

本文将结合平安壹钱包的实际使用情况，分享对 Alluxio 的理解以及使用 Alluxio 进行大数据建设的实践经验。

主要内容包括以下几大方面：

1. Alluxio介绍

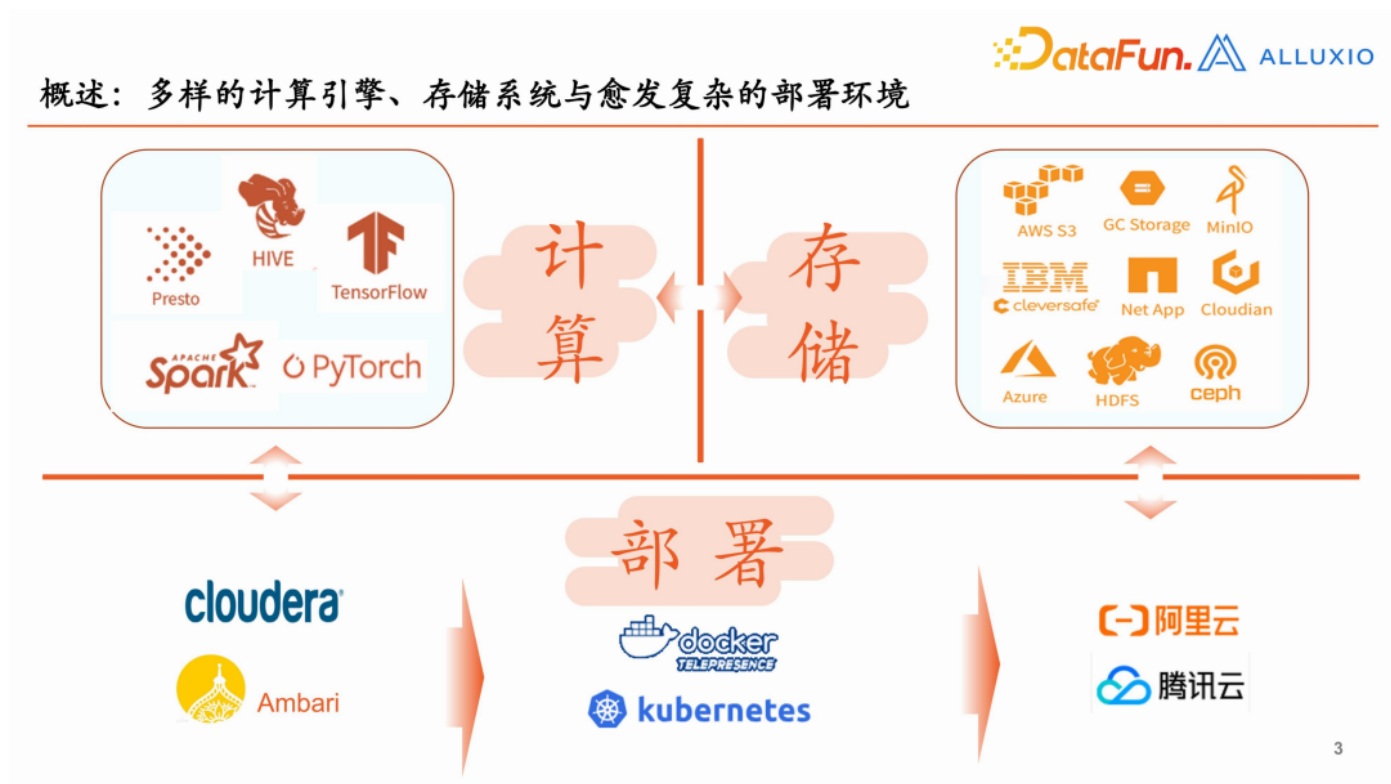
2. Alluxio实践分享

3. 展望

01

Alluxio 介绍

1. 背景介绍

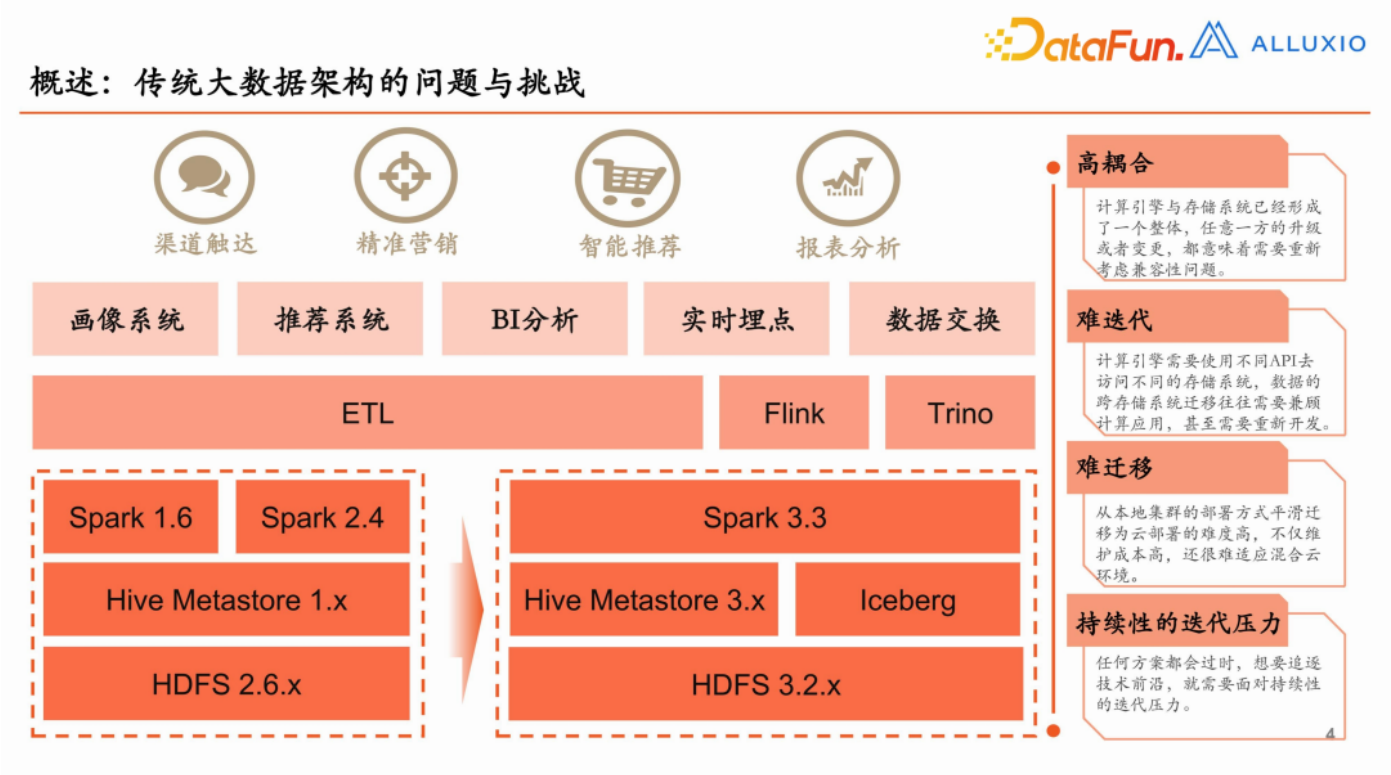


当前计算引擎越来越多样化，存储系统和部署环境也越来越复杂。不同的计算和存储适用于不同的场景。比如 Presto 的即席查询性能比较好，比 Spark 要快很多，而 Spark 又更适合于批量处理。存储系统也是一样的，常用的 HDFS 是一个可以存取海量数据的存储系统，但有些场景下又希望能够使用存储密度更高的存储系统。与此同时，部署方式也在不断变化，比如之前一般使用的是 CDH，将来可能会用

Ambari 这样一个开源的方式，还可能会考虑使用容器化的部署，为将来的云化做一个铺垫。

环境的日趋复杂化，带来了很多问题与挑战。比如我们在实践过程中，为了建设数据湖仓，必须要引入 Iceberg，但是已有的 HDFS 是 2.6.x，Spark 是 1.6.x，对 Iceberg 是不支持的。为了引入 Iceberg，HDFS 要升级到 3.x，Spark 也要升级，才能建设成一个整体。

在这样一个过程中，我们面对了四点比较突出的问题：



- 高耦合：计算引擎与存储系统已经形成了一个整体，任意一方的升级或者变更，都意味着需要重新考虑兼容性的问题。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/177043162146006040>