私有数据

什么是私有数据?

如果一个通道上的一组组织需要对该通道上的其他组织保持数据私有，则可以选择创建一个新通道，其中只包含需要访问数据的组织。但是，在每种情况下创建单独的通道会产生额外的管理开销（维护链码版本、策略、MSP等），并且不能在保留一部分数据私有的同时，可以让所有通道参与者看到该事务。

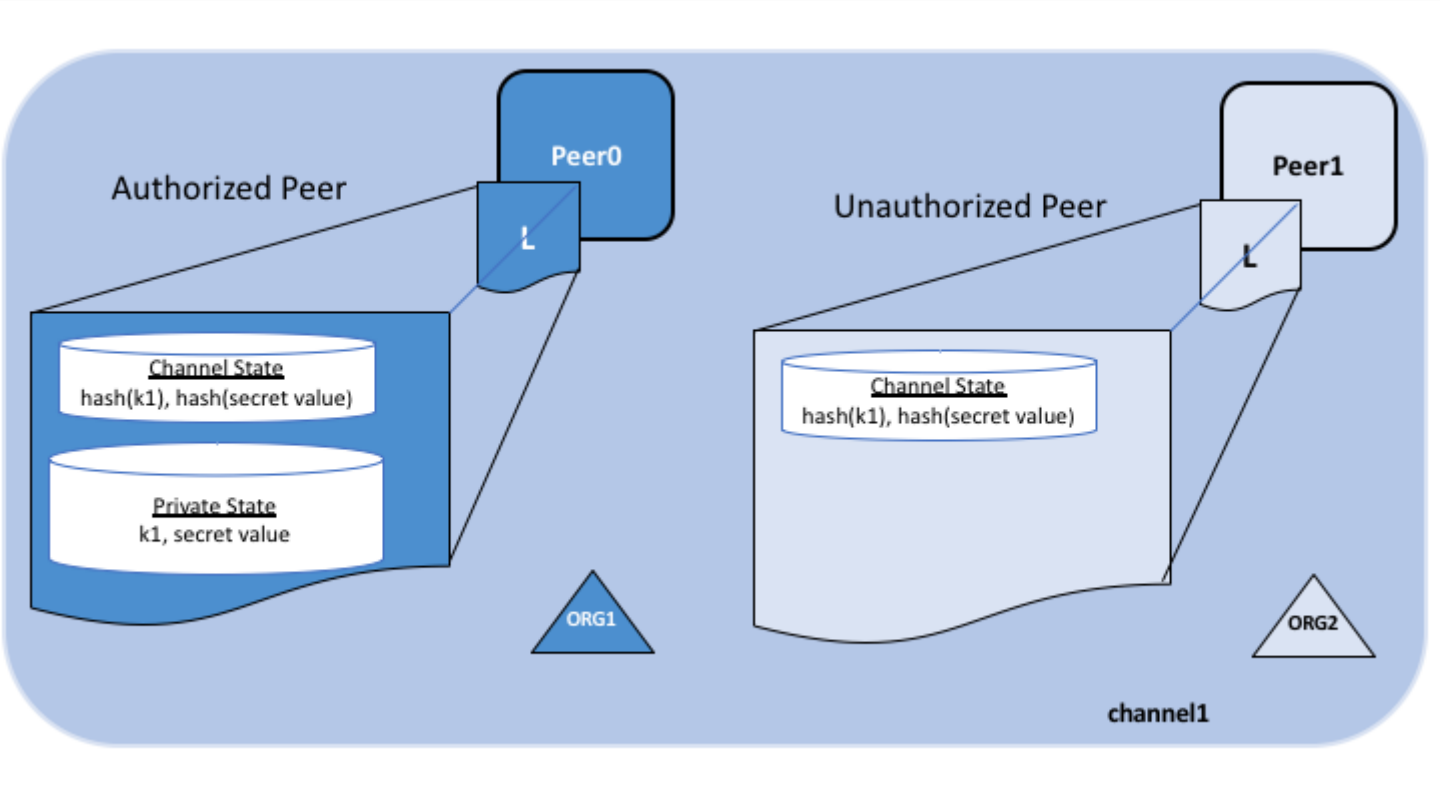
这就是为什么从v1.2开始，Fabric 提供了创建**私有数据集合**的功能，它允许在通道上定义的组织子集能够背书、提交或查询私有数据，而无需创建单独的通道。

什么是私有数据集合?

