
| | | |
|----|------------------|----|
| 1. | 范围..... | 4 |
| 2. | 规范性引用文件..... | 5 |
| 3. | 术语及缩略语..... | 6 |
| 4. | 总体概述..... | 9 |
| | 4.1 测试综述..... | 9 |
| | 4.2 测试原则..... | 9 |
| | 4.3 测试纪律管理..... | 10 |
| | 4.4 评价标准及评审..... | 10 |
| 5. | 测试方案..... | 11 |
| | 5.1 测试环境..... | 11 |
| | 5.2 环境准备..... | 11 |
| | 5.2.1 硬件环境..... | 11 |
| | 5.2.2 软件环境..... | 12 |
| | 5.2.3 测试数据..... | 12 |
| | 5.2.4 数据范围..... | 13 |
| | 5.3 测试内容..... | 13 |
| | 5.3.1 功能测试..... | 15 |
| | 5.3.2 性能测试..... | 34 |
| | 5.4 测试安排..... | 36 |
| 6. | 附录..... | 38 |
| | 6.1 组件部署列表..... | 38 |

前 言

近年来大数据技术和应用迅猛发展，通过横向扩展，分布式集群部署方式比传统集中式架构性能更优，在数据平台架构云化重构、实时应用支撑等方面发挥了重要作用。

中国电信基于“集中、开放、云化”原则，应用大数据技术，构建企业级大数据平台核心处理系统，以提升系统扩展性、系统可用性和分析性能。

本规范是中国电信大数据平台核心处理系统检测总体规范，为系统测试实施的指导性文件，整个系统测试方案主要包括：测试环境、环境准备、数据说明、功能测试、性能测试等。

本技术方案由中国电信股份有限公司提出并归口。

本方案起草单位：中国电信股份有限公司上海研究院

本方案主要起草人： 、 、

本方案解释单位：中国电信股份有限公司上海研究院

1. 范围

本规范规定了中国电信大数据平台核心处理系统检测的总体情况、检测要求、测试环境、数据情况等情况、功能要求以及性能要求等。

本标准适用于中国电信大数据平台核心处理系统。

2. 规范性引用文件

中国电信信息〔2016〕12号《中国电信大数据平台技术要求》

3. 术语及缩略语

| 缩略语 | 全拼 | 说明 |
|--------|--------------------------------|---|
| Hadoop | Hadoop | Hadoop 是一个由 Apache 基金会所开发的分布式系统基础架构，用户可以在不了解分布式底层细节的情况下，开发分布式程序，充分利用集群的威力进行高速运算和存储。Hadoop 实现了一个分布式文件系统，简称 HDFS，HDFS 有高容错性的特点，并且设计用来部署在低廉的硬件上，而且它提供高吞吐量来访问应用程序的数据，适合那些有这超大数据集的应用程序。Hadoop 的框架最核心的设计就是 HDFS 和 MapReduce。 |
| HBASE | HBASE | HBASE 是一个分布式的、面向列的开源数据库，该技术来源于 Fay Chang 所撰写的 Google 论文“Bigtable: 一个结构化数据的分布式存储系统”。就像 Bigtable 利用了 Google 文件系统（File System）所提供的分布式数据存储一样，HBASE 在 Hadoop 之上提供了类似于 Bigtable 的能力。HBASE 是 Apache 的 Hadoop 项目的子项目。HBASE 不同于一般的关系数据库，它是一个适合于非结构化数据存储的数据库。另一个不同的是 HBASE 基于列的而不是基于行的模式。 |
| HDFS | Hadoop Distributed File System | HDFS 被设计成适合运行在通用硬件上的分布式文件系统。它和现有的分布式文件系统有很多共同点。但同时，它和其他的分布式文件系统的区别也是很明显的。HDFS 是一个高度容错性的系统，适合部署在廉价的机器上。HDFS 能提供高吞吐量的数据访问，非常适合大规模数据集上的应用。HDFS 放宽了一部分 POSIX 约束，来实现流式读取文件系统数据的目的。HDFS 在最开始是作为 Apache Nutch 搜索引擎项目的基础架构而开发的。HDFS 是 Apache Hadoop Core 项目的一部分。 |
| Hive | Hive | 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。 |
| Kafka | Kafka | Kafka 是一种高吞吐量的分布式发布订阅消息系统，它可以处理消费者规模的网站中的所有动作流数据。对于像 Hadoop 的一样的日志数据和离线分析系统，但又要求实时处理的限制，Kafka 是一个可行的解决方案。Kafka 的目的是通过 Hadoop 的并行加载机制来统一线上和离线的消息处理，也是为了通过集群机来提供实时的消费。 |
| Redis | Redis | Redis 是一个开源的使用 ANSI C 语言编写、支持网络、可基于内存亦可持久化的日志型、Key-Value 数据库，并提供多种语言的 API。从 2010 年 3 月 15 日起，Redis 的开发工作由 VMware 主持。从 2013 年 5 月开始，Redis 的开发由 Pivotal 赞助。 |
| MapReduce | MapReduce | MapReduce 是一种编程模型，用于大规模数据集（大于 1TB）的并行运算。概念“Map（映射）”和“Reduce（归约）”，是它们的主要思想，都是从函数式编程语言里借来的，还有从矢量编程语言里借来的特性。它极大地方便了编程人员在不会分布式并行编程的情况下，将自己的程序运行在分布式系统上。当前的软件实现是指定一个 Map（映射）函数，用来把一组键值对映射成一组新的键值对，指定并发的 Reduce（归约）函数，用来保证所有映射的键值对中的每一个共享相同的键组。 |
| Spark | Spark | Spark 是 UC Berkeley AMP lab 所开源的类 Hadoop MapReduce 的通用并行框架，Spark，拥有 Hadoop MapReduce 所 |

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/808116035045006044>