



### 亮点

- 处理一系列静态数据用例
- 支持实时客户应用
- 交付强大的分析能力

## Hortonworks 数据平台

用于管理动态和静态数据的开放架构平台

现在，任何业务都离不开数据。数据是企业的未来，是企业最宝贵的资产。Hortonworks 数据平台 (HDP) 是基于集中式架构的高度安全且企业就绪的开源 Apache Hadoop 版本 (YARN)。HDP 可以满足静态数据需求，支持实时客户应用，交付强大的分析能力，帮助加快决策和创新。

### Hortonworks 优势

HDP 凭借以下优势，释放大数据的全部潜力，帮助企业实现转型：

开放	集中	可互操作	企业就绪
HDP 由多个 Apache 软件基金会 (ASF) 项目构成，支持企业部署、整合和处理海量的结构化和非结构化数据。ASF 的方法是交付企业级软件，用于促进创新和防止供应商锁定。	YARN 是开放式企业级 Hadoop 的架构中心。它在处理数据的不同应用程序之间分配资源。YARN 协调集群范围的运营、数据治理和安全性服务。YARN 通过支持企业分析数据，最大限度增加数据采集，从而支持不同的用例。这个过程使 Hadoop 运营商能够满怀信心地让企业中尽可能多的受众接触自己的大数据资产。	它 100% 的开源架构支持 HDP 与广泛的数据中心和商业智能应用程序互操作。HDP 的互操作性能够帮助以最低的成本和最少的工作量，将客户的 IT 基础架构与 HDP 的数据和处理能力连接起来。有了 HDP，客户在采用 Hadoop 时就可以维持其对现有 IT 架构的投资。	HDP 专为企业而构建。开放式企业级 Hadoop 提供一致的操作，通过单一管理平台集中管理和监控集群。有了 HDP，该平台就具备了安全性和治理能力。该功能帮助提供高度安全的环境，并在各种数据访问引擎中实现一致的管理。



## Hortonworks 数据平台

HDP 提供基于集中式架构的高度安全且企业就绪的开源 Hadoop 版本。HDP 可以满足各种静态数据用例需求，支持实时客户应用，交付强大的分析能力，从而加快决策和创新。

### 数据管理

HDP 的基础组件是 Apache Hadoop YARN 和 Hadoop 分布式文件系统 (HDFS)。HDFS 面向大数据湖提供可扩展的、具备容错能力和成本效益的存储，而 YARN 提供集中式架构，支持企业同步处理多个工作负载。YARN

还提供资源管理和可插式架构，用于支持广泛的数据访问方法。

### 数据访问

通过以 YARN 作为架构中心，HDP 提供一系列处理引擎，允许用户以多种方式与数据进行同步互动。YARN 支持在同一个集群中存在多种针对共享数据集的访问方法。这一特性可以避免不必要且代价高昂的数据孤岛。HDP 支持多种数据处理引擎，包括交互式结构化查询语言 (SQL) 以及数据科学和批处理实时流，都可以使用单个平台中存储的数据。

