

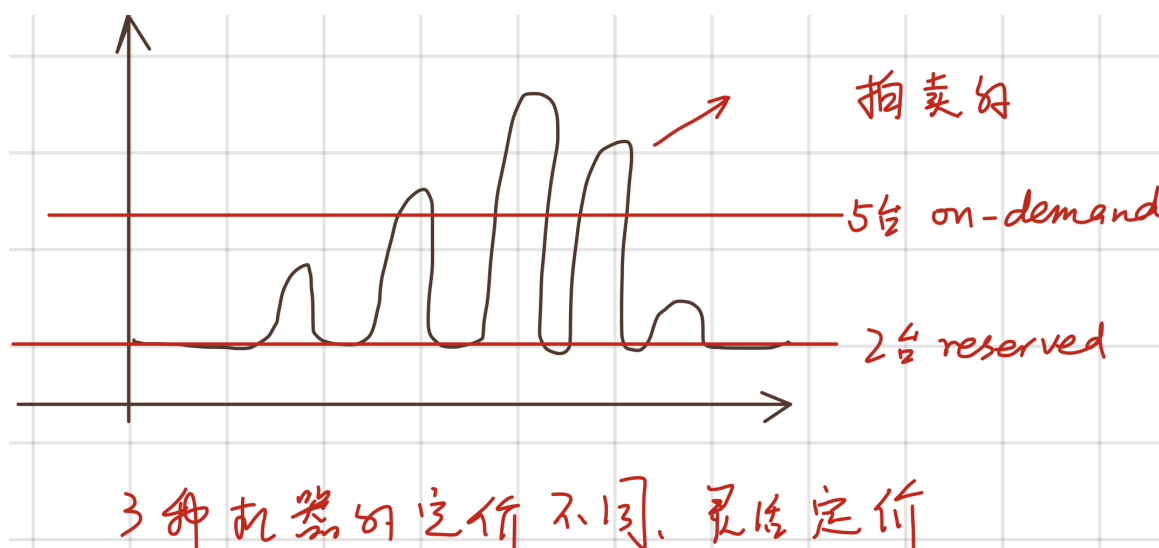
# lect24: Cloud Computing

云计算的特点：

1. 按需收费
2. 弹性扩展，可以增加资源
3. 管理的复杂性交给了云的提供商
4. 虚拟化，大家的需求（操作系统，内核数，资源）不一样，要模拟出这样的机器。

– Common attributes of clouds

- Flexible pricing（价格灵活）



- Elastic scaling（弹性伸缩）
- Rapid provisioning（快速供给）
  - 拿到新的服务器之后马上部署，就像容器分层，一些基础的层大家的机器都有，一些依赖下载一下就可以。
- Advanced virtualization

云服务的种类

SaaS	Software-as-a-Service	Google Apps, Microsoft "Software+Services" 百度地图, 支付宝API
PaaS	Platform-as-a-Service	IBM IT Factory, Google AppEngine, Force.com
IaaS	Infrastructure-as-a-Service	基础设施 Amazon EC2, IBM Blue Cloud, Sun Grid
dSaaS	data-Storage-as-a-Service	Nirvanix SDN, Amazon S3, Cleversafe dsNet 网盘

