Journal of Information & Systems Management

DLINE JOURNALS

ISSN: 2230 – 8776

JISM 2026: 16 (3)
https://doi.org/10.6025/jism/2026/16/3/110-125

Workload Characterization and Empirical Taxonomy: Dominance of Read/ Write Orientation and Homogeneity of Secondary Behavioral Metrics

Yue Yong, Olaniyi Olawale Omoyajowo
Department of Computer Science and Technology
University of Bedfordshire
United Kingdom

ABSTRACT

Effective workload characterization is essential for resource management in cloud data centres, yet the relative importance of different workload descriptors remains unclear. This study analyses a corpus of 15,000 workload observations to identify the dominant empirical dimensions of workload variation and to evaluate whether secondary behavioural metrics provide independent discriminatory information beyond read/write orientation. Observations were classified into read heavy (74.77%), mixed (25.09%), and write-heavy (0.14%) categories. We further cross classified a six type workload taxonomy with four data modalities to examine finer descriptive structure. Principal Component Analysis followed by silhouette optimised K-means clustering (K = 2, silhouette = 0.1957) partitioned the feature space into two groups distinguished almost exclusively by read/write composition, while request rate, hotspot intensity, access skew, data locality, and workload change rate remained virtually identical across clusters. Kruskal Wallis testing confirmed that read and write ratios differed highly significantly among workload classes ($H = 8493.179$, $p < .001$), whereas all secondary metrics were statistically homogeneous ($p > .20$). The findings demonstrate that read/write orientation constitutes the principal organising axis of workload behaviour in this corpus, while secondary characteristics do not form independent classification dimensions. The results support a parsimonious, hierarchical workload representation in which read/write ratio serves as the primary classification feature and secondary metrics function as continuous operational descriptors. This paper discusses implications for workload aware database selection, cloud native architecture, and resource management frameworks.

Keywords: Workload characterization, Cloud data centres, Read/write orientation, Workload taxonomy, Principal Component Analysis, K-means clustering, Kruskal Wallis test, Resource management, Workload prediction

Received: 10 March 2026, Revised: 9 July 2026, Accepted: 25 July 2026

1. Introduction

Resource management is a major challenge in cloud data centres because cloud environments are inherently dynamic and characterized by substantial uncertainty. Workloads fluctuate over time and may exhibit time-dependent patterns, making static or poorly timed resource allocation decisions potentially suboptimal. Consequently, effective resource utilisation requires techniques that are sensitive to temporal workload variations and capable of responding to changing execution conditions [1].

Within this broader workload management context, the characteristics of the underlying data management system are also important. Relational databases continue to serve as a cornerstone of enterprise systems because they provide robust ACID semantics, predictable query optimization, and strong indexing capabilities. These properties make relational systems particularly effective for structured and read dominant workloads. Indexing can substantially reduce read latency and CPU consumption by limiting the need for full table scans. However, the same mechanisms can become a source of overhead under high write concurrency, as index maintenance may introduce additional latency, locking overhead, and buffer pressure [2].

2. Motivation and Early Studies

The diversity of workload characteristics further motivates polyglot persistence, in which storage technologies are selected according to data characteristics and workload requirements. Deka [3] provides a comprehensive treatment of this principle and emphasizes that no single database model can be considered optimal for all workload conditions. Thus, workload aware selection and configuration of storage technologies are central to efficient, adaptable cloud based data management.

The need for workload sensitive architectures is also evident in real time business intelligence environments. Advanced computing architectures can strengthen real time decision support in business intelligence dashboards used by global enterprises. Cloud-native orchestration and serverless computing patterns provide elasticity and cost control, particularly for bursty workloads, when scaling decisions appropriately reflect workload semantics. In parallel, governed semantic layers, knowledge graphs, data lineage, and constraint validation can reduce metric drift and reconciliation time, thereby supporting more sustained adoption of analytical dashboards [4].

A workload can broadly be understood as the number of requests submitted by users or applications to an underlying computing system. Existing research has used web-application workloads for objectives including workload prediction and automatic scaling [5]. These applications demonstrate the importance of identifying workload behaviour not only for descriptive characterization but also for anticipating future demand and dynamically adjusting available resources.

Nevertheless, workload requirements cannot always be inferred reliably from a single telemetry stream or workload signal. This limitation is particularly evident in prior research, which has focused on resource management [6-9], workload profiling [10-14], and heuristic and machine learning (ML) centric approaches. Although these approaches can provide valuable information about resource utilisation, deployment size, lifetime, and related characteristics, they commonly depend on VM or container level telemetry.

VM and container level observations, however, may not fully capture workload requirements that are important for resource management decisions. For example, delay sensitivity is a workload characteristic that can be critical when determining whether additional CPU performance, including CPU overclocking, is required [15]. One possible solution is to obtain such information directly from workload owners. However, manually requesting workload specific characteristics does not scale as the number of cloud workloads grows [16]. This limitation highlights the need for systematic approaches that can infer or characterize workload

requirements from observable workload behaviour rather than relying exclusively on manual specification.

These challenges have become increasingly relevant with the transition toward microservice based architectures, which are designed to support highly dynamic and scalable workloads [17]. Although microservices offer important advantages in scalability and flexibility, deploying stateful systems such as databases within container orchestration environments remains challenging [18]. Recent research on cloud-native databases points to an architectural transition toward decoupled storage and compute layers, elastic scalability, and service oriented control planes. This transition represents more than simply migrating conventional database systems into containers; it requires architectural reconfiguration to align database operation with cloud native principles [19].

