Hadoop环境安全加固

### 一.Hadoop介绍

Hadoop是一个由Apache基金会所开发的一个开源 高可靠 可扩展的分布式计算框架。 Hadoop的框架最核心的设计就是：HDFS和MapReduce。

HDFS为海量的数据提供了存储，MapReduce则为海量的数据提供了计算。

HDFS是Google File System(GFS)的开源实现。

MapReduce是一种编程模型，用于大规模数据集(大于1TB)的并行运算。

### 二.Hadoop安全问题

#### 1.WebUI敏感信息泄漏：

　　Hadoop默认情况开放了很多端口提供WebUI， 下面列举了相关端口信息：

* HDFS

　　1).NameNode 默认端口 50070

　　2).SecondNameNode 默认端口 50090

　　3).DataNode 默认端口 50075

　　4).Backup/Checkpoint node 默认端口 50105

* MapReduce

　　1).JobTracker 默认端口 50030

　　2).TaskTracker 默认端口 50060

通过访问NameNode WebUI管理界面 50070端口，可以下载任意文件；

DataNode 默认端口50075开放的话，攻击者可以通过hdsf提供的restful api对hdfs存储数据进行操作。

#### 2.MapReduce代码执行漏洞

#### 3.Hadoop的第三方插件安全漏洞

* Cloudera Manager <=5.5

　　1).[Cloudera Manager CVE-2016-4949 Information Disclosure Vulnerability](http://www.securityfocus.com/bid/93882)

　　2).[Template rename stored XSS (CVE-2016-4948)](https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2016-4948)

　　3).[Kerberos wizard stored XSS (CVE-2016-4948)](https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2016-4948)

　　4).[Host addition reflected XSS (CVE-2016-4948)](https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2016-4948)

* Cloudera HUE =< 3.9.0

　　1).[Enumerating users with an unprivileged account (CVE-2016-4947)](https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2016-4947)

　　2).[Stored XSS (CVE-2016-4946)](https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2016-4946)

　　3).Open redirect

* Apache Ranger =< 0.5

　　1).Unauthenticated policy download

　　2).[Authenticated SQL injection (CVE-2016-2174)](https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2016-2174)

* Apache Group Hadoop 2.6.x

　　1).[Apache Hadoop MapReduce信息泄露漏洞(CVE-2015-1776)](https://www.cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2015-1776)

#### 4.Hive任意命令/代码执行漏洞

Hive是建立在 Hadoop 上的数据仓库基础构架。它提供了一系列的工具，可以用来进行数据提取转化加载（ETL），这是一种可以存储、查询和分析存储在 Hadoop 中的大规模数据的机制。Hive 定义了简单的类 SQL 查询语言，称为HQL，它允许熟悉 SQL 的用户查询数据。同时，这个语言也允许熟悉 MapReduce 开发者的开发自定义的 mapper 和 reducer 来处理内建的 mapper 和 reducer 无法完成的复杂的分析工作。

HQL可以通过transform自定义Hive使用的 Map/Reduce 脚本，从而调用shell/python等语言，导致攻击者可以通过hive接口等相关操作方式直接获取服务器权限。

从以上信息可以看出，对外暴露服务端口会存在严重的安全风险。

### 三.安全加固方案

#### 1.网络访问控制

使用[安全组防火墙](https://help.aliyun.com/document_detail/25475.html?spm=5176.doc25470.6.204.SlE11A)或本地操作系统防火墙对访问源IP进行控制，如果仅对内网服务器提供服务，建议禁止将Hadoop服务所有端口发布到互联网上。

#### 2.启用认证功能

启用kerberos认证功能

#### 3.更新补丁

不定期关注Hadoop官方发布的最新版本，并及时更新补丁

**更多信息：**

* Hadoop所有端口信息
* 
* <http://archive.hack.lu/2016/Wavestone%20-%20Hack.lu%202016%20-%20Hadoop%20safari%20-%20Hunting%20for%20vulnerabilities%20-%20v1.0.pdf>
* <http://blog.cloudera.com/blog/2009/08/hadoop-default-ports-quick-reference/>