

满足新一代数据管理需求的  
TxData 企业级大数据一体机

### 硬件环境概览

- 高达 1536 核 CPU 及 16TB 内存的大数据运算集群
- 1 到 16 台 2U 4 节点或 2 节点高密度服务器
- 每机架高达 768T 的 SATA 物理磁盘存储空间
- 高达 76.8T 的 SSD 智能计算闪存
- 完全冗余实现高可用性
- 可选用高速 Infiniband 网络，最高可配 128 个万兆口

### 软件环境概览

- 无限水平扩展
- 软硬一体化数据处理平台
- 超融合架构软件定义集群
- 支持容器虚拟化、多租户、硬件资源池
- 高速数据分析
- 完整的 SQL2003、PL/SQL、R 语言支持
- 分布式机器学习算法
- 灵活数据处理
- 实时流计算
- 超高性价比

Transwarp TxData Appliance (简称 TxData) 通过软件一体化装置实现计算、存储、网络、虚拟化和应用融合的紧密集成的大数据平台。TxData 基于 Docker 和 Kubernetes, 支持在物理机集群中以极小的系统开销创建多个 TDH 大数据集群, 提供大数据集群的动态伸缩和故障恢复能力。TxData 进行了独特的工程优化, 使得 TxData 性能得到飞速提升, 包括 SSD 闪存加速、系统优化等。TxData 可将您所有的数据处理负载性能最大化, 包括联机事务处理 (OLTP)、数据仓库 (DW) 和混合工作负载, 是海量大数据仓库应用的最佳平台。TxData 实施简便, 可以即时处理最大型的业务应用, 数据处理速度通常提升 2 倍甚至更多。



### 便捷实施、性能加速的企业级 Hadoop 一体化系统

TxData 是提供运行 TDH 企业级 Hadoop 平台所有硬件环境的一个简单快捷易部署的系统。TxData 通过软件一体化装置实现计算、存储、网络、虚拟化和应用融合的紧密集成的大数据平台。大数据平台的关键组件 Namenode、Zookeeper、HMaster 服务器以及网络环境已经预配置、预优化、预测试, 免去了通常规划、设计、部署 Hadoop 系统时的复杂性。所有组件无缝集成地工作, 不存在影响整个系统的性能瓶颈或单点故障。

TxData 进行了独特的工程优化, 使得 TxData 性能得到飞速提升。TxData 采用 SSD 闪存加速, TxData 允许用户将频繁访问的表数据保存在 PCIE-SSD 缓存中, TxData 还采用独有专利技术 Holodesk, 在 SSD 闪存中实现列式存储, 更进一步提高了 SSD 闪存加速效果, 在数据分析和挖掘场景下整体性能提升 10 倍以上。TxData 通过中断优化、内核优化、BIOS 参数优化、NUMA 架构优化, 使系统运行状态处于最优化状态。

TxData 运行的是 TDH 企业级 Hadoop 大数据综合平台, 因此任何运行在其他 TDH 安装点上的应用均可无缝移植到 TxData 而无需做任何改动。

### 超融合架构软件定义 Hadoop 集群

软件定义的模式正在改变数据中心的运维经济学。TxData 将客户简化 IT 运维, 降低相关成本, 同时提供更大的灵活性, 以满足迅速变化的业务需求。TxData 通过融合基础架构实现大数据平台标准化, 将它作为其下一代数据中心的基础。TxData 旨在建立自动化的、适应业务需求的大数据动态基础架构, 全面支持 Docker 的现有管理、自动化、监控和可用性工具, 可显著简化 IT 运维, 同时降低成本。

TxData 有物理机和容器虚拟化两个版本。TxData 容器虚拟化版本基于 Transwarp Operating System (TOS) 构建。TOS 采用 Docker 和 Kubernetes 虚拟化技术, 在 Docker 容器之上构建多个 Hadoop 集群, 实现了软件定义 Hadoop 集群, 进行容器生命周期管理, 并将 Docker 单机管理扩展到集群管理, 进行基于优先级的抢占式资源调度和细粒度资源分配。TDH 基于 Dockers 进行了定制开发, 保证 TDH 各个软件组件能正常稳定运行在 Docker 容器中。用户以多租户形式使用容器资源, 创建属于自己组织的 TDH 大数据集群, 实现真正的隔离并提高物理资源的利用率。