现实中：

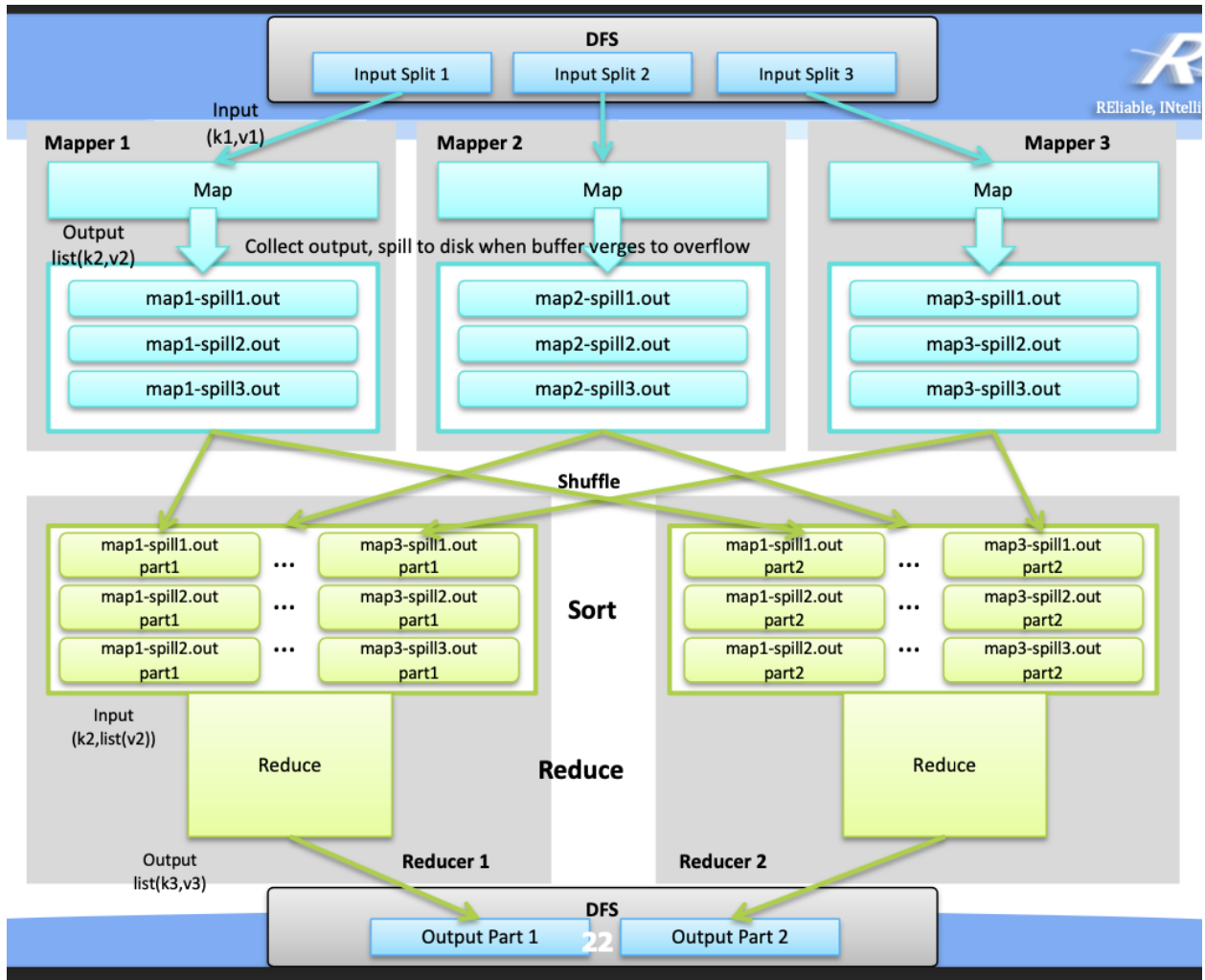
不用的时候要节电。

服务检测它不行了，要迁移，怎么迁移，到哪里。

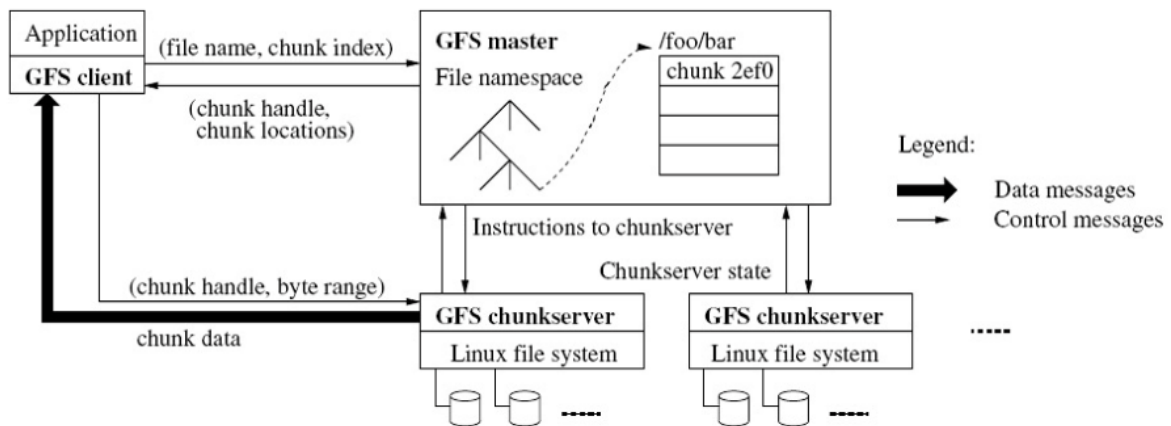
server挂了， 分布式要保证客户端看不到，要容错。

## 作业调度 → mapreduce

batch data 处理的很好（数据已经有了），但是流数据不行。有频繁的IO磁盘操作，性能很差。map1-spill1.out, 是内存放不下，写到小文件。之后就放弃了java, 用scala, 函数式编程（一个变量被定义了之后不能改写了）



