

## Blog

### Hive – Order By vs Sort By vs Cluster By vs Distribute By

JUNE 14, 2019    **ADMIN ([HTTP://WWW.BIGDATAINTERVIEW.COM/AUTHOR/ADMIN/](http://www.bigdatainterview.com/author/admin/))**    **2 COMMENTS ([HTTP://WWW.BIGDATAINTERVIEW.COM/HIVE-ORDER-BY-VS-SORT-BY-VS-CLUSTER-BY-VS-DISTRIBUTE-BY/#COMMENTS](http://www.bigdatainterview.com/hive-order-by-vs-sort-by-vs-cluster-by-vs-distribute-by/#comments))**

Hive has so many clubbing operations like Order By, Sort By etc. Each clause it's own uses, advantages and disadvantages. These will be most frequently asked interview questions. These may not be asked together but they will be pairs wise. So we thought of covering all of them together in a single for comparability. Let's look at them individually..

#### **Order By**

Order clause is used to sort the data globally. We can specify DESC to sort the data in decreasing order and ASC to sort in ascending order. Order By clause has some limitations. In strict mode i.e., when we set hive.mapred.mode to strict, then the Hive query must have limit at the end. This is because Order By sorts the data globally, so there should be only one reducer to produce the output. If the input has huge data then one reducer might take lot of time. To avoid that we have to use Limit clause at the end.

#### **Sort By**

Sort By clause sorts the data per reducer but not globally. The mail difference between Sort By and Order By is the latter one guarantees global sort of data whereas the former guarantees per reducer sorting of data.

## Distribute By

Distribute By clause is used to distribute the values columns among the reducers. All the distribute columns will go to the same reducer. It doesn't sort the data per reducer and not even globally.

## Cluster By

This is the resultant command of performing Distribute By and Sort By clauses together. This command is used to distribute the rows among the reducers. Columns mentioned in the Clustered By clause will go to the multiple reducers and ensures that rows are sorted per reducer.

Please give us your valuable feedback..

POSTED IN: **HIVE** ([HTTP://WWW.BIGDATAINTERVIEW.COM/CATEGORY/SPARK/HIVE/](http://www.bigdatainterview.com/category/spark/hive/)).

[← What are the differences between Spark 1.x and Spark 2.x?](http://www.bigdatainterview.com/what-are-the-differences-between-spark-1-x-and-spark-2-x/)  
(<http://www.bigdatainterview.com/what-are-the-differences-between-spark-1-x-and-spark-2-x/>).

[map vs mapValues →](http://www.bigdatainterview.com/map-vs-mapvalues/)  
(<http://www.bigdatainterview.com/map-vs-mapvalues/>)

2 thoughts on “Hive – Order By vs Sort By vs Cluster By vs Distribute By”



**Gavin Gentleman** (<https://Proxies123.com>) says:

**February 1, 2020 at 5:42 pm** (<http://www.bigdatainterview.com/hive-order-by-vs-sort-by-vs-cluster-by-vs-distribute-by/#comment-188>).

I do love the manner in which you have presented this particular matter and it does supply us a lot of fodder for consideration. However, through just what I have personally seen, I really trust when the actual reviews pack on that people continue to be on issue and in no way get started on a tirade involving the news du jour. Yet, thank you for this excellent piece and although I can not agree with this in totality, I respect your point of view.

**Reply** →



**jordan 4** (<http://www.jordan4.us.com>) says:

**August 9, 2020 at 6:50 am** (<http://www.bigdatainterview.com/hive-order-by-vs-sort-by-vs-cluster-by-vs-distribute-by/#comment-332>).

You must participate in a contest for probably the greatest blogs on the web. I will suggest this site!

**Reply** →

## Leave a Reply

Your email address will not be published. Required fields are marked \*

### Comment

**Name \***

**Email \***

**Website**

☐ **Save my name, email, and website in this browser for the next time I comment.**

POST COMMENT

Search ...

SEARCH

## Test Ride India's Quickest Scooter

Book Now



### Recent Posts

[Option, Some, None in Scala \(OR\) How to handle null values in Scala?](http://www.bigdatainterview.com/option-some-none-in-scala-or-how-to-handle-null-values-in-scala/)

[\(http://www.bigdatainterview.com/option-some-none-in-scala-or-how-to-handle-null-values-in-scala/\)](http://www.bigdatainterview.com/option-some-none-in-scala-or-how-to-handle-null-values-in-scala/)

[What is Singleton object in Scala? \(http://www.bigdatainterview.com/what-is-singleton-object-in-scala/\)](http://www.bigdatainterview.com/what-is-singleton-object-in-scala/)

