

【技术帖】Apache Kylin高级设置：必要维度（Mandatory Dimension）原理解析

原创 2017-04-21 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层级维度（Hierarchy Dimension）的实现原理与场景实例。本系列最后一篇现如约而至，将着重介绍必要维度（Mandatory 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。

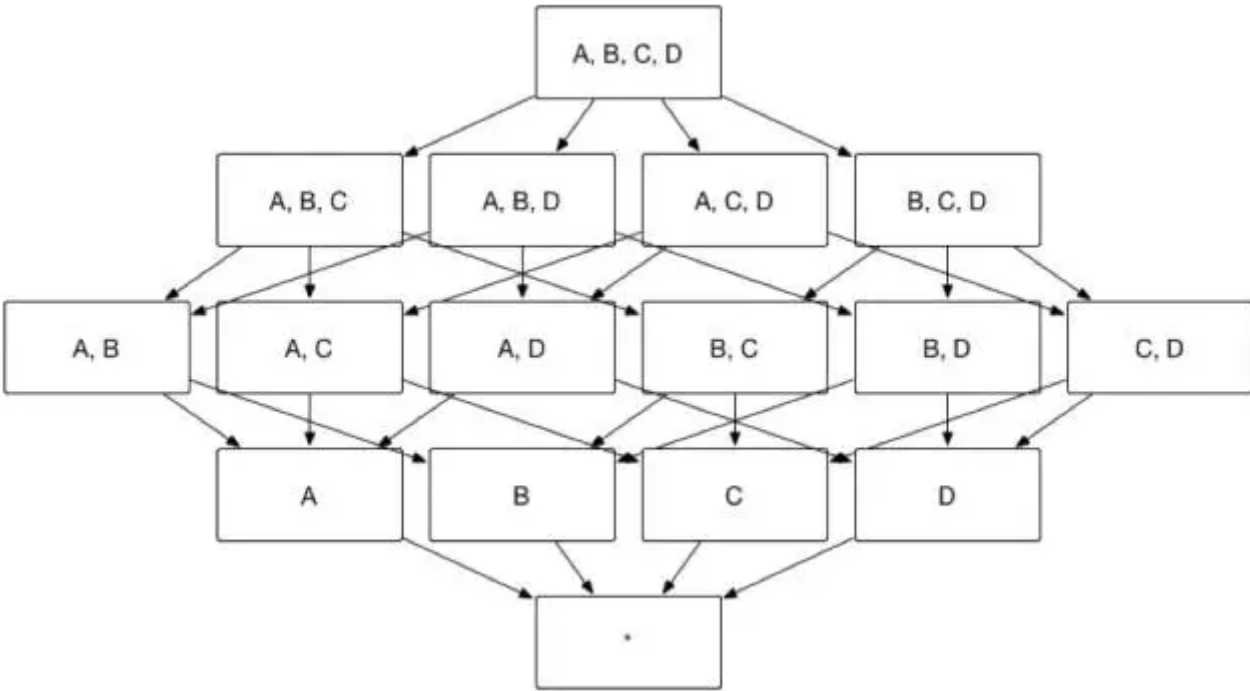


图1

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，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。

本文将着重介绍必要维度（Mandatory Dimension）的实现原理与适用的场景实例。

必要维度
.....

用户有时会对某一个或几个维度特别感兴趣，所有的查询请求中都存在group by这个维度，那么这个维度就被称为必要维度，只有包含此维度的Cuboid会被生成（如图2）。

