

# 一种海量数据长期保存TCO的计算模型

严文瑞 曹强 姚杰 谢长生

武汉光电国家实验室

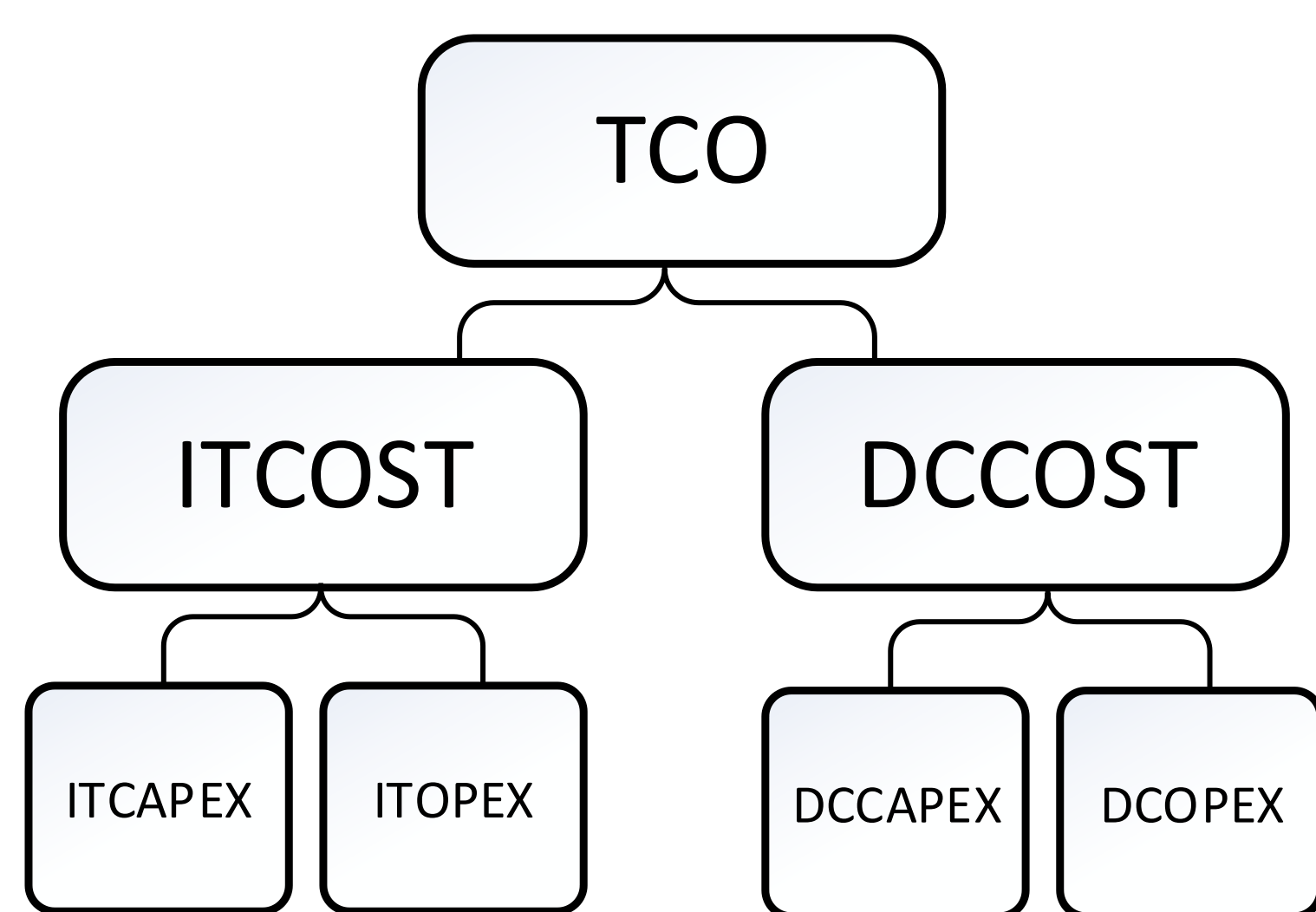
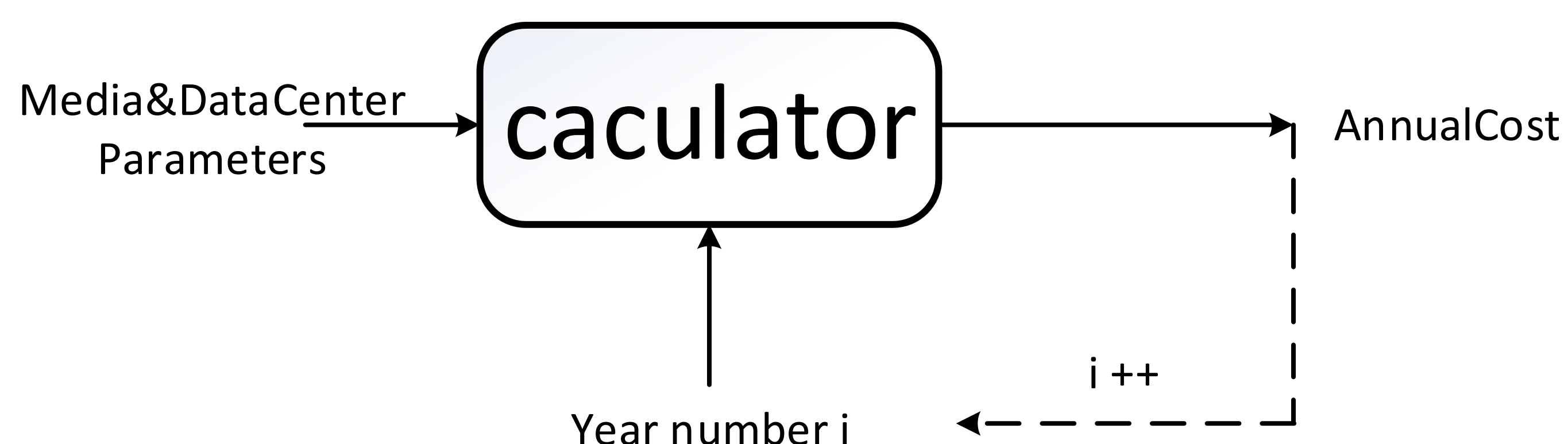
## 摘要

