UDH可以帮助企业应对大数据带来的哪些挑战

目录

UDH可以帮助企业应对大数据带来的哪些挑战 1

目录 2

一、 背景 3

二、 企业遇到的挑战 3

挑战一：业务部门没有清晰的大数据需求 3

挑战二：企业内部数据孤岛严重 4

挑战三：数据可用性低，数据质量差 4

挑战四：数据相关管理技术和架构 4

挑战五：数据安全 4

挑战六：大数据人才缺乏 4

三、 UDH能够带来的帮助 5

四、 总结 5

#### 背景

随着企业信息化的不断深入，各种不同的系统每分钟都在产生着大量的数据，这些数据包括业务数据、经营数据、人员数据、客户数据、日志数据、传感器数据等等。企业在面对这些不断增长的海量数据时往往感到力不从心，传统的技术体系并不足以应对海量数据增长的需求，大数据技术正是在这种背景下走入了企业的视野之中，用友集团iUAP中心的大数据产品UDH也正是为了满足企业处理大数据的迫切需求应运而生的。

#### 企业遇到的挑战

 那么，企业在面对大数据浪潮时会遇到哪些问题呢？在需要对大数据进行处理时又会遇到哪些挑战呢？在本文中我们将会逐一进行探讨和总结。

 企业希望对大数据进行处理时往往会遇到以下问题和挑战：



###### 挑战一：业务部门没有清晰的大数据需求

目前很多企业的业务部门不了解大数据，也不了解大数据的应用场景和价值，更不了解大数据能帮他们解决什么问题，因此难以提出针对大数据的准确需求。由于业务部门需求不清晰，大数据部门往往又是非盈利部门，企业决策层担心投入比较多的成本，导致很多企业在组建大数据部门时犹豫不决，或者很多企业都抱着观望尝试的态度，从根本上影响了企业在大数据方向的发展，也阻碍了企业积累和挖掘自身的数据资产。甚至由于数据没有应用场景，被迫删除了很多有价值的历史数据，导致企业数据资产流失。

因此，需要大数据从业者和专家一起推动和分享大数据的应用场景，让更多的业务人员了解大数据的价值。

###### 挑战二：企业内部数据孤岛严重

企业想要启动大数据工作时面对的最重要的挑战就是数据的碎片化。在很多企业中，尤其是大型企业，数据常常散落在不同的部门，而且这些数据存储在不同的数据仓库中，不同部门所采用的数据技术也有可能不一样，这导致企业内部自己的数据都没法打通。如果不打通这些数据，大数据的价值则非常难以挖掘。大数据需要不同数据之间的关联和整合才能更好的发挥理解客户和理解业务的优势。将不同部门的数据打通，并且实现技术和工具共享，才能更好的发挥企业大数据的价值。

###### 挑战三：数据可用性低，数据质量差

现代的企业，不管规模大小，几乎每时每刻都在产生大量数据，这些数据中有结构化数据，更多的是非结构化数据，很多企业对结构化数据比较重视，但对非结构化数据的重视程度不够，导致数据处理很不规范，甚至一些有价值的数据会被丢弃。在大数据预处理阶段需要把原始数据转化为方便处理的数据类型，对数据进行清洗和去除噪声，以便进行数据抽取。大数据的意义不仅仅是要收集规模庞大的数据信息，还要对收集到的数据进行很好的预处理，才有可能让数据分析和数据挖掘人员从可用性高的数据中提取出有价值的信息。

###### 挑战四：数据相关管理技术和架构

技术架构的挑战包含以下几个方面：

1. 传统的数据库部署很难处理TB级别的数据，快速增长的数据量超越了传统数据库的管理能力。如何构建分布式的数据仓库，并使数据仓库具备灵活、方便的横向扩展能力成为很多传统企业的挑战。
2. 很多企业采用传统的数据库技术，在设计之初就没有考虑数据类别的多样性，尤其是对结构化数据、半结构化数据和非结构化数据的兼容。
3. 海量数据需要良好的网络架构，需要强大的数据中心来支撑服务器的运行，数据中心的运维工作也将成为挑战。如何在保证数据稳定、支持高并发的同时，减少服务器的低负载情况，成为海量数据中心运维的一个重点工作。

###### 挑战五：数据安全

现代社会已经高度的数字化和网络化，带给人们便利的同时，也带来了更多的风险和隐患。一些知名网站的密码泄露、系统漏洞等问题导致大量用户资料外泄，对企业和用户都造成了巨大的损失。如何保证数据的信息安全成为大数据时代非常重要的课题，对数据的隔离、访问权限控制，以及数据的多副本和容灾机制都提出了更高的要求。

###### 挑战六：大数据人才缺乏

 企业在应用大数据技术时所面对的最大困难之一就是大数据人才的缺乏。大数据技术的应用是一整套系统工程，从底层的集群搭建，到上层的数据分析，涉及到非常多的专业领域，其中每个环节都需要依靠专业人员来完成。因此，必须培养和打造一支掌握大数据技术、懂管理、有大数据应用经验的专业队伍。

#### UDH能够带来的帮助

UDH（UAP Distribution for Hadoop）是用友企业级大数据处理平台，是用友集团的第一款大数据产品。UDH基于Hadoop开源的分布式系统架构实现了可视化的集群管理器和数据处理服务，能够让用户在不了解底层运行细节的情况下，通过集群管理器对Hadoop集群中的节点、服务、数据进行便捷的管理和配置，支持节点和服务的实时监控及邮件预警，支持在可视化界面中对hadoop平台的结构化或非结构化数据进行SQL查询，支持自定义工作流和数据处理流程。

UDH目前已支持大部分主流的Hadoop数据处理组件。极大的降低了管理Hadoop集群的技术门槛，为用户提供了可横向扩展的，便捷、稳定、高效的大数据处理能力。

针对上文中提到的各种挑战，UDH又能为企业带来哪些具体的帮助呢？

1. 企业可以将已有的各种类型的数据存放至UDH的分布式文件系统中，使用UDH提供的大数据探查工具来查看数据的结构和内容，从中发现数据的规律或潜在的价值。
2. 企业可以将UDH作为统一的数据仓库，利用UDH的兼容性和可扩展性将各部门中散落的数据汇总、聚合，最终使数据打通，形成可被企业重复利用、进行分析的有价值数据。
3. UDH既能存储结构化数据，也能够存储非结构化数据，企业可以利用UDH提供的强大数据处理能力对各种数据进行预处理，从繁杂的数据中提取出最核心的价值。
4. UDH天生具备优良的可扩展性，可以随时根据数据处理能力和存储能力的需要对集群节点数量进行调整，可以轻松应对持续增长的数据量对系统架构带来的压力。
5. UDH集成了基于Kerberos网络认证协议的安全机制，对用户访问权限进行了严格限制，可以有效防止恶意用户的非法访问，极大提高了大数据集群的安全级别。
6. 使用UDH可以方便的通过Web界面对现有大数据集群进行管理，实现对集群添加节点、定义新节点上运行的服务、修改各个服务的配置等等操作，可以极大的降低大数据集群管理和维护的工作量和对维护人员的技术门槛要求，企业无需关心底层的运行原理，可以将更多的精力集中在数据分析等方面。

#### 总结

 综上所述，UDH依托于Hadoop强大的数据处理和分析能力，为用户提供了便捷、实用的及集群管理、集群监控、数据管理、数据探查、数据分析等功能于一身的综合平台，能够为企业的大数据集群运维、数据分析等工作提供强有力的支撑。