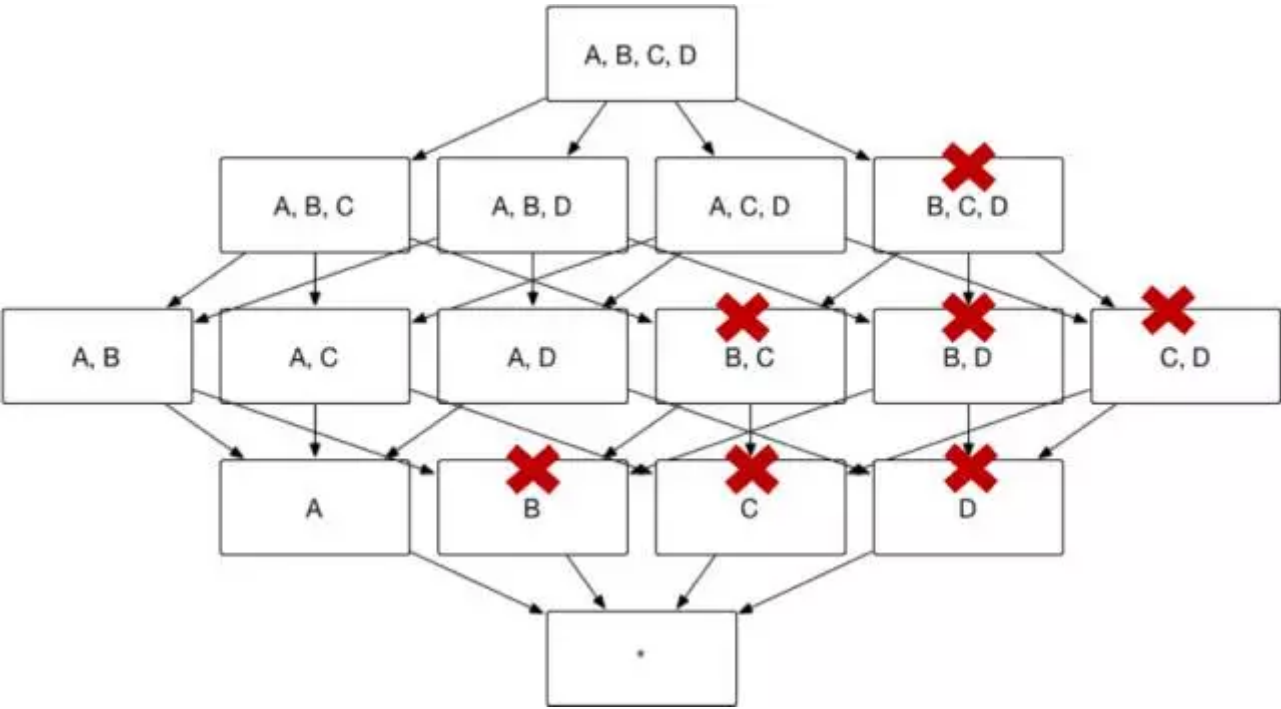


图2

以图 1中的Cube为例，假设维度A是必要维度，那么生成的Cube则如图3所示，维度数目从16变为9。

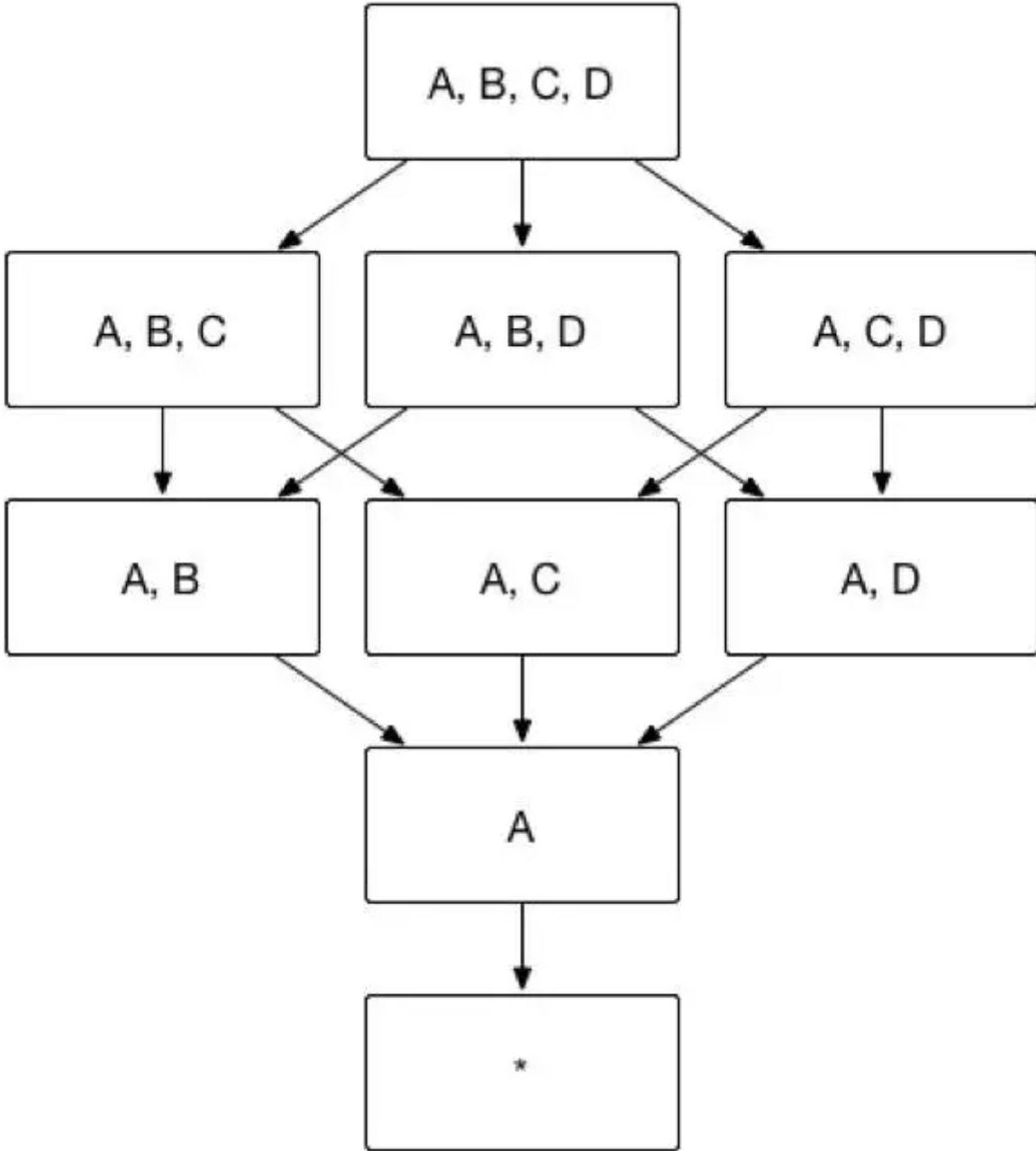


图3

应用实例

假设一个交易数据的Cube，它具有很多普通的维度，像是交易时间order_dt，交易的地点location，交易的商品product和支付类型pay_type等。其中，交易时间就是一个被高频作为分组条件（group by）的维度。如果将交易时间order_dt设置为必要维度，包含的维度和组合方式如图4：