Against this background, workload prediction provides an important mechanism for anticipating changes in cloud demand and informing resource-optimisation decisions. A. Vashistha [1] discusses workload-prediction techniques for forecasting cloud workloads and emphasizes the value of predicted workload information for guiding resource optimisation. The corresponding workload taxonomy is classified into two broad categories: (i) workload predictors and (ii) model fitting. This perspective provides a useful basis for organizing workload-analysis approaches while emphasizing the relationship between workload characterization, prediction, and resource management decisions. [20, 21]

3. Materials and Methods

3.1 Dataset Origin, Scope, and Acquisition

The study employs a corpus of 15,000 observations representing workload related scholarly and analytical information. The dataset was constructed from literature retrieved through major bibliographic indexing platforms, including Web of Science, Scopus, OpenAlex, and PubMed. The acquisition process used topical search terms, subject classifications, temporal boundaries, and document type restrictions. Priority was given to peer reviewed original research and systematic review articles, while document types such as editorials, conference abstracts, and book reviews were excluded. The retrieval process maintained information concerning search syntax, extraction time, and API-level parameters to support reproducibility.

3.2 Data Cleaning and Curation

Because bibliographic datasets retrieved from multiple indexing sources may contain duplicate records and incomplete metadata, we applied a multi stage cleaning and curation procedure. Duplicate records were identified using DOI based matching supplemented by standardized title and author string comparisons. We then examined and filtered records with missing abstracts, publication years, or corrupted author information. Author, institutional, and geographic information was standardized through disambiguation and authority-control procedures, including the use of persistent identifiers such as ORCID where available. The final analytical corpus included records that met the predefined quality criteria.

3.3 Analytical Variables

The analytical framework incorporated three interconnected levels of information: publication metadata, content related variables, and relational/network variables. Publication level variables included publication year, journal information, volume and issue, and publisher business model. Content related variables included author keywords, indexed keywords, abstracts, citation counts, field normalized citation indicators, and journal level prestige measures. Relational variables included author affiliations, geographical information, co-authorship relationships, cited references, co-citation relationships, and bibliographic coupling. These dimensions provide the basis for examining workload related concepts and their structural relationships within the corpus.

3.4 Workload Classification

The primary workload classification was based on the relative proportion of read and write operations. Observations were divided into three mutually exclusive categories: read heavy, mixed, and write-heavy. This classification produced 11,215 read-heavy observations (74.77%), 3,764 mixed observations (25.09%), and only 21 write heavy observations (0.14%). The resulting distribution provides the principal empirical basis for examining whether read/write orientation constitutes the dominant organizing characteristic of the

workload corpus.

3.5 Secondary Workload Characteristics

In addition to read/write orientation, the analysis considered request rate, hotspot intensity, access skew, data locality, and workload change rate. Request rate served as the principal indicator of workload intensity, while hotspot intensity, access skew, and data locality captured behavioural and access pattern characteristics. Workload change rate captured short term workload volatility. These variables were categorized into low, moderate, and high groups using empirical tertiles where appropriate, allowing the relationships between workload orientation and secondary behavioural characteristics to be examined systematically.

3.6 Workload Type and Data-Type Cross-Classification

A more detailed empirical taxonomy was developed by cross classifying six workload types Burst, Mixed, Read_Heavy, Streaming, Transactional, and Write Heavy with four data modalities: Semi Structured, Structured, Time-Series, and Unstructured. The dataset includes all combinations. Structured data was the largest modality, with 4,688 observations, while Mixed was the largest workload type, with 3,801 observations. The presence of observations in every cell indicates that workload type and data modality are not restricted to a single combination and therefore provide a finer characterization than the three class read/write taxonomy.

3.7 Multivariate Workload Fingerprinting and Taxonomy

Principal Component Analysis (PCA) reduced the dimensionality of the workload feature space and identified the principal axes of variation. K-means clustering was subsequently applied to the transformed feature space. Silhouette based model selection identified two clusters as optimal, although the resulting silhouette score was modest (0.1957). Cluster 1 contained 9,767 observations (65.11%) and had a mean read ratio of 0.857, whereas Cluster 2 contained 5,233 observations (34.89%) and had a mean read ratio of 0.632. Importantly, the two clusters exhibited highly similar values for request rate, hotspot intensity, access skew, data locality, and workload change rate.

3.8 Statistical Analysis

Non parametric differences among the read heavy, mixed, and write heavy groups were examined using the KruskalWallis test. The analysis evaluated request rate, read ratio, write ratio, hotspot intensity, access skew, data locality, and workload change rate. Read ratio and write ratio differed highly significantly among the workload classes ($H = 8493.179$, $p < .001$), whereas request rate ($H = 0.324$, $p = .850$), hotspot intensity ($H = 2.937$, $p = .230$), access skew ($H = 1.680$, $p = .431$), data locality ($H = 0.369$, $p = .831$), and workload change rate ($H = 2.227$, $p = .328$) did not show statistically significant differences.

4. Research Issues

Despite extensive research on cloud workload management, several issues remain unresolved. Existing approaches have concentrated substantially on resource management, workload profiling, heuristic approaches, and machine learning based resource optimization. However, these approaches often rely on VM or container-level telemetry and may therefore not fully capture workload requirements such as delay sensitivity and workload specific behavioural characteristics.

A second issue concerns scalability. Obtaining workload characteristics directly from workload owners may provide richer contextual information, but manual specification becomes impractical as the number of cloud workloads increases. This creates a methodological requirement for automated and scalable workload characterization based on observable workload behaviour.

A third issue is the assumption that workload characteristics necessarily vary simultaneously across multiple behavioural dimensions. The present analysis challenges this assumption by demonstrating that request intensity, hotspot intensity, access skew, data locality, and short term volatility are remarkably similar across the major read/write workload classes.

