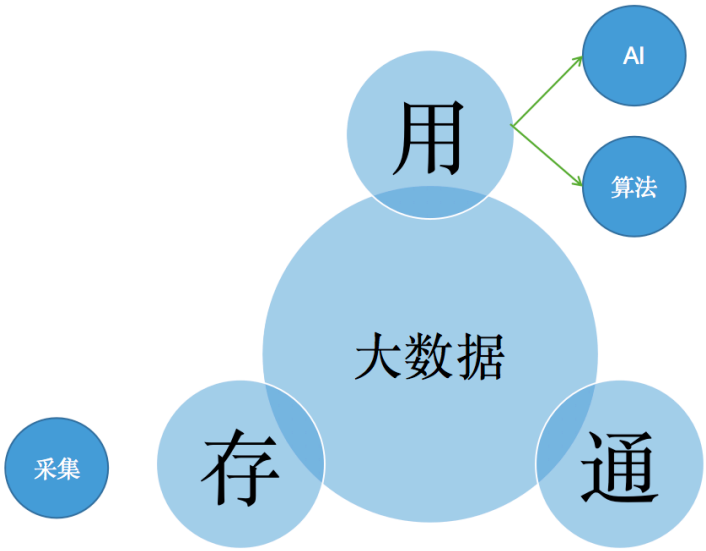
`中台`这个概念是阿里15年提出来的，首先阿里提出这个概念，我想更多是从阿里的组织结构和内部产品以及业务出发考虑的，我觉得这只是一个名词，一个概念，只是当前很多传统IT公司转型或者升级的另一个说法而已，比如有些公司可能叫做数字化平台建设或者战略，当然由`中台`，我们也很容易联想到`数据平台`，`业务中台`或者`大数据平台`，由此想到的大数据平台厂商，比如cloudera、Hortonworks、mapR为代表的三家大数据公司，包括大数据最常用hadoop、spark等为代表技术框架工具及其背后的生态和公司，总之能联想到的概念，技术冗杂繁多，在这里我将尝试从以下几个方面来阐述数据中台。

拿到一个新概念，我们一般会尝试使用3w的方式去思考并理解它？首先数据中台的聚合根是`数据`，那自然而然就会想到：
1)、为什么要使用数据，使用数据的元意义是什么？数据中台在这里有什么价值？
2)、怎么建设并使用好数据，有哪些理论支撑或者方法论？ 数据中台是如何帮助我们使用好数据的？

至于第一个问题，大数据被大家都声嘶力竭地呼喊了至少10年，让大家误以为数据天生就是有价值的，这里引用阿里提出的三个字来概括的大数据使用（如下图）



关于大数据核心价值，网上的回答比较官方如下：
1、户群体细分，然后为每个群体量定制特别的服务。

2、模拟现实环境，发掘新的需求同时提高投资的回报率。
3、加强部门联系，提高整条管理链条和产业链条的效率。
4、降低服务成本，发现隐藏线索进行产品和服务的创新。
或者如同马云所讲的，大数据是水电煤，其实最终的目前无非就是能够帮助我们开源节流。

为什么大数据这个概念是最近10年才会爆发出来，莫非数据是最近10年才存在的，肯定不是这样的，抛开远的不讲，从图灵机问世至少有半个多世纪了，然后经历了差不多20多年的互联网技术，再到10多年的云计算基建和铺垫，才打开了现在的一个大数据爆发的窗口，所以大数据天生是不具有价值的，只有有了他们成长的土壤，社会价值，商业场景等一个比较完整的生态，它才会被挖掘出来，凸显价值。
大数据也不是天生就存在的，价值密度也是分布不均匀的，只有被采集，打通，存储，然后基于相关的技术框架，平台产品才存在的。
“世上数据本没有价值的，只有被发现，采集，存储，挖掘，应用才会有价值”。

至于第二个问题，又可以概括为从如下几个方面来阐述的。
我们在理解一个新概念，也会尝试从已有的经验和知识去找一个和他相关或者类似的体系做对比，那这个就是已经存在并且多年实践应用的就是数据仓库，在我看来数据中台更多的是从一个多维的立方体的数据链路的解决方案和建设思路，而数仓更多的是从一个面或者一些点来解决大数据的问题，当然这里并不是说传统数仓就没价值，毕竟它背后沉淀的一些建设思路，模型设计，实施流程在很多地方依旧可以被数据中台所吸收。

