

Aggregation and Grouping – Key Points

- aggregation operators apply an operation to a group of rows
 - e.g. COUNT, SUM, MAX
- grouping allows rows to be grouped together so that an aggregation operator can be applied to each group instead of (only) the table as a whole
 - GROUP BY to specify column(s) used for grouping
 - HAVING allows the selection of only certain groups
 - ORDER BY sorts the rows in the result
- be aware of how NULL is handled

Aggregation

Aggregation operators apply an operation to a group of rows.

operator	notes
SUM(s)	operator is applied to the values in the column s
AVG(s)	
MIN(s)	
MAX(s)	
COUNT(s)	
COUNT(*)	computes the number of rows

Duplicates are included, unless DISTINCT is used.

- e.g. COUNT(DISTINCT s)

Aggregation

```
SELECT COUNT(*) AS numMovies
FROM MOVIES
WHERE year >= 2000
```

```
SELECT MIN(length) AS shortest,
       MAX(length) AS longest
FROM MOVIES
WHERE year = 2012
```

```
SELECT COUNT(DISTINCT year)
FROM ( MOVIES M JOIN STARSIN SI
      ON M.title=SI.movieTitle ) JOIN
     MOVIESTAR MS ON SI.starName=MS.name
WHERE MS.name = 'Harrison Ford'
```

What to Count

```
SELECT COUNT(*)
FROM SAILOR JOIN RESERVATION USING (Sid)
– counts the number of reservations
```

```
SELECT COUNT(Sid)
FROM SAILOR JOIN RESERVATION USING (Sid)
– counts the number of reservations (Sid is PK, so cannot be NULL)
```

```
SELECT COUNT(Sname)
FROM SAILOR JOIN RESERVATION USING (Sid)
– counts the number of reservations made by sailors whose Sname is not NULL
```

```
SELECT COUNT(DISTINCT Sid)
FROM SAILOR JOIN RESERVATION USING (Sid)
– counts the number of different sailors who have made reservations
```

```
SELECT COUNT(DISTINCT Sname)
FROM SAILOR JOIN RESERVATION USING (Sid)
– counts the number of different names of sailors who have made reservations
```

Grouping

Grouping allows rows to be grouped together so that an aggregation operator can be applied to each group instead of (only) the table as a whole.

```
SELECT studioName, AVG(length)
FROM MOVIES
GROUP BY studioName
```

- GROUP BY groups rows which have the same value for the specified attribute(s)
- aggregation operations in the SELECT clause are applied to each group
- other columns in the SELECT clause can only be those used for grouping
 - or those dependent on the PK when grouping columns include the PK