超融合架构 TxData 一体机具有如下优点:

运维自动化: 实现大数据集群的弹性部署、弹性扩容、自动修复, 显著简化 IT 运维, 使客户有更多精力关注业务, 并进行创新。

应用标准化: 采用统一的容器镜像进行开发、测试和部署, 一次开发, 到处部署。通过应用商店来管理标准化的应用。采用微服务架构来构造大型应用。

大数据支持: 采用容器这种轻量级虚拟化技术, 实现计算资源和存储资源的高效利用, TDH 在 TOS 之上完成了定制开发, 可以在容器之上部署大数据集群, 实现了软件定义大数据集群, 使大数据云服务成为现实。

## 优秀的系统扩展性和扩容性

TxData 采用了横向扩展的架构，可随着 TxData 节点的增加而使得 CPU、内存、存储、网络等均在一种平衡模式上增加，确保扩展没有瓶颈。TxData 是高密度计算集群，在 2U 空间承载 4 个高性能计算存储节点，节省机房空间和能源消耗。

旧有的数据库和数据仓库系统是以磁盘为中心，使用基于磁盘寻道的优化算法来加快数据库执行速度。TxData 系统是采用大容量内存和智能闪存来加快工作负载执行。为了用最低的成本达到最高水平的性能，TxData 系统实现了一个大容量存储分级体系，可以将所有活跃数据移动到闪存和 RAM 内存，而将不活跃数据保持在低成本的磁盘上。

横向扩展架构可以与任何规模的工作负载相匹配，并允许从很小到极大的配置无缝扩展，同时免除了性能瓶颈和单点故障。

TxData 中的所有组件可选配万兆以太网或者 Infiniband 网络或者千兆以太网互联。Infiniband 具有高带宽低延迟 56Gb/s 速率的特点，且采用特殊的网络互联协议，效率更高。

TxData 分两个产品型号：高密度型号和中高密度型号。

- 高密度型号一个单元具有 4 个计算节点、最高 96 个 CPU 核心、1024GB 内存、4.8TB 闪存容量、48TB 磁盘容量。
- 中高密度型号的一个单元具有 2 个计算节点、最高 48 个 CPU 核心、512GB 内存、2.4TB 闪存容量、48TB 磁盘容量。

TxData 支持在线升级，保证随数据存储和计算需求的增长而平滑升级。

除了在一个机柜内升级的扩容方法以外，通过 Infiniband 网络或者万兆以太网互联可以将多个机柜互联形成更大的配置。举例来说，由四个满配机柜组成的系统是一个满配机柜系统性能的四倍——四倍的 I/O 吞吐量、存储容量、处理器个数及内存容量供计算所用。它可以作为一个大型的独立系统，也可以逻辑上分割，作为多个 Hadoop 集群的集合体。TxData 的横向扩展很容易：TDH 本身就提供了动态增加数据处理和存储的能力。

TxData 可以在现有 TxData 中无缝集成新一代的处理器和存储，从而保护投资。同样的，新发布的软件可以和旧款 TxData 兼容。

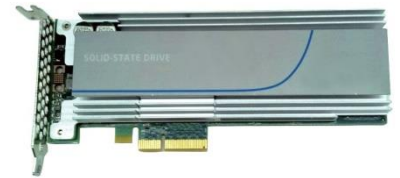


## SSD 闪存的超高性价比

每个高密度服务器的节点包含 1 块 1.2T PCIE SSD 盘。满配机柜的 TxData 包含 64 块 SSD 盘，76.8TB SSD 空间。

TxData 允许用户将频繁访问的表数据保存在 SSD 中缓存中，同时不常访问的数据仍保存在磁盘中。这种方式兼具 SSD 的高性能与磁盘的大容量和低成本。TDH 允许用户任意指定数据库表、分区和段的数据驻留在 SSD 上。

通过这种方式，TxData 减轻了内存计算的成本压力，在没有显著损失内存计算性能的情况下，帮助用户节省了购买成本，将可用于内存计算的表数据容量提高了一个数量级。



## 统计分析和数据挖掘的极速性能

TxData 允许用户将二维数据表缓存入独立的分布式内存或 SSD 中，建立列式存储、分区/分块和索引，采用改进后的内存计算引擎 Spark 作为执行引擎，SQL 执行性能比 Apache Hadoop/Hive 快数十倍，性能超过主流 MPP 数据库 2 倍到 10 倍。

