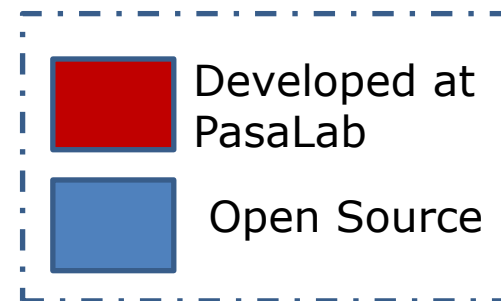
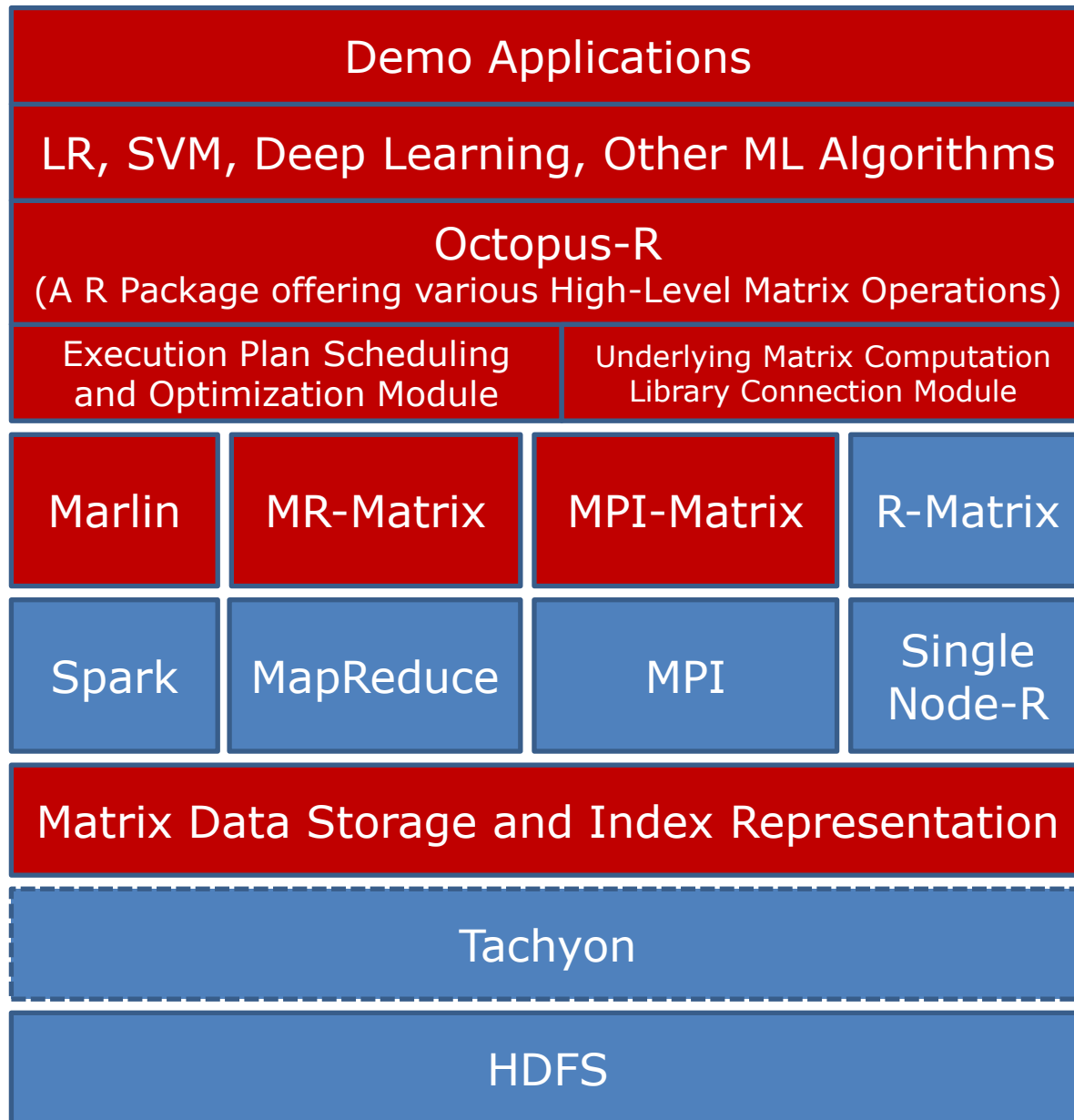


A very brief introduction to Octopus

Rong Gu
PASALab

Architecture



Octopus的特点

- 提供一个支持大规模矩阵操作的R包，能与标准的R生态环境无缝结合；
- 包含的分布式矩阵操作的接口都是高层的，无需用户具备分布式系统的基本概念或知识
 - 提供常见的分布式矩阵操作，包括矩阵相加、相乘、转置、多种分解等；
 - 支持对分布式矩阵的元素并行地执行R的标准函数和用户自定义函数的计算；（类似于单机R里面的lapply函数）

Octopus的特点

- 底层支持选用多种分布式计算框架和存储平台，能与现有的大数据生态系统无缝对接
 - 支持的分布式计算引擎包括Spark, MapReduce, MPI，同时也提供单机的R计算引擎；
 - 数据可以存储在HDFS、Tachyon上，并可以通过Tachyon在多个计算引擎间快速共享矩阵数据；
- 同一套程序可以无需修改地运行在多个底层计算平台上，从而减轻用户开发和维护代码的负担。
 - 支持对底层计算引擎的灵活配置、智能选择；
 - 避免在各个计算平台间迁移重实现application的成本；