

| Topics (Lecture note) | Database systems: the complete book (Chap no.) | Big Data Fundamentals (Chap & page no. ***) |
|---|--|---|
| LN1: Big data characteristics: 5Vs | | Chap 1: 13-16 |
| LN2: data models | Chap 2.1 | Chap 1: 17-20 |
| LN3: relational data model & algebra | Chap 2.2, 2.4, 2.5.3 | Chap 7: 147-149 |
| LN4: FDs & normal forms | Chap 3.1-3.3, 3.5 | |
| LN5: XML basics; DTD& XML schema | Chap 11.1-11.4 | |
| LN6: RDF and RDF schema | | |
| LN7: Relational queries; query processing | Chap 2.4, Chap 14.1.2-14.1.5, 14.2.1 (index) Chap 15.3, 15.5, 15.6 | |
| LN8: Graph queries & operators | | |
| LN9: Approximate query processing (APQ): query driven | | |
| LN10: APQ: data driven | | |
| LN11: APQ: views | Chap 8.1- 8.2, 8.5 | |
| LN12: querying data streams | Chap 23.4-23.5 | Chap 6: 137-142 |
| LN13: parallel DBMS & operators | Chap 20.1 | Chap 6: 120-121 |
| LN14: MapReduce | Chap 20.2 | Chap 6:122-134 |
| LN15: MapReduce & Beyond: parallel graph processing | Chap 20.3 | Chap 2: 40-42 |
| LN16: Hadoop Ecosystem | | |
| LN17: NoSQL: ACID/EASE; KV & Column | | Chap 5: 94-117 Chap 7: 152-157 |
| LN18: NoSQL: Document DB & Graph DB | | Chap 7: 159-160 |
| LN19: NewSQL | | Chap 7: 163 |
| LN20: NewSQL & In-memory DBs | | Chap 7: 163-179 |
| LN21: BD-tractability | | |
| LN22: BD processing: review | | |
| LN23: BD Analytics: classification | | Chap 1: 6-11, Chap 4: 78-79; Chap 8: 184, 190 |
| LN24: BD Analytics: Clustering | | Chap 8: 191-192 |
| LN25: BD Analytics: Pattern mining | Chap 22.1-22.2 | |
| LN26: BD Quality: issues | | |
| LN27: BD Quality: Data Dependencies | | |
| LN28: BD Quality: Data Cleaning | | |
| LN29: BD Privacy & Security | | Chap 3, 49-51 |
| LN30: Conclusion & Vision | | |

*** "Big Data Fundamentals" Page -> section complete index:

<http://ptgmedia.pearsoncmg.com/images/9780134291079/samplepages/9780134291079.pdf>

the page numbers may vary due to different versions of the book.

Additional resources (color coded regions):

Hadoop: the definitive guide <http://hadoopbook.com/>

Mining of Massive Datasets. *Jure Leskovek, Anand Rajaraman and Jeffrey Ullman. v2.1, Cambridge University Press. 2014 (available online)*

Ethics of Big Data.

<https://eecs.wsu.edu/~yinghui/mat/courses/fall%202015/resources/Ethics%20of%20Big%20Data.pdf>

Regions not covered: refer to papers, reading list and surveys listed in the lecture notes.