随着大数据时代的到来，数据总量急剧增加。其中大部分数据需要进行长时间保存来不断挖掘其价值。现有以磁盘和固态硬盘为存储介质的数据中心进行数据长期保存不是最优选择，需要一种归档型数据中心来长期存储数据。长期存储数据中心的总拥有成本是由数据中心采用的存储介质类型和散热、电力等基础设施决定的，此外存储介质密度增长率也会影响数据中心总拥有成本。本文提出了一种针对长期数据保存的数据中心总拥有成本的计算模型，其基本思想是以年为单位计算数据中心的每年的开销，最后累加得到总拥有成本。数据中心的总开销分为IT设施开销和基础设施开销，分别计算它们的投资开销和运维开销。选取蓝光光盘、磁盘、固态硬盘和磁带四种存储介质，通过计算得出蓝光光盘为介质的数据长期保存数据中心总成本最小。

## TCO计算模型设计

数据中心分为基础设施(DC)和存储设施(IT)设施的开销包括投资开销(CAPEX)和维护开销(OPEX)  
计算方法

- 计算第i年的开销AnnualCost<sub>i</sub>
- 最后TCO= $\sum_i$  AnnualCost<sub>i</sub>



### 基础设施

- 投资开销：建筑开销、散热设施开销、电力设施开销
- 维护开销
  - 能耗
  - 设施更换开销

### 存储设施

- 投资开销：存储介质、驱动、服务器、机柜
- 维护开销
  - 能耗、人力、设备失效更换（根据AFR计算）
  - 设施更换开销（到达介质寿命，强制更换）

## 实验设置

选取存储介质

- Optical disc library
- SSD
- HDD
- Tape library

数据中心参数

- 容量100PB
- 时间100年
- 介质密度年化增长率K<sub>r</sub>