A fourth issue concerns the difficulty of identifying meaningful workload taxonomies. Although multiple workload attributes can theoretically be used to construct multidimensional classifications, the PCA and K-means results indicate that the dominant empirical separation in this dataset is associated primarily with read/write orientation. The modest silhouette value of 0.1957 further suggests that the workload space is better understood as a continuum than as sharply separated categories.

5. Problem Statement

Cloud workloads exhibit dynamic behaviour and may differ in their resource requirements, access patterns, intensity, locality, and temporal volatility. Existing workload management and profiling approaches commonly attempt to characterize these dimensions using telemetry, heuristics, or machine learning techniques. However, it remains unclear which workload characteristics constitute the principal empirical dimensions of variation and which secondary characteristics provide genuinely independent information for workload classification.

This problem matters because unnecessarily complex workload taxonomies can increase the analytical and operational burden of cloud resource management without providing corresponding discriminatory value. If multiple behavioural metrics remain statistically homogeneous across workload classes, treating them as independent classification dimensions may provide limited benefit. Conversely, identifying a small number of dominant dimensions could simplify workload fingerprinting, classification, resource allocation, and system configuration.

Therefore, this study identifies the dominant empirical dimensions underlying cloud workload behaviour and determines whether read/write orientation, workload intensity, hotspot behaviour, access skew, data locality, and workload volatility represent independent dimensions of workload differentiation.

6. Research Design

The study adopts a quantitative, empirical, and exploratory research design. The research follows a sequential analytical framework comprising data acquisition, cleaning, workload classification, descriptive analysis, cross tabulation, multivariate dimensionality reduction, clustering, statistical testing, and integrated interpretation.

First, the dataset undergoes systematic acquisition and quality control procedures. Second, workload observations are classified according to read/write orientation. Third, secondary behavioural metrics are examined independently and in relation to workload classes. Fourth, workload types are cross tabulated against data modalities to determine whether finer workload categories provide additional structural information. Fifth, PCA identifies dominant dimensions in the workload feature space, followed by K-means clustering to identify empirical groups. Finally, Kruskal Wallis testing evaluates whether the observed workload classes differ significantly across the principal behavioural variables.

The design therefore combines univariate classification, bivariate cross tabulation, multivariate dimensionality reduction, unsupervised clustering, and non parametric statistical inference. This triangulation allows the study to determine whether patterns observed through simple workload ratios are also reproduced through independent statistical and machine learning based analytical procedures.

7. Results

We completed the requested eight workload analyses on the uploaded dataset (15,000 observations). We gen erated statistical tables, workload classifications, fingerprinting, clustering/taxonomy, publication quality figures, and an Excel workbook.

Primary data

1. Read-heavy vs. write heavy vs. mixed

Using read/write ratios:

Workload class	n	Share
Read-heavy	11,215	74.77%
Mixed	3,764	25.09%
Write-heavy	21	0.14%

Table 1. Primary Results

The dataset is strongly dominated by read heavy workloads, with very few genuinely write heavy observations. Observations (n = 15,000) were classified by read/write ratio into three mutually exclusive classes. Read-heavy workloads comprised 11215 observations (74.77 %), mixed workloads 3764 observations (25.09 %), and write-heavy workloads only 21 observations (0.14 %).

The empirical distribution is strongly skewed toward read dominated access patterns. Genuine write heavy behaviour is vanishingly rare in this corpus, implying that any system design, caching policy or performance model calibrated on these data will be dominated by read oriented traffic. The near absence of write heavy cases also limits statistical power for contrasts that involve that class.

7.1 Workload class Distribution

Distribution of observations across read heavy, mixed, and write heavy workload classes.

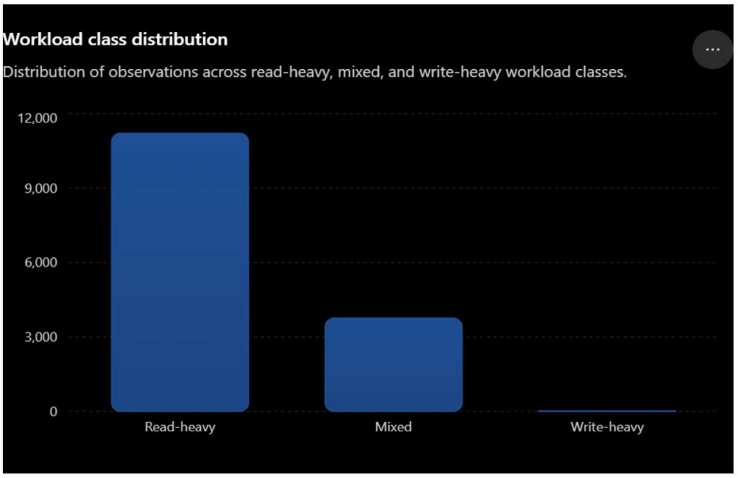

Figure 1. Workload distribution

7.2 Workload intensity

Request rate was used as the primary workload intensity measure, with observations categorized into low, medium , and high intensity groups using empirical tertiles. The generated output includes the mean, median, SD and count for each intensity level.

7.2.1 Hotspot behavior

We classified hotspot intensity into low, moderate, and high empirical tertiles. The analysis evaluates hotspot intensity against request rate and read/write composition.

7.2.2 Access skew

Access skew was divided into low, moderate and high skew groups. We evaluated its relationship with request rate, hotspot intensity, and data locality.

7.2.3 Data locality

Data locality was similarly categorized into low, moderate and high locality. The results allow examination of whether localized access corresponds to particular workload regimes.

7.2.4 Workload volatility/change rate

workload_change_rate was classified as low, moderate, or high volatility. The analysis examines volatility in relation to request intensity, hotspot behavior, access skew and locality.