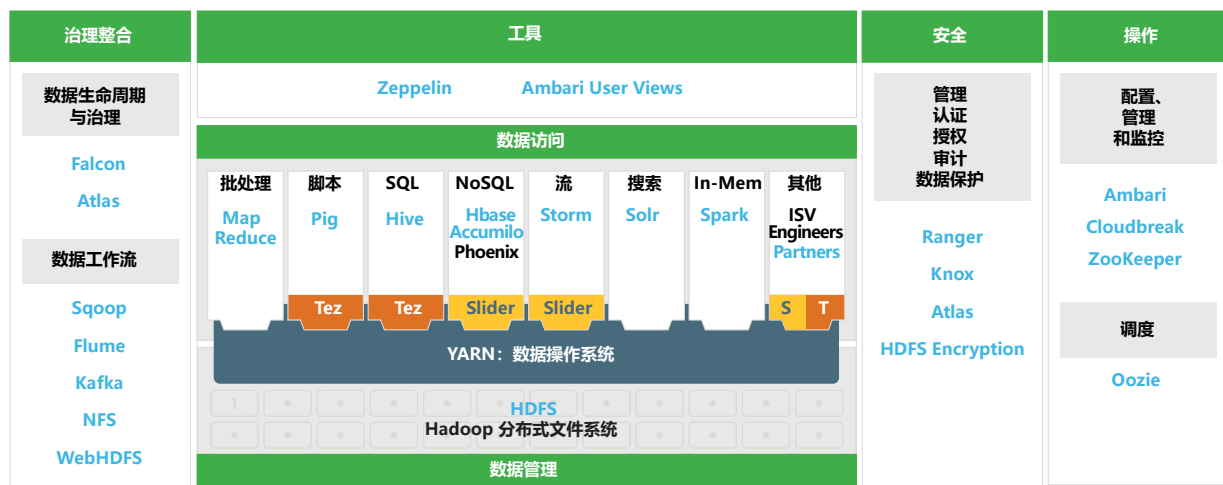


图 1：新一代 Hadoop 安全性

### 安全性与治理

随着企业纷纷采用 Hadoop 计划，以期抓住机遇获得数据驱动的洞察，数据治理和安全性需求可能会构成一项重大挑战。为应对这种挑战，多个行业的领导者组成“数据治理计划” (DGI) 联盟，其目的是满足开源治理解决方案管理数据分类、数据沿袭、数据安全性和数据生命周期的需求。

Apache Atlas 作为 DGI 的一部分而创建，支持企业在数据生态系统中应用一致的数据分类标准。Apache Ranger 为 Hadoop 提供集中的安全性管理。Hortonworks 通过将 Atlas 与 Ranger 相集成，支持企业在运行时制定动态访问策略，从而帮助主动预防违规事件。

这种整合成使企业能够实施基于分类的动态安全策略。Ranger 的集中式平台支持数据管理员根据 Atlas 元数据标签或属性确定安全策略。然后，他们可以对数据集的整个层级（包括数据库、表和列）实时应用该策略。

### 安全性

Hadoop 支持的数据湖可以为新一代分析和洞察提供强大的基础。然而，在启动或扩展 Hadoop 计划之前，保护数据非常重要。通过确保将数据保护和治理功能嵌入到其大数据环境中，企业可以充分利用高级分析的价值，避免将业务置于新的风险之中。

### 治理

随着企业纷纷采用 Hadoop 计划，以期抓住机遇获得数据驱动的洞察，数据治理需求可能会构成一项重大挑战。对于传统架构和大数据架构来说，管理信息，从而发现其中的价值并实现对客户和企业数据的有效性控制、安全性和合规性，都是核心要求。

### 操作

HDP 操作旨在通过消除手动流程中的猜测，以预配置的自动化最佳实践、引导配置和完整操作控制取而代之，支持 IT 企业将 Hadoop 快速上线。HDP 操作帮助简化分布式多用户、多租户和多数据访问引擎的操作，通过一体化 Web 用户界面或单一管理平台大规模管理 HDP 集群。

HDP 使用开源管理平台 Apache Ambari 来配置、管理、监控和保护 Hadoop 集群。Ambari 消除了与操作 Hadoop 相关的手动且容易出错的任务。它还提供所需的整合点，以便无缝整合到企业当中。