1)、如何建设并使用数据中台，有哪些方法论，工具，平台或者产品?
国外公司以cloudera、hortonworks、mapR为代表，提供了一些列大数据平台和工具，这里以hadoop、spark等为技术能力的底层平台与传统的oracle/db2为底层平台的传统数仓有什么区别呢？其实从某些层面上来讲，他们都可以解决问题，但是在这里我认为以hadoop为代表的构建的新型数仓较传统数仓提供了更多能力和应用场景以及更少的成本投入，这里我尝试从以下几个方面分别阐述下:
a)、首先是能力要求层面，我们知道大数据应用的4v+1c特性。
variety: 数据种类繁多，编码格式，数据格式，应用特征等多个方面的差异性，多信息源并发形成的异构数据。
volume:通过各种设备，终端产生的海量数据，数据规模庞大，PB级别是常

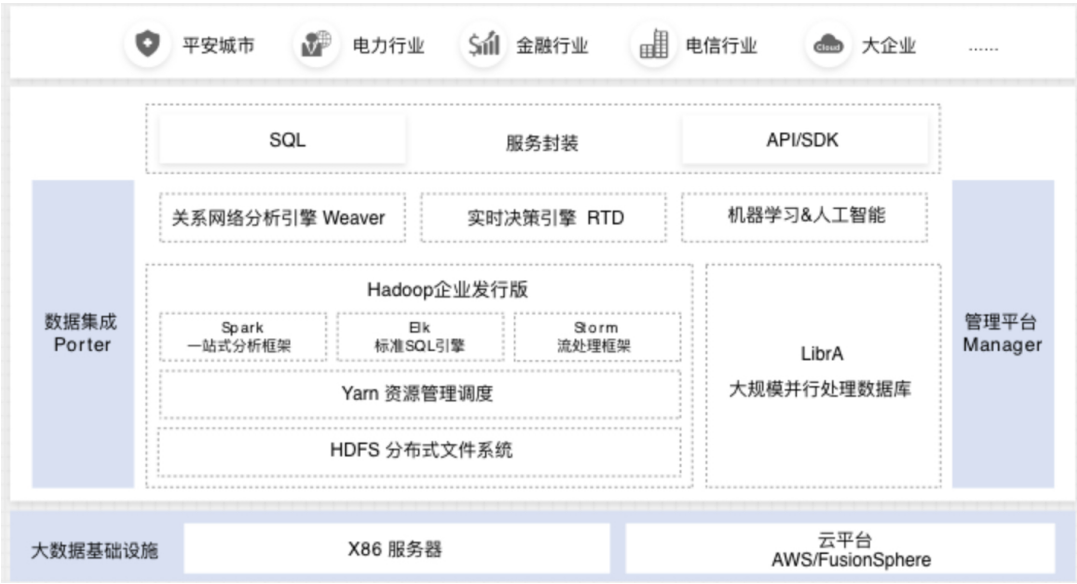
态。

velocity:涉及到感知，传输，决策等大数据，对数据实时处理有着极高的要求，通过传统数据库查询方式得到的当前结果可能已经失效

vitality: 数据持续得到，并且只有在特定时间空间才有意义

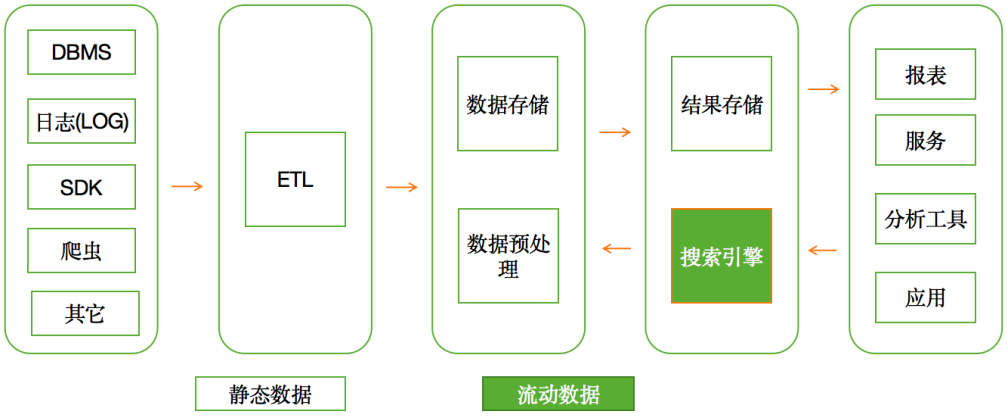
complexity: 最后要强调的一点是复杂度，新的环境和场景下对于数据的存储和处理的复杂度都在上升，依靠传统的数据库存储方式不　　再适用，依照上面的4v+1c特性，我们再反观传统的数仓建设是否能够胜任，以oracle/db2为代表的数据库能力已经很难解决新环境下数据适用要求我们的能力了。