[How to process JSON data or file in HIVE without using JsonSerDe?](http://www.bigdatainterview.com/how-to-process-json-data-or-file-in-hive-without-using-jsonserde/)

[\(http://www.bigdatainterview.com/how-to-process-json-data-or-file-in-hive-without-using-jsonserde/\)](http://www.bigdatainterview.com/how-to-process-json-data-or-file-in-hive-without-using-jsonserde/)

**B.Com @ PES  
Electronic City**

### Recent Comments

[CURRY 7 SOUR PATCH \(HTTP://WWW.CURRY7.US\)](http://www.curry7.us) ON [SPARK GROUPBYKEY VS REDUCEBYKEY VS AGGREGATEBYKEY \(HTTP://WWW.BIGDATAINTERVIEW.COM/SPARK-GROUPBYKEY-VS-REDUCEBYKEY-VS-](http://www.bigdatainterview.com/spark-groupbykey-vs-reducebykey-vs-aggregatebykey)

[AGGREGATEBYKEY/#COMMENT-412\)](#)

[JORDAN 4 \(HTTP://WWW.JORDAN4.US.COM\) ON HIVE - ORDER BY VS SORT BY VS CLUSTER BY VS DISTRIBUTE BY \(HTTP://WWW.BIGDATAINTERVIEW.COM/HIVE-ORDER-BY-VS-SORT-BY-VS-CLUSTER-BY-VS-DISTRIBUTE-BY/#COMMENT-332\)](#)

[LOUBOUTIN SHOES \(HTTP://WWW.LOUBOUTIN-SHOES.US.COM\) ON SPARK RDD VS DATAFRAME VS DATASET \(HTTP://WWW.BIGDATAINTERVIEW.COM/SPARK-RDD-VS-DATAFRAME-VS-DATASET/#COMMENT-331\)](#)

 **ATHER**

**Test ride the  
Ather 450X.**

**Book Now**



## Archives

[JANUARY 2021 \(HTTP://WWW.BIGDATAINTERVIEW.COM/2021/01/\)](#)

[DECEMBER 2020 \(HTTP://WWW.BIGDATAINTERVIEW.COM/2020/12/\)](#)

[OCTOBER 2020 \(HTTP://WWW.BIGDATAINTERVIEW.COM/2020/10/\)](#)

[JULY 2020 \(HTTP://WWW.BIGDATAINTERVIEW.COM/2020/07/\)](#)

[MAY 2020 \(HTTP://WWW.BIGDATAINTERVIEW.COM/2020/05/\)](#)

[APRIL 2020 \(HTTP://WWW.BIGDATAINTERVIEW.COM/2020/04/\)](#)

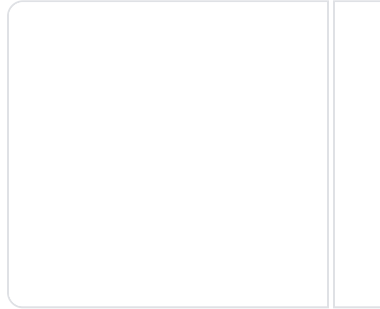
[MARCH 2020 \(HTTP://WWW.BIGDATAINTERVIEW.COM/2020/03/\)](#)

**NOVEMBER 2019 ([HTTP://WWW.BIGDATAINTERVIEW.COM/2019/11/](http://www.bigdatainterview.com/2019/11/))**

**JULY 2019 ([HTTP://WWW.BIGDATAINTERVIEW.COM/2019/07/](http://www.bigdatainterview.com/2019/07/))**

**JUNE 2019 ([HTTP://WWW.BIGDATAINTERVIEW.COM/2019/06/](http://www.bigdatainterview.com/2019/06/))**

**MAY 2019 ([HTTP://WWW.BIGDATAINTERVIEW.COM/2019/05/](http://www.bigdatainterview.com/2019/05/))**



Anti-Odor and Anti  
Innersense

## **FOLLOW US**



[\\_\(<https://www.facebook.com/Bigdatainterviewcom-112413567199733/>\)](https://www.facebook.com/Bigdatainterviewcom-112413567199733/)



[\\_\(<https://twitter.com/harisha23700231>\)](https://twitter.com/harisha23700231)

## **CONTACT US**

### **Email**

[sparkandbigdatainterview@gmail.com](mailto:sparkandbigdatainterview@gmail.com) (<mailto:sparkandbigdatainterview@gmail.com>)

**Never take your  
phone out.**

—  
On-board **Google maps navigation.**  
**Ather 450X.** Quick. Electric. Intelligent.

**Book Now**



[Privacy Policy \(http://www.bigdatainterview.com/privacy-policy/\).](http://www.bigdatainterview.com/privacy-policy/)

Copyright © 2021 Big Data Interview