

# 【技术帖】Apache Kylin 高级设置：层级维度（Hierarchy Dimension）原理解析

原创 2017-04-14 施继成 apachekylin

点击上方蓝色 [apachekylin](#) 可以关注我哟

“ 为了缓解 Cube 的构建压力，减少生成的 Cuboid 数目，Apache Kylin 引入了一系列的高级设置，帮助用户筛选出真正需要的 Cuboid。这些高级设置包括聚合组（Aggregation Group）、联合维度（Joint Dimension）、层级维度（Hierarchy Dimension）和必要维度（Mandatory Dimension）等。

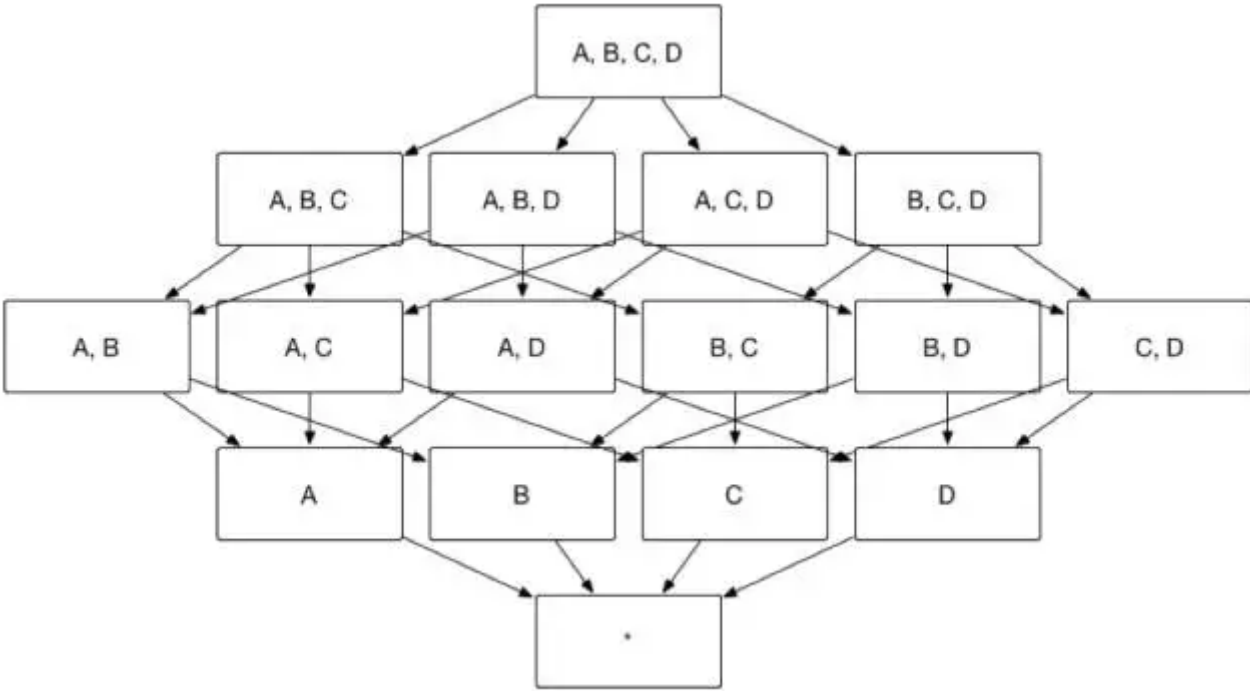
上一篇 Apache Kylin 高级设置技术帖介绍了联合维度（Joint Dimension）的实现原理与场景实例。本系列第三篇现如约而至，**将着重介绍层级维度（Hierarchy Dimension）的实现原理与应用场景实例。**

作者 | 施继成 翟鹿渊

编辑 | Zoe



众所周知，Apache Kylin 的主要工作就是为源数据构建 N 个维度的 Cube，实现聚合的预计算。理论上而言，构建 N 个维度的 Cube 会生成  $2^N$  个 Cuboid，如图 1 所示，构建一个 4 个维度（A，B，C，D）的 Cube，需要生成 16 个 Cuboid。