集合是两个元素的组合:

1. **实际的私有数据**，通过 Gossip 协议点对点地发送给授权可以看到它的组织。私有数据保存在被授权的组织的节点上的私有数据库上，它们可以被授权节点的链码访问。排序节点不能影响这里也不能看到私有数据。注意，由于 gossip 以点对点的方式向授权组织分发私有数据，所以必须设置通道上的锚节点，也就是每个节点上的 CORE_PEER_GOSSIP_EXTERNALENDPOINT 配置，以此来引导跨组织的通信。
2. **该数据的 hash 值**，该 hash 值被背书、排序之后写入通道上每个节点的账本。Hash 值作为交易的证明用于状态验证，并可用于审计。

下面的图表分别说明了被授权和未被授权拥有私有数据的节点的账本内容。



如果集合成员陷入争端，或者他们想把资产转让给第三方，他们可能决定与其他参与方共享私有数据。然后，第三方可以计算私有数据的 hash，并查看它是否与通道账本上的状态匹配，从而证明在某个时间点，集合成员之间存在该状态。

有些情况，你可能选择每个组织会有一套集合。比如一个组织可能会将私有数据记录到他们自己的集合中，之后可以共享给其他的通道成员并且在链码中引用。我们会在下边的共享私有数据部分看到一些例子。

什么时候使用一个通道内的组织集合，什么时候使用单独的通道

- 当必须将整个账本在属于通道成员的组织中保密时，使用**通道**比较合适。
- 当账本必须共享给一些组织，但是只有其中的部分组织可以在交易中使用这些数据的一部分或者全部时，使用**集合**比较合适。此外，由于私有数据是点对点传播的，而不是通过块传播的，所以在交易数据必须对排序服务节点保密时，应该使用私有数据集合。

解释集合的用例

设想一个通道上的五个组织，他们从事农产品贸易：

- **农民**在国外售卖他的货物
- **分销商**将货物运往国外
- **托运商**负责参与方之间的货物运输
- **批发商**向分销商采购商品
- **零售商**向托运人和批发商采购商品

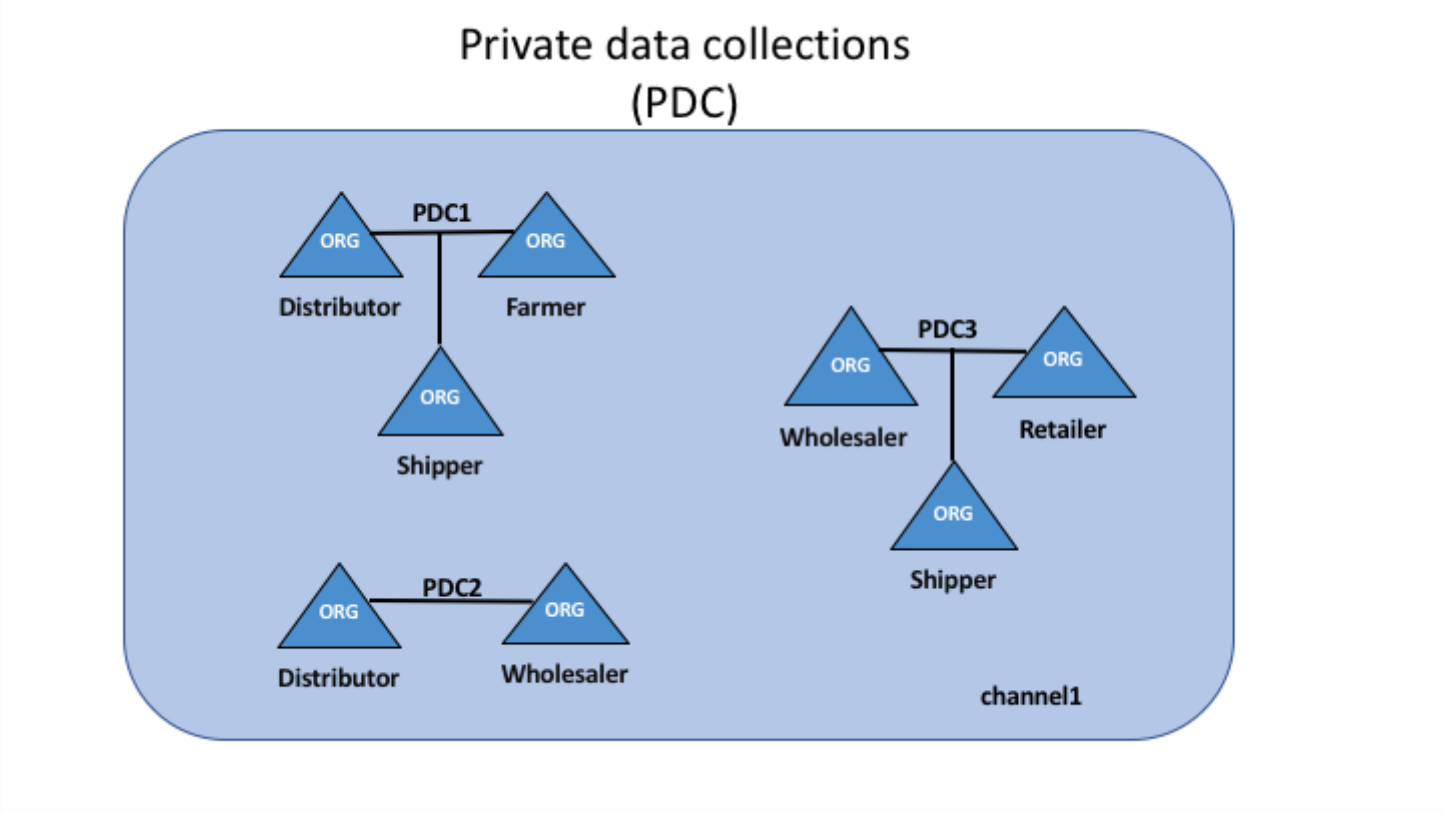
分销商可能希望与**农民**和**托运商**进行私下交易，以对**批发商**和**零售商**保密交易条款（以免暴露他们收取的加价）。

