

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

原创 2017-04-07 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。

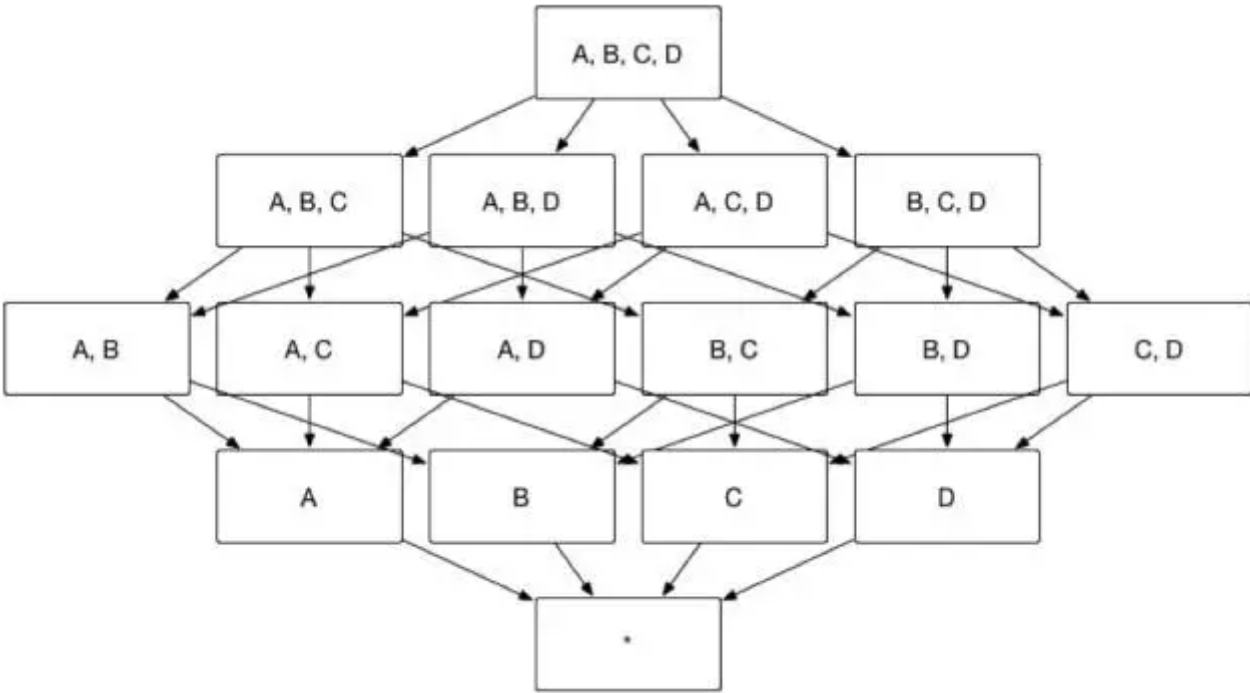
上一篇 Apache Kylin 高级设置技术帖介绍了聚合组（Aggregation Group）的实现原理与场景实例。本系列第二篇现如约而至，将**着重介绍联合维度（Joint Dimension）的实现原理与应用场景实例**。