(图1)

随着维度数目的增加 Cuboid 的数量会爆炸式地增长，不仅占用大量的存储空间还会延长 Cube 的构建时间。为了缓解 Cube 的构建压力，减少生成的 Cuboid 数目，Apache Kylin 引入了一系列的高级设置，帮助用户筛选出真正需要的 Cuboid。这些高级设置包括聚合组（Aggregation Group）、联合维度（Joint Dimension）、层级维度（Hierarchy Dimension）和必要维度（Mandatory Dimension）等。

本文将着重介绍层级维度（Hierarchy Dimension）的实现原理与适用的场景实例。

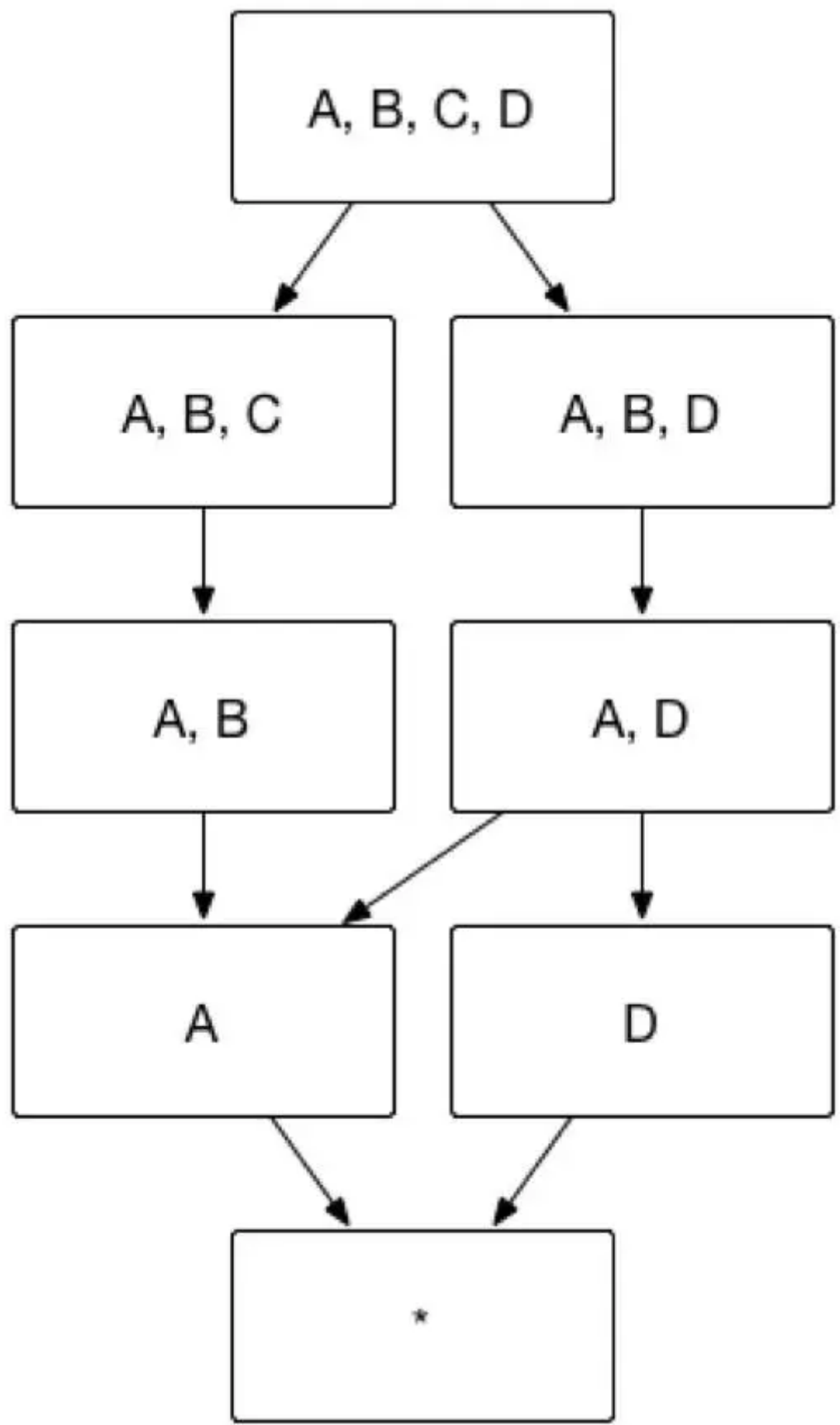
+

层级维度