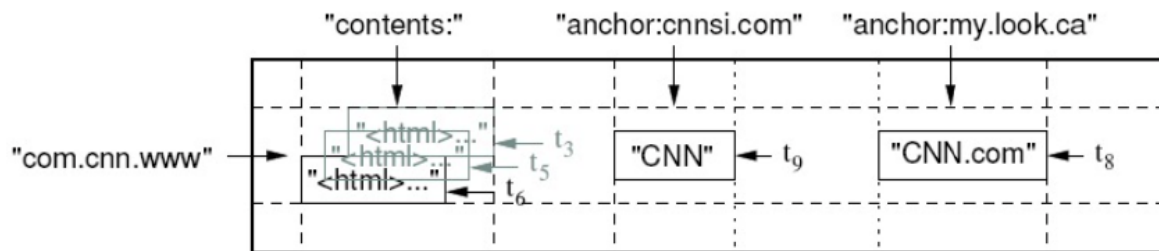
文件系统 → GFS



master存储这个文件在哪个chunk, 类似原来的文件系统说这个文件的block在哪里, 就是大文件切成小文件, 然后维护文件的存放位置。还有多备份, data flow 和 control flow

## Bigtable

列有二级化的处理。列族, 可以按照时间戳



## hadoop



- The Apache Hadoop project develops open-source software for reliable, scalable, distributed computing. Hadoop includes these subprojects:
  - **Hadoop Common**: The common utilities that support the other Hadoop subprojects.
  - **HDFS**: A distributed file system that provides high throughput access to application data.
  - **MapReduce**: A software framework for distributed processing of large data sets on compute clusters.
- IBM, Amazon, Yahoo
  - Base stone

