**Aggregate Functions**

SQL can not only retrieve data but it can calculate statistics about the data in the tables. For example:

SELECT COUNT(\*)

FROM mccMatches;

The aggregation function COUNT counts the number of rows, and (\*) signifies all rows. So this query will count the number of matches that have been played.

If you specify a column instead of \* it will count the number of data values in the column, ignoring any NULL values as these are not true values. For example:

SELECT COUNT(age)

FROM mccPlayer;

This will tell us how many players there are, whose ages are known.

Using DISTINCT removes duplicate values before the aggregation function is applied.

SELECT COUNT(DISTINCT g\_town)

FROM mccGround;

This will tell us how many different towns have cricket grounds.

Other aggregation functions are SUM (the sum), AVG (the average), MIN (the minimum value) and MAX (the maximum value). These functions must specify a numeric attribute in the aggregation clause. For example:

SELECT MAX(batting\_score)

FROM mccMatch\_performance;

This will tell us the highest score achieved.

However the following query could produce misleading results, and in some SQL implementations will produce an error.

SELECT registration\_number, MAX(batting\_score)

FROM mccMatch\_performance;

This is because an aggregate function returns one value for the table as a whole, and you cannot also specify attributes that could have more than one value, for example two or more players, with different registration\_numbers, could have the same maximum score. SQL doesn't bother to test whether this is true or not, it just returns the first one (or an error).

You can restrict the rows that are considered by the aggregation function in the WHERE clause. For example, what is the highest batting score made since 2012?

SELECT MAX(batting\_score)

FROM mccMatch\_performance

WHERE match\_date > 120100;

The condition, match\_date > 120100, is applied first and the aggregate function is applied to the result.

More complex queries can be performed with multiple tables. The following assumes you are familiar with sub-queries (mcc tutorial 4).

If we want to know the registration number, or numbers, of players who have achieved the maximum score we have to use an indirect method.

SELECT registration\_number, batting\_score

FROM mccMatch\_performance

WHERE batting\_score IN (SELECT MAX(batting\_score)

FROM mccMatch\_performance);

We could also use '=' instead of IN. You can use comparison operators (=, >, <) with any sub-query provided that the sub-query only retrieves one value to be used in the comparison. Aggregate functions retrieve just one value.

SELECT registration\_number, batting\_score

FROM mccMatch\_performance

WHERE batting\_score = (SELECT MAX(batting\_score)

FROM mccMatch\_performance);

Some queries for you to try.

i) What is the average batting score of all match performances?

ii) What is the age of the youngest player?

iii) How many matches have been played?

iv) How many different teams have MCC played against?

The following assumes you are familiar with sub-queries (mcc tutorial 4).

v) List the registration numbers and batting scores of players who have the lowest batting score.

vi) List the registration number of the player who has the highest score against the opposing team 'Whitley Bay'.

vii) List the registration numbers and batting scores of players who have scored more than the average batting score (see query xxiv) on at least one occasion.

viii) List the names and batting scores of players who have scored more than the average batting score (see query xxiv) on at least one occasion.

SELECT registration\_number, batting\_score FROM mccMatch\_performance WHERE batting\_score = (SELECT MIN(batting\_score) FROM mccMatch\_performance);

SELECT DISTINCT registration\_number FROM mccMatch\_performance WHERE batting\_score = (SELECT MAX(batting\_score) FROM mccMatch\_performance WHERE match\_date IN (SELECT match\_date FROM mccMatches WHERE opposing\_team = "Whitley Bay"));

SELECT registration\_number, batting\_score FROM mccMatch\_performance WHERE batting\_score > (SELECT AVG(batting\_score) FROM mccMatch\_performance);

SELECT name, batting\_score FROM mccPlayer,mccMatch\_performance WHERE batting\_score > (SELECT AVG(batting\_score) FROM mccMatch\_performance);