PROJECT2 REPORT PART-1

Question 1: Number of enrollments per semester for each course

The solution implemented for this problem gives as output the number of student enrollments for each course in a given semester. One mapper and one reducer is utilized to arrive at a solution. The reducer is also used as the combiner.

Problem Description:

In the dataset given, each row contains the Semester ID, Semester, Room number, Days the course is being held, Timings of the course, Course ID, Course Name, Capacity of the Room, and the number of students enrolled.

To get valuable insights into the data represented in the dataset, we have analyzed the data to get meaningful deductions. The problem at hand was to find out how many students were enrolled in each semester for every course offered in that particular semester. By getting this number, we can find out the most popular and the most not taken courses in each semester over the years. This can help us in removing or discontinuing all the unpopular courses to save resources. Such unpopular courses can be given less priority when allocating rooms to maximize the use of classrooms. On the contrary, by knowing which are the popular courses, we can give them priority and allocate suitable classrooms and faculty.

The dataset described above is taken as the input and the number of students enrolled for each semester in each course is returned as the output.

This is done by giving the input to the map reduce program written.

MAPPER

The mapper takes in the dataset as the input and reads one row at a time and stores them in a variable called values. This variable is then split on "," to get the columns in each row separately as an element of a string array. We need only 3 columns, Semester, Course name and number of students enrolled. For each semester, the number of enrollments should be output for every course. Therefore, we give the Semester concatenated with the Course name as the Mapper output key. The concatenation is done by using the StringBuilder class. The output value of the Mapper is the 9th column, the number of students enrolled of type IntWritable.

REDUCER

The output of the Mapper for each key is then passed to the Reducer. The Reducer then takes as input each key and iterates over the values for that key and combines all the values. Therefore, the input key for the reducer is just one key which here is <Course name_Semester> and the input value is the same as the output value of the Mapper which is the number of enrollments. The reducer sums all the values for that key sent by the mapper. The Reducer then outputs the value which is the key as key and the sum of all the values for that particular as the value.

Thus, the number of students enrolled for each course in each semester is found using the mapper-reducer.

Question 2: Number of enrollments per year

The solution implemented for this problem gives as output the number of student enrollments per year One mapper and one reducer is utilized to arrive at a solution. The reducer is also used as the combiner.

Problem Description:
In the dataset given, each row contains the Semester ID, Semester, Room number, Days the course is being held, Timings of the course, Course ID, Course Name, Capacity of the Room, and the number of students enrolled.
To get valuable insights into the data represented in the dataset, we have analyzed the data to get meaningful deductions. The problem at hand was to find out how many students were enrolled in each year. By getting this number, we can know the increasing trend of the University over the years. We observe that, as the reputation of the University increased, the number of students enrolled increased. For the year with the most enrollment, the activities and the courses offered that year can be taken and seen if those activities contributed to the increased enrollments. If yes, the same can be done to increase the enrollments in future.

The dataset described above is taken as the input and the number of students enrolled each year is returned as the output.
This is done by giving the input to the map reduce program written.

MAPPER
The mapper takes in the dataset as the input and reads one row at a time and stores them in a variable called values. This variable is then split on "," to get the columns in each row separately as an element of a string array. We need only 2 columns, year and number of students enrolled. To get the year, we split on the column containing the Semester and its year separated by a space. For each year, the number of enrollments should be output. Therefore, we give the year as the Mapper output key. The output value of the Mapper is the 9th column, the number of students enrolled of type IntWritable.

REDUCER
The output of the Mapper for each key is then passed to the Reducer. The Reducer then takes as input each key and iterates over the values for that key and combines all the values. Therefore, the input key for the reducer is just one key which here is <Year> and the input value is the same as the output value of the Mapper which is the number of enrollments. The reducer sums all the values for that key sent by the mapper. The Reducer then outputs the value which is the key as key and the sum of all the values for that particular as the value.

Thus, the number of students enrolled in each year is found using the mapper-reducer.

Question 3: Courses with zero enrollment in each semester

The solution implemented for this problem gives as output all the courses that have zero enrollment. One mapper and one reducer is utilized to arrive at a solution. The reducer is also used as the combiner.

Problem Description:

In the dataset given, each row contains the Semester ID, Semester, Room number, Days the course is being held, Timings of the course, Course ID, Course Name, Capacity of the Room, and the number of students enrolled.

To get valuable insights into the data represented in the dataset, we have analyzed the data to get meaningful deductions. The problem at hand was to find out the courses with zero enrollment in any given semester. By getting this number, we can know which courses had nkdkdfkfo enrollments at all and thus these courses can be removed if it has zero enrollments for more than one consecutive semester. Or, if the courses have to be offered, they can be given minimum priority to ensure that the popular courses get the most of the resources and the best of faculty.