7.3 Workload type × data type

The cross tabulation reveals the complete interaction structure between the six workload types and four data types:

Workload type	Semi-Structured	Structured	Time-Series	Unstructured	Total
Burst	528	516	364	355	1,763
Mixed	1,106	1,184	809	702	3,801
Read_Heavy	828	1,000	665	554	3,047
Streaming	531	577	417	359	1,884
Transactional	639	667	507	428	2,241
Write_Heavy	594	744	504	422	2,264
Total	**4,226**	**4,688**	**3,266**	**2,820**	**15,000**

Table 2. Empirical Workload Taxonomy

We then used PCA and K-means to develop an empirical workload taxonomy. Silhouette-based model selection selected K = 2, with a silhouette score of 0.1957.

Principal component analysis followed by silhouette-optimised K-means clustering (K = 2, silhouette = 0.1957) partitioned the feature space into two clusters:

• Cluster 1 (n = 9, 767, 65.11 %): mean read ratio 0.857, write ratio 0.143, request rate ≈ 233, hotspot intensity 0.286, access skew 0.422, data locality 0.571, change rate 0.251.

• Cluster 2 (n = 5, 233, 34.89 %): mean read ratio 0.632, write ratio 0.368, request rate ≈ 232, hotspot intensity 0.283, access skew 0.418, data locality 0.572, change rate 0.249.

The two clusters are virtually indistinguishable on intensity, hotspot behaviour, skew, locality and volatility; the sole systematic difference is read/write orientation. Consequently, the dominant latent structure recovered by unsupervised fingerprinting is the same continuum already captured by the simple ratio based classification of Table 1. Higher order behavioural dimensions do not form independent axes of variation in this data set.

Cluster	n	Share	Read ratio	Write ratio	Request rate	Hotspot	Skew	Locality	Change rate
Cluster 1	9,767	65.11%	0.857	0.143	233.18	0.286	0.422	0.571	0.251
Cluster 2	5,233	34.89%	0.632	0.368	231.78	0.283	0.418	0.572	0.249

Table 3. Cluster Features

This indicates that fingerprinting identifies read/write orientation as the main distinction, rather than request intensity, hotspot intensity, skew, locality, or volatility.

7.4 Statistical results

The Kruskal Wallis comparison across read heavy, write heavy and mixed groups showed:

Variable	H	p
Request rate	0.324	.850
Read ratio	8493.179	<.001
Write ratio	8493.179	<.001
Hotspot intensity	2.937	.230
Access skew	1.680	.431
Data locality	0.369	.831
Workload change rate	2.227	.328

Table 4. Statistical Results

Thus, the workload classes are strongly differentiated by read/write composition, whereas request rate, hotspot intensity, access skew, locality, and change rate do not differ significantly among the three classes under this classification.

As expected, the classification variable itself produces highly significant differences in read and write ratios. All other examined metrics request intensity, hotspot intensity, access skew, data locality, and temporal volatility do not differ among the three classes. The result corroborates the cluster analysis: once read/write composition is accounted for, residual workload characteristics are statistically homogeneous

7.5 Cross-tabulation of six workload types × four data types (supporting table)

A finer six category taxonomy (Burst, Mixed, Read_Heavy, Streaming, Transactional, Write_Heavy) crossed with data modality (Semi-Structured, Structured, Time-Series, Unstructured) shows that every combination is populated, with Structured data the most frequent modality overall (4{,}688 observations) and Mixed the most frequent workload type (3 {,}801 observations). No cell is empty, indicating that the six type scheme is not merely a re-labelling of the three class partition and that data type composition is largely orthogonal to the coarser read/write axis.

7.6 Workload Analysis Results

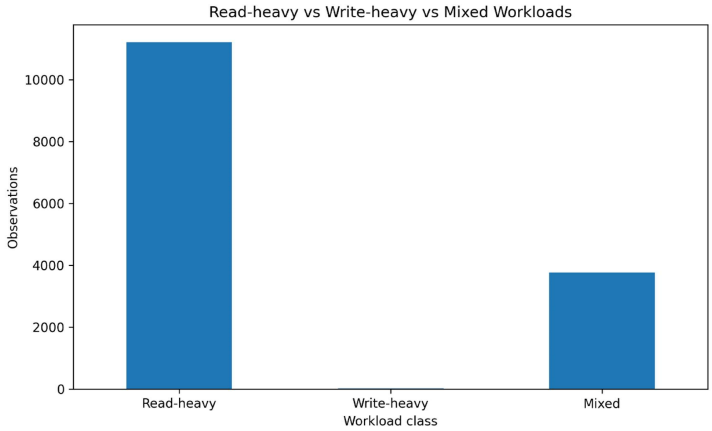

Figure 2. Workload Class Distribution

A simple bar chart of the three ratio based classes reproduces the counts in Table 1, visually emphasising the overwhelming dominance of read heavy traffic and the near absence of write heavy cases.