当源数据无法全部放进内存时，TxData 支持换入换出，保证 SQL 执行不中断。业务不受限于内存大小，可对大于内存容量的数据进行分析处理。同时 SSD 可作为数据的二级缓存，从而大幅扩展内存存储的容量。当将数据放于 SSD 上进行 SQL 统计分析或数据挖掘时，性能比机械硬盘仍要快 10 倍以上。

每个 TxData 服务器节点包含最高 256G 内存和 1.2T SSD 空间。满配机架的 TxData 包含 16T 内存和 76.8T SSD 空间，超大的一级和二级缓存容量，最大程度保障了大数据环境下用户仍能获得统计分析和数据挖掘的极限性能。

## 关键应用的高可用性

TxData 采用工程化设计，可提供最高级别的高可用性。所有故障类型均可得到防护，包括磁盘、服务器、网络等简单故障，以及站点故障和人为失误等复杂故障。TxData 拥有完整的冗余的硬件保护，包括冗余的配电单元 (PDU)、冗余的供电电源、冗余的 NameNode 服务器、冗余的 DataNode 服务器。Zookeeper 用于防范 HMaster 故障。HDFS 提供磁盘块备份机制以防范磁盘或存储服务器故障。TDH Erasure Code 提供了更为高效和安全的存储机制。

TDH 具有行业领先的高可用性，已经在多家大型企业的关键应用中进行了部署，包括政府的企业关联查询系统、运营商的流量经营和经营分析系统、金融机构的统计分析系统、省级交通实时监控和分析系统、电商的实时榜单和智能推荐系统等。在这些关键业务的应用中，TDH 不局限于 OLAP 系统的统计分析负载，也包含数据库的 OLTP 负载。

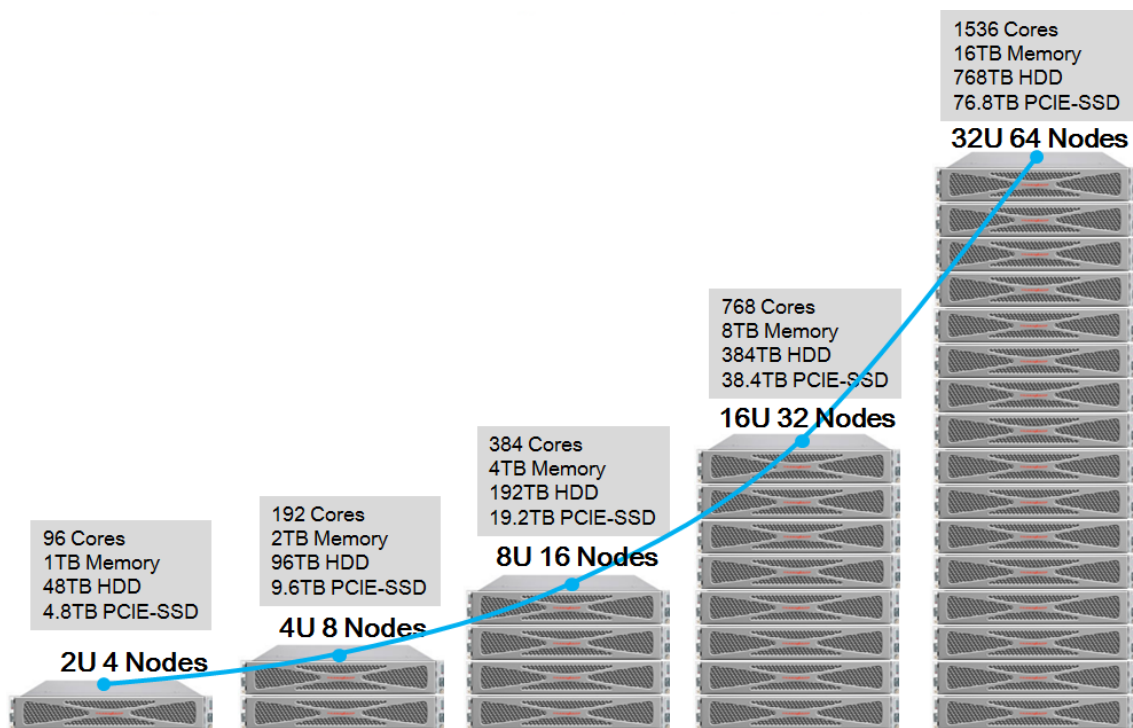
## 安全性

TxData 在多个层面提供了安全机制来保障大数据服务器的数据安全和通信安全。TxData 支持 Kerberos 和 LDAP 进行身份认证和访问权限控制。支持对文件、目录、表、列、行等细粒度的访问控制，支持数据加密和传输加密。

