

# 汉水大数据操作系统产品手册

版本	作者	修改时间
V1.0	梁泽运	2022年10月3日

## 一. 概要

汉水大数据操作系统（HanWate BigData OS）V1.0 是一款基于开放源代码 CentOS 7.0 和 Hadoop 生态系统软件经过定制化改造的国产服务器专用操作系统，主要用于大数据计算集群的构建和大数据应用软件开发。软件著作权登记证书号：软著登字第 5689517 号。

## 二. 主要功能

- 可用于分布式 HDFS 存储集群构建
- 可用于分布式 MapReduce 计算集群构建
- 可用于分布式 Yarn 资源调度计算集群构建
- 可用于 Hive2 大数据数据仓库构建
- 可用于 HBase 数据库构建
- 可用于 Spark2.2 分布式计算集群构建
- 集成 Kafka 和 Spark-streaming 流计算软件
- 集成 MySQL 数据库服务器
- 集成 Java、Python 和 Scala 运行环境
- 集成 Java 和 Scala 开发环境

## 三. 主要特性

- 支持高可用分布式 hadoop 集群
- 支持元数据治理
- 支持 Kerberos 加密认证
- 支持 Jupyter 远程 spark 控制台
- 支持 flume 和 kafka 数据采集

## 四. 核心软件配置目录

### 1. Hadoop

/etc/hadoop/conf

## 2. Hive2

/etc/hive2/conf

## 3. Spark

/etc/spark2/conf

# 五. 大数据相关的系统服务

## 1. HDFS 服务

hadoop-hdfs-namenode

