNTY_NAME
LOT#
AREA
FD1
FD2
FD5
BCNF Normalization
LOTS1AX
PROPERTY_ID#
AREA
LOT#
LOTS1AY
AREA
COUNTY_NAME


(b)
R
A
B
C
FD1
FD2

# BCNF

- RC AF Qrs bcl r Ams pqc Cl qrps armp

FD1: {Student, Course} → Instructor
FD2:[12] Instructor → Course

**TEACH**

| Student | Course | Instructor |
|---|---|---|
| Narayan | Database | Mark |
| Smith | Database | Navathe |
| Smith | Operating Systems | Ammar |
| Smith | Theory | Schulman |
| Wallace | Database | Mark |
| Wallace | Operating Systems | Ahamad |
| Wong | Database | Omiecinski |
| Zelaya | Database | Navathe |
| Narayan | Operating Systems | Ammar |

# BCNF

- Rf pcc nmqqg jc n gpq

1. {Student, Instructor} and {Student, Course}.
2. {Course, Instructor} and {Course, Student}.
3. {Instructor, Course} and {Instructor, Student}.

- jj rf pcc bcamk nmqggml q *lose the functional dependency FD1* Rf c bcqgp jc bcamk nmqggml mdrf mqc hs qr qf mu l gq ca s qc gr u gjj l mr ecl cp rc qns pgms q rs njcq dcp hmgl

# Nonadditive Join Decomposition into BCNF Schemas

**Algorithm 16.5.** Pcj rgml jBcamk nmqgrgml gl rm ALDu gf Lml bbgrgtc Hmgl Npmmcprw

**Input:** slgtcpq jpcj rgml $R$ lb qcrmddslargml j bcncl bclagcq $F$ ml rfc rrpg srcqmd $R$

**1.** Qcr $D$ $R$

**2.** Ufgjc rfcpc gq pcj rgml qafck $Q$ gl $D$ rf rggqlmrgl ALDbm

afmmqc pcj rgml qafck $Q$ gl $D$ rf rggqlmrgl ALD
dgl b dslargml jbcncl bclaw $X$ $Y$ gl $Q$ rf rtgmj rcq ALD
pcnj ac $Q$ gl $D$ wru mpcj rgml qafck q $Q$ $Y$ lb $X \cup Y$

# Contents




Q
A

# Exercise

Aml qgbcprf c s l gtcpq j pcj rgml P $A, B, C, D, E, F, G, H, I, J$

l b rf c qcr mdds l argml j bcncl bcl agcq

1) $A, B \rightarrow C$
2) $B, D \rightarrow E, F$
3) $A, D \rightarrow G, H$
4) $A \rightarrow I$
5) $H \rightarrow J$ Bcamk nmqc $R$ gl rm L D rf cl L D pcj rgml q

| | | |
|---|---|---|
| Second (2NF) | For relations where primary key contains multiple attributes, no nonkey attribute should be functionally dependent on a part of the primary key. | Decompose and set up a new relation for each partial key with its dependent attribute(s). Make sure to keep a relation with the original primary key and any attributes that are fully functionally dependent on it. |
| Third (3NF) | Relation should not have a nonkey attribute functionally determined by another nonkey attribute (or by a set of nonkey attributes). That is, there should be no transitive dependency of a nonkey attribute on the primary key. | Decompose and set up a relation that includes the nonkey attribute(s) that functionally determine(s) other nonkey attribute(s). |

# Exercise

- Aml qgbcprf c pcj rgml

  **BOOK (Book_Name, Author, Edition, Year)**

- qcb ml amk k ml qcl qcsl bcpqr l bgl e mdrf c b r u f r pc rf c nmqqg jc a l bgb rc i cwq mdrf gq pcj rgml

| Book_Name | Author | Edition | Copyright_Year |
|---|---|---|---|
| DB_fundamentals | Navathe | 4 | 2004 |
| DB_fundamentals | Elmasri | 4 | 2004 |
| DB_fundamentals | Elmasri | 5 | 2007 |
| DB_fundamentals | Navathe | 5 | 2007 |

# Exercise

Rp l e F zaf I f m RNF AK

**PHIẾU ĐIỂM - Học kỳ 1 năm 2008**

- *MSSV:*
- *Tên sinh viên:* L es w l T l
- *Địa chỉ:* I RV zaf I f m
- *Khoa:* I f m f a i rf s r k zwr l f

| Mã môn học | Tên môn học | Nhóm | Số TC | Điểm KT | Điểm thi | Điểm tổng kết |
|---|---|---|---|---|---|---|
| | ARBJ ER | RL | | | | |
| | AQBJ | A | | | | |

- *Ghi chú:* R wr l e k l f aq a f q a g k i g k rp t y g k rf gi f za l f s g k r l e i r ar l f b rp l g k I R g k rf gt y aza f q l yw

# Key finding algorithms

Extended part

# Key and super key

- **Superkey** of R: A set of attributes **SK** of **R** such that no two tuples in any valid relation instance r(R) will have the same value for SK. That is, for any distinct tuples t1 and t2 in r(R), t1[SK] ≠ t2[SK]
- **Key** of R: A **"minimal" superkey**; that is, a superkey K such that removal of any attribute from K results in a set of attributes that is not a superkey
- If K is a **key** of R, then K **functionally determines all attributes** in R

# Key-finding algorithm (1)

***Input:*** *A relation R and a set of functional dependencies F on the attributes of R.*

***Output:*** *a key K of R*

1. Set K to contain all attributes in R
2. For each attribute A in K {

   compute $(K - A)^+$ with respect to F;

   if $(K - A)^+$ contains all attributes in R, then

   set K := K – {A}

   };

# Key-finding algorithm (1)

- In algorithm (1), we start by setting K to all the attributes of R; we then remove one attribute at a time and check whether the remaining attributes still form a superkey.
- The algorithm (1) determines only **one key** out of the possible candidate keys for R; the key returned depends on the order in which attributes are removed from R in step 2.