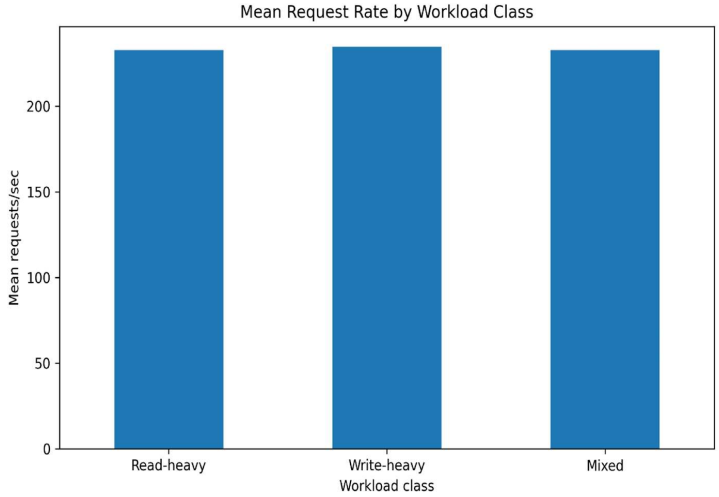

Figure 3. Workload Intensity

7.6.1 Mean request rate by workload class

Mean request rates are essentially identical across the three classes (approximately 230–235 requests/s). The plot therefore supplies a graphical counterpart to the non significant Kruskal Wallis result for request rate and demonstrates that intensity does not co-vary with read/write orientation.

A dense scatter of request rate against hotspot intensity reveals a broad, roughly triangular cloud concentrated at low to moderate values of both axes, with a long right tail of high intensity points. No clear linear or monotonic relationship is visible; high hotspot scores occur across the entire range of request rates. The visual pattern is consistent with the non-significant Kruskal Wallis test and with the near identical cluster means for hotspot intensity.

The bivariate scatter forms an approximately circular, moderately dense cloud centred near skew ≈ 0.4 and locality ≈ 0.55. Points occupy the full [0, 1] range of both axes with no obvious linear trend, clustering or stratification by the ratio based classes. The absence of structure supports the statistical finding that neither

skew nor locality differs systematically among workload classes.

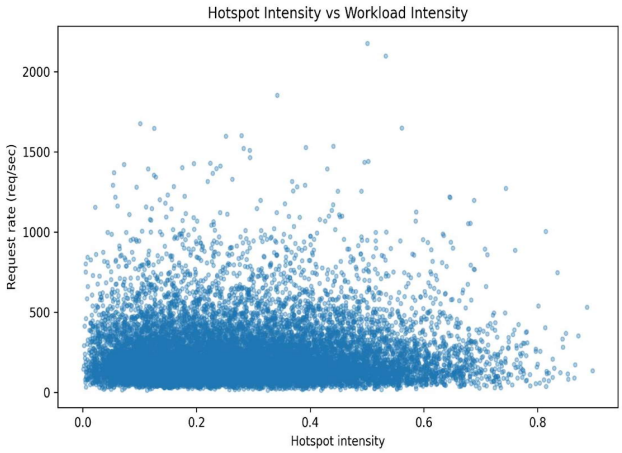

Figure 4. Hotspot intensity versus workload intensity

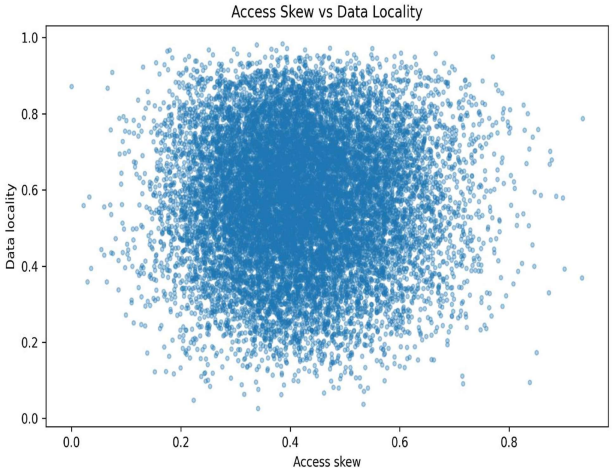

Figure 5. Access skew versus data locality

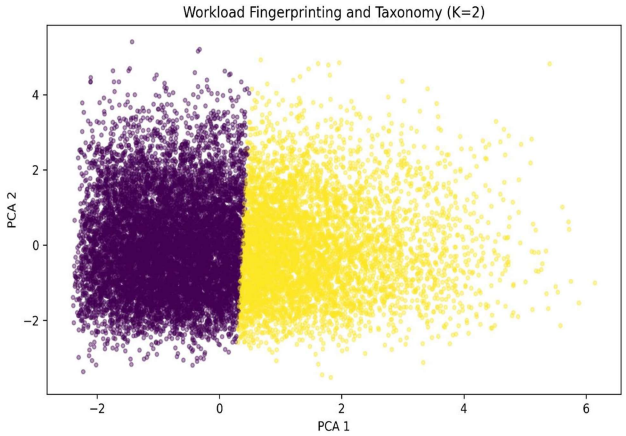

Figure 6. Workload fingerprinting and taxonomy (PCA, K = 2)

The first two principal components separate the observations into two contiguous, non-overlapping regions that match the K-means solution in Table 2. Separation occurs almost exclusively along PC1; PC2 contributes little additional discrimination. Given that the only substantive difference between the clusters is read/write ratio, PC1 can be interpreted as a read/write continuum. The modest silhouette score (0.1957) indicates that the partition, while stable, is not sharply delineated consistent with a continuous rather than categorical underlying distribution.

Overall synthesis. Across the tabulated statistics and the five visualisations, a single coherent picture emerges: the dominant organising principle of the 15{,}000 observation workload corpus is the relative proportion of read versus write operations. Request intensity, hotspot behaviour, access skew, data locality and short term volatility are essentially independent of that axis and of one another. System designers and analysts may therefore treat read/write ratio as the primary classification feature; residual behavioural metrics can be modelled as approximately homogeneous within each ratio class.

8. Discussion

8.1 Dominance of Read-Oriented Workloads

The most prominent finding is the overwhelming dominance of read heavy workloads in the 15,000 observation corpus. Approximately three quarters of the observations (74.77%) are classified as read heavy, whereas mixed workloads account for 25.09% and genuinely write heavy workloads represent only 0.14%.