目前基于开源或者自研的大数据框架平台来搭建数据中台的技术框架平台部分，应该不是什么难事，这里需要感谢这个开源的时代赋予我们每个人的能力，这里也罗列部分大数据平台建设的架构图:

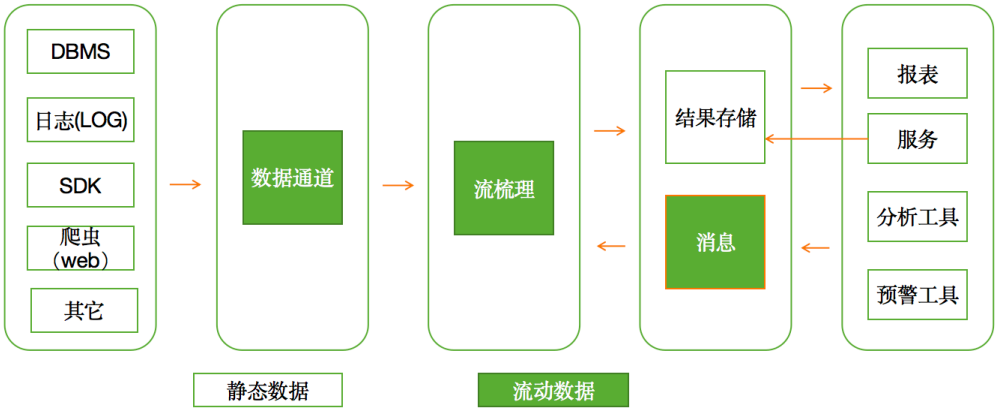


不管是传统的**大数据架构**，或者**Lambda**架构，还是**Kappa**架构，抑或是**Unified**架构，或者是Mpp架构的在线分析数据库，都是为了解决不同场景下对数据处理的要求不同而已。