hadoop-hdfs-secondarynamenode

hadoop-hdfs-datanode

## 2. MapReduce 和 YARN 服务

hadoop-mapreduce-historyserver

hadoop-yarn-resourcemanager

hadoop-yarn-nodemanager

## 3. MySQL 服务

mysql

## 4. Hive2 服务

hive2-metastore

hive2-server

## 5. HBase 服务

hbase-master

hbase-regionserver

hbase-rest

hbase-thrift

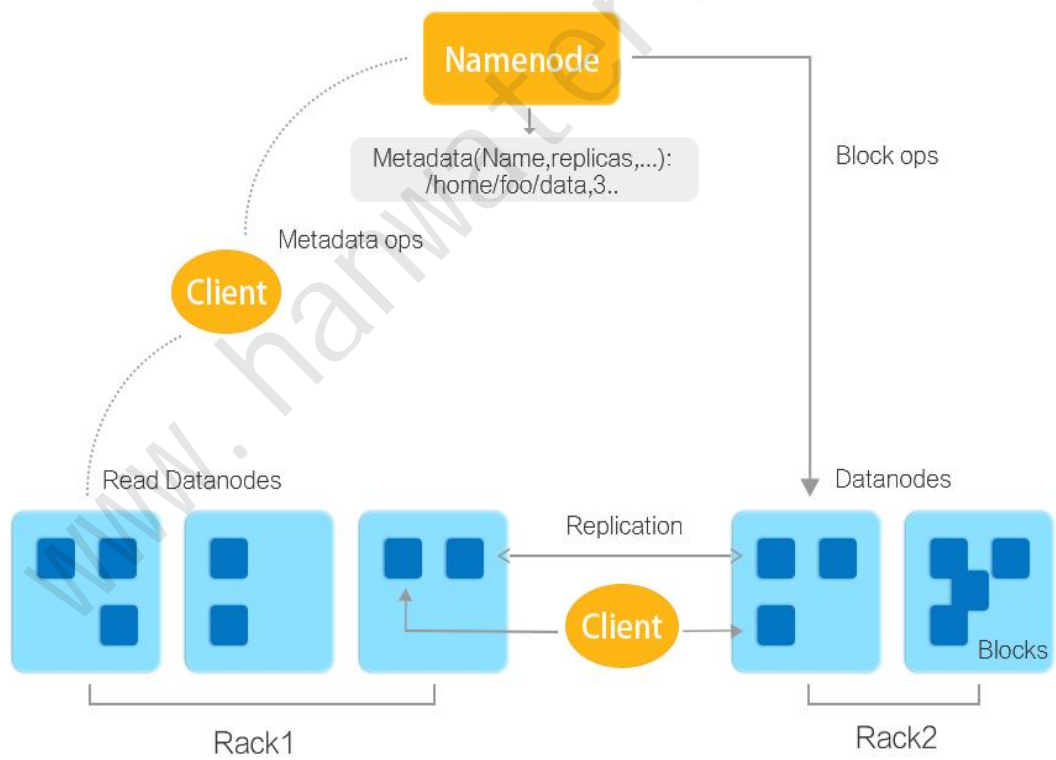
hbase-thrift2