The dataset described above is taken as the input and the courses with zero enrollment is returned as the output.

This is done by giving the input to the map reduce program written.

MAPPER

The mapper takes in the dataset as the input and reads one row at a time and stores them in a variable called values. This variable is then split on "," to get the columns in each row separately as an element of a string array. We need only 3 columns, Semester, Course name and number of students enrolled. For each semester, the courses with zero enrollment should be output. Therefore, we give the Semester concatenated with the Course name as the Mapper output key. The concatenation is done by using the StringBuilder class. The output value of the Mapper is the 9th column, the number of students enrolled of type IntWritable. We check if the number of enrollments is zero and then pass the keys and value corresponding to that since we want the rows which have zero in the number of enrollments columns.

REDUCER

The output of the Mapper for each key is then passed to the Reducer. The Reducer then takes as input each key and iterates over the values for that key and combines all the values. Therefore, the input key for the reducer is just one key which here is <Course name_Semester> and the input value is the same as the output value of the Mapper which is the number of enrollments. Here since we want to display the courses which have zero enrollment, the reducer does nothing except write null values as the output since we have handles the zero enrollment condition in the mapper.

Thus, the courses which have zero enrollment for each semester is found using the mapper-reducer.

Question 4: Courses whose enrollment exceeds the capacity of the rooms they are held in and their enrollments

The solution implemented for this problem gives as output the courses for which the number of enrollments exceeds the capacity of the room alloted. One mapper and one reducer is utilized to arrive at a solution. The reducer is also used as the combiner.

Problem Description:
In the dataset given, each row contains the Semester ID, Semester, Room number, Days the course is being held, Timings of the course, Course ID, Course Name, Capacity of the Room, and the number of students enrolled.
To get valuable insights into the data represented in the dataset, we have analyzed the data to get meaningful deductions. The problem at hand was to find out the courses for which the number of students enrolled exceeds the capacity of the room. By getting the courses for which the capacity of the room does not satisfy the required capacity, we can know how to increase the efficiency of the allocation of rooms for the courses.

The dataset described above is taken as the input and the courses for which the enrollments exceed the capacity of the room alloted for the course is returned as the output.
This is done by giving the input to the map reduce program written.

MAPPER
The mapper takes in the dataset as the input and reads one row at a time and stores them in a variable called values. This variable is then split on "," to get the columns in each row separately as an element of a string array. We need only 4 columns, Room number, Course name, room capacity and number of students enrolled. Each course which satisfies the above criteria, should be output . Therefore, we give the Course name concatenated with the Room number as the Mapper output key. The concatenation is done by using the StringBuilder class. The output value of the Mapper is the 9th column, the number of students enrolled of type IntWritable. We also implement a condition to check if the number of enrollments is lesser than the capacity of the room. Only the rows that satisfy the condition are selected to be sent to the reducer.

REDUCER
The output of the Mapper for each key is then passed to the Reducer. The Reducer then takes as input each key and iterates over the values for that key and combines all the values. Therefore, the input key for the reducer is just one key which here is <Course name_Room number> and the input value is the same as the output value of the Mapper which is the number of enrollments. The reducer sums all the values for that key sent by the mapper. The Reducer then outputs the value which is the key as key and the sum of all the values for that particular as the value.

Thus, the number of students enrolled for each course in each semester is found using the mapper-reducer.

Question 5: Courses that are held on Monday, Wednesday and Friday

The solution implemented for this problem gives as output the courses for which the classes are held on 3 days of a week. One mapper and one reducer is utilized to arrive at a solution. The reducer is also used as the combiner.

Problem Description:

In the dataset given, each row contains the Semester ID, Semester, Room number, Days the course is being held, Timings of the course, Course ID, Course Name, Capacity of the Room, and the number of students enrolled.

To get valuable insights into the data represented in the dataset, we have analyzed the data to get meaningful deductions. The problem at hand was to find out the courses for which classes are held on 3 days of a week. By getting such courses and their enrollments, we can know the popular opinion of students if they would enroll in a course that holds classes on 3 days or would they prefer some other course which has classes only on 1 day of the week and the courses can be planned accordingly.

The dataset described above is taken as the input and the courses for which the classes are held on 3 days is returned as the output..

This is done by giving the input to the map reduce program written.

MAPPER

The mapper takes in the dataset as the input and reads one row at a time and stores them in a variable called values. This variable is then split on “,” to get the columns in each row separately as an element of a string array. We need only 3 columns, Course name, Days the classes are held and number of students enrolled. Each course which satisfies the above criteria, should be output . Therefore, we give the Course name as the Mapper output key. The output value of the Mapper is the 9th column, the number of students enrolled of type IntWritable. We also implement a condition to check if the classes are helo on Monday,Wednesday and Friday. Only the rows that satisfy the condition are selected to be sent to the reducer.

