## Sorting on Binary files and the relation to databases.

This handout will use this data as an example (shown in Excel as text with headers for reference) of the layout of a datafile. New objects are typically are placed in a new slot at the end (append in C++) of the file. As an example, Jeff would have been the last employee added by using an append operation, increasing the number of objects in the data file by one. The records are in the physical order they were added but not sorted in a logical order based on a characteristic or combination of characteristics. Actual binary files would not be have headers or text format.

Nai	me	Gross	Net	Federal	State	FICA	Taxes	Hours	Rate
0 Geo	orge	200.00	136.70	40.00	8.00	15.30	63.30	40.00	5.00
1 San	n	180.00	123.03	36.00	7.20	13.77	56.97	30.00	6.00
2 Sue	2	300.00	205.05	60.00	12.00	22.95	94.95	50.00	6.00
3 Bet	th	300.00	205.05	60.00	12.00	22.95	94.95	50.00	6.00
4 Am	nos	270.00	184.55	54.00	10.80	20.66	85.46	30.00	9.00
5 Jef	f	440.00	330.74	88.00	17.60	33.66	139.26	40.00	11.00

An alternative to physically sorting would be create what is sometimes referred to as an index file, which in this example uses only the record number of the object(s) for sorting rather than the entire object. The example below show the order the original file would be accessed as sorted by the field.

Name Gross Hours Each 4 1 1 index 0 3 4 file 0 4 0 contain 2 5 5 only the 3 2 1 number 5 3 2 not the entire object.

Record

Number

The files could be physically sorted on a characteristic (or combination such as last name first name) and stored in physical order often in a separate file. However this approach creates a separate file for each sort and occupies the same amount of space on a storage device as the original file. In addition there is the problem of updating all files if the original file has records added, deleted or changed.

	Rate	Hours	Taxes	FICA	State	Federal	Net	Gross	Name	
	9.00	30.00	85.46	20.66	10.80	54.00	184.55	270.00	Amos	4
Physicall sorted b	6.00	50.00	94.95	22.95	12.00	60.00	205.05	300.00	Beth	3
	5.00	40.00	63.30	15.30	8.00	40.00	136.70	200.00	George	0
	11.00	40.00	139.26	33.66	17.60	88.00	330.74	440.00	Jeff	5
	6.00	30.00	56.97	13.77	7.20	36.00	123.03	180.00	Sam	1
	6.00	50.00	94.95	22.95	12.00	60.00	205.05	300.00	Sue	2
	Rate	Hours	Taxes	FICA	State	Federal	Net	Gross	Name	
	6.00	30.00	56.97	13.77	7.20	36.00	123.03	180.00	Sam	1
Physicall sorted b	5.00	40.00	63.30	15.30	8.00	40.00	136.70	200.00	George	0
	9.00	30.00	85.46	20.66	10.80	54.00	184.55	270.00	Amos	4
	6.00	50.00	94.95	22.95	12.00	60.00	205.05	300.00	Sue	2
gross	6.00	50.00	94.95	22.95	12.00	60.00	205.05	300.00	Beth	3
	11.00	40.00	139.26	33.66	17.60	88.00	330.74	440.00	Jeff	5
	Rate	Hours	Taxes	FICA	State	Federal	Net	Gross	Name	
Physicall	6.00	30.00	56.97	13.77	7.20	36.00	123.03	180.00	Sam	1
	9.00	30.00	85.46	20.66	10.80	54.00	184.55	270.00	Amos	4
sorted b	5.00	40.00	63.30	15.30	8.00	40.00	136.70	200.00	George	1 0
hours	11.00	40.00	139.26	33.66	17.60	88.00	330.74	440.00	Jeff	5
	6.00	50.00	94.95	22.95	12.00	60.00	205.05	300.00	Sue	2
	6.00	50.00	94.95	22.95	12.00	60.00	205.05	300.00	Beth	3

Each file sorted by a characteristic would occupy the same amount of space as the original file.