用户选择的维度中常常会出现具有层级关系的维度。例如对于国家（country）、省份（province）和城市（city）这三个维度，从上而下来说国家 / 省份 / 城市之间分别是一对多的关系。也就是说，用户对于这三个维度的查询可以归类为以下三类：

- 1. group by country
- 2. group by country, province ( 等同于 group by province )
- 3. group by country, province, city ( 等同于 group by country, city 或者 group by city )

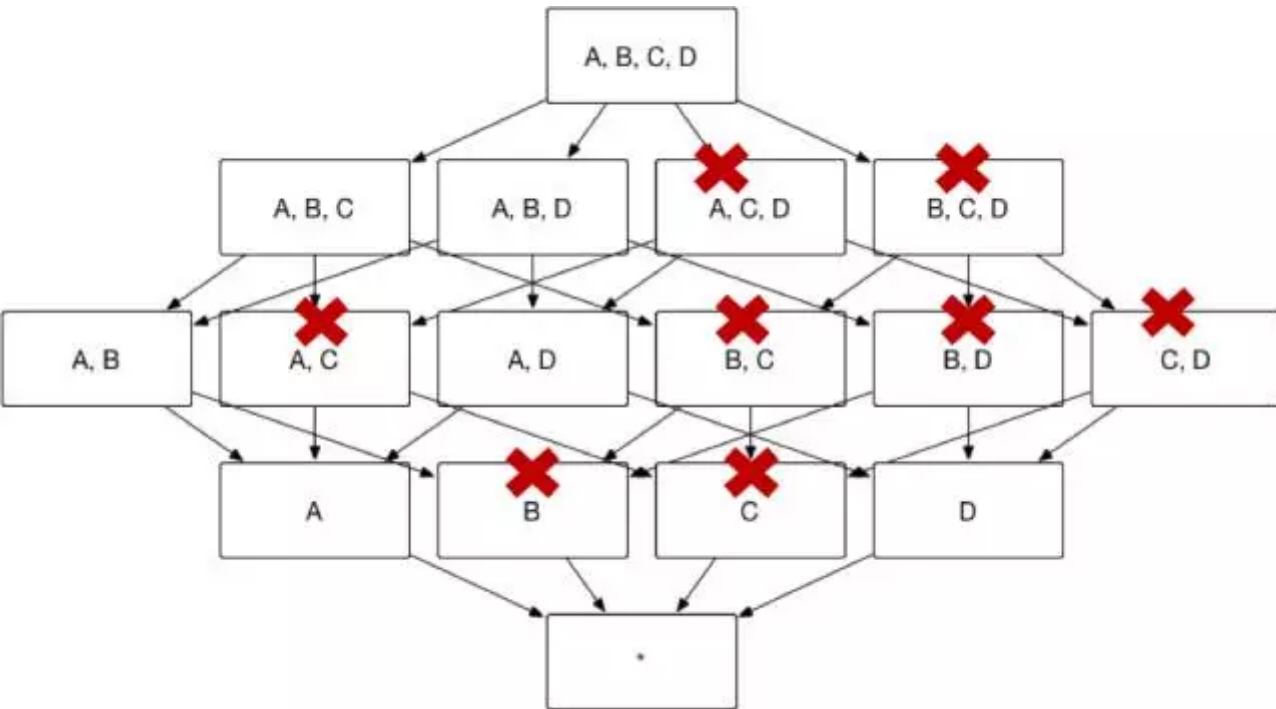
以图 2 所示的 Cube 为例，假设维度 A 代表国家，维度 B 代表省份，维度 C 代表城市，那么 ABC 三个维度可以被设置为层级维度，生成的Cube 如图 2 所示。