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。



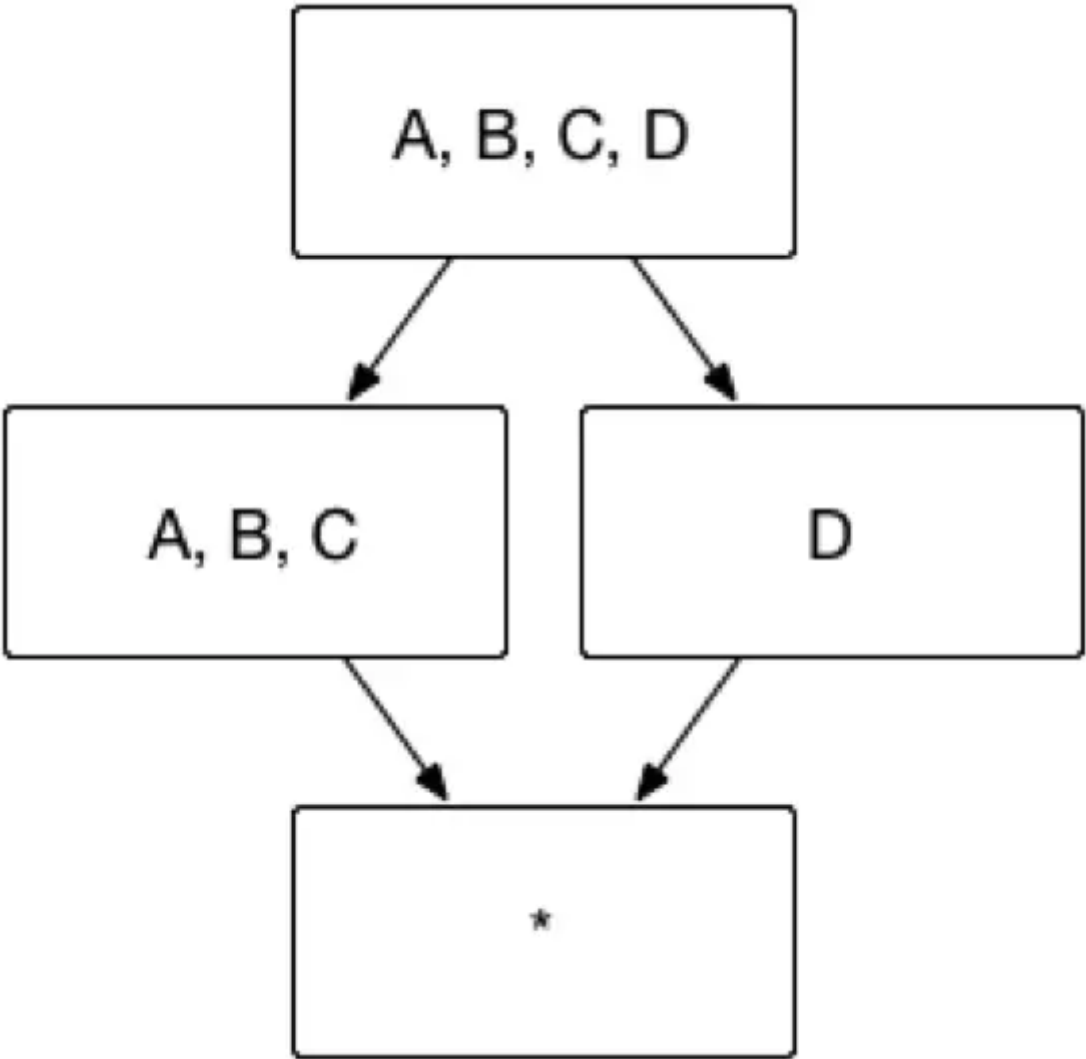
（图1）

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，本系列将深入讲解这些高级设置的含义及其适用的场景。

本文将着重介绍联合维度的实现原理与应用场景实例。

联合维度  
.....

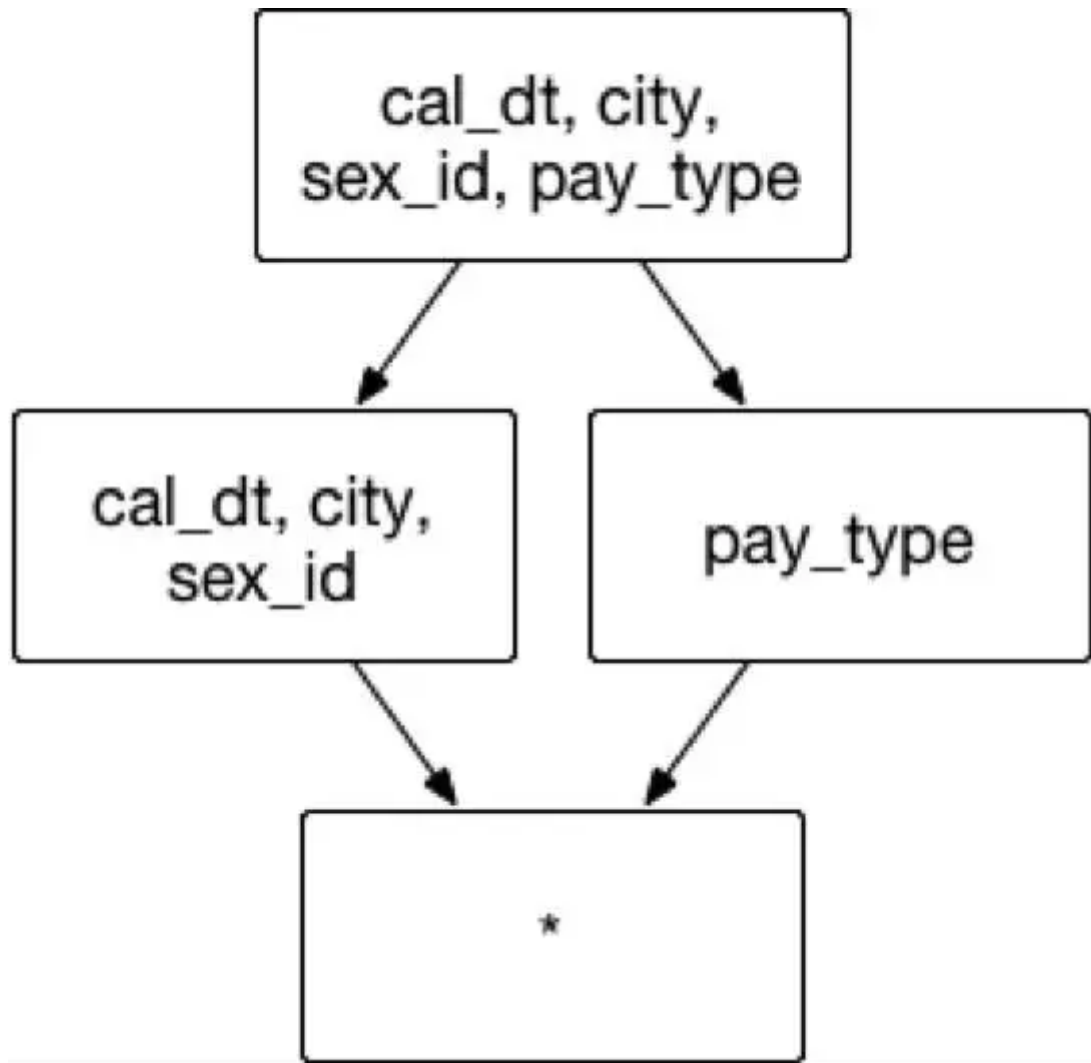
用户有时并不关心维度之间各种细节的组合方式，例如用户的查询语句中仅仅会出现 group by A, B, C，而不会出现 group by A, B 或者 group by C 等等这些细化的维度组合。这一类问题就是联合维度所解决的问题。例如将维度 A、B 和 C 定义为联合维度，Apache Kylin 就仅仅会构建 Cuboid ABC，而 Cuboid AB、BC、A 等等 Cuboid 都不会被生成。最终的 Cube 结果如图 2 所示，Cuboid 数目从 16 减少到 4。



( 图2 )

应用实例

假设创建一个交易数据的Cube，它具有很多普通的维度，像是交易日期 cal\_dt，交易的城市 city，顾客性别 sex\_id 和支付类型 pay\_type 等。分析师常用的分析方法为通过按照交易时间、交易地点和顾客性别来聚合，获取不同城市男女顾客间不同的消费偏好，例如同时聚合交易日期 cal\_dt、交易的城市 city 和顾客性别 sex\_id来分组。在上述的实例中，推荐在已有的聚合组中建立一组联合维度，包含的维度和组合方式如图3：



( 图3 )

聚合组：[cal\_dt, city, sex\_id, pay\_type]

联合维度：[cal\_dt, city, sex\_id]

### Case 1：

*SELECT cal\_dt, city, sex\_id, count(\*) FROM table GROUP BY cal\_dt, city, sex\_id* 则它将从 Cuboid [cal\_dt, city, sex\_id] 中获取数据

### Case2如果有一条不常用的查询：

*SELECT cal\_dt, city, count(\*) FROM table GROUP BY cal\_dt, city* 则没有现成的完全匹配的 Cuboid，Apache Kylin 会通过在线计算的方式，从现有的 Cuboid 中计算出最终结果。

### 小结

.....

Apache Kylin 作为一种多维分析工具，其采用预计算的方法，利用空间换取时间，提高查询效率。本文介绍了 Apache Kylin 的高级设置中联合维度的部分，联合维度适用于固定用来分组查

询的维度。之后的文章我们还将就 Apache Kylin 其他的高级设置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**

MEETUP@北京

# 今日头条+Kyligence

北京线下交流      网上平台直播

2017年4月22日 13:00 ~ 17:30

北京市海淀区北三环西路43号

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

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：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

【技术贴】揭秘Apache Kylin V2.0新特性：字典编码模块的优化



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