REDUCER

The output of the Mapper for each key is then passed to the Reducer. The Reducer then takes as input each key and iterates over the values for that key and combines all the values. Therefore, the input key for the reducer is just one key which here is <Course name> and the input value is the same as the output value of the Mapper which is the number of enrollments. The reducer sums all the values for that key sent by the mapper. The Reducer then outputs the value which is the key as key and the sum of all the values for that particular as the value.

Thus, the courses held on 3 days in a week is found using the mapper-reducer.

Question 6: Number of courses held at a given time per semester

The solution implemented for this problem gives as output the number of courses for each time period in a semester. One mapper and one reducer is utilized to arrive at a solution. The reducer is also used as the combiner.

Problem Description:

In the dataset given, each row contains the Semester ID, Semester, Room number, Days the course is being held, Timings of the course, Course ID, Course Name, Capacity of the Room, and the number of students enrolled.

To get valuable insights into the data represented in the dataset, we have analyzed the data to get meaningful deductions. The problem at hand was to find out the courses for which the number of students enrolled exceeds the capacity of the room. By getting the number of courses for which the times are same in each semester we can come to know all the courses that have the same timings and also know which is the popular time for the courses.

The dataset described above is taken as the input and the number of courses which share a common time per semester is returned as the output.

This is done by giving the input to the map reduce program written.

MAPPER

The mapper takes in the dataset as the input and reads one row at a time and stores them in a variable called values. This variable is then split on "," to get the columns in each row separately as an element of a string array. We need only 3 columns, Semester, Course name, and times the classes are held. The number of courses which satisfies the above criteria, should be output . Therefore, we give the Semester concatenated with the Time as the Mapper output key. The concatenation is done by using the StringBuilder class. The output value of the Mapper is the emission of 1 for each key encountered , of type IntWritable.

REDUCER

The output of the Mapper for each key is then passed to the Reducer. The Reducer then takes as input each key and iterates over the values for that key and combines all the values. Therefore, the input key for the reducer is just one key which here is <Semester_Time> and the input value is the same as the output value of the Mapper which is the number of enrollments. The reducer sums all the values for that key sent by the mapper. The Reducer then outputs the value which is the key as key and the sum of all the values for that particular as the value.

Thus, the number of courses at a time for each semester is found using the mapper-reducer.

Question 7: Number of courses per year

The solution implemented for this problem gives as output the number of courses per year. One mapper and one reducer is utilized to arrive at a solution. The reducer is also used as the combiner.

Problem Description:

In the dataset given, each row contains the Semester ID, Semester, Room number, Days the course is being held, Timings of the course, Course ID, Course Name, Capacity of the Room, and the number of students enrolled.

To get valuable insights into the data represented in the dataset, we have analyzed the data to get meaningful deductions. The problem at hand was to find out the number of courses per year. By getting the number of courses per year, we can know the increase in the number of courses each year and can keep track of the growth of the University as more and more courses get added.

The dataset described above is taken as the input and the number of courses per year is returned as the output.

This is done by giving the input to the map reduce program written.

MAPPER

The mapper takes in the dataset as the input and reads one row at a time and stores them in a variable called values. This variable is then split on "," to get the columns in each row separately as an element of a string array. We need only 2 columns, Year which is got by splitting the Semester column on space and the course name. Each row which satisfies the above criteria, should be output . Therefore, we give Year as the Mapper output key. The output value of the Mapper is the emission of 1 for each key encountered , of type IntWritable.

REDUCER

The output of the Mapper for each key is then passed to the Reducer. The Reducer then takes as input each key and iterates over the values for that key and combines all the values. Therefore, the input key for the reducer is just one key which here is <Year> and the input value is the same as the output value of the Mapper which is the number of enrollments. The reducer sums all the values for that key sent by the mapper. The Reducer then outputs the value which is the key as key and the sum of all the values for that particular as the value.

Thus, the number of courses per year is found using the mapper-reducer.

Question 8: Top 10 courses based on enrollment

The solution implemented for this problem gives as output the top 10 courses for which the highest 10 number of enrollments are found. One mapper and one reducer is utilized to arrive at a solution. The reducer is also used as the combiner.

Problem Description:
In the dataset given, each row contains the Semester ID, Semester, Room number, Days the course is being held, Timings of the course, Course ID, Course Name, Capacity of the Room, and the number of students enrolled.
To get valuable insights into the data represented in the dataset, we have analyzed the data to get meaningful deductions. The problem at hand was to find out the top 10 courses for which have the highest number of enrollments. By getting this, we can see the popular courses and hence assign proper and appropriate resources for the betterment of these courses.

