International Journal of Emerging Research in Engineering and Technology
Pearl Blue Research Group| ICRCEDA2025-Conference Proceeding
ISSN: 3050-922X | https://doi.org/10.63282/3050-922X.ICRCEDA25-145

Original Article

Tackling Power Consumption Challenges in the Tech Industry: Apache Ozone's Role in Greener Data Centers

Venkateswara Varma Srivatsavaya
Principal Solutions Architect, Cloudera, USA.

Abstract *- As the digital economy expands, so does the infrastructure needed to support it. The tech industry has witnessed unprecedented growth in big data, cloud computing and AI. Underpinning all of these are massive data centers, the nerve centers of modern information technology. However, these data centers come at a cost a growing demand for power. Managing energy consumption has become one of the most pressing challenges in maintaining sustainable operations. This paper explores how the tech industry is grappling with the problem of increasing power consumption in data centers. It delves into how Apache Ozone, a scalable object store for big data workloads, is uniquely positioned to alleviate some of these issues. By supporting high-density storage and erasure coding, Apache Ozone can significantly reduce the data center footprint and, consequently, the power consumption and carbon emissions associated with running large-scale data infrastructure.*

Keywords *- Apache Ozone, data center power, erasure coding, storage efficiency, sustainable infrastructure*.

1. Current State of the Problem: Power Consumption in Data Centers

Over the past decade, the number of data centers has exploded. According to the International Energy Agency (IEA), data centres globally consumed an estimated 460 terawatt-hours (TWh) in 2022, and consumption could reach more than 1000 TWh in 2026. With the explosion of AI workloads, this demand is expected to rise sharply. Power consumption is no longer a marginal cost factor but a central concern in data center design and operation.

1.1. The Drivers of Power Consumption

Key contributors to energy consumption in data centers include:

- Compute Resources: High-performance CPUs and GPUs used for AI, analytics, and services
- Storage Systems: Large-scale data storage demands, particularly in object and file storage
- Cooling Requirements: Cooling infrastructure to maintain optimal temperatures
- Power Conversion and Distribution Losses: Energy lost during transformation and distribution

1.2. Environmental and Financial Impact

Energy consumption translates directly into operational costs and carbon emissions. With increasing regulatory and societal pressure to reduce carbon footprints, tech companies are under the microscope for their environmental practices. Data centers, particularly those using traditional storage architectures, represent a significant portion of their emissions footprint.

2. How to Reduce Data Center Footprint with Apache Ozone

Apache Ozone is a scalable, distributed object store designed to handle the immense data storage needs of modern enterprises. It originated from the Hadoop ecosystem and **is** built to support workloads that previously relied on HDFS (Hadoop Distributed File System), but with improved scalability and manageability. Apache Ozone helps with reducing the data center foot print by efficiently handling Storage Density and Data Redundancy.

2.1. Storage Density

Storage density per server refers to the amount of data that a single physical server can store. Traditional storage like HDFS supports 8TB per disk and up to 100Tb per server where as Ozone supports up to 16TB per disk and up to 400TB per server .The more data you can store in a single server/rack, the less physical infrastructure you need. This reduction in hardware translates to lower power consumption, both in terms of power used to operate the disks, servers and power used for cooling.

2.2. Data Redundancy

Data Redundancy is the process of storing multiple copies of data across different storage nodes to ensure fault tolerance, data availability, and reliability in the event of hardware failures.

Traditional storage like HDFS stores 3 Replicas of each data block on different nodes with 33% Storage Efficiency with 3x storage overhead. Ozone supports multiple Data Replication options i.e. 3X (RATIS) and Erasure Coding. Apache Ozone has native support for erasure coding. Administrators can choose policies based on required fault tolerance and performance characteristics, allowing them to balance redundancy with storage and energy efficiency.

Table 1: Comparison of Storage Efficiency and Overhead in Replication and Erasure Coding Methods

Method	Storage Efficiency	Multiplier = 1 / Efficiency	Notes
3x replication	33%	3×	Traditional HDFS storage
EC RS(3,2)	60%	1.67×	3 data + 2 parity
EC RS(6,3)	67%	1.49×	6 data + 3 parity

2.3. Erasure Coding

Erasure coding is a method of data protection in which data is broken into fragments, expanded and encoded with redundant data pieces, and stored across different locations or storage media. Unlike traditional replication, which stores multiple copies of data, erasure coding allows the same fault tolerance with far less storage overhead. For example, a 3x replicated file with 6 blocks will consume 6*3 = 18 blocks of disk space. But with EC (6 data, 3 parity) deployment, it will only consume 9 blocks of disk space which increases the storage efficiency. In EC, data is broken into data chunks (smaller units of data) and supplemented with parity chunks (calculated from the data) to enable fault tolerance.