批处理有基础开销，输入多的时候就适合用批处理。

## two phase commit —— google percolator

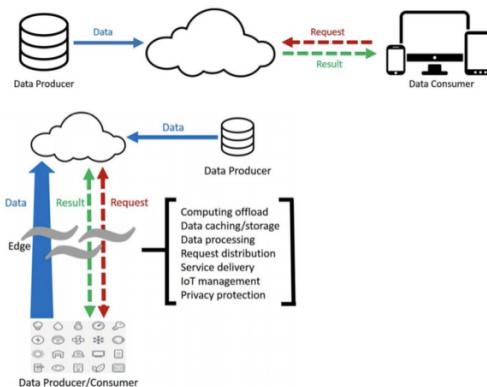
在表里加行锁，而不是锁住整个表

## 边缘计算

让摄像头直接连接的服务器计算，再之后摄像头就可以有计算能力。把计算的任务推到了网络的边缘：

- **Cloud computing**

- **Edge computing**



终端到云的每一层都可处理。

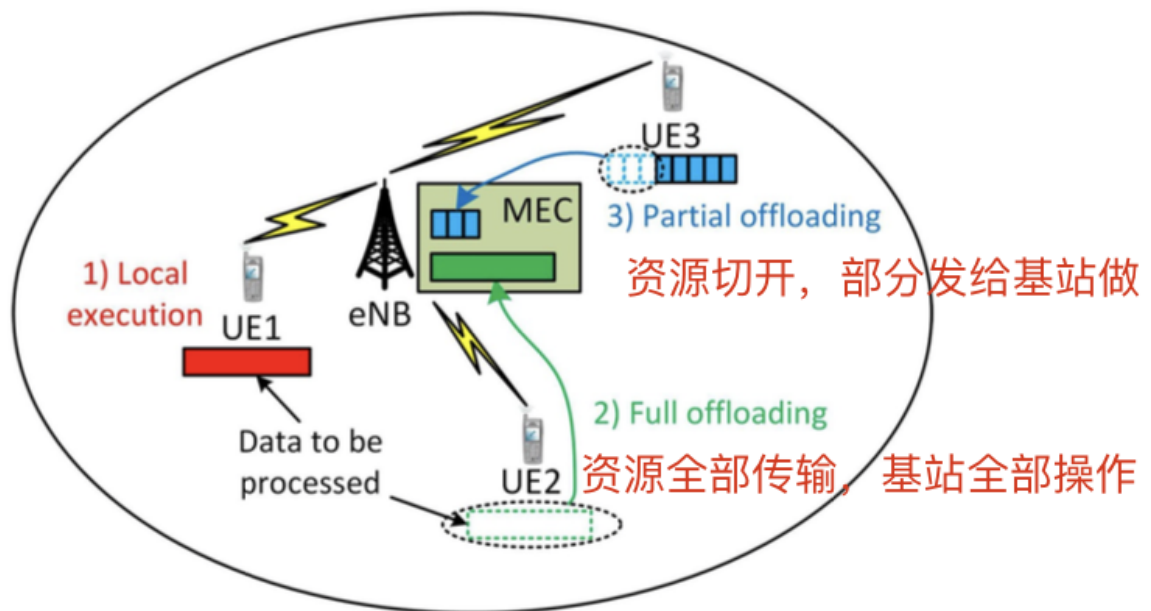
数据在靠近他的地方处理效率高。就近存储，就近处理。

好处：通过在数据源处或附近执行分析和知识生成，减少了传感器和中央数据中心之间所需的通信带宽。

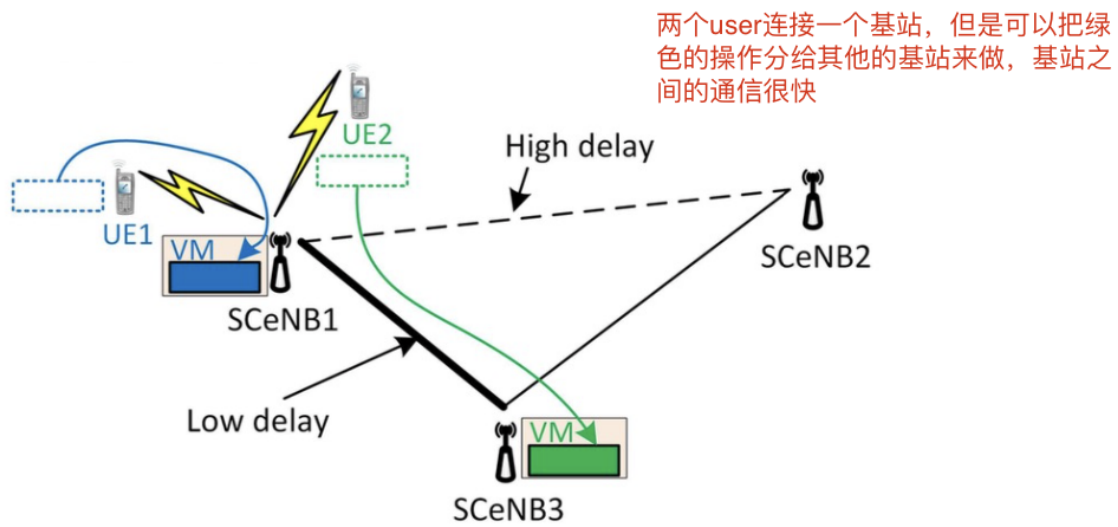
坏处：这种方法需要利用可能无法持续连接到网络（如笔记本电脑、智能手机、平板电脑和传感器）的资源

