

# FAQ Kylin 4.X

If you did not find the answer of your question, [feel free to leave your comment under this wiki](#).

---

## Question List

- How do RowKey affect storage & performance in Kylin 4.0?

RowKey determined sort-by column, check detail at [How to improve cube building and query performance](#) .

- What is Sparder(SpargerContext)? And how should I take care of it ?

Sparder is the implemenatation of new distributed query engine which backend by a spark application. If Sparder is dead, all your query will failed. And you can check sparder's liveness after Kylin instance(Query Server) was started in application list of Resource Manager Web UI. SparderCanary tool has been added in 4.0-beta to monitor the inventory status of sparder. When the sparder dies, SparderCanary will try to pull it up automatically.

- Is Hadoop3 supported ?

Hadoop 3 is supported in kylin 4.0-beta, and has been verified on CDH 5.7, CDH 6.2, EMR 5.31, EMR 6.0.0, HDP2.4. Hadoop3 and EMR environments require additional configuration, please check the [Installation Guide](#).

- If you faced Exception with message like this : "Cannot find hive-site.xml in kylin\_hadoop\_conf\_dir", please:

1. Copy all files under `/etc/hadoop/conf` to one directory ("`/path/to/hadoop_conf`").
2. Copy `hive-site.xml` to "`/path/to/hadoop_conf`".
3. Edit `kylin.properties`, modify `kylin.env.hadoop-conf-dir=/path/to/hadoop_conf`, restart Kylin.

- How to achieve Read/Write Separation Deployment

Please refer to [Read-Write Separation Deployment for Kylin 4.0](#).

- How to refresh the lookup table snapshot?

It will be automatically refreshed the next time build.

- How to use the new garbage cleaning tool, which garbage will be cleaned up?

Please refer to [How to clean up storage in Kylin 4](#).

- Can Cube Planner be used?

Cube Planner Phase1 is supported in 4.0.0-beta. Please refer to [How to use Cube Planner in Kylin 4](#).

- Where is the dimension dictionary stored?

Dimension dictionary is removed. The only dictionary remained in Kylin 4.0 is Global Dictionary.

- What are the best practice of optimization for build engine?

Check [How to improve cube building and query performance](#) .

- What are the best practice of optimization for query engine(sparder)?

Please refer to [How to improve cube building and query performance](#) and [Improve query performance by setting shard by column](#) .

- Is Kylin 3.x and Kylin 4.x metadata compatible?

Almost fully compatible, except please purge segments of your cube because HBase Storage is removed now. Kylin 4.0 remommend to use RDBMS as Metadata, please refer to [Use MySQL as Metastore](#) and [How to use HBase metastore in Kylin 4.0](#).

- Is Kylin 3.x and Kylin 4.x pre-calculated cuboid data compatible? If not, will there be a migration plan?

The pre-calculated cuboid data is completely incompatible, and there is no migration plan for the time being, due to relatively large effort in development.

- Is the Spark used by Kylin the community version?

Spark 2.4.6 is currently supported. Other spark distribution is not supported offically.

- What features will no longer be supported in Kylin 4? And what do Kylin 4 provided ?

Please refer to [Kylin 4.X Feature List](#).

- What is the performance of query engine and build engine in Kylin 4?

To be updated

- Will query results in Kylin 4 be consistent with the previous version?

To be updated

- "Caused by: com.amazonaws.services.s3.model.AmazonS3Exception: Please reduce your request rate. (Service: Amazon S3; Status Code: 503; Error Code: SlowDown;"
  - <https://aws.amazon.com/premiumsupport/knowledge-center/emr-s3-503-slow-down/>
  - We can reverse the Kylin's working dir.
- How to develop UDF and UDAF

To be updated

- Does Kylin 4 support AWS Glue?

It is not supported in Kylin 4.0.0-alpha and Kylin 4.0.0-beta.

- Does query on Spark support Spark Scheduler Pool setting(resource isolation)?

[Use different spark pool for different query](#)

- What is the implementation of the new global dictionary?

Please refer to [Global Dictionary on Spark](#).

- Are all the query results of Cube the same as query results from Push down engine(Spark SQL) ?

**No.** There are two cases will be different, show below:

1. When cube contains 'COUNT\_DISTINCT' from HLL measure, Spark SQL will still calculate the accurate measure values from source data;
2. When cube contains 'PERCENTILE' measure, the algorithm used to calculate the values in Kylin 4.0 is different from the one of Spark SQL;

- Is it recommended to use the TopN measure in Kylin 4.0 ?

**No.** In Kylin 4.0, if there is a TopN measure in cube, the data of 'TopN' measure will be saved in parquet file as 'ArrayType', which will lead to low reading performance,

because Spark can't use 'VectorizedParquetRecordReader' to read parquet file when the returned schemas include 'ArrayType'. Please use the original design (dimension + sum measure) directly to execute TopN-style SQL.