

**File Organizations and Size Estimations**  
**Shivam Pathak**

1. Using the entities and attributes shown for FIGHTERS creating MATCH STATISTICS

a. determine the size of a record in each file. (Use realistic data type and size parameters for each attribute.)

**FIGHTERS**

Attribute	Data Type	Bytes	Notes
FighterID	Integer	4 Bytes	Atomic, single-valued
LastName	Char(30)	30 Bytes	Atomic, single-valued
FirstName	Char(30)	30 Bytes	Atomic, single-valued
Nationality	Char(30)	30 Bytes	Atomic, single-valued
WeightClass	Char(20)	20 Bytes	Atomic, single-valued
FightStyles	Char(100)	100 Bytes	Multi-valued, estimated max of 100 characters
Email	Char(100)	100 Bytes	Multi-valued, estimated max of 100 characters
Phone	Integer, 15 digits	15 Bytes	Atomic, single-valued,
Status	Boolean	1 Byte	Atomic, True/False
Age	Integer	4 Bytes	Atomic, single-valued
BirthDate	Date	8 Bytes	Atomic, single-valued, date format
#Wins	Integer	4 Bytes	Atomic, single-valued
#Loses	Integer	4 bytes	Atomic, single-valued
<b>Total Size for Fighters</b>		<b>350 Bytes</b>	

## MATCH STATISTICS

Attribute	Data Type	Bytes	Notes
MFighterID	Integer	4 Bytes	Atomic, single-valued
MatchNumber	Integer	4 Bytes	Atomic, single-valued
Score	Float	4 Bytes	Atomic, single-valued
Win/Loss	Char(4)	4 Bytes	Atomic, single-valued, typically “win”, “loss”, or “draw”
CurrentRank	Integer	4 Bytes	Atomic, single-valued
FightStylesUsed	Varchar(30)	1-30 Bytes	Multivalued, est max 30 characters for longest fight style
TakedownUsed	Varchar(30)	1-30 Bytes	Multi-Valued est max 30 characters for longest takedown
ChampionStatus	Boolean	1 Byte	Atomic, True/False
MatchWeighIn	Float	4 Bytes	Atomic, single-valued
WeightDeviation	Float, Optional	4 Bytes	Optional, single-valued
WeightClass	Char(20)	20 Bytes	Atomic, single-valued
FinePaid	Float, Optional	4 Bytes	Optional, single-valued
<b>Total For Match Statistics</b>	<b>Without Optional</b>	<b>47- 105 Bytes</b>	
<b>Total for Match Statistics</b>	<b>With Optional</b>	<b>55- 113 Bytes</b>	

b. For each entity, tell if the records will be a fixed or variable length and explain how we know that.

**-FIGHTER RECORDS** is a **fixed** variable record field.

The fields in Fighters have fixed sizes using Integer data types or Char(n) which assign a certain number of bytes for the data. There is no range for multivalued attributes that would change the length of this record.

**-MATCH STATISTICS** is a **variable** record size because there are Varchar fields as well as optional attributes.

The Varchar fields are FightstylesUsed, and TakedownUsed. This means that they have different lengths based on the data stored. Also there exists optional fields: WeightDeviation and FinePaid which are not present in every record. This adds to the variety of sizes in the record since many fields will change based on the actual data.

2. Consider a disk with the following characteristics (these are not parameters of any particular disk unit): block size  $B = 512$  bytes; interblock gap size  $G = 128$  bytes; number of blocks per track = 20; number of tracks per surface = 400. A disk pack consists of 15 double-sided disks.

The MATCH STATISTICS file has  $r = 2000$  records. Only 5% of the MATCH STATISTICS have a WeightDeviation recorded. Only 60% of the MATCHES are won by a TakedownUsed and only 1 FIGHTER in the MATCH performed that winning TakedownUsed. In addition to the size and type of a record you estimated above, there is an additional byte added to each record which is used as a deletion marker. The file is stored on the disk whose parameters are given above.

a. Recalculate the record size  $R$  in bytes.

**Match Statistics**

Attribute	Data Type	Bytes	Notes
MFighterID	Integer	4 Bytes	Atomic, single-valued

MatchNumber	Integer	4 Bytes	Atomic, single-valued
Score	Float	4 Bytes	Atomic, single-valued
Win/Loss	Char(4)	4 Bytes	Atomic, single-valued, typically “win”, “loss”, or “draw”
CurrentRank	Integer	4 Bytes	Atomic, single-valued
FightStylesUsed	Varchar(30)	1-30 Bytes	Multivalued, est max 30 characters for longest fight style
TakedownUsed	Varchar(30)	1-30 Bytes	Multi-Valued est max 30 characters for longest takedown
ChampionStatus	Boolean	1 Byte	Atomic, True/False
MatchWeighIn	Float	4 Bytes	Atomic, single-valued
WeightDeviation	Float, Optional	4 Bytes	Optional, single-valued
WeightClass	Char(20)	20 Bytes	Atomic, single-valued
FinePaid	Float, Optional	4 Bytes	Optional, single-valued
Deletion Marker	Hexadecimal	1	Byte used to mark records for deletion
<b>Total For Match Statistics</b>	<b>Without Optional</b>	<b>48- 106 Bytes</b>	
<b>Total for Match Statistics</b>	<b>With Optional</b>	<b>56- 114 Bytes</b>	

b. Calculate the blocking factor bfr and state in English what the number represents for this type of record.

Since this variable the bfr is calculated using the average size:

-The average for the without optional is **77 bytes**

-The average with optional is **85 bytes**

Bfr is the average portion of a record in a block

- Without optional:  $512/77 = 6.65$

This means 6 records can be held in one block the rest is spanned

-With optional:  $512/85 = 6.02$

This means 6 records can be held in one block the rest is spanned

c. Calculate the number of file blocks b needed to store the file, assuming a spanned organization.

$r = 2000$  records

$R = 77$  bytes (without optional) or 85 bytes (with optional)

$B = 512$  bytes

optional not included:  $2000 * 77 / 512 = 300$  blocks

optional included:  $2000 * 83 / 512 = 325$  blocks

d. Calculate the file size r in bytes.

$r = R \text{ bytes} * r$

$r = 2000 \text{ bytes} * 77 = 154,000$  bytes with optional attributes not included

$r = 2000 \text{ bytes} * 83 = 166,000$  bytes with the optional attributes included

e. Calculate the unused space in b.

There will be no unused space because we are using a spanned organization

**U = 0 bytes**

f. If the MATCH STATISTICS file grows each year because there are 200 more MATCHES to record MATCH STATISTICS for every year, what will r be at the end of 3 years?

The file grows by 200 or more matches per year

1. Year 1: 2200

2. Year 2: 2400

3. Year 3: 2600

**The file will contain 2600 records in 3 years**

New Size

Without Optional:  $2600 * 77 = 200,200$  bytes

With Optional:  $2600 * 83 = 215,800$  bytes