J'ai seulement accès au manuel en anglais et j'ai pris la base de données I en anglais, alors j'ai fait ce travail en anglais. Je peux certainement donner une traduction en français si nécessaire. Je sais que le professeur Beldjehem enseigne aussi la section anglaise de ce cours et que le TA est anglophone, alors j'ai pensé que ce serait bien.

1,

a) An access path is consists of either a file scan or an index plus a matching selection condition. An access path is a way of retrieving tuples from a table.

   An index corresponds to an access path when the tuples that satisfy the selection condition can retrieved using only the index. A conjunctive normal form (CNF) selection matches a hash index if there is a term in the attribute=value form in the selection for each attribute in the index's search key. A CNG selection matches a tree index if the in a term in the attribute op value from for each attribute in a prefix of the index's search key.

   A primary conjunct is an index that matches some subset of the conjunctions in a CNP selection condition although it does not match the enfire condition. Primary conjuncts are important as the selectivity of an access path depends on the primary conjunctions in the selection condition (with respect to the index involved). A reduction factor is the fraction of tuples in the table that satisfy a given conjunct. When there are several primary conjuncts, a widely used practice is to to approximate the fraction of tuples that satisfy all conjuncts by using the product of their reduction factors.

b) The advantages of a system catalog for relationships is the system catalog contains statistics used to estimate the cost of operations and the size of the results.

2.

a) Pipelining used when a query is composed of several operators and the result of one operator is pipelined to another without creating a temporary table. Pipelining is enabled by left-deep plans. The advantage of pipelining an output of an operation into the next operator is is it saves the cost of writing out the immediate results and reading it back in. This can be substantial cost savings.
b) The statistics gathered from the database are used by the query optimizer to improve the chances of selecting an optimal query plan. Reductions factors are calculated using the statistics. The reduction factors determine the results the optimizer expects from the different inputs and indexes.

3.

a) The goal of query optimisation is to resolve two problems. The first is what plans should be considered. The second is how the cost of the plans is estimated. The goal is the find the lowest cost plan that achieves the desired result. Optinimation is important as its goal is to find the lowest cost strategy.

b) Les plans gauche-profond permettent l'évitement des relations temportaire car avec les plans gauche-profond permettent à la sortie de chaque opérateur d'être canalisée.

4.

1. a) Match, primary keys is Sailors.sid<50000
   b) Match, primary keys is Sailors.sid=50000
2. a) Not a match, hash indexes do not work with range queries
   b) Match, primary keys is Sailors.sid=50000

5.

1. There are 40000 tuples (80 Sailor tuples times 500 pages). Assuming uniform distribution around half the tuples will match the selection condition. The cost of finding the first leaf node (4 I/O's) plus the cost of retrieving matching tuples is then equal to the total cost. There are two cases:
   Clustered index: The total cost is equal to 254 I/Os (4 + 250 I/O's).
   Unclustered index: The cost of retrieving each matching tuple potentially as high as one I/O for each matching tuple (20000 I/O's). This results in a total cost of 20004 I/Os.
2. For range queries a hash index is cannot be used. This means we have to do a file scan at the cost of reading the entire relation (500 pages I/O's).