**CS 49995 & CS 63016 ST: Big Data Analytics**

**Homework 1**

**Instructor:** Xiang Lian

**Due Date:** Please refer to the course website

1. What are the 3V features of big data? Please briefly describe them. [15 points]

2. In the following figure, which is sparse index? Which is dense index? [20 points]

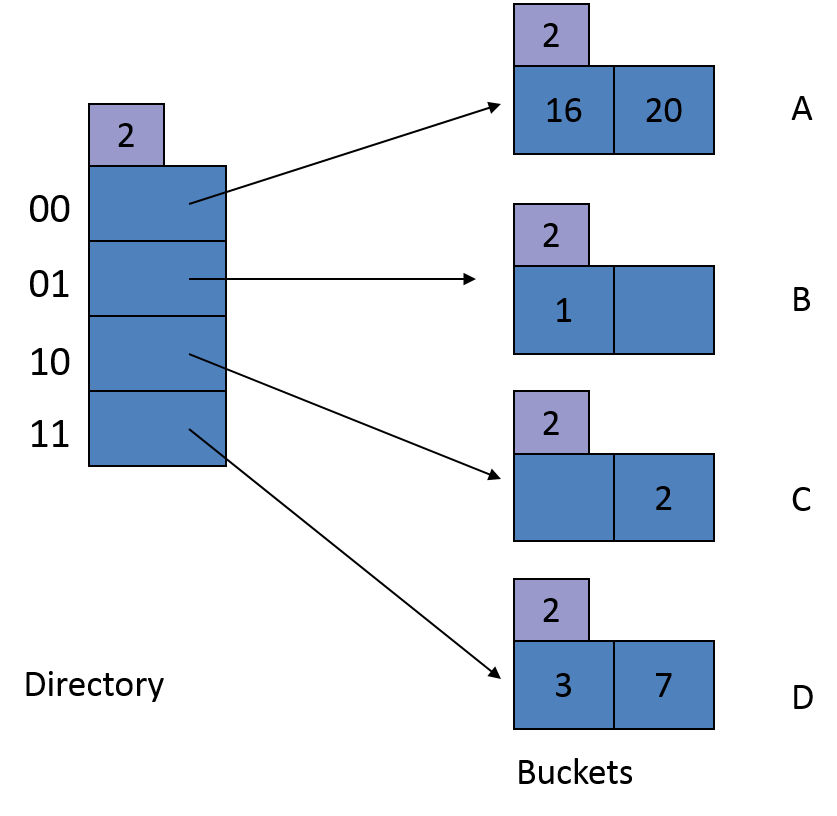
**Index A:**

**Index B:**



3. Please tell the differences between B-tree and B+-tree. (**Hint:** please use Google to find the structure of B-tree index) [20 points]

4. In the extensible hash index (each bucket can hold at most 2 items) below, if we want to insert 19, what will be the structure of the hash index after the insertion? [20 points]



*h*(1) = 000**01**

*h*(2) = 000**10**

*h*(3) = 000**11**

*h*(7) = 001**11**

*h*(16) = 100**00**

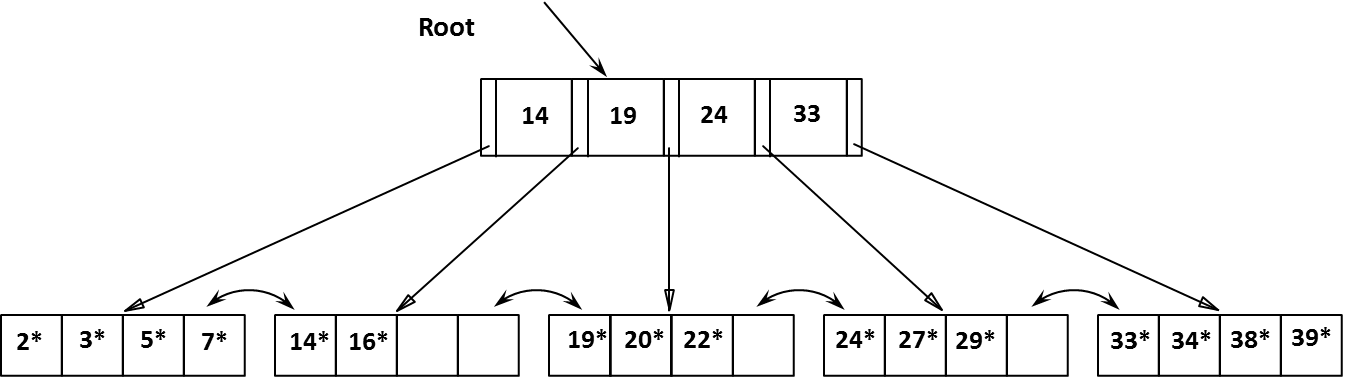
*h*(18) = 100**10**

*h*(19) = 100**11**

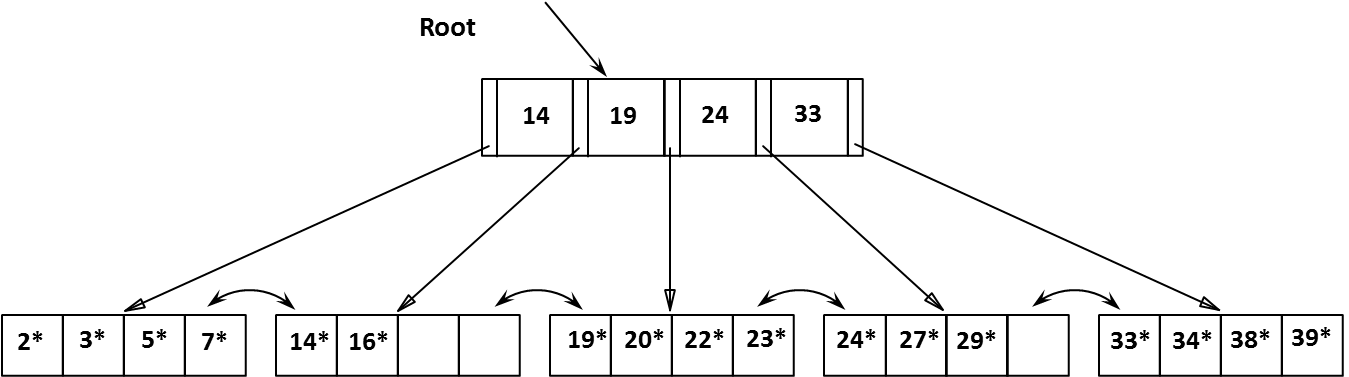
*h*(20) = 101**00**

5. Assuming that the node capacity of B+-trees is 4, please draw the updated B+-trees upon insertions. [25 points]

5(a). Insert tuples with key 23\* into the following B+-tree. [10 points]



5(b). Insert tuples with key 8\* into the following B+-tree. [15 points]



**Submission**

Submit an electronic copy of your homework solution to the [Blackboard](https://learn.kent.edu/).

18