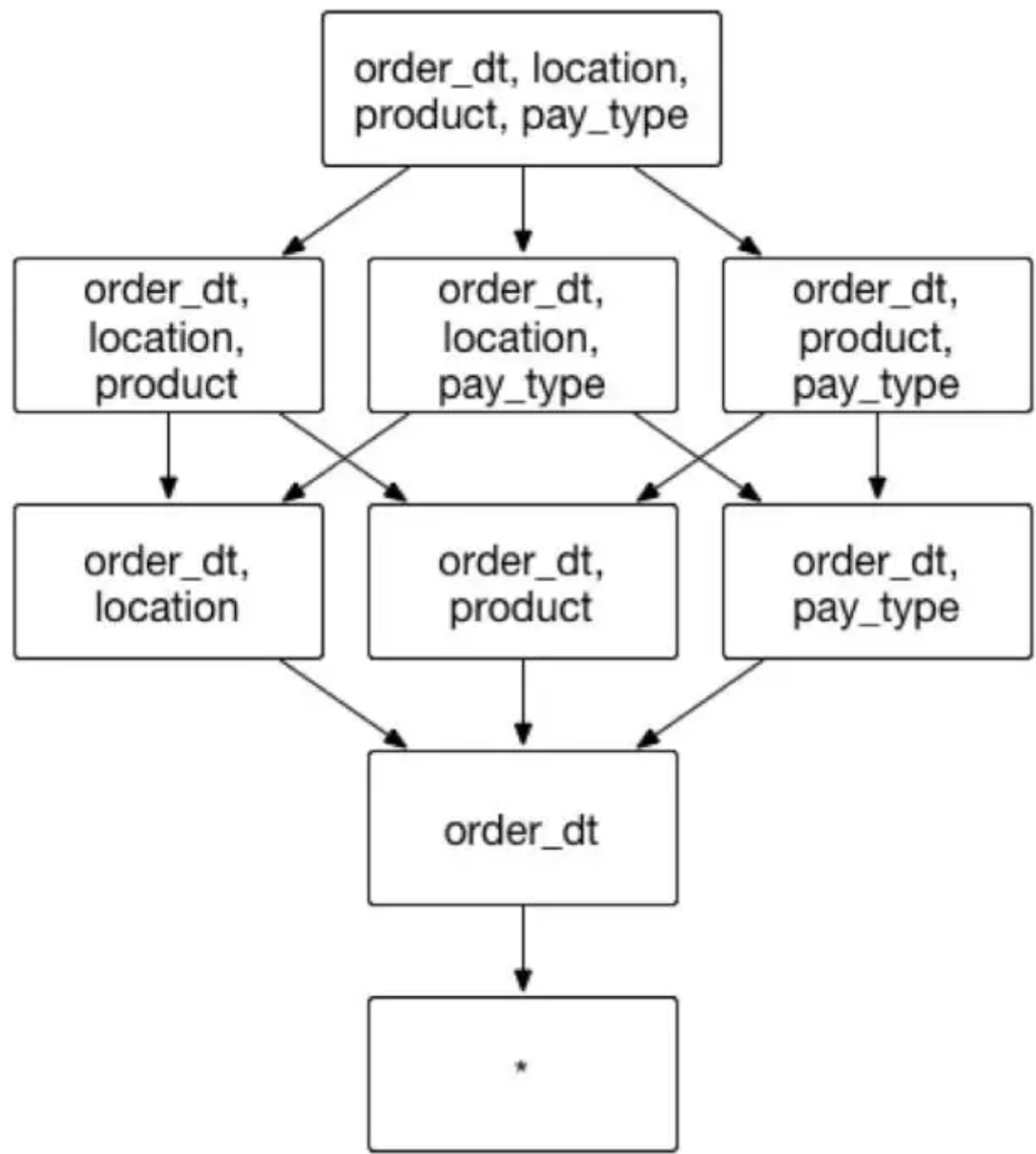


图4

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的高级设置中必要维度的部分，必要维度适用于某些维度被高频使用的情景下。

系列总结

.....

根据本系列的原理介绍，在Kylin的高级设置中，用户可以根据查询需求对Cube构建预计算的结果进行优化（剪枝），从而减少占用的存储空间。而优化得当的Cube可以在占用尽量少的存储空间的同时提供极强的查询性能。

您可能还会想看

.....

【技术帖】入门：论一个好Cube的养成

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

【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

【技术帖】Apache Kylin v2.0.0 Beta尝鲜版上线！！



长按图片识别二维码关注
Apache Kylin官方公众号