



Chapter 2. Data Models and Query Languages



Chapter 3. Storage and Retrieval



Chapter 8. The Trouble with Distributed Systems



Chapter 7. Transactions



Chapter 4. Encoding and Evolution



Chapter 9. Consistency and Consensus



Chapter 5. Replication

Distributed Data Management DDM-Spark homework documentation

Viktoriia Hvozdk, Timo Büchert

Responsibility of the main actors

The code begins by reading input data and flattening it into key-value pairs. These pairs, representing column names and cell values, are then merged from multiple sources into a single dataset. The merged data is grouped by cell values to identify distinct groups.

Inclusion dependencies (INDs) are then found within each group by intersecting attribute values. The resulting dependencies are sorted by attribute names and printed out. Pairs with empty sets of values are excluded from the output for clarity and relevance.