## 完善的系统管理

Transwarp Manager 提供对整个集群系统从监控、管理到维护全面的生命周期管理手段。通过提供的 Web 图形化界面和快速向导，帮助用户在短时间内部署一个或多个集群，同时还配上丰富的性能指标监控和功能故障监控，并且可定制邮件告警。

TxData 高密度型号 TxData-4 以 4 个节点构成一个基本单元，每个单元高 2U，用户最少需要购买 4 个节点才能正常使用，采用堆叠方式最多支持横向扩展到 64 个节点。多配置方案提供灵活性让用户只购买真正需要的，满足从业务负载整合到极速性能需求等各种场景。使用 TxData 可以利用现有的 IT 投资和流程即可扩展大数据计算环境。TxData 采用可信赖、业界主流技术同步发展、持续进化更新的先进技术，让用户的业务在行业处于创新领先地位。



TxData-4 产品型号配置如下：

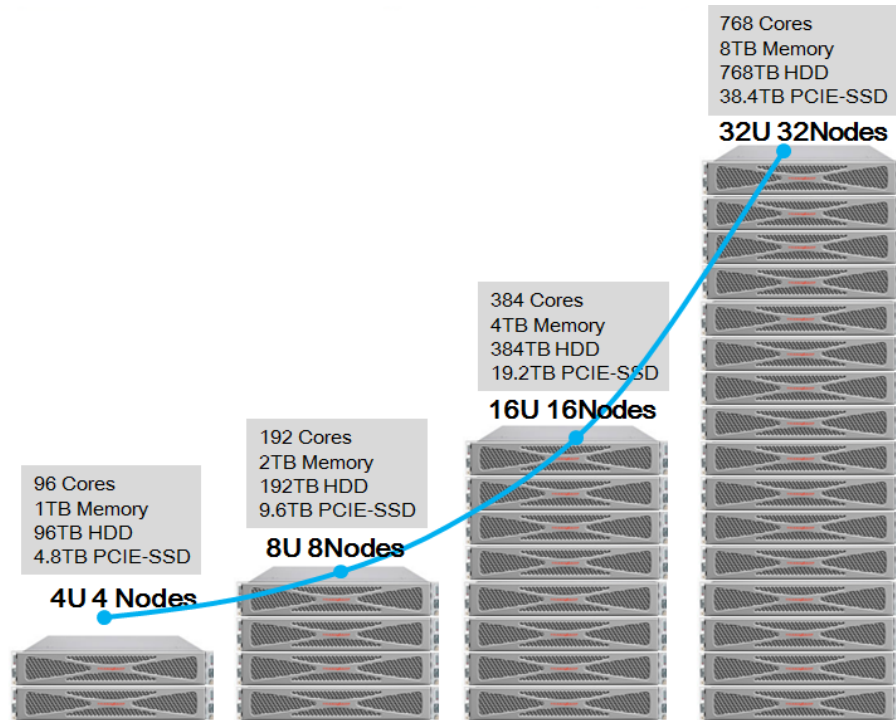
TxData 超融合架构企业级大数据一体机 TxData-4 规格参数

COMPONENTS	TxData-4S	TxData-4M	TxData-4H
PROCESSOR CORES (PER NODE)	12/16	16/20	20/24
PROCESSOR (PER NODE)	2 Intel E5-2620 v3 2.4 GHz OR 2 Intel E5-2620 v4 2.1 GHz	2 Intel E5-2630 v3 2.4 GHz OR 2 Intel E5-2630 v4 2.2 GHz	2 Intel E5-2650 v3 2.30 GHz OR 2 Intel E5-2650 v4 2.2 GHz
MEMORY/RAM (PER NODE)	64GB /128GB /256 GB	64GB /128GB /256 GB	64GB /128GB /256 GB
CACHING PCIE-SSD (PER NODE)	800GB/1.2TB(option)	800GB/1.2TB(option)	800GB/1.2TB(option)
STORAGE-RAW (PER NODE)	6TB/12TB	6TB/12TB	6TB/12TB
NETWORK CONNECTION	Dual Port 10GE(option) Dual Port 1GE	Dual Port 10GE(option) Dual Port 1GE	InfiniBand 56G Dual Port 10GE(option) Dual Port 1GE

MINIMUM NODES PER CLUSTER	4	4	4	
MAXIMUM NODES PER CLUSTER	64	64	64	
SCALING INCREMENTS (IN NODES)	4	4	4	
CHASSIS	2U rack-mounted chassis supporting 4 hot swappable nodes and 2 hot swappable power supplies			
POWER SUPPLIES	2 1600W high- efficiency redundant PSUs, 220V AC 50/60Hz			
COOLING	Dedicated cooling/node (no single point of failure) – 4x 8cm heavy duty fans with PWM fan speed control			
<b>物理规格</b>				
COMPONENTS	HEIGHT (MM/IN)	WIDTH (MM/IN)	DEPTH (MM/IN)	WEIGHT (MAX KG/LB)
APPLIANCE	3.47" (88mm)	17.25" (438mm)	26.75" (679mm)	103 lbs (46.7kg)
<b>操作环境</b>				
运行温度/湿度：	5°C 到 32°C (41°F to 89.6°F)，相对湿度 10% 到 90%，不可冷凝。			
运行高度：	最高 3,048 米。在高于 900 米的地方，高度每上升 300 米温度下降 1°C。			
认证体系 <sup>1</sup>	<ul style="list-style-type: none"> <li>•安全：UL 60950-1 第二版，EN60950-1:2006 年第 2 版，CB Scheme 随国家不同而各异</li> <li>•射频/电磁干扰：FCC CFR47 A 类 B 子条目第 15 部分，EN 55022:2006+ A1：2007A 类，EN 61000-3-11:2000，EN61000-3-12:2005，ETSI EN 300 386 V1.4.1(2008 年)</li> <li>•抗干扰：EN 55024:1998+A1:2001:+A2:2003</li> </ul>			
资质认证 <sup>1</sup>	<ul style="list-style-type: none"> <li>•安全：UL/cUL，CE，BSMI，GOST R，S-Mark，CSA C22.2 No.60950-1-07 2nd Ed，CCC</li> <li>•电磁兼容性：CE，FCC，VCCI，ICES，KCC，GOST R，BSMI Class A，AS/NZ 3548，CCC</li> <li>•其他：符合 WEEE 指令 ( 2002/96/EC ) 和 RoHS 指令 ( 2002/95/EC )</li> </ul>			

<sup>1</sup> 在某些情况下，标准和认证为组件级别的。

TxData 中高密度型号 TxData-2 以 2 个节点构成一个基本单元，每个单元高 2U，用户最少需要购买 4 个节点才能正常使用，采用堆叠方式最多支持横向扩展到 32 个节点。TxData-4 和 TxData-2 多配置方案提供灵活性让用户只购买真正需要的，满足从业务负载整合到极速性能需求等各种场景。



TxData-2 产品型号配置如下：

**TxData 超融合架构企业级大数据一体机 TxData-2 规格参数**

COMPONENTS	TxData-2S	TxData-2M	TxData-2H
PROCESSOR CORES (PER NODE)	12/16	16/20	20/24
PROCESSOR (PER NODE)	2 Intel E5-2620 v3 2.4 GHz OR 2 Intel E5-2620 v4 2.1 GHz	2 Intel E5-2630 v3 2.4 GHz OR 2 Intel E5-2630 v4 2.2 GHz	2 Intel E5-2650 v3 2.30 GHz OR 2 Intel E5-2650 v4 2.2 GHz
MEMORY/RAM (PER NODE)	64GB /128GB /256 GB	64GB /128GB /256 GB	64GB /128GB /256 GB
CACHING PCIE-SSD (PER NODE)	800GB/1.2TB(option)	800GB/1.2TB(option)	800GB/1.2TB(option)
STORAGE-RAW (PER NODE)	12TB/18TB/24TB	12TB/18TB/24TB	12TB/18TB/24TB
NETWORK CONNECTION	InfiniBand 56G(option) Dual Port 10GE(option) Dual Port 1GE	InfiniBand 56G(option) Dual Port 10GE(option) Dual Port 1GE	InfiniBand 56G(option) Dual Port 10GE(option) Dual Port 1GE
MINIMUM NODES PER CLUSTER	4	4	4
MAXIMUM NODES PER	32	32	32

## CLUSTER

<b>SCALING INCREMENTS (IN NODES)</b>	2	2	2	
<b>CHASSIS</b>	2U rack-mounted chassis supporting 2 hot swappable nodes and 2 hot swappable power supplies			
<b>POWER SUPPLIES</b>	2 1600W high- efficiency redundant PSUs, 220V AC 50/60Hz			
<b>COOLING</b>	Dedicated cooling/node (no single point of failure) – 4x 8cm heavy duty fans with PWM fan speed control			
<b>物理规格</b>				
<b>COMPONENTS</b>	<b>HEIGHT (MM/IN)</b>	<b>WIDTH (MM/IN)</b>	<b>DEPTH (MM/IN)</b>	<b>WEIGHT (MAX KG/LB)</b>
<b>APPLIANCE</b>	3.47" (88mm)	17.25" (438mm)	26.75" (679mm)	85lbs (38.6kg)
<b>操作环境</b>				
<b>运行温度/湿度：</b>	5°C 到 32°C (41°F to 89.6°F)，相对湿度 10% 到 90%，不可冷凝。			
<b>运行高度：</b>	最高 3,048 米。在高于 900 米的地方，高度每上升 300 米温度下降 1°C。			
<b>认证体系<sup>1</sup></b>	<ul style="list-style-type: none"> <li>•安全：UL 60950-1 第二版，EN60950-1:2006 年第 2 版，CB Scheme 随国家不同而各异</li> <li>•射频/电磁干扰：FCC CFR47 A 类 B 子条目第 15 部分，EN 55022:2006+ A1：2007A 类，EN 61000-3-11:2000，EN61000-3-12:2005，ETSI EN 300 386 V1.4.1(2008 年)</li> <li>•抗干扰：EN 55024:1998+A1:2001:+A2:2003</li> </ul>			
<b>资质认证<sup>1</sup></b>	<ul style="list-style-type: none"> <li>•安全：UL/cUL，CE，BSMI，GOST R，S-Mark，CSA C22.2 No.60950-1-07 2nd Ed，CCC</li> <li>•电磁兼容性：CE，FCC，VCCI，ICES，KCC，GOST R，BSMI Class A，AS/NZ 3548，CCC</li> <li>•其他：符合 WEEE 指令 ( 2002/96/EC ) 和 RoHS 指令 ( 2002/95/EC )</li> </ul>			

<sup>1</sup> 在某些情况下，标准和认证为组件级别的。

## Transwarp TxData Appliance 物理机版本软件说明

TDH 软件	
对于数据存储计算服务器	Transwarp Hadoop 企业版 Transwarp Inceptor 分析型数据库企业版 Transwarp Hyperbase 实时在线数据库企业版 Transwarp Discover 分析软件企业版
对于流处理服务器	Transwarp Hadoop 企业版 Transwarp Inceptor 分析型数据库标准版 Transwarp Stream 流处理引擎企业版
操作系统	
对于所有服务器	在安装时可选择： CentOS (64 位) 6.1, 6.2, 6.3, 6.4, 6.5
TDH 的核心功能特性	
<p>高速 SQL 分析和 R 语言数据挖掘能力</p> <p>完整的 SQL99 语法支持以及 PL/SQL 存储过程扩展支持</p> <p>灵活的存储索引技术，支持多维索引、组合索引，加速查询效率</p> <p>高效的 OLTP 应用、高并发 OLAP 应用、批处理应用、全文搜索以及高并发图形数据库检索应用</p> <p>支持结构化、半结构化、非结构化等多种数据结构的存储和检索</p> <p>支持实时批量的流处理计算</p> <p>支持计算集群的多租户概念</p> <p>支持在同一份数据集上运行多种计算框架</p>	
TDH 的高可用性	
<p>所有服务器的冗余电源、冗余风扇、冗余千兆以太网或万兆以太网或 InfiniBand 网络</p> <p>Transwarp 分布式文件系统：所有数据块都有备份块或校验块；磁盘故障不中断数据查询过程</p> <p>Namenode HA：可防范 Namenode 故障</p> <p>Zookeeper 机制：可防范 HMaster 故障</p>	
TDH 的可管理性	
TDH 内嵌 Transwarp Manager (自带报警监控组件)	