分销商还可能希望与**批发商**建立单独的私人数据关系，因为它收取的价格低于**零售商**。

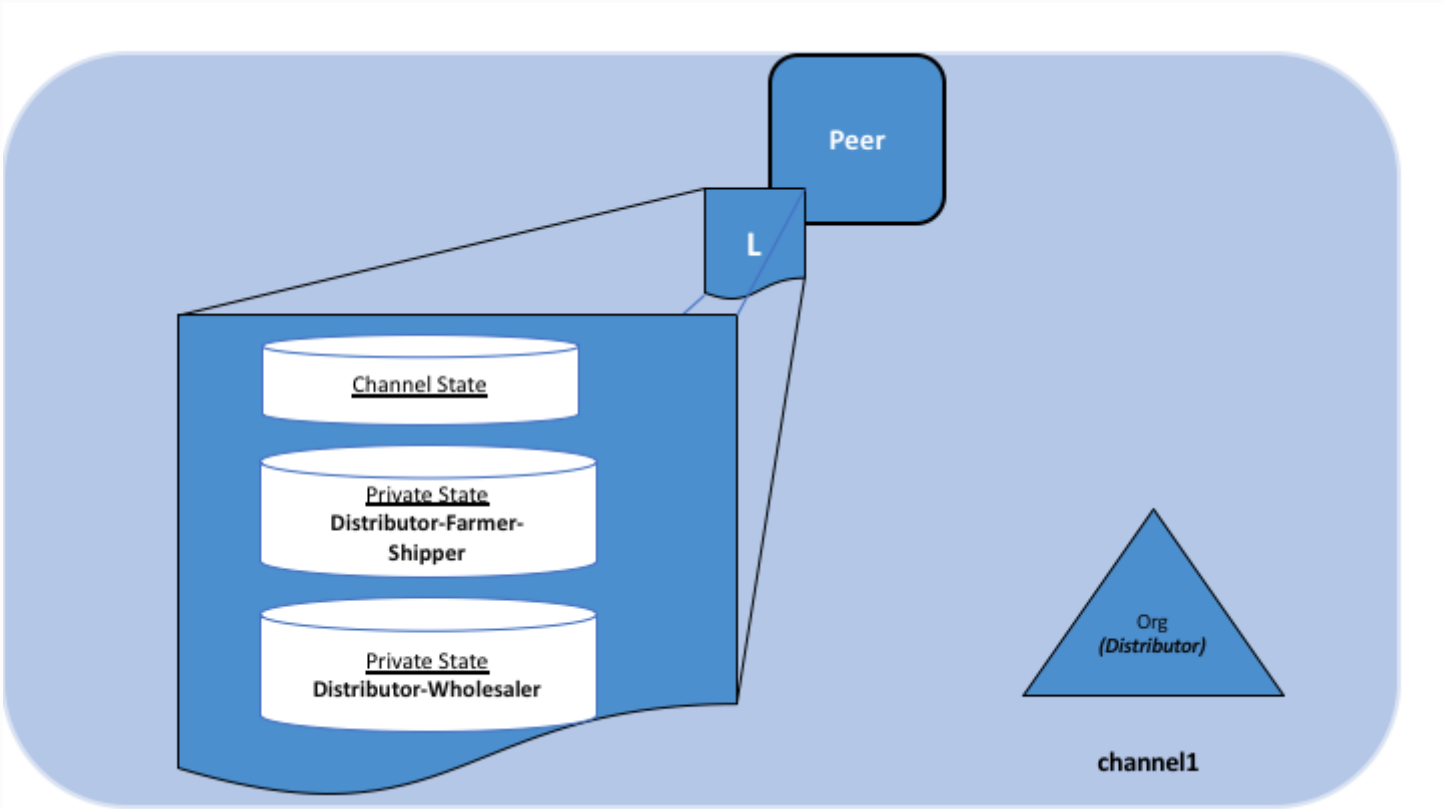
批发商可能还想与**零售商**和**托运商**建立私有数据关系。

相比于为每一个特殊关系建立各种小的通道来说，更好的做法是，可以定义多个私有数据集合（**PDC**），在以下情况下共享私有数据:

1. PDC1： **分销商**, **农民** 和 **托运商**
2. PDC2：**分销商** 和 **批发商**
3. PDC3：**批发商**, **零售商** 和 **托运商**



使用此示例，属于**分销商**的节点将在其账本中包含多个私有的私有数据库，其中包括来自**分销商**、**农民**和**托运商**子集合关系和**分销商**和**批发商**子集合关系的私有数据。