The dataset described above is taken as the input and the top 10 courses for which the enrollments are maximum is returned as the output.
This is done by giving the input to the map reduce program written.

MAPPER
The mapper takes in the dataset as the input and reads one row at a time and stores them in a variable called values. This variable is then split on "," to get the columns in each row separately as an element of a string array. We need only 2 columns, Course name,and number of students enrolled. Each row which satisfies the above criteria, should be output . Therefore, we give the Course name as the Mapper output key. The output value of the Mapper is the 9th column, the number of students enrolled of type IntWritable.

REDUCER
The output of the Mapper for each key is then passed to the Reducer. The Reducer then takes as input each key and iterates over the values for that key and combines all the values. Therefore, the input key for the reducer is just one key which here is <Course name> and the input value is the same as the output value of the Mapper which is the number of enrollments. The reducer sums all the values for that key sent by the mapper. The Reducer then outputs the value which is the key as key and the sum of all the values for that particular as the value. We also implement a TreeMap and a cleanup method to get the top 10.

Thus, the top 10 courses is found using the mapper-reducer.

Question 9: Top 10 classrooms based on capacity

The solution implemented for this problem gives as output the top 10 classrooms which have the highest capacity. One mapper and one reducer is utilized to arrive at a solution. The reducer is also used as the combiner.

Problem Description:

In the dataset given, each row contains the Semester ID, Semester, Room number, Days the course is being held, Timings of the course, Course ID, Course Name, Capacity of the Room, and the number of students enrolled.

To get valuable insights into the data represented in the dataset, we have analyzed the data to get meaningful deductions. The problem at hand was to find out the top 10 classrooms which have the highest capacity. By getting this, we can assign proper courses which need the highest capacity to the rooms with maximum capacity by comparing the number of enrollments and the capacity.

The dataset described above is taken as the input and the top 10 classrooms f which have the highest capacity is returned as the output.

This is done by giving the input to the map reduce program written.

MAPPER

The mapper takes in the dataset as the input and reads one row at a time and stores them in a variable called values. This variable is then split on "," to get the columns in each row separately as an element of a string array. We need only 2 columns, Room number,and capacity served. Each row which satisfies the above criteria, should be output . Therefore, we give the Room number as the Mapper output key. The output value of the Mapper is the capacity of the room of type IntWritable.

REDUCER

The output of the Mapper for each key is then passed to the Reducer. The Reducer then takes as input each key and iterates over the values for that key and combines all the values. Therefore, the input key for the reducer is just one key which here is <Room number> and the input value is the same as the output value of the Mapper which is the capacity served. The reducer sums all the values for that key sent by the mapper. The Reducer then outputs the value which is the key as key and the sum of all the values for that particular as the value. We also implement a TreeMap and a cleanup method to get the top 10.

Thus, the top 10 classrooms is found using the mapper-reducer.

Question 10: Top 10 semesters based on enrollment

The solution implemented for this problem gives as output the top 10 semesters for which the highest 10 number of enrollments are found. One mapper and one reducer is utilized to arrive at a solution. The reducer is also used as the combiner.

Problem Description:

In the dataset given, each row contains the Semester ID, Semester, Room number, Days the course is being held, Timings of the course, Course ID, Course Name, Capacity of the Room, and the number of students enrolled.

To get valuable insights into the data represented in the dataset, we have analyzed the data to get meaningful deductions. The problem at hand was to find out the top 10 semesters for which have the highest number of enrollments. By getting this, we can see the popular semesters and the courses offered in that semester and hence get the courses to increase the enrollment in the future.

The dataset described above is taken as the input and the top 10 semesters for which the enrollments are maximum is returned as the output.

This is done by giving the input to the map reduce program written.

MAPPER

The mapper takes in the dataset as the input and reads one row at a time and stores them in a variable called values. This variable is then split on "," to get the columns in each row separately as an element of a string array. We need only 2 columns, Semester, and number of students enrolled. Each row which satisfies the above criteria, should be output . Therefore, we give the Semester as the Mapper output key. The output value of the Mapper is the 9th column, the number of students enrolled of type IntWritable.

REDUCER

The output of the Mapper for each key is then passed to the Reducer. The Reducer then takes as input each key and iterates over the values for that key and combines all the values. Therefore, the input key for the reducer is just one key which here is <Semester> and the input value is the same as the output value of the Mapper which is the number of enrollments. The reducer sums all the values for that key sent by the mapper. The Reducer then outputs the value which is the key as key and the sum of all the values for that particular as the value. We also implement a TreeMap and a cleanup method to get the top 10.

Thus, the top 10 semesters is found using the mapper-reducer.