一句话：分布式是并联工作的，集群是串联工作的。

**1.分布式是指将不同的业务分布在不同的地方。 而集群指的是将几台服务器集中在一起，实现同一业务。**

分布式中的每一个节点，都可以做集群。 而集群并不一定就是分布式的。

举例：就比如新浪网，访问的人多了，他可以做一个群集，前面放一个响应服务器，后面几台服务器完成同一业务，如果有业务访问的时候，响应服务器看哪台服务器的负载不是很重，就将给哪一台去完成。

而分布式，从窄意上理解，也跟集群差不多， 但是它的组织比较松散，不像集群，有一个组织性，一台服务器垮了，其它的服务器可以顶上来。

分布式的每一个节点，都完成不同的业务，一个节点垮了，哪这个业务就不可访问了。

**2.简单说，分布式是以缩短单个任务的执行时间来提升效率的，而集群则是通过提高单位时间内执行的任务数来提升效率。**

简单的例子说明一下：

如果一个任务由10个子任务组成，每个子任务单独执行需1小时，则在一台服务器上执行该任务需10小时。

采用分布式方案，提供10台服务器，每台服务器只负责处理一个子任务，不考虑子任务间的依赖关系，执行完这个任务只需一个小时。(这种工作模式的一个典型代表就是Hadoop的Map/Reduce分布式计算模型）

而采用集群方案，同样提供10台服务器，每台服务器都能独立处理这个任务。假设有10个任务同时到达，10个服务器将同时工作，1小时后，10个任务同时完成，这样，整身来看，还是1小时内完成一个任务！

**二.集群概念**

**1. 两大关键特性**

集群是一组协同工作的服务实体，用以提供比单一服务实体更具扩展性与可用性的服务平台。在客户端看来，一个集群就象是一个服务实体，但事实上集群由一组服务实体组成。与单一服务实体相比较，集群提供了以下两个关键特性：

**a.可扩展性：**集群的性能不限于单一的服务实体，新的服务实体可以动态地加入到集群，从而增强集群的性能。

**b.高可用性：**集群通过服务实体冗余使客户端免于轻易遇到out of service的警告。在集群中，同样的服务可以由多个服务实体提供。如果一个服务实体失败了，另一个服务实体会接管失败的服务实体。集群提供的从一个出 错的服务实体恢复到另一个服务实体的功能增强了应用的可用性。

**2. 两大能力**

为了具有可扩展性和高可用性特点，集群的必须具备以下两大能力：

**a.负载均衡：**负载均衡能把任务比较均衡地分布到集群环境下的计算和网络资源。

**b.错误恢复：**由于某种原因，执行某个任务的资源出现故障，另一服务实体中执行同一任务的资源接着完成任务。这种由于一个实体中的资源不能工作，另一个实体中的资源透明的继续完成任务的过程叫错误恢复。

负载均衡和错误恢复都要求各服务实体中有执行同一任务的资源存在，而且对于同一任务的各个资源来说，执行任务所需的信息视图（信息上下文）必须是一样的。

**3. 两大技术**

实现集群务必要有以下两大技术：

**a.集群地址：**集群由多个服务实体组成，集群客户端通过访问集群的集群地址获取集群内部各服务实体的功能。具有单一集群地址（也叫单一影像）是集群的一个基 本特征。维护集群地址的设置被称为负载均衡器。负载均衡器内部负责管理各个服务实体的加入和退出，外部负责集群地址向内部服务实体地址的转换。有的负载均 衡器实现真正的负载均衡算法，有的只支持任务的转换。只实现任务转换的负载均衡器适用于支持ACTIVE-STANDBY的集群环境，在那里，集群中只有 一个服务实体工作，当正在工作的服务实体发生故障时，负载均衡器把后来的任务转向另外一个服务实体。

**b.内部通信：**为了能协同工作、实现负载均衡和错误恢复，集群各实体间必须时常通信，比如负载均衡器对服务实体心跳测试信息、服务实体间任务执行上下文信息的通信。

具有同一个集群地址使得客户端能访问集群提供的计算服务，一个集群地址下隐藏了各个服务实体的内部地址，使得客户要求的计算服务能在各个服务实体之间分布。内部通信是集群能正常运转的基础，它使得集群具有均衡负载和错误恢复的能力。

**三.集群分类**

Linux集群主要分成三大类( 高可用集群， 负载均衡集群，科学计算集群)

**1. 高可用集群(High Availability Cluster)**

常见的就是2个节点做成的HA集群，有很多通俗的不科学的名称，比如"双机热备", "双机互备", "双机".

高可用集群解决的是保障用户的应用程序持续对外提供服务的能力。 (请注意高可用集群既不是用来保护业务数据的，保护的是用户的业务程序对外不间断提供服务，把因软件/硬件/人为造成的故障对业务的影响降低到最小程度)。

**2. 负载均衡集群(Load Balance Cluster)**

负载均衡系统：集群中所有的节点都处于活动状态，它们分摊系统的工作负载。一般Web服务器集群、数据库集群和应用服务器集群都属于这种类型。

负载均衡集群一般用于相应网络请求的网页服务器，数据库服务器。这种集群可以在接到请求时，检查接受请求较少，不繁忙的服务器，并把请求转到这些服务器上。从检查其他服务器状态这一点上看，负载均衡和容错集群很接近，不同之处是数量上更多。

**3. 科学计算集群(High Performance Computing Cluster)**

高性能计算(High Perfermance Computing)集群，简称HPC集群。这类集群致力于提供单个计算机所不能提供的强大的计算能力。

**四. 分布式（集群）与集群的联系与区别**

**分布式是指将不同的业务分布在不同的地方。**

**而集群指的是将几台服务器集中在一起，实现同一业务。**

**分布式中的每一个节点，都可以做集群。**

**而集群并不一定就是分布式的。**

举例：就比如新浪网，访问的人多了，他可以做一个群集，前面放一个响应服务器，后面几台服务器完成同一业务，如果有业务访问的时候，响应服务器看哪台服务器的负载不是很重，就将给哪一台去完成。

而分布式，从窄意上理解，也跟集群差不多， 但是它的组织比较松散，不像集群，有一个组织性，一台服务器垮了，其它的服务器可以顶上来。

分布式的每一个节点，都完成不同的业务，一个节点垮了，哪这个业务就不可访问了。

以上就是本文的全部内容，希望本文的内容对大家的学习或者工作能带来一定的帮助，同时也希望多多支持脚本之家！