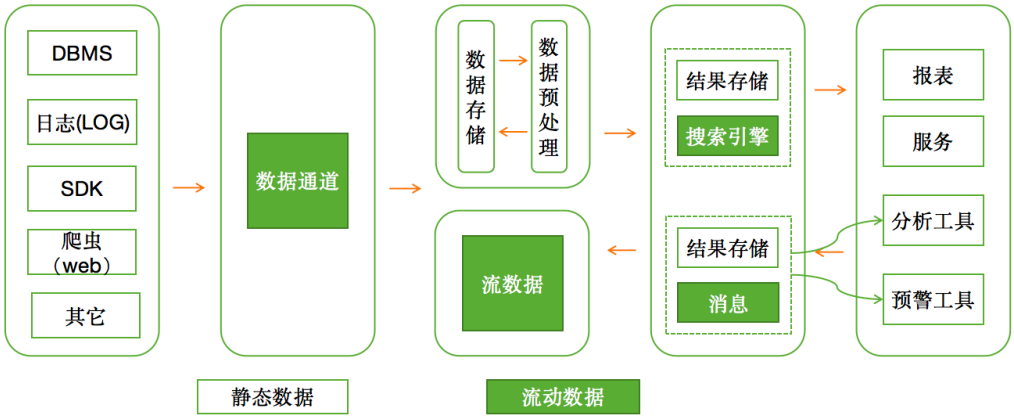
传统大数据架构



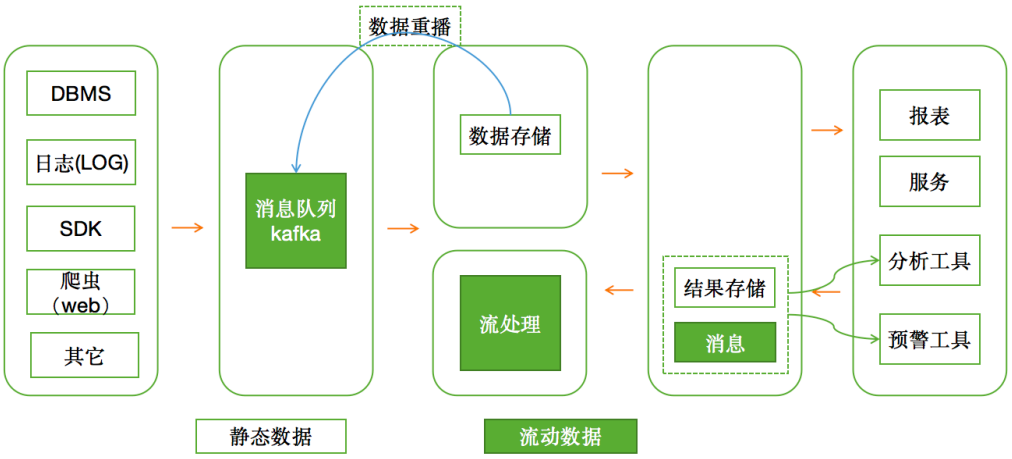
流式架构



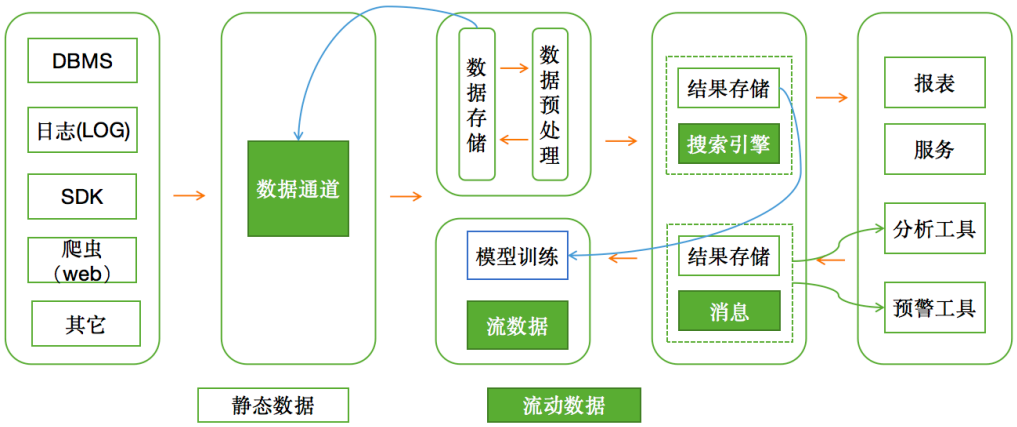
Lambda架构（实时流+离线）



Kappa架构（实时流为主）



Unified架构（以lambda为基础）



以上无论何种架构，相对的都有一系列的开源工具框架或者商业产品来做支撑，我觉得这些都不足以形成公司在大数据层面的核心竞争力，而真正能构建在大数据方面的壁垒反而是数据本身以及基于数据挖掘出来的应用，比如说通过机器学习形成了几万套标签或者一些列核心算法。当然除了底层技术平台本身，以此衍生了一些帮助技术人员提升数据开发效率和管理数据的工具产品，比如数据质量管理，数据api，数据地图，血缘管理等。

2)、有了底层技术平台，只是有了地基或者骨架，基于上面能够生长出什么应

用，或者通用的数据产品，比如用户画像，DMP，智慧选址，IDMapping等，为什么大数据发展（呐喊吆喝）了10来年，为什么还是有些人不能认可它带来的价值（骗人），这里很大一部分原因
恐怕就是太多的概念和底层设计，而缺乏能够给客户带来价值的上层产品应用吧，如何让客户买单，这里恐怕非深入客户实际业务，以及对行业，对客户数据的理解而不能做到。

最后再对以上内容做一个总结：
数据中台是一个从数据采集，同步集成，存储搜索，开发，分析，应用的全面多维立体的解决方案和产品，不仅仅是数据仓库的点线面的延伸和扩展，更是一种从客户业务出发，数据运营，战略决策到上层组织架构的延伸配套，不仅仅是传统数仓沉淀的理论，设计，实施的实践经验，乃至方法论，更是一种矩阵式的平台+框架+产品+应用的高阶组合。