## 6. Zookeeper

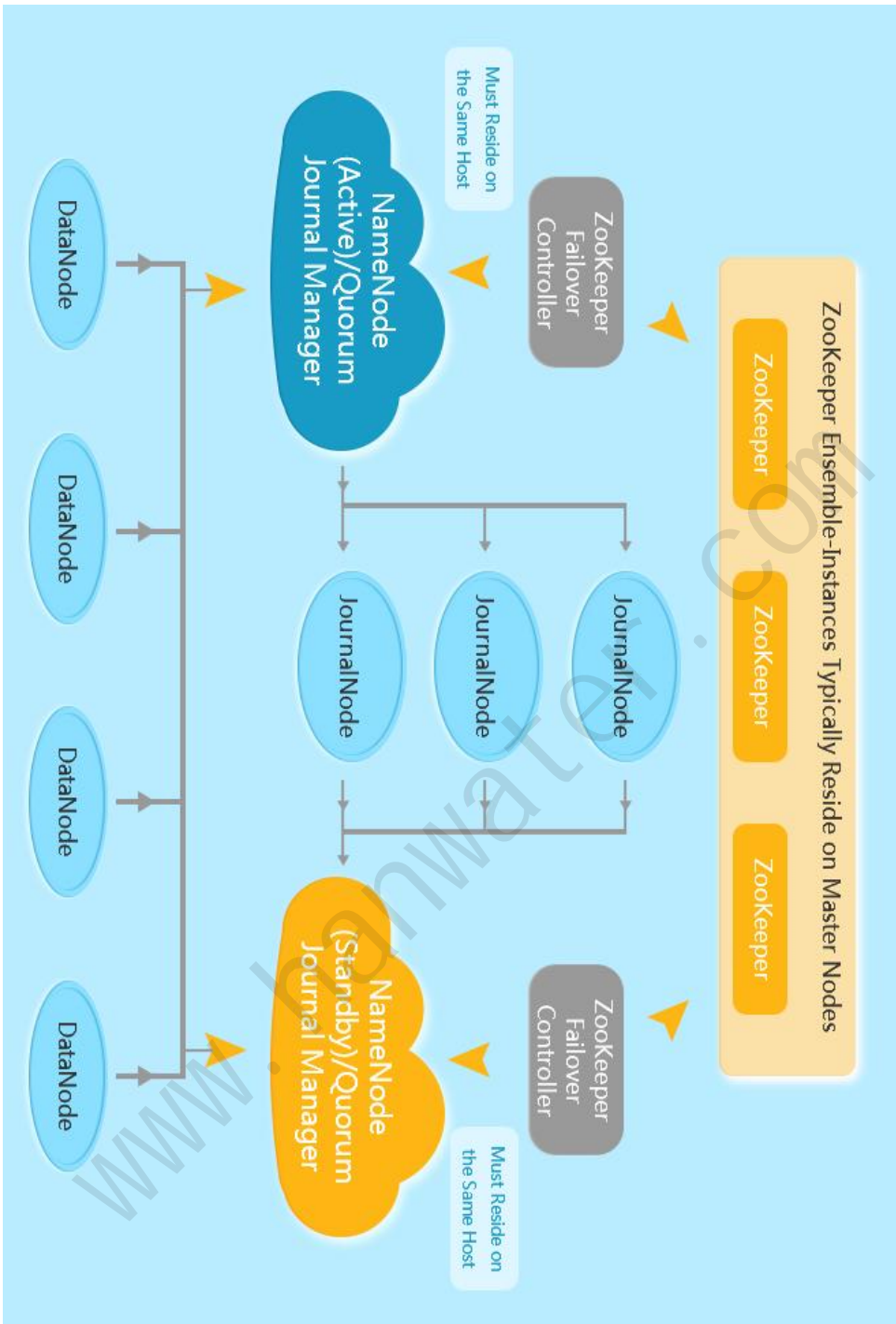
zookeeper-server

## 六. 参考架构图

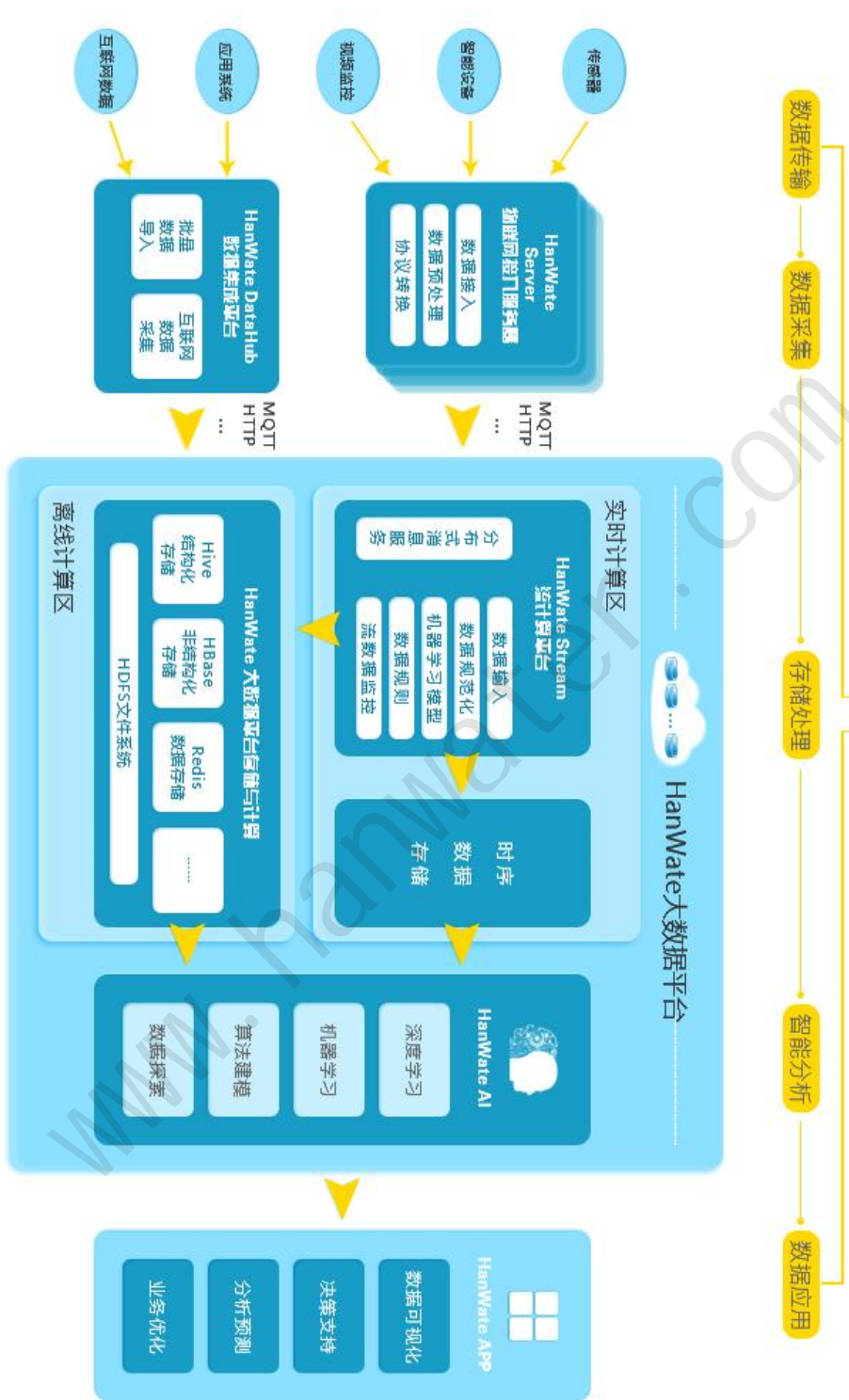
### 1. HDFS 分布式存储架构图



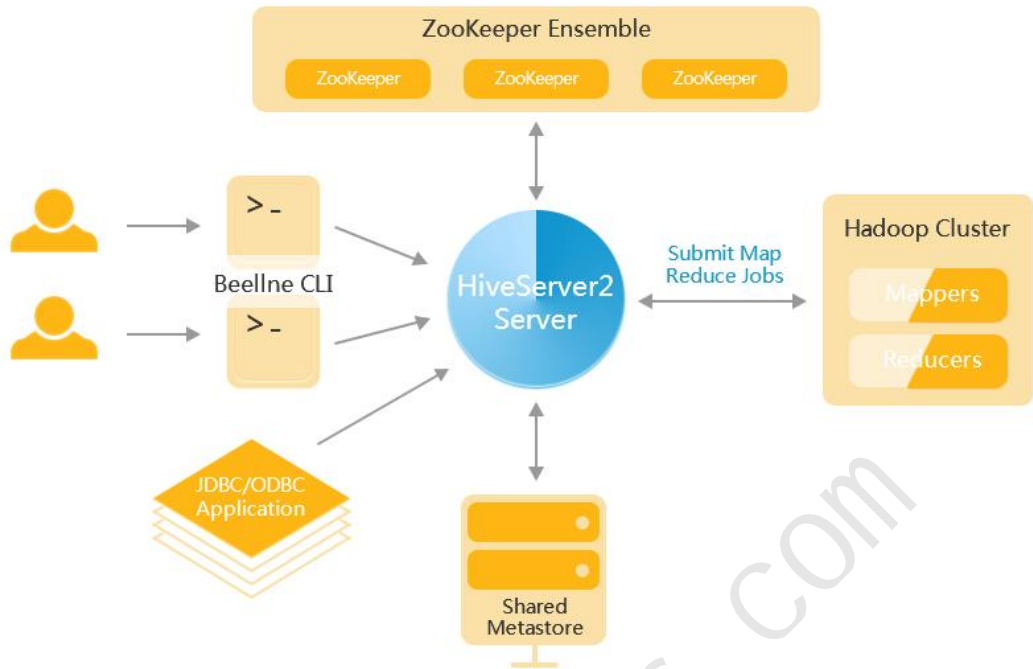
## 2. 生产环境高可用 Hadoop 架构



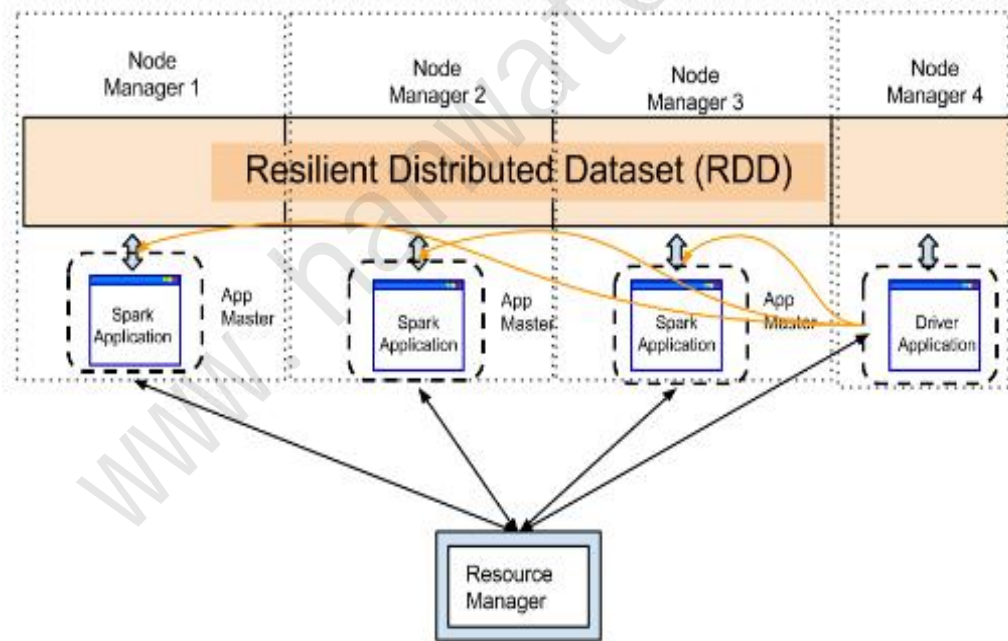
### 3. 大数据业务系统架构参考



#### 4. Hive2 数据仓库架构



#### 5. Spark 分布式计算集群架构



附：产品网站

<http://hanwater.com>