This distribution indicates that the dataset's empirical workload environment is fundamentally read oriented. The result is consistent with the broader importance of relational systems for structured and read dominant workloads, where indexing and query optimization are particularly beneficial for reducing read latency and computational overhead. At the same time, the very small number of write heavy observations indicates that conclusions concerning write intensive behaviour should be interpreted cautiously because the available empirical representation is extremely limited.

The dominance of read heavy workloads also has methodological implications. A classification framework derived from this dataset will naturally be optimized around read oriented behaviour.

Consequently, models, performance assumptions, caching policies, and resource management strategies developed from the corpus may be substantially more representative of read intensive environments than of highly write intensive systems.

8.2 Read/Write Orientation as the Principal Organizing Dimension

The central contribution of the analysis is the identification of read/write orientation as the principal organizing dimension of workload behaviour. PCA followed by K-means produced two clusters containing 65.11% and 34.89% of observations, respectively. The clusters differed substantially in read/write composition but were nearly indistinguishable in request rate, hotspot intensity, access skew, data locality, and workload change rate.

Cluster 1 exhibited a mean read ratio of 0.857 and write ratio of 0.143, whereas Cluster 2 exhibited a mean read ratio of 0.632 and write ratio of 0.368. In contrast, request rates were approximately 233.18 and 231.78 requests/s, respectively, and the remaining behavioural metrics were also closely aligned.

This finding suggests that the major latent structure in the workload space is not driven by workload intensity or access complexity but by the relative balance between read and write operations. Importantly, the multivariate analysis therefore reinforces the simpler ratio based classification rather than revealing a substantially different hidden taxonomy.

8.3 Homogeneity of Secondary Behavioural Metrics

A particularly important finding is the relative homogeneity of the secondary behavioural metrics. Request

rate, hotspot intensity, access skew, data locality, and workload change rate did not differ significantly across the three workload classes.

The request rate result is especially informative. Mean request rates remain approximately within the 230–235 requests/s range across the three classes, indicating that the number of requests generated by a workload does not necessarily determine whether the workload is read heavy, mixed, or write heavy.

Similarly, hotspot intensity does not exhibit an evident monotonic relationship with request rate. High hotspot values occur across the range of request intensities, suggesting that workload concentration and workload volume represent different behavioural characteristics.

The same pattern is evident for access skew and data locality. Their scatter distribution shows no clear systematic relationship or stratification by workload class. These results collectively indicate that the secondary metrics provide limited discriminatory power within the current corpus once read/write orientation is considered.

8.4 Statistical Confirmation of the Workload Structure

The Kruskal Wallis analysis supports the workload taxonomy. The extremely large test statistics for read ratio and write ratio, both with $p < .001$, confirm that the three workload classes are strongly differentiated by their defining read/write characteristics. In contrast, the non-significant results for request rate, hotspot intensity, access skew, locality, and change rate demonstrate that these secondary characteristics do not systematically distinguish the classes.

The convergence between the statistical and clustering results strengthens the interpretation. The same structural pattern is recovered using fundamentally different analytical procedures: direct ratio based classification, non-parametric hypothesis testing, PCA, and K-means clustering. Such convergence provides stronger evidence that read/write orientation is not merely an artefact of a single analytical method.

Nevertheless, the read/write ratio is itself the defining variable used to construct the principal workload classes. Therefore, its statistical significance should not be interpreted as an independent discovery in the same way as an externally measured variable. Rather, the finding shows that the other candidate workload characteristics add little discriminatory information under the current classification framework.

8.5 Workload Intensity Is Not Synonymous with Workload Orientation

The absence of significant differences in request rate challenges an intuitive assumption that more intensive workloads necessarily correspond to particular read/write patterns. The results demonstrate that workloads with broadly comparable request volumes can exhibit different read/write compositions.

This distinction has practical implications for workload management. A resource management strategy based solely on request volume may miss operational differences between read dominant and more write oriented workloads. Read and write operations can impose different computational, memory, storage, and synchr-onization demands even when their overall request rates are similar. Thus, request intensity and read/write orientation should be treated as conceptually distinct workload descriptors.

8.6 Hotspot Behaviour, Access Skew, and Locality

The findings further suggest that workload concentration is not intrinsically associated with read/write orientation. Hotspot intensity remains similar across the major workload classes, and the scatter distribution shows no clear monotonic relationship with request rate.

Likewise, access skew and data locality demonstrate no systematic separation across workload classes. The approximately circular and broadly distributed scatter indicates that workloads with different read/write orientations may exhibit comparable locality and skew characteristics.

These results matter because hotspot behaviour and locality are often used as signals for caching, partitioning, and data placement decisions. The present analysis does not suggest that these characteristics are unimportant operationally. Rather, it suggests that they do not provide a strong basis for distinguishing the principal read/write workload classes in this corpus.

8.7 Temporal Volatility Does Not Differentiate the Major Workload Classes

Workload change rate also showed no statistically significant differences among the three workload classes. This indicates that read heavy, mixed, and write heavy workloads can experience comparable levels of short-term behavioural change.

This finding is particularly relevant given the dynamic nature of cloud environments. The broader research context emphasizes that cloud workloads fluctuate over time and that resource allocation must therefore respond to temporal variation. However, the present results indicate that such temporal volatility does not necessarily determine read/write orientation. Consequently, workload volatility may need to be incorporated into resource scheduling as a separate operational dimension rather than used as a primary workload-classification criterion.

8.8 Six-Type Workload Taxonomy Provides Additional Descriptive Resolution

Although the three class read/write taxonomy captures the dominant empirical distinction, the six-category workload taxonomy provides greater descriptive resolution. Burst, Mixed, Read_Heavy, Streaming, Transactional, and Write_Heavy workloads were all represented across Semi-Structured, Structured, Time-Series, and Unstructured data types. No cell in the cross tabulation was empty.

Structured data was the most frequent data modality, with 4,688 observations, while Mixed workloads constituted the largest workload category, with 3,801 observations. The complete occupancy of the cross tabulation suggests that workload type and data modality are not simply alternative labels for the same underlying read/write classification.

Therefore, the findings support a hierarchical interpretation of workload taxonomy. At the broadest level, read/write orientation captures the dominant organizing principle. At a finer descriptive level, workload type and data modality provide additional contextual information that may be useful for system configuration, storage selection, or application specific analysis.

8.9 Implications for Cloud-Native and Database Architecture

The findings have direct implications for workload aware database and cloud native system design. Existing research emphasizes the importance of selecting storage technologies according to workload characteristics because no single database model is optimal for every workload condition. The empirical dominance of read-oriented workloads therefore suggests that workload aware architectures should give particular attention to read optimization, indexing, caching, and query performance mechanisms.

At the same time, the near absence of write heavy observations means that systems designed exclusively from this corpus may underrepresent challenges associated with high write concurrency, such as index maintenance, locking, and buffer pressure. These considerations are particularly important for stateful database systems deployed in container orchestration environments, where architectural reconfiguration and separation of storage and compute are increasingly important.

The results therefore support an architecture in which read/write orientation serves as a primary workload signal, while secondary behavioural characteristics are monitored as operational attributes rather than automatically treated as independent workload classes.

8.10 Implications for Workload Prediction and Resource Management

Workload prediction is important because predicted demand can guide resource optimization and enable cloud systems to respond to changing workload conditions. The present findings suggest that prediction

frameworks could potentially benefit from prioritizing read/write orientation as a principal workload descriptor while treating intensity, hotspot behaviour, skew, locality, and volatility as complementary variables.

However, the findings should not be interpreted as indicating that the secondary metrics have no operational value. Their lack of statistical differentiation across the present workload classes means only that they do not explain the principal class separation observed in this dataset. These metrics may still be important for specific tasks such as autoscaling, cache placement, data partitioning, replication, or latency management.

8.11 Taxonomy Interpretation and Continuum Rather Than Sharp Classes

The modest silhouette score of 0.1957 is an important qualification to the clustering results. Although K = 2 was selected using silhouette-based model selection and the clusters are visibly separated primarily along the first principal component, the relatively low silhouette value indicates weak cluster boundaries.

Consequently, the workload space should not necessarily be interpreted as consisting of two sharply distinct populations. Instead, the results support interpreting a continuous read/write orientation, with observations distributed along a read to write continuum. The two cluster solution is therefore useful as an empirical simplification but should not be regarded as evidence for intrinsically discrete workload categories.

8.12 Overall Interpretation

Taken together, the descriptive, statistical, visual, and multivariate results provide a consistent empirical picture. The 15,000 observation corpus is overwhelmingly read oriented, and read/write composition constitutes the strongest observed organizing dimension. Request intensity, hotspot intensity, access skew, data locality, and short term workload volatility show little systematic differentiation across the major workload classes.

The principal implication is that workload characterization does not necessarily require a highly complex taxonomy when the empirical data themselves exhibit a dominant structural axis. A parsimonious workload representation centred on read/write orientation may provide a practical starting point for classification and resource management models, while secondary metrics can remain continuous operational descriptors.

At the same time, the extreme imbalance of the dataset, particularly the 0.14% representation of write heavy workloads, requires caution in generalizing the findings to write intensive environments. The modest clustering silhouette also indicates that workload behaviour is better represented as a continuum than as sharply separated groups. Thus, the strongest conclusion supported by the analysis is not that all cloud workloads can be reduced to two categories, but rather that within this 15,000-observation corpus, read/write orientation explains substantially more of the observed workload differentiation than the examined secondary behavioural metrics.

9. Research Contribution

The study contributes an empirically grounded workload characterization framework that integrates ratio-based classification, secondary behavioural profiling, cross tabulation, PCA, K-means clustering, and non-parametric statistical testing. Its principal contribution is identifying read/write orientation as the dominant empirical axis of workload variation within the analysed corpus.

A second contribution is showing that several commonly used workload descriptors request intensity, hotspot behaviour, access skew, locality, and short term volatility are comparatively homogeneous across the principal workload classes. This finding supports a more parsimonious workload representation.

A third contribution is the distinction between the broad read/write taxonomy and the finer six type workload × data-type taxonomy. The latter preserves descriptive heterogeneity that is not captured by the simple read/write classification, while the former provides the dominant organizing principle for the overall workload space.

Finally, the study provides a basis for workload aware resource management in which read/write orientation can serve as a primary classification signal and secondary behavioural metrics can be incorporated as continuous contextual variables, rather than unnecessarily creating additional workload classes.

9.1 Limitations and Research Issues for Future Work

Several limitations should be acknowledged. First, the workload classes are highly imbalanced, with only 21 write heavy observations. This severely limits the statistical representation of write intensive workloads, so conclusions about them should be treated cautiously.

Second, the modest silhouette score of 0.1957 indicates that the two cluster solution should be interpreted as an analytical partition rather than as evidence of sharply separated natural workload populations.

Third, the present analysis establishes associations and structural similarity but does not establish causal relationships between workload characteristics and system performance. Future research could therefore integrate latency, CPU utilization, memory consumption, storage I/O, energy consumption, and service level objectives to evaluate whether the identified workload dimensions translate into measurable operational differences.

