

ALIBABA CLOUD

阿里云

Databricks 数据洞察
常见问题

文档版本：20201014

(-) 阿里云

法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的“现状”、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或惩罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。未经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含“阿里云”、“Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

格式	说明	样例
 危险	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	 危险 重置操作将丢失用户配置数据。
 警告	该类警示信息可能会导致系统重大变更甚至故障，或者导致人身伤害等结果。	 警告 重启操作将导致业务中断，恢复业务时间约十分钟。
 注意	用于警示信息、补充说明等，是用户必须了解的内容。	 注意 权重设置为0，该服务器不会再接受新请求。
 说明	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	 说明 您也可以通过按Ctrl+A选中全部文件。
>	多级菜单递进。	单击设置>网络>设置网络类型。
粗体	表示按键、菜单、页面名称等UI元素。	在结果确认页面，单击 确定 。
Courier字体	命令或代码。	执行 <code>cd /d C:/window</code> 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	<code>bae log list --instanceid</code> <i>Instance_ID</i>
[] 或者 [a b]	表示可选项，至多选择一个。	<code>ipconfig [-all -t]</code>
{} 或者 {a b}	表示必选项，至多选择一个。	<code>switch {active stand}</code>

目录

1. 集群	05
2. Notebook	06
3. 数据源	07

1. 集群

本文将解答与集群有关的问题。

Q: 如何购买Databricks数据洞察集群

请使用阿里云账号进行购买，如需折扣优惠，请提工单或在产品群中联系我们享受折扣优惠。

Q: 可以使用阿里云子账号创建集群么？

可以。但子账号必须通过主账号进行授权方可购买，权限包括：AliyunEMRFullAccess、AliyunEMRDevelopAccess，请使用主账号到ram进行权限授予<https://ram.console.aliyun.com/users>。

Q: 集群创建后，可以登录集群的ECS么？

Databricks数据洞察是全托管的服务。用户在创建集群后，不能登录集群的ECS。用户不需要承担对集群环境的运维。

Q: Spark库、表可以在多集群之间共享么？

可以。Databricks数据洞察中的元数据是账号维度的，可以在同一个阿里云账号下的多集群之间共享。用户无需重复建表。

Q: 集群规模有限制么？

Databricks数据洞察的最小集群规模为4台。

2. Notebook

本文将解答与Notebook相关的问题

Q: Notebook里执行任务，不同用户之间是串行的么？

如果是SQL，是可以并行执行的；如果是Scala代码，是串行执行的，因为默认所有Note共享一个SparkContext。

相关配置：

参数	默认取值
zeppelin.spark.concurrentSQL	true
zeppelin.spark.currentSQL.max	10

Q: 如何使Note之间相互独立？

设置 per note isolated，使每个Note是一个独立的SparkContext。



Q: 如何配置依赖 (jars, python packages)

通过设置spark.jars 来设置依赖的jar包。jar包必须放在oss上，多个jar包之间通过逗号分隔。

通过设置spark.jars.packages 来设置依赖的package，格式：groupId:artifactId:version, 多个package逗号分隔。

Python package 需要通过Databricks数据洞察控制页面选择相应的python包

Q: 如何配置Spark Application 资源

通过以下属性配置driver的资源：

`spark.driver.memory`

`spark.driver.cores`

通过以下属性配置executor的资源：

`spark.executor.memory`

`spark.executor.cores`

通过以下属性配置executor数目：

`spark.executor.instances`

3.数据源

本文将解答与数据源相关的问题

Q: Databricks数据洞察的支持哪些数据源?

目前支持读写阿里云对象存储(OSS)的数据。

Q: 是否支持当前的DDI集群访问其他集群的HDFS数据?

支持访问同账号下的EMR Hadoop集群上的HDFS数据。