( 图2 )

例如 , Cuboid [A,C,D]=Cuboid[A, B, C, D] , Cuboid[B, D]=Cuboid[A, B, D] , 因而 Cuboid[A, C, D] 和 Cuboid[B, D] 就不必重复存储。

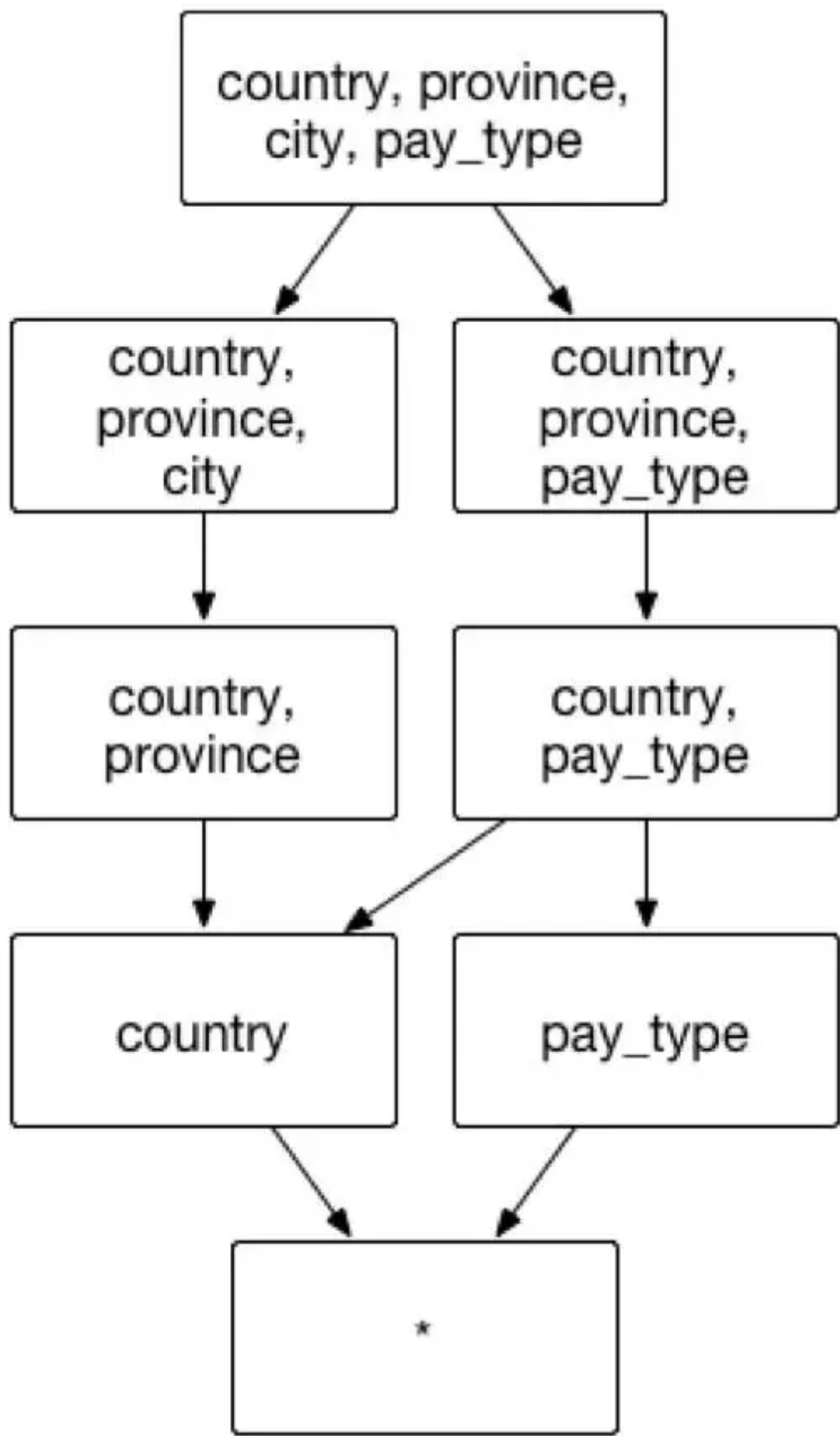
图 3 展示了 Kylin 按照前文的方法将冗余的Cuboid 剪枝从而形成图 2 的 Cube 结构 , Cuboid 数目从 16 减小到 8。



