

大数据：

HCIP(H13-723)

1.大数据的 4V 不包含哪一个？

- A.数据量大
- B.种类多
- C.价值密度低
- D.分布式**
- E.处理速度快

2.GaussDB 200 支持什么格式的数据存储？单选题

- A.行存储
- B.列存储
- C.行列混存
- D.以上都对**

3.ElasticSearch 进行全文检索一般需要哪几个步骤？

- A.清洗、分词、建立索引**
- B.清洗、分词
- C.清洗、建立索引
- D.清洗、建立索引、分词

4.假设现在要做一个可以根据线索指导运维人员进行排障的功能，你建议选择下列哪个工具实现该功能？

- A.Lucene
- B.HBase
- C.ElasticSearch
- D.GraphBase**

5.Flume 的高级组件不包含以下哪个？

- A.Sink Processor
- B.Channel Interceptor**
- C.Channel Selector
- D.Source Interceptor

6.关于 Gauss DB 200 的数据导入导出下列说法正确的是？

- A.使用 gsq1 元命令和 COPY 命令在使用和用法上是一样的
- B.GDS 在数据导入时可以做预处理，比如非法字符替换、容错处理、数据聚合等**
- C.COPY 方式使用简单，一般用在大量数据的导入导出中
- D.INSERT 数据写入适合数据量不大，并发度不高的场景

7.从数据库架构设计来看，主要有以下哪些设计思路？

- A.Shared-Disk

B.Shared-Everying

C.Shared-Nothing

D.以上全正确

8.Kettle 中有两种脚本文件 transformation（转换 ktr 结尾）和 job（任务 kjb 结尾）

A.正确

B.错误

9.哪些方式能够操作 HBase 的数据？

A.客户端命令

B.Phoenix SQL

C.Java API

D.Hives QL

10.下列关于 Explain 的关键字描述正确的是？

A.在 Explain 的基础上，通过添加 Analyze 同时显示该语句执行时每个算子的实际开销（单位 B 毫秒）和行数信息，对于判断优化器的估算是否准确提供依据

B.在 Explain 的基础上，添加 Verbose 显示出计划的额外详细信息，包括每个算子的输出列，重分布算子的分布列等

C.Explain 通过指定 option 选项参数 FORMAT 来指定输出格式，默认是 text 格式

D.在 Explain 的基础上，添加 Performance 显示出计划的额外详细信息，包括每个算子的输出列，重分布算子的分布列等