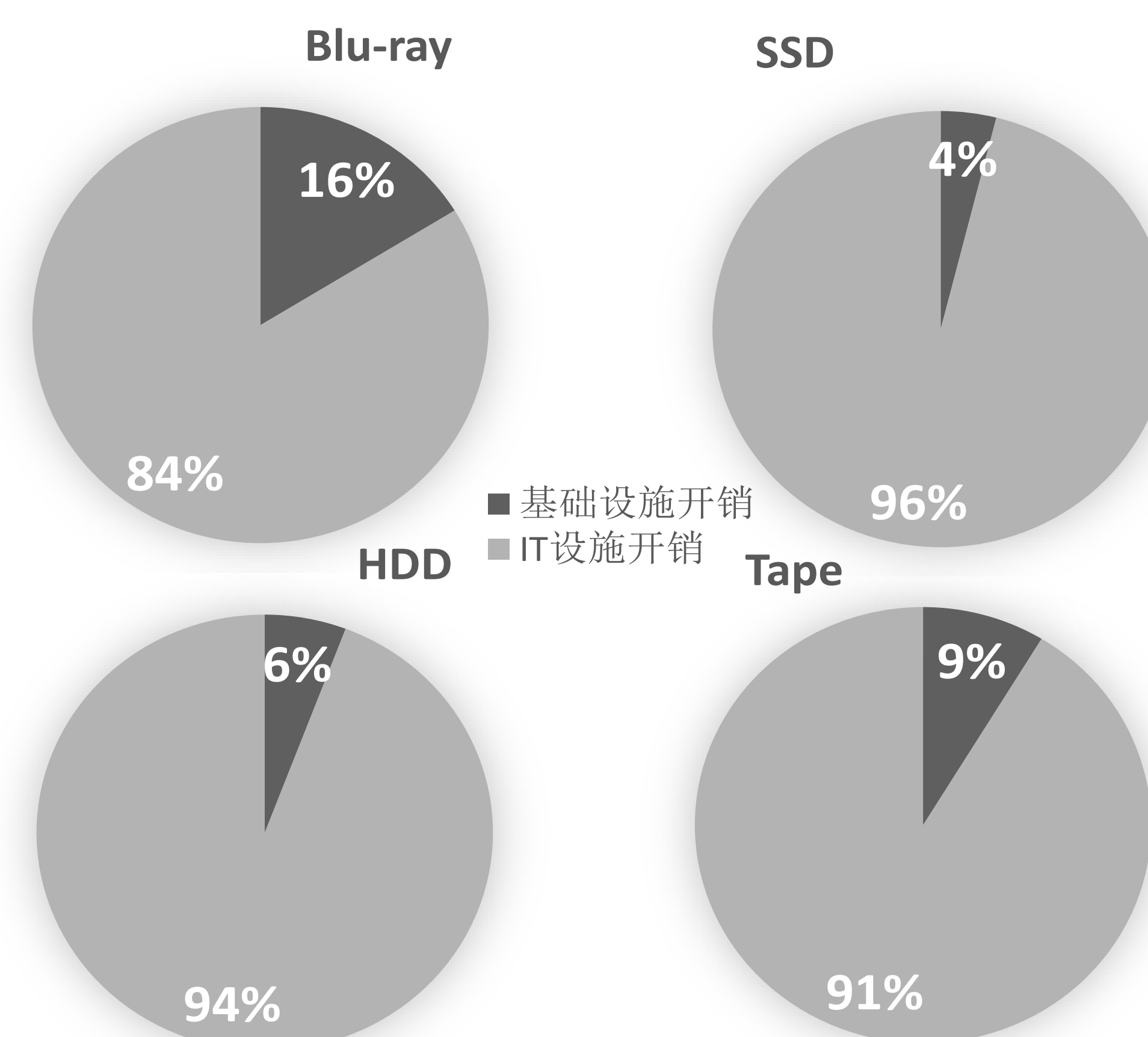
介质参数	Blu-ray	HDD	SSD	Tape
容量(TB)	0.1	4	0.25	2.5
价格(元)	20	1000	613	340
使用年限(年)	50	5	5	10
年化失效率(%)	0.05	4.85	0.61	1.5

## 实验结果

不同介质和不同介质密度增长率下的TCO计算结果（单位为元）

Kr	0	0.01	0.05
Blu-ray	119614519.573	94329700.083	71487629.522
HDD	899141333.333	601553190.571	259666746.271
SSD	8545783416.889	8369461468.205	8182853105.745
Tape	349829832.000	342579351.869	334905887.309

当Kr=0.01时，四种存储介质的基础设施开销和IT设施开销在总开销中所占的比例



## 结论和未来工作

本文提出了归档型数据中心的概念，设计并实现了一个数据中心总拥有成本TCO的计算模型。该模型将数据中心开销分为两部分：IT设施开销和基础设施开销。分别计算这两类设施每年的开销，最后累加得出数据中心的总开销。通过输入实际存储介质参数，得出使用不同介质构成数据中心的TCO。比较发现存储介质在不同的参数选择下对归档存储中心的可用性。

未来还需要做的工作有：1)长期归档中数据中心的基础设施的价格也会波动，本文未考虑长时间跨度中物质价值的影响；2)目前在介质换代时仅考虑了购买存储介质的开销，未考虑数据迁移的开销，后续需要在计算模型中细化数据迁移的开销计算部分。