( 图3 )

应用实例

假设一个交易数据的 Cube , 它具有很多普通的维度 , 像是交易的城市 city , 交易的省 province , 交易的国家 country , 和支付类型 pay\_type等。分析师可以通过按照交易城市、交易省份、交易国家和支付类型来聚合 , 获取不同层级的地理位置消费者的支付偏好。在上述的实例中 , 建议在已有的聚合组中建立一组层级维度 ( 国家country / 省province / 城市city ) , 包含的维度和组合方式如图 4 :



( 图4 )

聚合组：[country, province, city , pay\_type]  
层级维度：[country, province, city]

Case 1 当分析师想从城市维度获取消费偏好时：

*SELECT city, pay\_type, count(\*) FROM table GROUP BY city, pay\_type* 则它将从 Cuboid [country, province, city, pay\_type] 中获取数据。

#### Case 2 当分析师想从省级维度获取消费偏好时：

*SELECT province, pay\_type, count(\*) FROM table GROUP BY province, pay\_type* 则它将从 Cuboid [country, province, pay\_type] 中获取数据。

#### Case 3 当分析师想从国家维度获取消费偏好时：

*SELECT country, pay\_type, count(\*) FROM table GROUP BY country, pay\_type* 则它将从 Cuboid [country, pay\_type] 中获取数据。

#### Case 4 如果分析师想获取不同粒度地理维度的聚合结果时：

无一例外都可以由图 3 中的 cuboid 提供数据。

例如，*SELECT country, city, count(\*) FROM table GROUP BY country, city* 则它将从 Cuboid [country, province, city] 中获取数据。

小结

Kylin 作为一种多维分析工具，采用预计算的方法，利用空间换取时间，提高查询效率。本文介绍了 Kylin 的高级设置中层级维度的部分，层级维适用于维度间有一对多关系的场景，比如国家 / 省 / 城市，产品大类 / 产品子类等。而本系列最后一篇文章，我们将就高级设置中必要维度的使用方法和使用场景做详细介绍，敬请期待。

插播福利

技术干货没看够？来这里就对了！

今日头条 + Kyligence 两家技术团队强强联手独家解密最新大数据平台技术！

Meetup @ 北京线下交流  
更有全宇宙网上平台直播





# Apache Kylin & 今日头条 Meetup @ 北京

🕒 2017年4月22日 | 📍 北京

## 线上线下同步进行

🕒 2017年4月22日

线上

请先报名网络直播票，截图报名信息后，加工作人员微信：cicixing1206  
拉你进入活动直播群。活动开始后，会后直播链接发出。

线下

北京 海淀区北三环西路43号 中航 广场1号楼 今日头条

热爱技术的你不可错过  
快点击阅读原文报名吧

您可能还会想看

【技术帖】Apache Kylin 高级设置：联合维度（Joint Dimension）原理解析

【技术帖】Apache Kylin 高级设置：聚合组（Aggregation Group）原理解析

【技术贴】如何部署Apache Kylin集群实现负载均衡？

【技术帖】Apache Kylin v2.0.0 Beta尝鲜版上线！！！！

【福利帖】《Apache Kylin权威指南》正式发售



长按图片识别二维码关注  
Apache Kylin官方公众号



点击“阅读原文” 报名今日头条 + Kyligence Meetup@北京

阅读原文