1.a.

Data

A collection of facts in raw or unorganized form

- Data is comprised of the basic, unrefined, and generally unfiltered information
- Data are elements of analysis.
- Data must be organized to become information
- information output by a sensing device or organ that includes both useful and irrelevant or redundant information and must be processed to be meaningful

Information

Easier to measure, analyze and visualize data for specific purpose

- Information… is much more refined data… that has evolved to the point of being useful for some form of analysis
- Information is data put in context; it is related to other pieces of data. Information is about meaning, and it forms the basis for knowledge
- Information is data with context. Information must be put into context to become knowledge
- knowledge obtained from investigation, study, or instruction

Knowledge

Connected to other pieces help us to understand how to apply information to achieve our goal

- Knowledge resides in the user…happens only when human experience and insight is applied to data and information
- Knowledge is a body of information, technique, and experience that coalesces around a particular subject
- Knowledge is information with meaning
- the range of one's information or understanding

Insight

Knowledge applied in action

- What to do? What's best?

1.b.

Data

Olmpic games- Sports, Athlete, Countries
Soccer- League, session, competetion, match, squad, goal, player
UNWTO & Eurostat- tourism_Stats(monthly, annual, euro_resident)

Information

Combine- olympic games data about country, player and sports with soccer data
Combine games data by timestamp with tourism data to make useful data

Knowledge

European giant, Germany, Spain, France are good at tournaments, so their tourism at that time is popular

Insight

Airlines, tourism company can make some good business by offering trip planning in these events time of specific countries.

1.c.

Data Collection-

Collect olympic soccer info and national soccer team data and tourism monthly and annual information of country wise data.

Data integration-

Combine information about German football records in olympics with German national football data and Germany's tourism information on that specific time of the month when games plays

Data cleaning-
(Same name, USA, U.S.A or United States of America)

Combined information of Euro nations data, who plays football at olympic tournaments, and has good records, and their tourism industry success

Data Distribution-

Who can use these data? Tourism company, Airlines - who can make use of these data, giving specific offer to people during game time.

4.A.

Optimal ranking-
Objective- Top-k webTables
NaiveRanking - go through k-websites
FilterRanking - K relation
NaiveRank better
None of them is optimal because not all relevant docs are retrieved

4B.
More rows,columns, hasHeader=>True better,
Lower null, search value better

Other feature - Language, visitors,

4C.
Adding
Multiplying - when they are not related

4D.

4E.
O(n^2)

6.A.

Schema:
No. | Name | productioncompany | year
ID , String, String, String

Whitespace to differentiate field, slash to differ year with company
Parenthesis

Some cartoon doesn’t have date, filling this data is problematic.

6.B.
3words,
1+2+1=4
4words,
1+3+3+1=8
5words,
1+4+6+4+1 = 16
4c1+4c2+4c3+4c4
General: 2^(n-1)

6.C.
3words = 6
4words = 11
5words = 20

General: 2^(n-1) + (n-1)

6.D.
2nd,
Name for string
Int for year

7.A.

Manually-
Initial creation
Data collection

7.B.