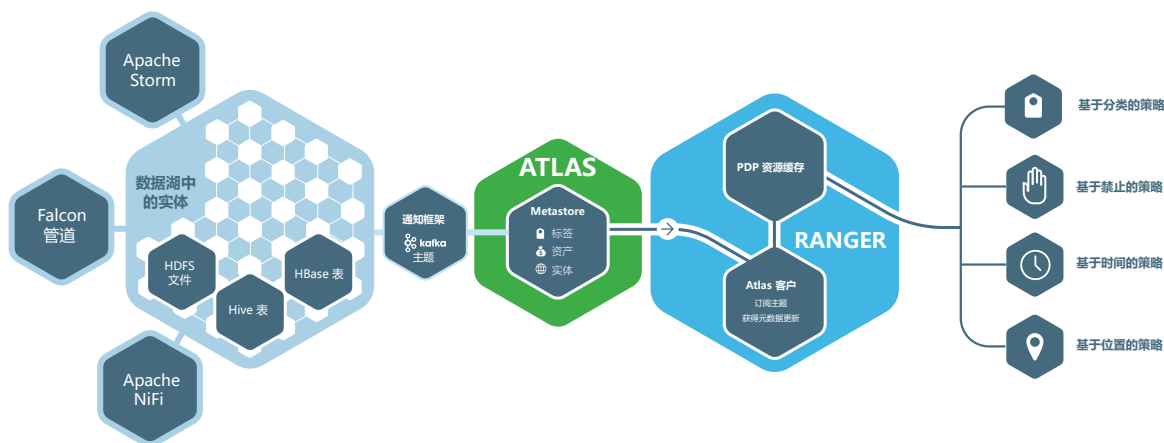


图 2：新一代 Hadoop 安全性

## 部署选项

HDP 提供一系列基础架构选项，用于部署开放且灵活的数据平台。用户可以灵活组合基础架构选项，做出最适合其独特用例的选择。

### 本地

多家企业已经投资于数据中心基础架构和外包服务，并且现在正考虑采用 Hadoop 的功能，他们将会发现本地实施是个可行的选项。HDP 可以轻松部署到本地，与现有数据中心实现整合。

### 云端

HDP 可以作为 Microsoft Azure HDInsight 的一部分部署到云端。Azure HDInsight 是 Microsoft Azure 云上的一款外包服务产品，由 HDP 提供支持。该部署选项支持企业通过随时启动任意数量个节点，按需将数据量从 TB 级扩展到 PB 级。借助 HDInsight，企业还可以将他们的本地 Hadoop 集群与云端连接起来。

### 混合云和 Cloudbreak

Cloudbreak 是在云基础架构上配置 Hadoop 集群的解决方案。Cloudbreak 作为 HDP 的一部分，由 Apache Ambari 提供支持，可帮助企业简化在云端的集群配置，通过弹性扩展优化云计算资源的使用。它面向采用本地 Hadoop 部署并且希望更轻松地在云端构建集群的客户。借助 Cloudbreak，客户可以选择自己的云供应商，让 Cloudbreak 在云端配置集群。

## 面向团队的 HDP

任何企业成功部署 Hadoop，都取决于利用现有技能和资源来采用大数据架构。HDP 可为大数据团队中的每个成员提供有用的工具和功能。

### 数据研究员

Apache Spark 作为 HDP 的一部分，在数据科学方面发挥着重要作用。数据研究员通常使用机器学习，即可以从数据中学习的一套技术和算法。这些算法往往具备迭代能力，而且 Spark 在内存中缓存数据的能力显著加快了迭代数据处理速度，使它成为实施此类算法的理想处理引擎。

### 业务分析师

HDP 支持业务分析师通过 Apache Hive、Spark SQL 和 Apache Phoenix 提供的 SQL on Hadoop 界面快速访问大量数据。通过这些界面，业务分析师可以利用他们最喜欢的商业智能和业务分析工具来创建报告、可视化功能、仪表板和计分卡，从而制定更加有效的洞察驱动型决策。

### 开发人员

HDP 为应用程序开发人员提供一系列丰富的工具，比如 Spark 和 Apache Zeppelin。HDP 还具有一组原生应用程序接口 (API)，有助于简化开发：WebHDFS 提供 REST 接口，用于编写、操作和删除 HDFS 文件，而 WebHCAT 提供关键整合点，用于访问 Hadoop 数据的元数据和模式。

### Hadoop 操作人员

使用 Ambari, Hadoop 操作人员可以获得以下优势, 用于简化 Hadoop 操作:

- 更加灵活的升级: Ambari 支持在集群关闭时通过自动执行维护和功能发布更快速地升级集群。
- 简化的安全操作: Ranger 的服务配置提供延续的新用户体验。Kerberos 凭证的可选存储和可定制的安全设置, 有助于简化管理并提供非常安全的环境。
- 改进的故障诊断: Ambari 提供可自定义的度量小部件图形显示, 并能够导出度量指标, 以快速识别和响应问题。

### 行业

HDP 支持各个行业的企业为其核心功能增加价值:

<b>金融服务</b> <ul style="list-style-type: none"><li>• 管理违约风险</li><li>• 提高客户交叉销售率</li><li>• 检测洗钱活动</li></ul>	<b>电信</b> <ul style="list-style-type: none"><li>• 分析通话详细记录 (CDR)</li><li>• 主动维护传输基础设施</li><li>• 使基础设施投资合理化</li><li>• 开发新产品和服务</li></ul>
<b>零售</b> <ul style="list-style-type: none"><li>• 构建全方位客户视图</li><li>• 提供本地化和个性化的客户体验</li><li>• 有效管理供应链</li><li>• 通过情绪分析了解品牌评价的变化</li><li>• 优化 Web 站点、营销活动和商店布局</li></ul>	<b>石油和天然气</b> <ul style="list-style-type: none"><li>• 监控远程位置的上游生产情况</li><li>• 减慢下降曲线</li><li>• 主动维护重要设备</li><li>• 报告环境运行状况和安全法规的合规情况</li></ul>

### 结语

数据是每个行业中每一家企业的基本工具。它决定了如何开发新产品和新服务, 如何有效运营等等。Hortonworks DataFlow (HDF) 帮助从 Hortonworks 数据平台 (HDP) 安全获取数据, 或向该平台传输数据, 进而管理动态数据。HDP 可以管理几乎所有类型的静态数据, 具备企业级的治理、安全性和操作能力, 可以帮助企业保持竞争优势。

### 为何选择 IBM ?

IBM 是美国一家跨国技术咨询企业, 总部位于纽约州阿蒙克市。IBM 制造并销售计算机硬件、中间件和软件, 提供涉及大型计算机、纳米技术等领域的托管和咨询服务。IBM 还是一家大型研究机构, 截至 2017 年, IBM 已连续 24 年保持在企业发明专利数量榜首。

### 了解更多信息

要了解有关 Hortonworks 数据平台的更多信息, 请联系您的 IBM 代表或 IBM 业务合作伙伴, 或者访问: [ibm.com/hadoop](http://ibm.com/hadoop)



---

© Copyright IBM Corporation 2017

IBM Corporation  
New Orchard Road  
Armonk, NY 10504

美国印刷  
2017 年 7 月

IBM、IBM 徽标及 [ibm.com](http://ibm.com) 是 International Business Machines Corporation 在全球许多司法管辖区的注册商标。其他产品和服务名称可能是 IBM 或其他公司的注册商标。Web 地址 [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml) 的“Copyright and trademark information”部分中包含了 IBM 商标的最新列表。

Microsoft、Azure 和 HDInsight 是 Microsoft Corporation 在美国和 / 或其他国家或地区的商标。

本文档为自最初公布日期起的最新版本，IBM 可能会随时对其进行更改。IBM 并不一定在开展业务的所有国家或地区提供所有这些产品或服务。

评估和验证任何其他产品或程序与 IBM 产品和程序一起运行的情况，则由用户自行负责。

本文档内的信息“按现状”提供，不附有任何种类（无论是明示的还是默示的）的保证，包括不附有关于适销性、适用于某种特定用途的任何保证以及非侵权的任何保证或条件。IBM 产品根据其提供时所依据的协议条款和条件获得保证。

Hortonworks 不是 IBM 产品或服务。Hortonworks 根据 Hortonworks 条款和条件向用户出售或提供许可（视情况而定），这些条款和条件随该产品或服务提供。Hortonworks 的可用性以及任何和所有保证、服务和支持是 Hortonworks 的直接责任，并由 Hortonworks 直接向用户提供。

客户应负责遵守适用的法律法规。IBM 不提供法律建议，也不表示或保证其服务或产品会确保客户遵守任何法律法规。

良好安全实践声明：IT 系统安全性涉及通过预防、检测和响应来自企业内外的不当访问，保护系统和信息的安全。不当访问可能导致信息被篡改、毁坏或挪用，或者可能导致您的系统被损坏或滥用，包括用来对他人进行攻击。没有任何 IT 系统或产品可视为是完全安全的，也没有任何单一产品、服务或安全措施能够完全有效地预防不当访问或使用。IBM 系统、产品和服务旨在成为全面的合法安全方法的组成部分，这必然会涉及额外的操作程序，还可能需其他系统、产品或服务才能达到最佳效果。IBM 不保证任何系统、产品或服务免受任何一方恶意或非法行为的影响，或使您的企业免受任何一方恶意或非法行为的影响。



请回收利用