## Transwarp TxData Appliance 容器虚拟化版本软件说明

Transwarp Operating System (TOS)	
<p>TOS 是为大数据应用量身订做的云操作系统。基于 Docker 和 Kubernetes, TOS 支持一键部署 TDH, 基于优先级的抢占式资源调度和细粒度资源分配, 让大数据应用轻松拥抱云服务。未来企业对于构建统一的企业大数据平台 (Data Hub) 来驱动各种业务具有强烈需求, 统一的企业大数据平台需要提供以下功能:</p> <p>一: 资源弹性共享 - 提高资源利用率</p> <p>灵活部署: Big Data + Application</p> <p>资源调度: Auto-scaling + Self-healing</p> <p>服务发现: Central Repository</p> <p>二: 隔离性 - 保障服务质量和安全性</p> <p>数据隔离: Data Sources, Access Pattern, Confidential Levels</p> <p>计算隔离: CPU, Memory, I/O</p> <p>应用隔离</p>	
TDH 软件	
对于数据存储计算服务器	Transwarp Hadoop 企业版 Transwarp Inceptor 分析型数据库企业版 Transwarp Hyperbase 实时在线数据库企业版 Transwarp Discover 分析软件企业版
对于流处理服务器	Transwarp Hadoop 企业版 Transwarp Inceptor 分析型数据库标准版 Transwarp Stream 流处理引擎企业版

## 操作系统

在安装时可选择：

对于所有服务器

CentOS (64 位) 6.1, 6.2, 6.3, 6.4, 6.5

## TDH的核心功能特性

高速 SQL 分析和 R 语言数据挖掘能力

完整的 SQL99 语法支持以及 PL/SQL 存储过程扩展支持

灵活的存储索引技术，支持多维索引、组合索引，加速查询效率

高效的 OLTP 应用、高并发 OLAP 应用、批处理应用、全文搜索以及高并发图形数据库检索应用

支持结构化、半结构化、非结构化等多种数据结构的存储和检索

支持实时批量的流处理计算

支持计算集群的多租户概念

支持在同一份数据集上运行多种计算框架

## TDH的高可用性

所有服务器的冗余电源、冗余风扇、冗余千兆以太网或万兆以太网或 InfiniBand 网络

Transwarp 分布式文件系统：所有数据块都有备份块或校验块；磁盘故障不中断数据查询过程 Namenode HA：可防范 Namenode 故障

## TDH的可管理性

TDH 内嵌 Transwarp Manager (自带报警监控组件)



## 支持与服务

星环科技对 TxData 产品系列提供一整套支持服务，包括 24x7 硬件服务、系统监控、其他标准的和客户化的软件安装和配置。如果出现问题，星环科技还将提供更快捷的响应并加速升级研发。该服务的软件维护和补丁由星环工程师执行。

星环科技专业级的企业级大数据一体机为客户可持续性架构的大数据平台提供坚强保证。客户因此可以轻松、高效地运用大数据技术，并为自己的企业带来新的增长动力。

### TxData 支持服务

- 硬件基本服务：3 年，工作时间内（周一至周五 早 8 点至下午 5 点），2 个工作日现场响应/更换部件
- 系统安装服务
- 软件配置服务
- TxData 一体机高级服务，包含一体机硬件、TOS、Inceptor、Hyperbase 和 Stream 支持。24x7，4 小时服务响应（根据现场距服务中心的距离，响应时间可能有所不同）
- TxData 一体化技术支持服务
- 系统升级支持服务，包含硬件安装和软件配置

### 关于我们：

星环信息科技(上海)有限公司是一家大数据领域的高科技公司，致力于大数据基础软件的研发。星环科技目前掌握的企业级 Hadoop 和 Spark 核心技术在国内独树一帜，其产品 Transwarp Data Hub (TDH) 的整体架构及功能特性堪比硅谷同行，在业界居于领先水平，性能大幅领先 Apache Hadoop，可处理从 GB 到 PB 级别的数据。星环科技的核心开发团队参与部署了国内最早的 Hadoop 集群，并在中国的电信、金融、交通、政府等领域的落地应用拥有丰富经验，是中国大数据核心技术企业化应用的开拓者和实践者。星环科技同时提供存储、分析和挖掘大数据的高效数据平台和服务，立志成为国内外领先的大数据核心技术厂商。