边缘计算兴起的背景：越来越多的终端计算能力增强。

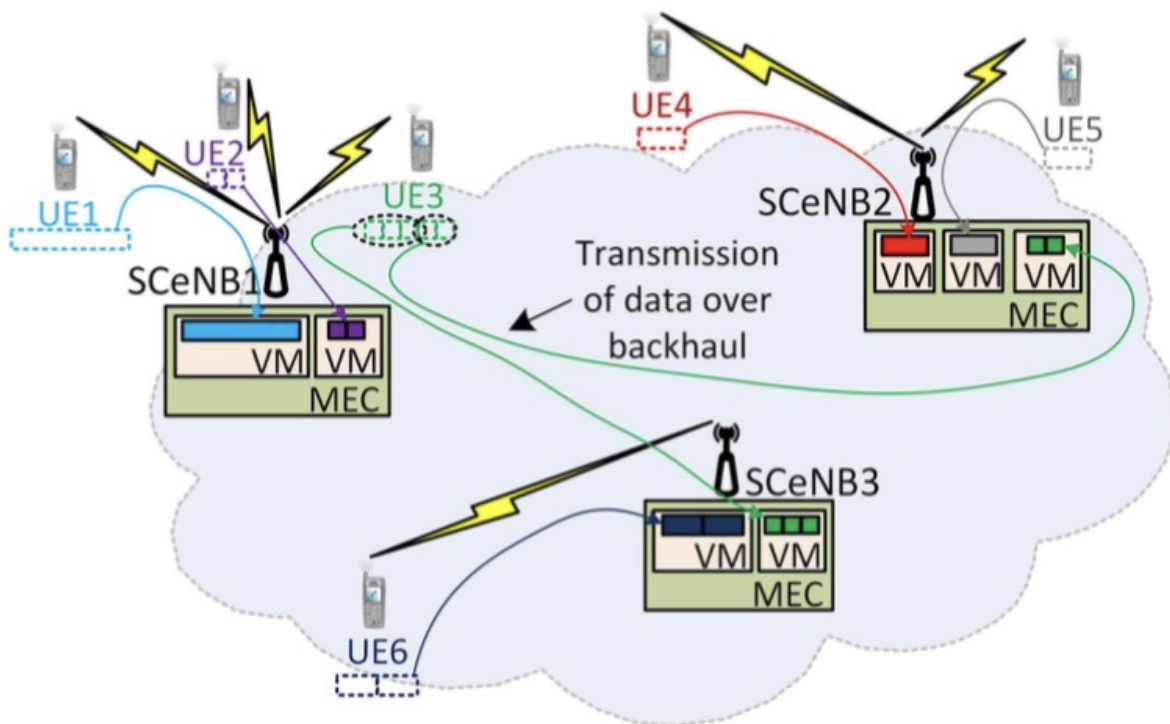
### Computation Offloading（计算迁移）



这个基站要是做不了，还可以投到更上的基站做，这样可以比较充分的利用资源。



需要5个单位，就在两个基站凑够了5个单位的资源来形成它

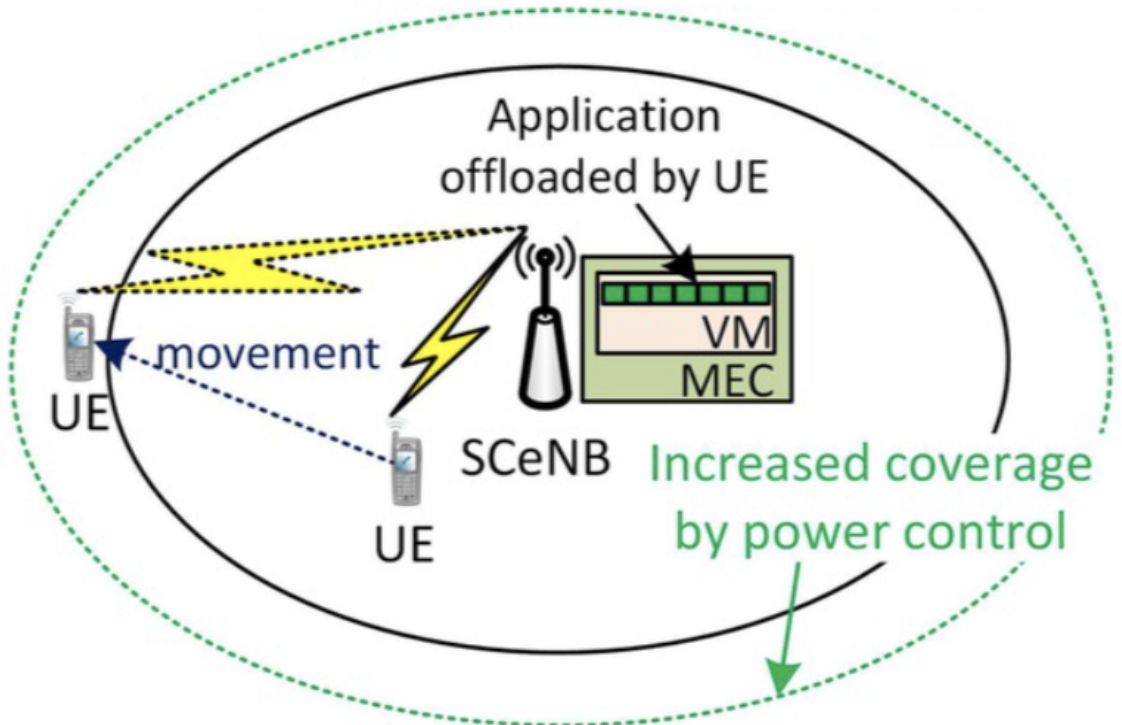


无限通信在于带宽是可变的，数据在传输的时候就带宽会不一样。

## 移动的管理

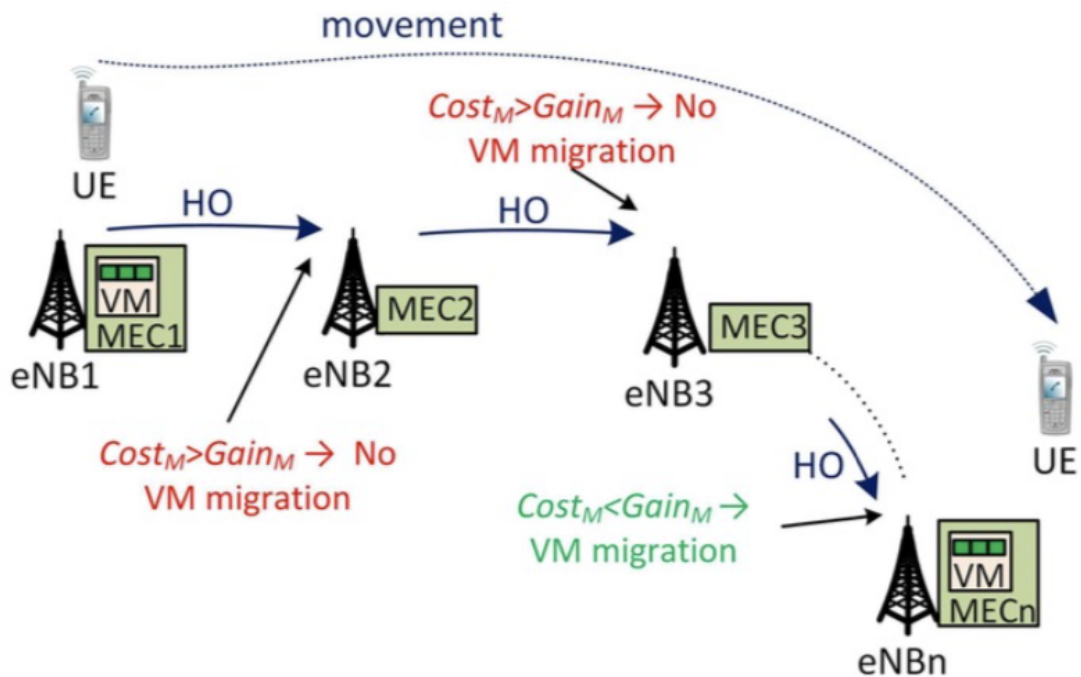
### 1. Power control

出了圈，能耗增大，带宽降低，这个时候是否需要换一个基站



### 2. VM migration





因为用户在移动，所以到了另外一个地区就要换了一个基站了，但是在城市里面，人的移动是有规律的，可以提前预测，就提前启动这个新的虚拟机。eg，人从家里到上班，在不同的基站做prefetch,基于你的移动速度，路径等等。

## summary

边缘计算就是雾计算，就是近在眼前的。

但是每一台的资源都是受限的，在云计算里面是看起来资源无限的。