Data Chunks:

- These are the original slices of your data
- The data is split into equal-sized parts before encoding

Parity Chunks:

- These are redundant chunks generated from the data chunks using mathematical formulas (XOR, Reed-Solomon, etc.)
- Used to reconstruct lost data chunks if some servers or disks fail

For example, with RS(6,3) (Reed-Solomon with 6 data, 3 parity) scheme,the 6 data chunks will be distributed to the first 6 data nodes in order and then the client generates the 3 parity chunks and transfers them to the remaining 3 nodes in order. These 9 chunks together are called a “Stripe”. Next 6 chunks will be distributed to the same first 6 data nodes again and the parity to remaining 3 nodes.

Imagine storing a file split into:

- Data Chunks: **D1 D2 D3 D4 D5 D6**
- Parity Chunks: **P1 P2 P3**

Even if any 3 chunks (e.g., D3, D6, P1) are lost or corrupted, the system can reconstruct the missing ones using the remaining chunks.

By combining high-density storage with erasure coding for data redundancy, Apache Ozone enables organizations to store more data in less physical space. This means:

- Fewer racks and servers.
- Smaller cooling zones.
- Reduced power delivery infrastructure

3. Case Example: Financial Services Organization

A Financial Services Organization with 50PB of Data on Traditional Storage(HDFS) with Low Density Storage (100TB per Server) and 3X Data Replication moving to Apache Ozone with High Density Storage (400TB per Server) and EC(6,3) Data Replication.

3.1. Hadoop Distributed File System - Low Density Storage with 3x Data Replication

Per Server Power:

- Disks 12x8TB HDDs at 7.2W each = 86.4W
- Server base power: ~200W (CPU, RAM, motherboard, etc.)
- Total: 86.4W + 200W = ~286W per server

Per Server Storage Capacity:

- Disks: 12x8TB HDDs = 96TB

Number of Servers needed for 50PB Raw Data:

- Raw Data – 50PB
- Replicated Data with 3x – 50 x 3 = 150PB = 153600TB
- Number of servers needed = 153600TB/96TB = ~1600 Servers
- Total Power Consumption = 1600 x 286W = ~458KW

3.2. Ozone Object Store - High Density Storage with Erasure Coding

Per Server Power:

- Disks 24x16TB HDDs at 7.2W each = ~180W
- Server base power: ~200W (CPU, RAM, motherboard, etc.)
- Total: 180W + 200W = ~380W per server

Per Server Storage Capacity:

- Disks: 24x16TB HDDs = 384TB

Number of Servers needed for 50PB Raw Data:

- Raw Data – 50PB
- Replicated Data with EC(6,3) = 50PB x 1.5 = 75PB =
- Number of servers needed = 76800TB/384TB = ~200 Servers
- Total Power Consumption = 200 x 380W = ~76KW

3.2. Power and Energy Savings

- Power Consumption Savings = 458KW – 76KW = 382KW
- Energy Savings per year = 382 KW × 24hrs × 365days = **3,346,320 kWh**

3.3. Estimated Cost Savings

- Commercial power cost (U.S.): $0.08–$0.12 per kWh

Annual savings range:

- Low estimate: $267,705/year (at $0.08/kWh)
- High estimate: $401,558/year (at $0.12/kWh)

3.4. CO_2 Emissions Avoided

- Estimated CO_2 intensity: ~0.4 kg CO_2 / kWh (U.S. grid average)

Annual CO_2 savings:

- 3,346,320×0.4=1,338,528kg CO_2=1,338.5metric tons CO_2/year

3.5. Real-World Equivalents (Per EPA Calculator)

- ~301 Gasoline cars off road per year
- ~1585 acres of forest carbon uptake per year
- ~316 U.S. homes powered per year
- ~623 NYC to London round trip flights avoided
- ~286 Million Smartphone Charges

4. How Ozone Can Help With Energy Efficiency and Sustainability

4.1. Aligning with ESG Goals

Environmental, Social, and Governance (ESG) goals are becoming essential for tech firms. Apache Ozone's energy-efficient architecture helps companies align with:

- Carbon reduction targets.
- Sustainable infrastructure initiatives
- Corporate social responsibility

4.2. TCO and ROI Benefits

Energy efficiency doesn't just help the planet; it also boosts the bottom line. By lowering:

- Hardware procurement costs
- Power and cooling costs
- Data center space needs

Organizations can realize substantial Total Cost of Ownership (TCO) savings with a favorable Return on Investment (ROI).

5. Conclusion

Power consumption is one of the most significant challenges facing the tech industry today, especially as data demands continue to soar. Apache Ozone presents a promising solution by enabling high-density, energy-efficient storage that reduces both physical and carbon footprints. With its support for erasure coding and optimized architecture, Ozone helps enterprises reduce operational costs and meet their sustainability goals. While migration and performance tuning require careful planning, the long-term benefits in terms of energy savings, environmental impact, and economic efficiency make a compelling case for its adoption. By embracing Apache Ozone and similar technologies, the tech industry can move toward a future where performance and sustainability go hand in hand. The road to greener data centers is not without its challenges, but with the right tools and strategies, it's a future well within reach.

References

[1] International Energy Agency (IEA): https://www.iea.org/reports/electricity-2024
[2] U.S. Energy Information Administration (EIA): https://www.eia.gov/electricity/monthly/
[3] EPA Greenhouse Gas Equivalencies Calculator: https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator