BARC

BARC Benchmark

Evaluating productivity and scalability of BI, Analytics & CPM software under real-world conditions.

Starting with Power BI and Qlik.

Authors: Thomas Zeutschler, Stefan Sexl

Publication: September 2025

Abstract

The BARC Benchmark (BB) introduces a reproducible method and score to evaluate, compare and decide upon BI and Corporate Performance Management software based on real-world enterprise scenarios. This paper presents the scoring model and approach, outlines its long-term vision and shares the first – and notably very insightful – benchmark results for Microsoft Power BI and Qlik. We invite enterprises, consultants and vendors to review the approach, results and tooling – and to contribute honest feedback and ideas as we develop the BB forward to become an open, versatile industry-wide standard.

Table of Contents

The Need for Better (Software) Decisions

So many vendors. Still no easy path to choose the right one.

Despite the massive proliferation of business intelligence (BI), analytics and corporate performance management (CPM) software in recent years, most enterprises still struggle with very basic questions:

- **Which software truly fits our needs?**
- **Which platform delivers – not on paper, but in practice?**

The answer is rarely obvious. It's often buried beneath endless feature checklists, glossy dashboard demos, polished reference calls, biased consultant advice, over-optimistic vendor promises and a chronic lack of time and resources for a proper in-house evaluation.

Top-level evaluations such as quadrants or our own BARC Scores – a combination of structured end-user feedback and deep analysis by experienced and specialized analysts – offer valuable guidance on capabilities, features and suitability for certain use cases. But they can't fully answer some of the most critical questions in software selection:

- Will it actually improve **our** BI, analytics and/or planning process?
- Can it handle the complexity of **our** business and data models?
- Will it scale with **our** data volumes and user base?

These questions can only be answered reliably under real-world conditions – using actual data volumes, concurrent users, often complex and highly specific modeling requirements and realistic user interactions. Comparing multiple solutions under such conditions is simply beyond the reach of most organizations due to time, budget and resource constraints.

The underestimated importance of performance as a success factor

In our yearly "The BI & Analytics Survey," performance consistently ranks among the top three factors influencing the overall success and business benefits of BI & analytics solutions. Yet in many software evaluations, performance is still treated as a "nice to have" – until users begin to drop off.

Studies such as Nielsen's usability heuristics, Miller's 2-second rule, and research from Google show that users begin to disengage if responses take longer than 2–3 seconds. In BI and CPM, even small delays can lead to significant drops in adoption – especially among non-technical or occasional users.

This effect is magnified by generational change: younger users expect instant results. What was tolerable 10 years ago in classic reporting now leads to rejection. The 2-second rules become a 0-second rule.

Performance also drives productivity – how many actions users can complete – and scalability – how systems handle increasing load. With AI and agent-based systems on the rise, load is no longer driven by humans alone. A single agent can generate hundreds of queries, pushing architectures to their limits.

Performance, productivity, and scalability are no longer optional – they've **become essential**. Especially in real-time business and AI-driven decision-making, they are key to operational success.

The BARC Benchmark Scores

BARC has supported thousands of enterprises in navigating their BI and CPM journeys. While we have a deep understanding of the BI & CPM market, the vendors and how their software compares and performs in practice, answering those critical scalability and productivity questions for a specific enterprise has – until now – largely remained a matter of informed estimation based on what we hear from the field. Only actual usage with real-world data and user volumes will reveal the full picture.

We want to fix that. The BARC Benchmark (BB) and the BARC Benchmark Score (BBS) are our attempt to provide a clear and structured way to objectively assess BI & CPM software – based on realistic enterprise scenarios, with real-world data volumes, modeling complexity and – most importantly – concurrent user loads.

It's not about declaring a winner or replacing traditional selection processes – but about providing a clearer **understanding of how different software performs under real-world conditions**, so enterprises can decide faster, better and with greater confidence on the productivity and scalability a specific software solution will offer.

In this document, we describe two of the world's most popular BI & analytics solutions as a starting point for our new benchmarking approach: **Microsoft Power BI** and **Qlik**.

What the BARC Benchmark Score tells you

The BARC Benchmark Score is a single, comparative performance indicator grounded in real-world software use. It is based on objective measurement results collected during standardized and automated benchmark executions that simulate typical front-end user interactions under realistic conditions. The best result in the two sub-score categories of the very first benchmarks run defines **the initial reference value of 100**. All other subsequent scores are calculated relative to that baseline – which therefore serves as a starting point.

As technologies evolve and vendors introduce innovations, future benchmark results may – and are expected to – exceed 100, reflecting measurable progress driven by real-world improvements in technology, architecture and design.

The BARC Benchmark Score is composed of two equally weighted sub-scores – Productivity and Scalability – each representing measurable aspects of software performance or behavior under real-world conditions and as experienced by real users of the respective software. Here's a sample:

	Weighting	Sample Score
Vendor XYZ – BARC Benchmark Total Score* A numeric score that indicates how well the BI and CPM software performs overall under real-world conditions. The higher the better.	100%	86
Vendor XYZ – BARC Productivity Score Measures how quickly and efficiently users can complete real-world BI or CPM tasks.	50%	92
Vendor XYZ – BARC Scalability Score Evaluates system performance under increasing data volumes, user load and solution complexity – including the ability to handle advanced business logic and demanding calculations beyond simple aggregations.	50%	80

** The BARC Score is normalized to 100 based on the best total result in the current benchmark round. Future results may – and likely will – exceed this value.*

Microsoft Power BI vs. Qlik

Benchmark setup and execution

For both Microsoft and Qlik, the standard entry-level cloud offerings of Power BI and Qlik Cloud were used. Power BI Premium or Qlik Premium (or Enterprise) – which both run on dedicated hardware – was therefore not considered in this first run. Neither Microsoft nor Qlik were aware that BARC was using/testing their software – to replicate the conditions of every company that uses the standard offering of these vendors.

As outlined later, in chapter BARC's Benchmark Approach, the respective Power BI and Qlik implementations were carried out by certified partners of the respective vendors, strictly adhering to the BARC Reference Data Model and implementation guidelines.

All benchmark runs were executed in parallel on an auto-scaling infrastructure set up on Microsoft Azure. The simulated clients were configured to reflect "average only" end-user hardware, featuring 8 GB RAM and just 2 to 4 virtual CPU cores (running on Intel Xeon Platinum-based CPUs).

Google Chrome was used as the browser. No other software or background processes were running on the benchmark clients during execution – apart from the user script runner itself – to ensure clean and undisturbed measurements. Since many modern browsers are based on Google technology, this represents a large proportion of the installed base of browsers.

BARC Benchmark Productivity Scores

The BARC Benchmark Productivity Score measures **how efficiently and productively** a user can complete real-world BI or CPM tasks. It reflects both the system's response times and the number of user interactions required. **In short:** the fewer clicks and the faster the responses, the higher the productivity.

The BARC Productivity Score is based on the **net execution time** over all system-side operations **for a single user session** – from login to final logout – excluding all human induced delays such as mouse movement, typing or idle time.

This net time reflects the pure system performance and ensures fair comparability between platforms – because otherwise, faster systems would be disproportionately penalized by slower human latencies. The final metric expresses how many such full user sessions can be completed per hour, serving as a practical indicator of user efficiency and system throughput.

Gross session durations – including human latency – are also reported to reflect typical real-world usage. As automation and bot-driven usage increases, net-based metrics will become even more relevant.

Productivity Score results

Vendor / Software	Metrics	Score
Microsoft Power BI	4,049 user sessions / hour (net) X / 13,058 * 100 = *≈ 2,372 human user sessions / hour (incl. human wait + latency)*	**31**
Qlik	13,058 user sessions / hour (net) X / 13,058 * 100 = *≈ 4,942 human user sessions / hour (incl. human wait + latency)*	**100***

** The best value in this very first benchmark round defines the baseline score of 100. All subsequent results will be relative to this 100.*

Productivity Score comments and observations

As reasons why Power BI achieved a lower Productivity Score than Qlik – or vice versa – the following measurable or expected factors were identified.

Expected factors are marked as ***BARC analysis*** and reflect BARC's current interpretation of the observed behavior. These may be subject to correction or clarification by the respective vendors – or revised by BARC considering new insights or better knowledge:

- **Response times** – defined as the delay between a user action and the corresponding screen update – **varied more strongly in Power BI**, ranging from 0.49 to 302.33 seconds (avg. 3.84 seconds), including one session dropout (no response). **Qlik showed more consistent response behavior**, with response times between 0.45 and 4.09 seconds (avg. 1.37 seconds), and no session dropouts.
- On average, over all actions with a user session, **Qlik delivered approximately 3 times faster response times** than Power BI – with a mean of 1.37 seconds for Qlik compared to 3.84 seconds for Power BI.
- ***BARC analysis:*** *The standard tier of Power BI seems to run on shared infrastructure that struggles to process higher interaction loads consistently. This could explain the response variability and inconsitencies we observed during the benchmark and that some Power BI customers reported on. Details of the technical setup of Qlik's standard tier are not publicly available.*
- **Qlik employs a real-time communication protocol for data transfer** (WebSockets) allowing continuous bidirectional server-client communication. **Power BI uses classic stateless unidirectional HTTP**, requiring a new request for each server response.
- ***BARC analysis:*** *The use of the WebSocket protocol – as commonly used in chat applications and online games – likely plays a key role in the snappy response times delivered by Qlik. This indicates a totally different and more modern client/server communication approach compared to Power BI.*
- ***BARC analysis:*** *Qlik also appears to cache and prefetch data more aggressively, so that information is often already available when the user requests it. In manual testing – not part of the benchmark, therefore not rated – this sometimes resulted in almost instant or none human-perceivable latency for full page refreshes. This caching behavior may be configurable at design time. In contrast, while Power BI generally responded quickly, full refreshes always showed a perceptible delay.*
- **Power BI required approximately 1.3 times (30%) more user interactions** – such as clicks and mouse movements – to complete the same task as in Qlik, particularly when navigating and adjusting hierarchical report filters. This factor was manually assessed. In future benchmark rounds, it is planned to capture and quantify this metric in a standardized, automated manner.

Evaluation of Microsoft Power BI productivity

Viewed in isolation, Microsoft Power BI delivers generally adequate and up-to-date performance under typical usage conditions. Most business users are unlikely to encounter critical response time limitations in everyday use. However, some users may experience inconsistent response times, and in rare cases, certain requests may not return results within an acceptable timeframe. Overall, the measured interaction patterns and response times indicate potential for optimization, particularly under increasing and sustained load.

Evaluation of Qlik productivity

Qlik exhibited very strong performance characteristics throughout the benchmark runs. The system delivered fast, stable response times and a smooth interaction flow, resulting in high session efficiency. These outcomes point to a technically mature, robust and modern implementation, particularly in terms of front-end responsiveness and user interaction handling and ergonomics.

Microsoft Power BI and Qlik productivity in comparison

Both systems successfully completed all benchmark scenarios. However, measurable differences in interaction design, communication protocols and system responsiveness resulted in clear variations in average performance, productivity and user experience.

Power BI required more user interactions to accomplish the same tasks as Qlik and showed higher response time variability. This indicates potential for further optimization, particularly with regard to business-user ergonomics and efficiency.

According to the gross productivity scores (48 for Power BI, 100 for Qlik), users working with Qlik achieved up to twice the throughput in comparable scenarios. It is expected that Premium tiers of both Microsoft Power BI and Qlik would deliver different, likely higher scores, but the relative differences observed in this benchmark are significant.

In summary, while Power BI offers adequate performance for typical usage, Qlik demonstrated consistently faster, more stable and ergonomically efficient execution, resulting in substantially higher end-user productivity in this benchmark round. Future benchmark iterations may provide deeper insights across broader configurations, vendor tiers and more demanding usage profiles.

BARC Benchmark Scalability Scores

The BARC Scalability Score is designed to assess how well BI and CPM systems perform under increasingly demanding workloads – evaluated across three distinct dimensions:

1. **Number of concurrent users** – Arguably the most critical factor for larger enterprises or high-traffic scenarios, such as field-based or customer-facing applications.
2. **Data volume** – Especially relevant for domains beyond finance, such as sales, supply chain, e-commerce or IoT, where data volumes have grown exponentially in recent years. Hundreds of millions – or even billions – of records are now a reality in many industries, including mid-sized businesses.
3. **Business and data model complexity** – While simple aggregations on large datasets are handled well by most platforms today, complex calculations – especially analytical, on-the-fly logic – remain a significant challenge. To account for this, the BARC reference data model includes a dynamic currency conversion scenario that cannot be easily bypassed through pre-calculated data.

In contrast to the BARC Productivity Score, the BARC Scalability Score is based on **gross execution time** – that is, the total duration of a complete user session from login to logout, including all human-induced delays such as mouse movement, typing and idle time. This approach ensures that the system is evaluated under **conditions that closely resemble real-world usage**, where human behavior adds variability and stress to system performance. Only this way can true end-to-end scalability – as experienced by actual users – be assessed.

The final metric is straightforward: it counts how many **complete user sessions a system can execute within one hour for a given level of concurrent users**.

In future benchmark rounds, we plan to additionally determine the **maximum number of concurrent users a system can handle** before performance degrades noticeably. This threshold will be defined as the point where the average user session duration doubles compared to the baseline measured at 10 concurrent users – indicating a significantly slower experience for end users under load.

Benchmark scope and limitations

Due to time constraints and a limited number of available licenses for both Power BI and Qlik, we were only able to test up to 50 concurrent users and a reasonably sized data model containing 10 million records. This equals a typical sales data use case of a midsized company or a departmental use case from a larger enterprise.

While this setup has already provided valuable insights, for future benchmark rounds, we intend and plan to scale up to 500 or more concurrent users (if supported/requested by vendors) and incorporate data models with 10 million (XL), 100 million (2XL) and even 1 billion records (3XL).

The latter would be particularly relevant for assessing the scalability of data warehouse and lakehouse back ends, as well as analytical query engines under real-world, high-volume conditions.

Scalability Score results

Vendor / Software	Metrics		Score
Microsoft Power BI	2,372 user sessions / hour (gross) incl. human wait-times and latency	X / 4,942 * 100 =	**48**
Qlik	4,942 user sessions / hour (gross) incl. human wait-times and latency	X / 4,942 * 100 =	**100***

** The best value in this very first benchmark round defines the baseline score of 100. All subsequent results will be relative to this 100.*

Scalability Score comments and observations

As reasons for the given results – for **50 concurrent users** over **1 hour straight** accessing the BARC reference **data model in size XL** (10 million records) – the following measurable or expected factors were identified.

Expected factors are marked as ***BARC analysis*** and reflect BARC's current interpretation of the observed behavior. These may be subject to correction or clarification by the respective vendors – or revised by BARC considering new insights or better knowledge:

- **Power BI was able to complete 2,372 full user sessions** within one hour, whereas **Qlik completed 4,942 user sessions** in the same period – corresponding to a throughput advantage of approximately 2× for Qlik in this test setup.
- As the user load was ramped from 1 to 50 concurrent sessions, neither platform exhibited material performance degradation. Qlik's median response times remained consistent across all tested load levels. Power BI did not slow down systematically; instead, its responsiveness became more variable, with occasional long-tail outliers. Within this range, both standard cloud tiers scale well, with Qlik demonstrating consistent behavior under increasing load. See also charts on page 14.

- Interestingly, Power BI exhibited **significant variability in average session durations** (up to a 4× difference) as concurrency increased, while Qlik maintained highly consistent session times, interestingly, even slightly improving under higher load.
- ***BARC analysis:*** *The consistent and slightly improving performance observed with Qlik was interpreted by our external software engineering experts as an indication of dynamic resource allocation within the back end, allowing the platform to scale effectively with user load. This was seen as a hallmark of a modern, adaptive software architecture.*
- ***BARC/expert analysis:*** *In contrast, the fluctuating performance pattern observed with Power BI raised questions among our experts. While no hard limitations were identified, the inconsistent behavior was interpreted either as a side effect of shared infrastructure, suboptimal workload orchestration or potentially even throttling policies. These interpretations remain speculative and would require confirmation from Microsoft.*
- Finally, it must be noted that **Qlik's advantage in core performance** and user productivity contributes directly to its higher scalability score. Fewer required interactions per task naturally lead to shorter session durations – independent of raw server-side execution performance.

Evaluation of Microsoft Power BI scalability

Viewed in isolation, Microsoft Power BI handled the gradually increasing concurrency level up to 50 users and a mid-sized dataset without critical performance degradation. Throughput remained within an acceptable range, but session durations fluctuated noticeably, and at 50 concurrent sessions a few extreme long-running outliers occurred. This indicates sensitivities in load handling. The assumed use of shared resources in the standard service plan could be a limiting factor.

Evaluation of Qlik scalability

Qlik delivered consistently fast and predictable performance across all benchmarked concurrency levels. Even under increased load, session durations remained very stable – in some cases even slightly improving – which indicates effective internal scaling / load balancing mechanisms.

In this benchmark, Qlik was not brought to any performance limit, and further scalability potential remains to be explored in test scenarios with more users and data.

Microsoft Power BI and Qlik scalability in comparison

Both platforms successfully completed the scalability benchmark without critical degradation as concurrency was ramped from 1 to 50 users on a mid-sized dataset (standard cloud tiers).

Clear differences in behavior under load were observable. Power BI delivered acceptable throughput but showed increasing variability in session durations; at 50 concurrent sessions a few extreme long-running outliers occurred, indicating sensitivities in load handling (potentially related to shared resources in the standard service plan). Qlik's median response times remained consistent across all tested load levels, and comparable long-tail outliers were not observed.

The tested setup did not exhaust either system's capacity, but it highlighted differences in predictability under load. Further rounds with higher concurrency, larger datasets and premium/dedicated tiers are required to assess true upper bounds. See also the charts on page 14 for a visualization of these results.

Viewed in isolation, both solutions handled the tested concurrency levels well. In direct comparison, however, clear differences emerged: Power BI showed greater variability in session durations and, at 50 concurrent users, a few extreme long-running outliers – indicating sensitivities in load handling, likely related to shared resources in the standard service tier. Qlik, by contrast, kept median response times very consistent across all tested load levels and exhibited no comparable outliers. Overall, both technologies can be considered scalable within the tested range up to 50 users. See also the charts on page 14.

Total BARC Benchmark Scores

Total Score results

Vendor / Software	Metrics	Total Score
Microsoft Power BI	50% Productivity Score 31 + 50% Scalability Score 48 = 39.5 -> 40	**40**
Qlik	50% Productivity Score 100 + 50% Scalability Score 100 = 100	**100***

** The best value in this very first benchmark round defines the baseline score of 100. All subsequent results will be relative to this 100.*

Microsoft Power BI and Qlik overall comparison

All relevant aspects and findings contributing to the overall BARC Benchmark Scores for Power BI and Qlik have been covered in the preceding evaluations of productivity and scalability.

The clarity and magnitude of the differences were unexpected even for the BARC team. Yet the outcome underlines a fundamental insight: when software is evaluated under real-world enterprise conditions, distinct vendor priorities become visible – and measurable differences in productivity, responsiveness and scalability emerge.

Despite Power BI's tremendous market popularity, driven by its broad feature set, vibrant community and integration with Microsoft 365 and Teams, the benchmark results show clear limitations in consistency and efficiency when compared directly to Qlik under identical conditions. Qlik, in contrast, demonstrated very consistent performance and interaction efficiency, leading to higher measured end-user productivity and scalability that proved more predictable under increasing load.

For decision-makers, the key takeaway is to look beyond features and marketing materials and place stronger emphasis on core qualities such as usability, productivity, efficiency and scalability – attributes with direct impact on business value and total cost of ownership.

Selected visual results

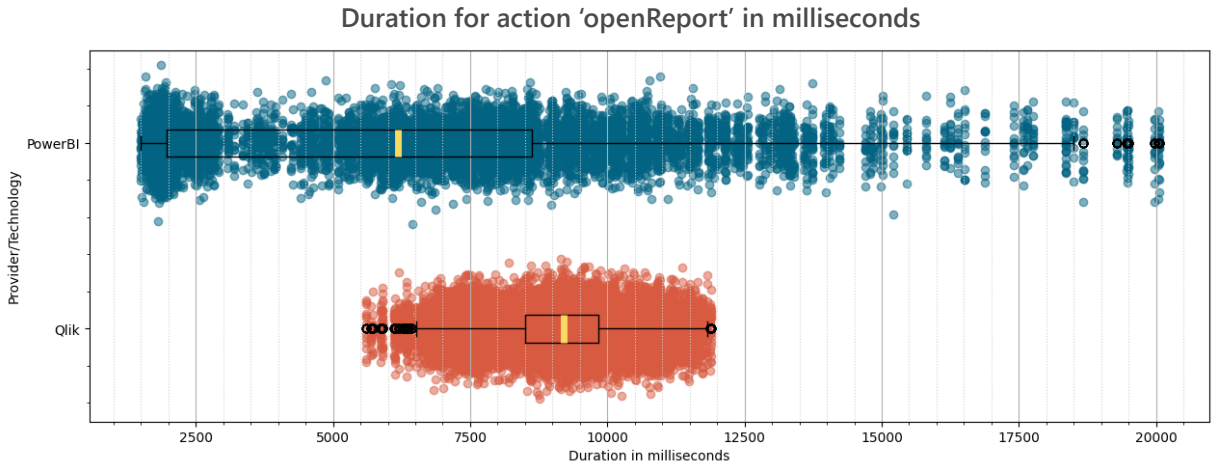
As charts often tell a story much better than words or numbers, the following visualizations highlight key findings from the benchmark study. Each chart has been selected to illustrate specific scoring results. To ensure maximum transparency, we present full datasets using box plots with all data points shown. The yellow bar indicates the median value – not the mean average – splitting the dataset into two equally sized halves.

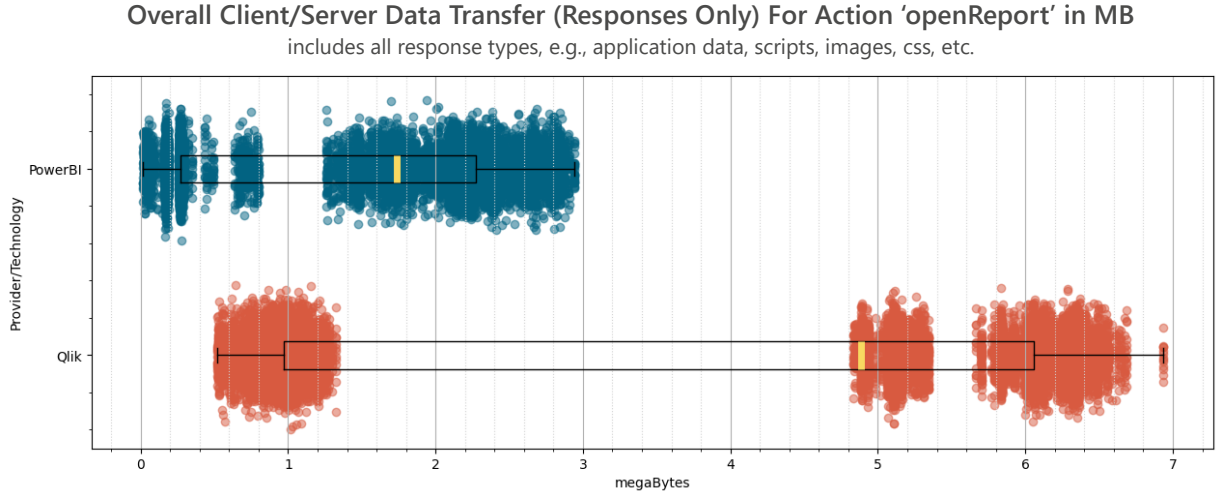
Productivity – user wait time to **open** a report or dashboard

In this test scenario, **Microsoft Power BI opens reports noticeably faster on average** (median: 6.2 seconds) compared to Qlik (median: 8.7 seconds). However, **Qlik loads significantly more data** (median: 4.9 MB) per report open than Power BI (median: 1.9 MB). Likely for a reason.

BARC analysis: *Qlik appears to invest more time during initial report loading in order to prefetch or cache relevant metadata (e.g., filter master data) – likely to enable smoother and faster navigation within the report afterward.*

How to read the charts: Each dot represents a single measurement. The yellow bars indicate the median, splitting all observations into two equally sized halves.





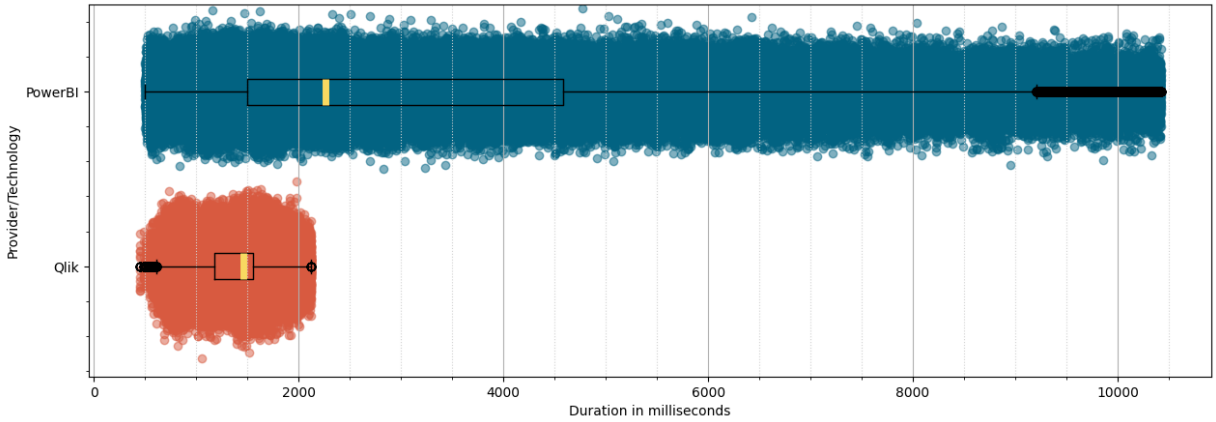
Interesting: The three distinct red clusters in the Qlik data transfer chart (in MB) closely correspond to the three benchmarked report types. The first cluster belongs to a P&L report with little filter metadata, while the second and third clusters represent sales reports with larger product dimensions and more data.

It is obvious that Qlik's front-end strategy accepts longer initial loading times in favor of faster navigation within reports later on.

Productivity – user wait time to **update** a report or dashboard

When navigating within a report – for example, by changing filters – **50% of all Microsoft Power BI responses return in under 2.1 seconds**. That is a solid result and close to the performance of **Qlik**, **where 50% of responses are faster than 1.5 seconds**. However, Power BI's slower half tells a different story: response times stretch up to (and beyond) 10 seconds. In contrast, Qlik consistently returns results in under 2.2 seconds.

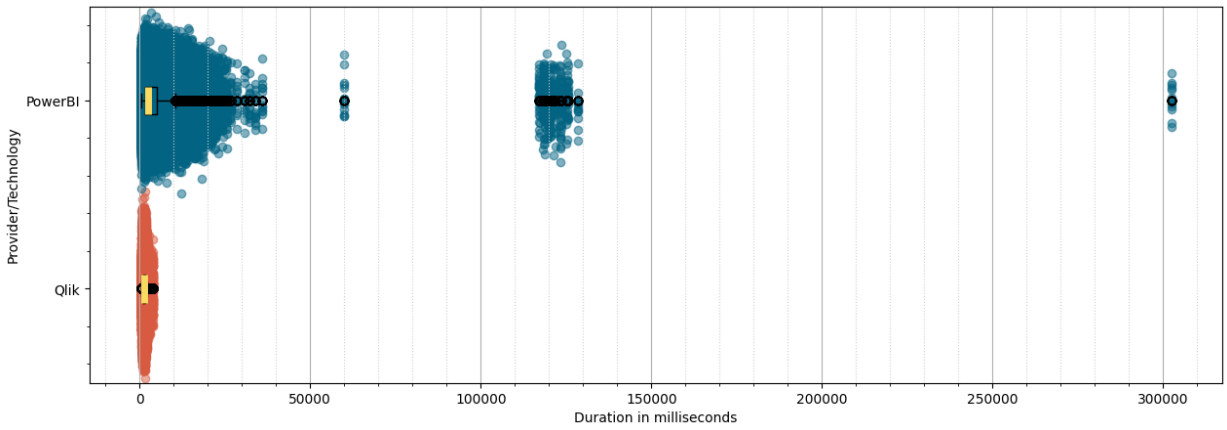
Duration for action 'changeReportFilter' in milliseconds (outliers removed)



To improve the readability of the chart and allow fair comparison, all outlier responses exceeding ±11 seconds were excluded from the first chart.

However, when these outliers are included (see next chart), a striking pattern emerges: some requests take almost exactly 60, 120 or 300 seconds – strongly aligned to minute boundaries. These long delays appear sporadically across the entire test duration, across all user sessions at low and high concurrent user loads.

Duration for action 'changeReportFilter' in milliseconds (including outliers)



BARC analysis: *This pattern may point to architectural limitations in Microsoft's back end or even deliberate throttling policies to balance the overall workload across all Power BI users of all customers on a shared Power BI instance or data center.*

While benchmarking bots will patiently wait several minutes, human users likely won't. For them, response times above 10 or 30 seconds are often disruptive and unacceptable in daily business use. Microsoft has an opportunity to improve end-user satisfaction, even for its base versions in this area.

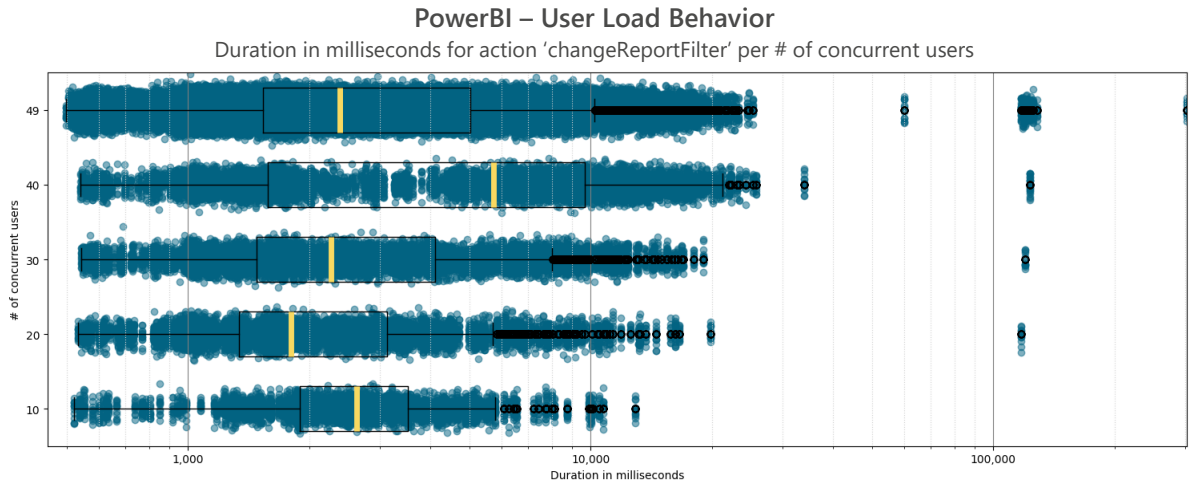
Scalability – response times under increasing user loads

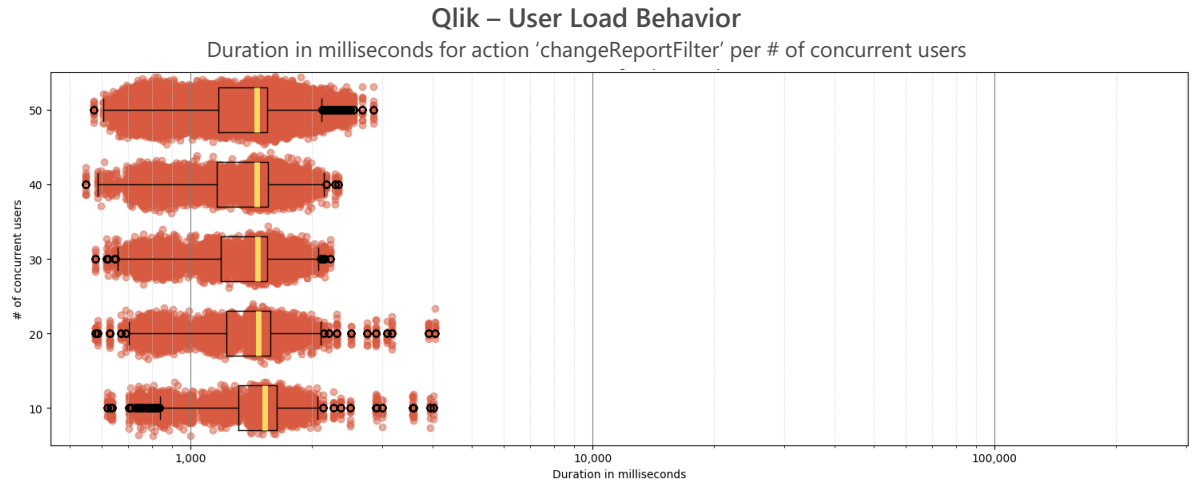
As previously stated, **our benchmark did not manage to push either Power BI or Qlik to their technical limits** – the infrastructure and data volumes were simply too small. Still, the charts below offer valuable insight into how both platforms behave under gradually increasing user loads.

The BoxPlots visualize the **response times for a common user action** – updating a report by changing a filter (e.g., year, month or product) – over increasing concurrent user numbers. Data, logic, filter hierarchy and call sequence were kept 100% identical between the two platforms to ensure a fair comparison. The yellow line indicates the median, not the mean average, which better reflects typical user experience in skewed distributions.

Power BI shows noticeable variance under higher loads, with a few isolated cases of extreme delays extending into minutes. **Qlik, in contrast, maintained very consistent response times across all tested levels of concurrency**.

Please **note the logarithmic X-axis**, which was chosen to accommodate these long-tail outliers in Power BI. Without the logarithmic scale, the chart would be visually dominated by these rare but extreme delays, compressing the typical response range and making Power BI appear even more erratic.





BARC observation: *The responsiveness pattern suggests that Qlik benefits from more efficient resource orchestration or a more elastic back end. Power BI appears less predictable under load, which could negatively affect user acceptance in medium- to large-scale or business-critical scenarios.*

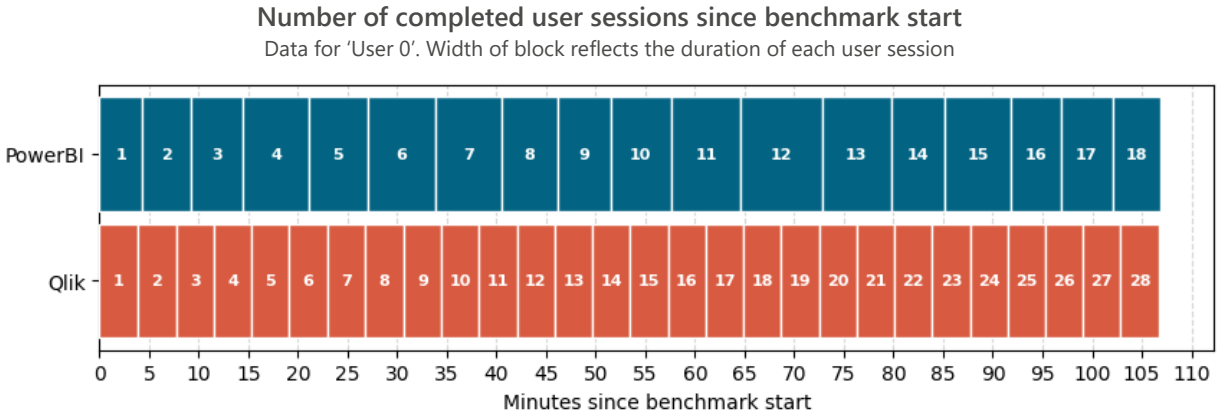
Productivity – session throughput = end-user productivity

If none of the previous charts told a clear story, this one certainly does.

Each block in the chart below represents a user session by the same automated user performing the same sequence of actions: logging in, opening the same reports, applying the exact same report filters and completing the same tasks – under identical conditions for Power BI and Qlik.

In the first ±107 minutes of the benchmark, 18 sessions were completed with Microsoft Power BI and 28 sessions with Qlik. This means the same user can **accomplish the same work in 64% of the time with Qlik compared to Power BI**.

Why is this ±50% higher productivity **not fully aligned with the Scalability Score** gap of 48 vs. 100, which implies a 2.1× or 108% advantage for Qlik? Simply because the user session shown below happened to be one of the Power BI user sessions with relatively stable performance during the test – although not entirely smooth, as the varying width of the bars indicates. **Other Power BI users experienced longer and more inconsistent wait times**, leading to an average session time twice as long as in Qlik.



Your feedback matters

Now that the first BARC Benchmark results are on the table, we'd love to hear your thoughts.

Did you find the concept, methodology, and results valuable? Could they support your software evaluations and decision-making – or, if you're a vendor, help strengthen your product positioning and demonstrate competitiveness?

Your feedback will help us refine the benchmark, expand its scope, and establish it as a transparent, standardized model for evaluating performance, productivity, and scalability in BI & CPM.

Please take part in our 2-minute feedback survey: Link

BARC's Benchmark Approach

After presenting the benchmark results, it's time to look behind the scenes. This chapter explains how the BARC Benchmark Score is determined – from the test scenarios and data models to the automation framework, measurement methodology and scoring logic. Our goal: transparency, reproducibility, and fairness across all tested platforms.

What is measured?

No glitter, no gimmicks. The BARC Benchmark Score focuses on essential, real-world scenarios that virtually every BI and CPM solution should be able to handle. We deliberately avoid edge cases, vendor-specific features and unnecessary formatting – and instead measure how reliably and efficiently core tasks are executed under realistic conditions, using the exact same analytical task on the exact same data executed across different software platforms.

Examples include:

- How long does it take to open or refresh a report or dashboard?
- How many interactions are needed to change report filters – and how long does it take?
- How many interactions are required to enter a forecast or trigger a planning step?
- Do response times remain stable as user load and data volume increase?

Accuracy is equally important. If a software returns incorrect values – for any reason – penalty points apply:

- 50% deduction of score for deviations greater than ±0.1% (1/1,000)
- 10% deduction of score for deviations greater than ±0.0001% (1/1,000,000)

Because let's face it: performance and productivity mean nothing if the numbers are wrong.

What data / data model is used?

To ensure comparability, BARC has developed the **BARC BI Reference Data Model** and the **BARC Data Model Generator** to simulate realistic enterprise structures and data volumes. The reference model is based on common business scenarios and will be made publicly available to promote transparency and reproducibility. All benchmarked software systems use the same underlying data structure and data volume – tailored to their respective modeling paradigms, but semantically and numerically identical.

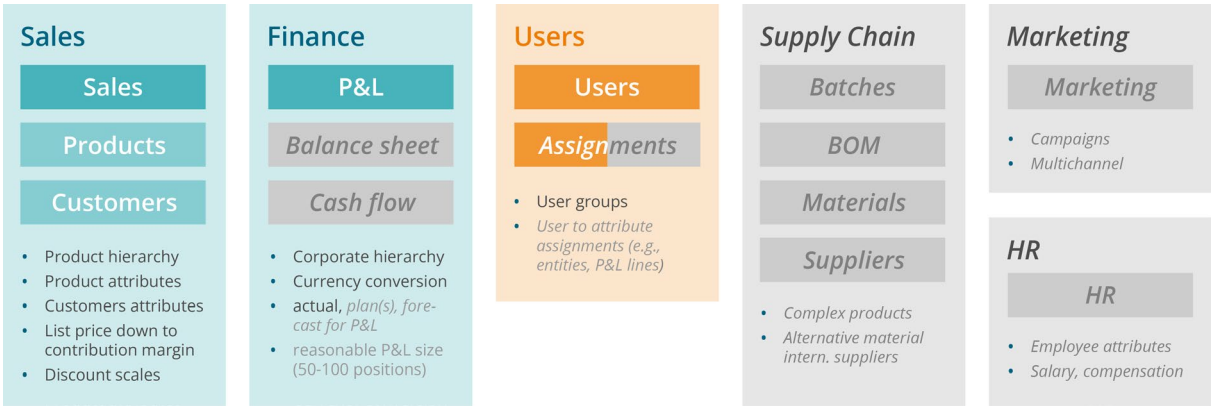


Figure 1: Actual scope and planned scope (gray) of the BARC Data Model and Data Model Generator

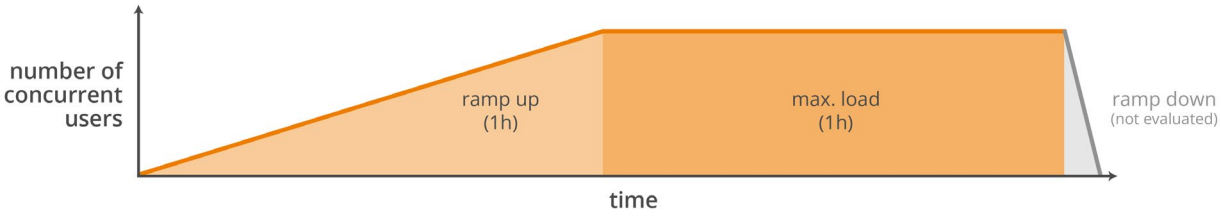
If the BARC Benchmark resonates, we also plan to further extend the model to support additional use cases such as planning, forecasting, analytics, AI and GenAI utilization. More on this can be found in the next chapter: BARC Benchmark – Future Directions.

How the BARC Benchmark is executed

All benchmark executions are fully automated, using **BARC's Benchmark Execution Framework** – a self-scaling cloud-native framework designed to simulate real-world user workloads at any scale.

For those interested in the technical details: A look under the hood

The Benchmark Execution Engine is cloud-native and built on industry standard tooling like Pulumi, Docker, Kubernetes (AKS) and Playwright – currently deployed on Azure Germany West Central. A central orchestrator manages the lifecycle of user sessions: gradually ramping up the number of concurrent users over the course of one hour, then running a full load of users for a sustained additional hour, as shown in the graph.



Benchmark bots simulate real user behavior on the tested web-based software – from login and navigation to filter changes, report refreshes and planning form entries. Like real users, they pause between steps and follow realistic cursor movements. The system scales from 1 to 10,000 concurrent users, supporting everything from small teams to global enterprise deployments.

Note: In this round, tests were limited to **50 concurrent users** due to licensing constraints. We hope that vendors will support future benchmark rounds by providing broader license access for more extensive testing with, for example, 100 or 1,000 concurrent users. Nevertheless, we know from practical experience that 50 "real" simultaneous users in real-life usage scenarios already correspond to many hundreds of end users – meaning that the benchmark already covers larger application scenarios that are common in medium-sized companies and many departmental use cases in larger companies.

While the technical implementation is platform-specific, the sequence of interactions and data requests is fully deterministic and identical across all benchmarked systems – ensuring fairness and reproducibility.

The approach follows the principles of black box testing and focuses exclusively on observable behavior and results, without checking internal structures, backend configurations or code.

Who prepared the software for benchmarking?

To ensure a realistic and vendor-neutral setup, all configurations were implemented by certified service partners with long-standing expertise in the respective technology. Neither BARC nor the vendors were involved at any stage. Both partners operated under strict confidentiality agreements; their identities remain undisclosed.

We used the standard tier cloud offerings of Power BI and Qlik and relied solely on input of the implementation partners and publicly available information – just like any enterprise would when working with experienced consultants.

Looking ahead: In future rounds, vendors will likely configure their own software – and that is perfectly fine. But it also means that tuning, tweaking and performance optimization will become part of the game.

As long as all benchmark rules are respected, any optimization is allowed – because in real-world projects, customers would benefit from the same expertise, fine-tuning and attention to detail.

Official BARC-validated benchmark results

To ensure reliability and cross-vendor comparability, only results validated by BARC are eligible for official scoring – providing confidence in every benchmark outcome.

Who executed the benchmark runs?

The benchmark tests were executed by a specialized and BARC trusted implementation partner, experts in custom and elastically scaling data & AI solutions, under direct BARC supervision.

The partner's engineers also designed and developed the BARC Benchmark Execution Framework based on BARC's technical and methodological specifications – and will continue to implement and execute future benchmark rounds and runs. Our efficient setup makes it easy and affordable for vendors, enterprises and BARC alike to prepare and execute future benchmark runs and rounds.

What data was collected and how was it evaluated?

During each benchmark run, the system captured detailed telemetry on every user interaction – including execution times, click paths, result accuracy, responses and network traffic. Data was logged per user session. To ensure comparability and statistical validity, each test scenario was repeated multiple times during the 2-hour benchmark run. Finally, all generated telemetry files were combined, validated, and any deviations were analyzed and verified before scoring.

All data analysis, evaluation and scoring logic – including the corresponding Python scripts – were developed by BARC and are executed exclusively by BARC. Only results from controlled, fully reproducible runs are considered for the official BARC Benchmark Score.

Benchmark challenges – or why bots suffer too

Before we jump into the results – running automated benchmarks on end-user web-based BI tools may sound easy, until two-factor authentication, welcome screen, random pop-ups, cookie banners, release update prompts or unexpected long wait times crash the party. Bots have no sense for surprise modals. So, our bots had to become smarter than we ever expected – almost human.

In addition, we are not exempt from physics: Every measurement slightly changes what is being measured. It's called the observer effect – and yes, the BARC benchmarks also suffer from it. To be clear:

- Benchmark results are always slightly slower than stopwatch-style manual observation.
- Hence performance differences are more pronounced on short-duration tasks.
- The exact "measurement impact" cannot be precisely quantified – but applies equally to all.

Bottom line: Benchmarking is not about absolute performance values – it's about fair, reproducible comparison under controlled, identical conditions.

BARC Benchmark – Future Directions

Toward an open evaluation standard for BI & CPM software

First, we need your feedback!

The BARC Benchmark Score is still in its early stages – and its future depends on whether enterprises and vendors find it useful. If there's enough interest from enterprise and vendors alike, we plan to publish more benchmark results and provide free access to the tooling – such as the data model generator – (via GitHub) on a dedicated public website.

Please let us know what you think about the benchmark initiative – and whether the tooling would be valuable for your use cases:

Share your feedback here: Link

These initial benchmark results only lay the groundwork. What comes next depends on the community. Enterprises, vendors and consulting partners are invited to run their own benchmarks, explore specific scenarios and help shape the future of this initiative – using and evolving BARC's public data model, concepts and tooling, and execution engine to become an industry standard.

Your feedback is essential – first, to test and kick off the next phase of the BARC Benchmark, and later, to continuously improve the BARC Benchmark Score as a shared, open platform for transparent, evidence-based BI, CPM, data, analytics and AI software evaluation.

Benefits for enterprises: Faster, safer software decisions

With the BARC benchmarking concepts and tools we aim to provide, enterprises will be able to simulate real-world use cases – including their own data structures and volumes – using standardized models and repeatable benchmark runs. This opens the door to more transparent, scenario-based software evaluations for enterprises:

How does this software perform under my conditions – at my scale?

Potential use cases include:

- Proof-of-concept evaluations under realistic workloads
- Comparative assessments based on consistent criteria
- Performance audits of existing BI and CPM platforms

Benefits for vendors: Compete, improve, demonstrate

Vendors can leverage the open BARC Data Model and Benchmark Execution Engine to:

- Benchmark new product features and versions under consistent conditions
- Fine-tune performance and scalability under mild to heavy load
- Demonstrate competitive strengths in sales and marketing scenarios

Re-runs after feature updates are supported – and even encouraged. The tooling can also be used for internal purposes such as optimization, issue tracking, training or partner enablement.

Most importantly, it helps vendors build trust with prospects and customers – through results that are independently validated and testified by BARC, and that clearly differentiate their products in a competitive market.

What's next – planned enhancements

Looking ahead, BARC plans to further expand the benchmark scope and relevance by:

- Including data and more advanced use cases such as integrated planning, predictive analytics, AI/ML and generative AI.
- Extending the benchmark to additional domains such as HR, supply chain and marketing.
- Enabling industry-specific scoring profiles tailored to sectors such as financial services, manufacturing, pharma and public sector.

Providing customizable data generation to reflect enterprise-specific structures – such as product hierarchies, legal entities or regional breakdowns – and to match data volumes and numbers of users to the specific scale of individual organizations precisely. Preview of usage:

```
pip install barcdatagen

from barcdatagen import *
model = SalesDataModel(size="XL", customers="my_customers.json")
generate(model, ParquetWriter())
```

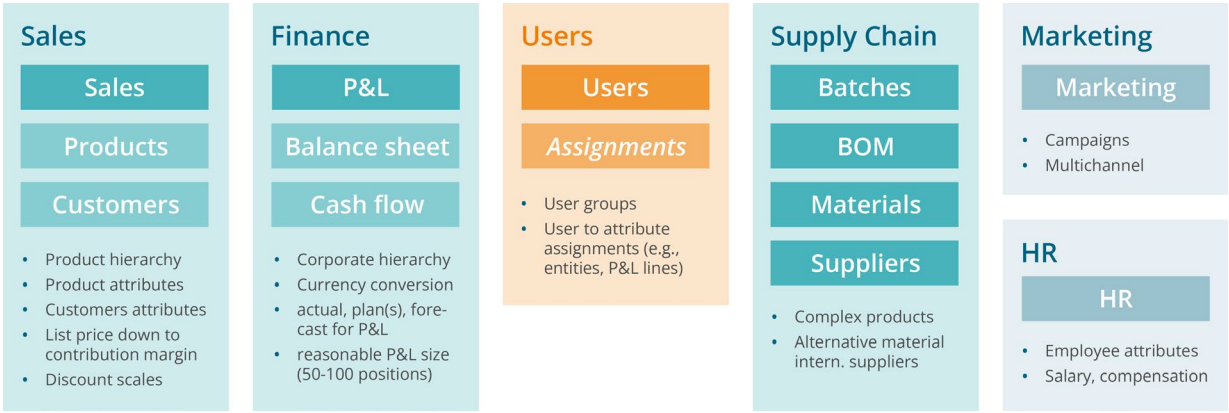


Figure 2: Full planned scope of the BARC Data Model and Data Model Generator

These enhancements aim to broaden applicability while maintaining comparability – ensuring that benchmarks remain relevant, realistic and valuable across diverse business contexts.

Last Words

The BARC Benchmark Score represents a major investment – not only in time, technology and infrastructure, but also in our core mission: to provide high-quality, neutral and actionable insights for decision-makers in the data-driven business of tomorrow.

This effort was driven by a simple goal: to enable enterprises and vendors to make smarter, more confident decisions in terms of productivity and scalability – based on measurable reality, not assumptions.

Call for action – Your feedback matters most

Please let us know whether you found this research and the benchmark approach valuable – and whether the concept and tools we propose would support your decisions, software evaluations or, as a vendor, your product strategy. Your feedback will help us to further develop the idea, concept and tooling for a standardized BI & CPM benchmark on performance, productivity and scalability.

Please take part in our 2-minute feedback survey: Link

About BARC

BARC is the leading analyst firm for data & analytics, AI, corporate performance management (CPM) and ESG with a reputation for unbiased and trusted advice. Our expert analysts deliver a wide range of research, events and consulting services for the data & analytics community. Our innovative research evaluates software, vendors and service providers rigorously and highlights market trends, delivering insights that enable our customers to innovate with data, analytics and AI. BARC's 25 years of experience with data strategy & culture, data architecture, organization and software selection helps clients transform into truly data-driven organizations.

Research

BARC user surveys, software evaluations and analyst advisory services along with expert driven content such as research notes, trend analysis and blogs give organizations the confidence to make the right decisions. Our independent research gets to the heart of market developments, evaluates software, vendors and service providers thoroughly and gives valuable ideas on how to turn data, analytics and AI into added value and successfully transform businesses.

Consulting

The BARC consulting practice is entirely focused on translating companies' requirements into future-proof decisions. The holistic advice we provide helps companies successfully implement their data & analytics strategy and culture as well as their architecture and technology.

BARC's research and experience-founded expert input sets organizations on the road to the successful use of data & analytics, from strategy to optimized data-driven business processes.

Events

At BARC events, leading minds and industry experts come together to share insights and drive innovation. Our conferences, roundtables and online webinars attract over 10,000 participants annually, offering a unique blend of information, inspiration and interactivity. These events provide a platform to exchange ideas with peers, explore emerging trends and gain expert perspectives on market developments.

By engaging with thought leaders and industry practitioners, participants discover actionable strategies to enhance their business and stay ahead in the evolving world of data & analytics.

BARC

www.barc.com

United States
BARC US
13463 Falls Drive
Broomfield, CO 80020
USA

Germany
BARC GmbH
Berliner Platz 7
D-97080 Würzburg
+49 931 880651-0

Austria
BARC GmbH
Hirschstettner Straße 19 / I / IS314
A-1220 Wien
+43 660 6366870

Switzerland
BARC Switzerland GmbH
Buchhaldenstraße 7
CH-5442 Fislisbach
+41 76 340 35 16