## 补充：云计算与边缘计算的区别

从**原理**上来看：边缘计算是指在数据源头的附近，采用开放平台，就近直接提供最近端的服务。而云计算，则是指通过网络，把众多数据计算处理程序分解，通过服务器组成的系统，把这些分解的小程序再处理分析来得到结果。

从**特点**上来看：边缘计算的应用程序是在数据源头边缘侧发起的，减少了数据在网络上转移的过程，那么所产生的网络服务也会更快，在一些行业中的实时业务、应用智能、安全与隐私保护等方面应用都很不错。云计算则融入了分布式计算、效用计算、负载均衡、并行计算、网络存储、热备份冗杂和虚拟化等计算机技术，具有虚拟化技术、动态可扩展、按需部署、灵活性高、可靠性高、性价比高、可扩展性的优点。

举个例子就会好理解很多。比如某个程序，用户在使用时会同步搜集他的地理位置、搜索记录、不同页面的停留时间等，以便筛选出他对什么更感兴趣，进而更精准地为用户投放

内容和广告。

如果使用云计算来完成这个操作，那么程序在搜集到信息后，会把所有的信息先上传到服务器中，然后由云服务器来完成算法，计算识别出用户的兴趣爱好、消费倾向等，服务器根据推算出的结果，再进一步为用户投放内容和广告。

如果使用边缘计算来完成这个操作，那么程序在搜集完信息后，直接自己计算和识别出用户的兴趣爱好，服务器只需要指定想给哪些用户投放内容和广告，程序就可以直接完成投放，在整个过程中，服务器没有参与搜集信息和计算，服务器只知道最后的结果，没有过程中的数据。