# Key-finding algorithm (2)

***Input:*** *A relation R and a set of functional dependencies F on the attributes of R.*

***Output:*** *all candidate keys of R*

Let:

- $\boldsymbol{U}$ contain **all** attributes of R
- $\boldsymbol{U_l}$ contain attributes of R that occur **only** on the **left-hand side** of FDs in F
- $\boldsymbol{U_r}$ contain attributes of R that occur **only** on the **right-hand side** of FDs in F
- $\boldsymbol{U_b}$ contain attributes of R that occur on **both sides** of FDs in F

Note:

- $\boldsymbol{U_l} \cap \boldsymbol{U_r} = \phi$, $\boldsymbol{U_l} \cap \boldsymbol{U_b} = \phi$ and $\boldsymbol{U_r} \cap \boldsymbol{U_b} = \phi$
- $\boldsymbol{U_l} \cup \boldsymbol{U_r} \cup \boldsymbol{U_b} = \boldsymbol{U}$
- For every attribute $\boldsymbol{A} \in \boldsymbol{U}$, if $\boldsymbol{A} \in \boldsymbol{U_l}$, then $\boldsymbol{A}$ **must be** part of every candidate key of **R**.
- For every attribute $\boldsymbol{A} \in \boldsymbol{U}$, if $\boldsymbol{A} \in \boldsymbol{U_r}$, then $\boldsymbol{A}$ will **not** be part of any candidate key of **R.**

# Key-finding algorithm (2)

***Input:*** *A relation R and a set of functional dependencies F on the attributes of R.*

***Output:*** *all candidate keys of R*

1. Determine $U_l$, $U_r$ and $U_b$
2. If $U_l^+ = U$ under F, then $U_l$ forms the only key of R and the algorithm stops here.

   Else: move to step 3 // $U_l^+ \neq U$ under F
3. Consider every subsets $U_{bi}$ of $U_b$: $U_{bi} \subset U_b$

   For each $U_{bi}$, if $(U_l \cup U_{bi})^+ = U$ under F, then $K_i = (U_l \cup U_{bi})$ is a candidate key of R (*)

(*) *If* $K_i = (U_l \cup U_{bi})$ *is a candidate key of R, then we need not to check* $U_{bj} \subset U_b$ *where* $U_{bi} \subset U_{bj}$

# Key-finding algorithm (2)

- A simple categorization of attributes into the sets $U_l$, $U_r$ and $U_b$ allows to distinguish between those attributes that will participate in the candidate keys of a relational database schema and those that do not.
- The algorithm (2) finds all candidate keys.

# Exercise 1

Aml qgbcprf c slgtcpq jpcj rgml P $A, B, C, D, E, F$ lb rf c qcr mddslargml jbcnclbclagcq

1) $A \rightarrow B$
2) $C, D \rightarrow A$
3) $B, C \rightarrow D$
4) $A, E \rightarrow F$
5) $C, E \rightarrow D$

U f rgqrf c i cwdmp $R$?

# Exercise 2

Aml qgbcprf c s l gtcpq j pcj rgml P $A, B, C, D, E, F$ l b rf c qcr mdds l argml j bcncl bcl agcq

1) $A, D \rightarrow B$
2) $A, B \rightarrow E$
3) $C \rightarrow D$
4) $B \rightarrow C$
5) $A, C \rightarrow F$

U f r gq rf c i cwdmp $R$? Bcamk nmqc $R$ gl rm L D rf cl L D pcj rgml q Ecl cp j L mpk j D mpk Bcdgl ggml q

# Exercise 3

Aml qgbcprfc slgtcpq jpcj rgml P $A, B, C, D, E, F$ lbrfc qcr mdds larg ml jbcncl bcl agcq

1) $A \rightarrow$
2) $C \rightarrow$ , $D$
3) $A, F \rightarrow A, E$

U f rgqrfc icwdmp $R?$ Bcamk nmqc $R$ glrm LD LD lb AL D pcj rgml q

→

A →

A → B

D → A

D → B

# Which of the following dependencies may hold in the above relation

| A | B | C | TUPLE# |
|---|---|---|---|
| 10 | b1 | c1 | 1 |
| 10 | b2 | c2 | 2 |
| 11 | b4 | c1 | 3 |
| 12 | b3 | c4 | 4 |
| 13 | b1 | c1 | 5 |

i. $A \rightarrow B$, ii. $B \rightarrow C$, iii. $C \rightarrow B$, iv. $B \rightarrow A$, v. $C \rightarrow A$

# QUIZ 2

- Leywigk rp
- Mljgc IcJ
  - Rpalefgk as
  - Rf gegl nf r
- L gbsle
  - Af le Pcj rgml j jec p
  - Af le QOJ
  - Af le Lmpk jgx rgml