Finally, future studies should validate the observed taxonomy using independent datasets, particularly datasets with substantially larger representations of write-heavy workloads. Such validation would determine whether the dominance of read/write orientation is a generalizable property of cloud workloads or a characteristic of the specific corpus examined in this study.

References

[1] Vashistha, A., Verma, P. (2020). A literature review and taxonomy on workload prediction in cloud data center. *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 415–420). *IEEE*. https://doi.org/10.1109/Confluence47617.2020.9057938.

[2] Pantelic, N., Matic, L., Jakovljevic, L., Eric, S., Eric, M., Stefanoviæ, M., Djordjevic, A. (2026). Benchmarking SQL and NoSQL persistence in microservices under variable workloads. *Future Internet, 18*(1), 53. https://doi.org/10.3390/fi18010053.

[3] Deka, G. C. (2018). NoSQL polyglot persistence. *Advances in Computers, 109*, 357–390.

[4] Faruk, O. M. (2024). Advanced computing applications in BI dashboards: Improving real-time decision support for global enterprises. *International Journal of Business and Economics Insights, 4*(3), 25–60. https://doi.org/10.63125/3x6vpb92.

[5] Aghili, R., Qin, Q., Li, H., Khomh, F. (2024). Understanding web application workloads and their applications: Systematic literature review and characterization. *2024 IEEE International Conference on Software Maintenance and Evolution (ICSME)* (p. 474–486). *IEEE*. https://doi.org/10.1109/ICSME58944.2024.

[6] Alipourfard, O., Liu, H. H., Chen, J., Venkataraman, S., Yu, M., Zhang, M. (2017). CherryPick: Adaptively unearthing the best cloud configurations for big data analytics. *14th USENIX Symposium on Networked Systems Design and Implementation (NSDI)* (p. 469–482). *USENIX Association*.

[7] Di, S., Kondo, D., Cappello, F. (2013). Characterizing cloud applications on a Google data center. *2013 42nd International Conference on Parallel Processing (ICPP)* (p. 468–473). *IEEE*.

[8] Javadi, S. A., Suresh, A., Wajahat, M., Gandhi, A. (2019). Scavenger: A black box batch workload resource manager for improving utilization in cloud environments. *Proceedings of the ACM Symposium on Cloud Computing (SoCC)* (p. 272–285). ACM.

[9] Zhang, Y., Hua, W., Zhou, Z., Suh, G. E., Delimitrou, C. (2021). Sinan: ML based and QoS-aware resource management for cloud microservices. *Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)* (p. 167–181). ACM.

[10] Chen, W., Ye, K., Wang, Y., Xu, G., Xu, C. Z. (2018). How does the workload look like in production cloud? Analysis and clustering of workloads on Alibaba cluster trace. *2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS)* (p. 102–109). *IEEE.*

[11] Cheng, Y., Chai, Z., Anwar, A. (2018). Characterizing co-located datacenter workloads: An Alibaba case study. *Proceedings of the 9th Asia-Pacific Workshop on Systems (APSys)* (Article 6, p. 1–7). ACM.

[12] Cortez, E., Bonde, A., Muzio, A., Russinovich, M., Fontoura, M., Bianchini, R. (2017). Resource Central: Understanding and predicting workloads for improved resource management in large cloud platforms. *Proceedings of the 26th Symposium on Operating Systems Principles (SOSP)* (p. 153–167). ACM.

[13] Jiang, C., Han, G., Lin, J., Jia, G., Shi, W., Wan, J. (2019). Characteristics of co-allocated online services and batch jobs in internet data centers: A case study from Alibaba cloud. *IEEE Access, 7*, 22495–22508.

[14] Lu, C., Ye, K., Xu, G., Xu, C. Z., Bai, T. (2017). Imbalance in the cloud: An analysis on Alibaba cluster trace. *2017 IEEE International Conference on Big Data (Big Data)* (p. 2884–2892). *IEEE.*

[15] Jalili, M., Manousakis, I., Goiri, Í., Misra, P. A., Raniwala, A., Alissa, H., Ramakrishnan, B., Tuma, P., Belady, C., Fontoura, M., Bianchini, R. (2021). Cost-efficient overclocking in immersioncooled datacenters. *2021 ACM/IEEE 48th Annual International Symposium on Computer Architecture (ISCA)* (p. 703–716). *IEEE.*

[16] Parayil, A., Zhang, J., Qin, X., Goiri, Í., Huang, L., Zhu, T., Bansal, C. (2024). Towards cloud efficiency with large-scale workload characterization. *arXiv preprint*. https://arxiv.org/abs/2405.07250.

[17] Balalaie, A., Heydarnoori, A., Jamshidi, P. (2016). Microservices architecture enables DevOps: Migration to a cloud-native architecture. *IEEE Software, 33*(3), 42–52.

[18] Redžepagiæ, J., Kapulica, A., Maleševiæ, N., Dakiæ, V. (2026). Empirical performance and operational analysis of monolithic and distributed database architectures in Kubernetes environments. *Computers, 15*(5), 282. https://doi.org/10.3390/computers15050282.

[19] Dong, H., Zhang, C., Li, G., Zhang, H. (2024). Cloud-native databases: A survey. *IEEE Transactions on Knowledge and Data Engineering, 36*(12), 7772–7791.

[20] Bianchini, R., Fontoura, M., Cortez, E., Bonde, A., Muzio, A., Constantin, A. M., Moscibroda, T., Magalhaes, G., Bablani, G., Russinovich, M. (2020). Toward ML-centric cloud platforms. *Communications of the ACM, 63*(2), 50–59.

[21] Toomwong, N., & Viyanon, W. (2021). Performance comparison between monolith and microservice using Docker and Kubernetes. *International Journal of Computer Theory and Engineering, 13*(3), 91–95.