使用私有数据的交易流

当在链码中引用私有数据集合时，交易流程略有不同，以便在交易被提案、背书并提交到账本时保持私有数据的机密性。

关于不使用私有数据的交易流程的详细信息，请参阅我们的交易流程的文档。

1. 客户端应用程序提交一个提案请求，让属于授权集合的背书节点执行链码函数（读取或写入私有数据）。 私有数据，或用于在链码中生成私有数据的数据，被发送到提案的 `transient` （瞬态）字段中。
2. 背书节点模拟交易，并将私有数据存储在 瞬态数据存储 （ transient data store ， 节点的本地临时存储）中。它们根据组织集合的策略将私有数据通过gossip分发给授权的 Peer 节点。
3. 背书节点将提案响应发送回客户端。提案响应中包含经过背书的读写集，这其中包含了公共数据，还包含任何私有数据键和值的 hash。*私有数据不会被发送回客户端*。更多关于带有私有数据的背书的信息，请查看这里。
4. 客户端应用程序将交易（包含带有私有数据 hash 的提案响应）提交给排序服务。带有私有数据 hash 的交易同样被包含在区块中。带有私有数据 hash 的区块被分发给所有节点。这样，通道中的所有节点就可以在不知道真实私有数据的情况下，用同样的方式来验证带有私有数据 hash 值的交易。
5. 在区块提交的时候，授权节点会根据集合策略来决定它们是否有权访问私有数据。如果节点有访问权，它们会先检查自己的本地 瞬态数据存储 ，以确定它们是否在链码背书的时候已经接收到了私有数据。如果没有的话，它们会尝试从其他已授权节点那里拉取私有数据，然后对照公共区块上的 hash 来验证私有数据并提交交易和区块。当验证或提交结束后，私有数据会被移动到这些节点私有数据库和私有读写存储的副本中。随后 瞬态数据存储 中存储的这些私有数据会被删除。

共享私有数据

在多数情况下，一个私有数据集合中的私有数据键或值可能需要与其他通道成员或者私有数据集合共享，例如，当你需要和一个通道成员或者一组通道成员交易私有数据，而初始私有数据集合中并没有这些成员时。私有数据的接收方一般都会在交易过程中对照链上的 hash 来验证私有数据。

私有数据集合的以下几个方面促成了私有数据的共享和验证：

- 首先，只要符合背书策略，尽管你不是私有数据集的成员也可以为集合写入键。可以在链码层面，键层面（用基于状态的背书）或者集合层面（始于 Fabric v2.0）上定义背书策略。
- 其次，从 v1.4.2 开始出现了链码 API GetPrivateDataHash() ，它支持非集合成员节点上的链码读取一个私有键的 hash。下文中你将发现这是一个十分重要的功能，因为它允许链码对照之前交易中私有数据生成的链上 hash 来验证私有数据。

当设计应用程序和相关私有数据集合时需要考虑这种共享和验证私有数据的能力。您当然可以创建出几组多边私有数据集合以供多个通道成员组合之间共享数据，但是这样做的话你就需要定义大量私有数据集合。或者你也可以考虑使用少量私有数据集合（例如，每个组织用一个集合，或者每对组织用一个集合），然后和其他通道成员共享私有数据，有需要时还可以和其他集合共享私有数据。从 Fabric v2.0 开始，隐含的组织特定集合可供所有链码使用，这样一来部署链码的时候你就不用定义每个组织中的私有数据集合。

私有数据共享模式

为各组织的私有数据集合构建模型时，有多种模式可被用来共享或传输私有数据，且无需费力定义多个多边集合。以下是链码应用程序中可以使用的一些共享模式：

- **使用相应的公钥来追踪公共状态**：您可以选择使用一个相应的公钥来追踪特定的公共状态（例如：资产性质，当前所有权等公共状态），对于每个需要拥有资产相应私有数据访问权的组织，您可以在它们的私有数据集合中创建一个私有秘钥或值。

- **链码访问控制**：您可以在您的链码中执行访问控制，指明什么客户端应用程序能够查询私有数据集合中的私有数据。例如，为一个或多个私有数据集合键存储一个访问控制列表，然后在链码中获取客户端应用程序提交者的证书（使用 GetCreator() 链码 API 或 CID 库 API GetID() or GetMSPID() 来获取），并在返回私有数据之前验证这些证书。同样，为了访问私有数据，您可以要求客户端将密码短语传送到链码中，且该短语必须与存储在秘钥级别的密码短语相匹配。注意，这种模式也可用于限制客户端对公共状态数据的访问权。
- **使用带外数据来共享私有数据**：这是一种链下选项，您可以同其他组织以带外数据的形式共享私有数据，这些组织可使用 GetPrivateDataHash() 链码 API 将键或值转换成 hash 以验证其是否与链上 hash 匹配。例如，如果一个组织想要购买一份你的资产，那么在同意购买之前它会检查链上 hash，以验证如下事项：该资产的属性；你是否为该资产的合法所有人。
- **与其他集合共享私有数据**：您可以同链码在链上共享私有数据，该链码在其他组织的私有数据集合中生成一个相应的键或值。你将通过临时字段把私有数据键或值传送给链码，收到私有数据后该链码使用 GetPrivateDataHash() 验证此私有数据的 hash 是否与您集合中的链上 hash 一致，随后将该私有数据写入其他组织的私有数据集合中。
- **将私有数据传送给其他集合**：您可以使用链码来”分发“私有数据，该链码把您集合中的私有数据键删除，然后在其他组织的集合中生成。与上述方法相同，这里在调用链码时使用临时字段传输私有数据，并且在链码中用 GetPrivateDataHash() 来验证你的私有数据集合中是否存在该数据，验证成功之后再把你的集合中该数据删除，并在其他组织的集合中生成该键。为确保每次操作都会从一个集合中删除一项交易并在另一个集合中添加该交易，你可能需要获得一些额外组织的背书，如监管者或审计者。
- **使用私有数据进行交易确认**：如果您想在交易完成之前获得竞争对手的批准（即竞争对手同意以某价钱购买一项资产这一链上记录），链码会在竞争对手或您自己的私有数据集合中写入一个私有数据（链码随后将使用 GetPrivateDataHash() 来验证该数据），从而要求竞争对手”率先批准“这项交易。事实上，嵌入式生命周期系统链码就是使用这种机制来确保链码定义在被提交到通道之前已经获得通道上各组织的同意。这种私有数据共享模式从从 Fabric v2.0 开始凭借集合层面的背书策略变得越来越强大，确保在集合拥有者自己的信任节点上执行、背书链码。或者，您也可以使用具有秘钥层面背书策略、共同商定的秘钥，随后该秘钥被预先批准的条款更新，并且在指定组织的节点上获得背书。
- **使交易者保密**：对上一种共享模式进行些许变化还可以实现指定交易的交易者不被暴露。例如，买方表示同意在自己的私有数据集合上进行购买，随后在接下来的交易中卖方会在自己的私有数据集合中引用该买方的私有数据。带有 hash 过的引用的交易证据被记录在链上，只有出售者和购买者知道这些是交易，但是如果需要需要知道的原则的话，他们可以暴漏原像，比如在一个接下来的跟其他的伙伴进行交易时，对方就能够验证这个 hash 了。

结合以上几种模式，我们可以发现，私有数据的交易和普通的通道状态数据交易情况类似，特别是以下几点：

- **重要级别交易访问控制** - 您可以在私有数据值中加入 所有权证书，这样一来，后续发生的交易就能够验证数据提交者是否具有共享和传输数据的所有权。在这种情况下，链码会获取数据提交者的证书（例如：使用 GetCreator() 链码 API or CID library API GetID() or GetMSPID() ），将此证书与链码收到的其他私有数据合并，
- **重要级别背书策略** - 和正常的通道状态数据一样，您可以使用基于状态的背书策略来指明哪些组织必须对共享或转移私有数据的交易做出背书，使用 SetPrivateDataValidationParameter() 链码 API 来进行一些操作，例如，指明必须对上述交易作出背书的仅包括一个拥有者的组织节点，托管组织的节点或者第三方组织。

样例场景：使用私有数据集合的资产交易

将上述私有数据共享模式结合起来能够赋能基于链码的应用程序。例如，思考一下如何使用各组织私有数据集合来实现资产转移场景：

- 在公共链码状态使用 UUID 键可以追踪一项资产。所记录信息只包括资产的所属权，没有其他信息。
- 链码要求：所有资产转移请求必须源自拥有者客户端，并且关键在于需要有基于状态的背书，要求从所有者的组织和一个监管者的组织的一个 Peer 必须要为所有的交易请求背书。
- 资产拥有者的私有数据集合包含了有关该资产的私有信息，用一个 UUID 的 hash 作为键值。其他的组织和排序服务将会只能看到资产详细的 hash 。
- 假定监管者是各私有数据集合的一员，因此它一直维护着私有数据，即使他并没必要这么做。

一项资产交易的进行情况如下：

1. 链下，资产所有者和意向买家同意以某一特定价格交易该资产。
2. 卖家通过以下方式提供资产所有权证明：利用带外数据传输私有信息；或者出示证书以供买家在其自身节点或管理员节点上查询私有数据。
3. 买家验证私有信息的 hash 是否匹配链上公共 hash。
4. 买家通过调取链码来在其自身私有数据集合中记录投标细节信息。一般在买家自己的节点上调取链码，但如果集合背书策略有相关规定，则也可能会在管理员节点上调取链码。
5. 当前的资产所有者（卖家）调取链码来卖出、转移资产，传递私有信息和投标信息。在卖家、买家和管理员的节点上调用链码，以满足公钥的背书策略以及买家和买家私有数据集合的背书策略。
6. 链码验证交易发送方是否为资产拥有者，根据卖家私有数据集合中的 hash 来验证私有细节信息，同时还会根据买家私有数据集合中的 hash 来验证投标细节信息。随后链码为公钥书写提案更新（将资产拥有者设定为买家，将背书策略设定为买家和管理员），把私有细节信息写入买家的私有数据集合中，以上步骤成功完成后链码会把卖家私有数据集合中的相关私有细节信息删除。在最终背书之前，背书节点会确保已将私有数据分布给卖家和管理员的所有授权节点。
7. 卖家提交交易等待排序，其中包括了公钥和私钥 hash，随后该交易被分布到区块中的所有通道节点上。
8. 每个节点的区块验证逻辑都将一致性地验证背书策略是否得到满足（买家、卖家和管理员都作出背书），同时还验证链码中已读取的公钥和私钥自链码执行以来未被任何其他交易更改。
9. 所有节点提交的交易都是有效的，因为交易已经通过验证。如果买家和管理员节点在背书时未收到私有数据的话，它们将会从其他授权节点那里获取这些私有数据，并将这些数据保存在自己的私有数据状态数据库中（假定私有数据与交易的 hash 匹配）。
10. 交易完成后，资产被成功转移，其他对该资产感兴趣的通道成员可能会查询公钥的历史以了解该资产的来历，但是它们无法访问任何私有细节信息，除非资产拥有者在须知的基础上共享这些信息。

以上是最基本的资产转移场景，我么可以对其进行扩展，例如，资产转移链码可能会验证一项付款记录是否可用于满足付款和交付要求，或者可能会验证一个银行是否在执行资产转移链码之前已经提交了信用证。交易各方并不直接维护节点，而是通过那些运行节点的托管组织来进行交易。

清除私有数据

对于非常敏感的数据，即使是共享私有数据的各方可能也希望（或者应政府相关法规要求必须）定期“清除”节点上的数据，仅把这些敏感数据的 hash 留在区块链上，作为私有数据不可篡改的证据。

在某些情况下，私有数据只需要在其被复制到节点区块链外部的数据库之前存在于该节点的私有数据库中。而数据或许也只需要在链码业务流程结束之前存在于节点上（交易结算、合约履行等）。

为了支持这些用户案例，如果私有数据已经持续 N 个块都没有被修改，则可以清除该私有数据，N 是可配置的。链码中无法查询已被清除的私有数据，并且其他节点也请求不到。

私有数据集合怎么定义

有关集合定义的更多详细信息，以及关于私有数据和集合的其他更底层的信息，请参阅私有数据主题。