

# Adapt or Be Outpaced: The Competitive Edge of Real-Time Data

Introduction: The Real-Time Imperative	3
Understanding the Time-Value of Data	5
Real-Time Data Across Industries	6
Enabling Technologies for Real-Time Data	7
Building a Real-Time Data Strategy	8
Enabling Data Immediacy	9
Apache Pinot	10
Measuring the ROI of Real-Time Data Initiatives	11
StarTree - Enabling Data Immediacy with Real-Time User-Facing Analytics	12
Conclusion	14
Next Steps	15

## INTRODUCTION

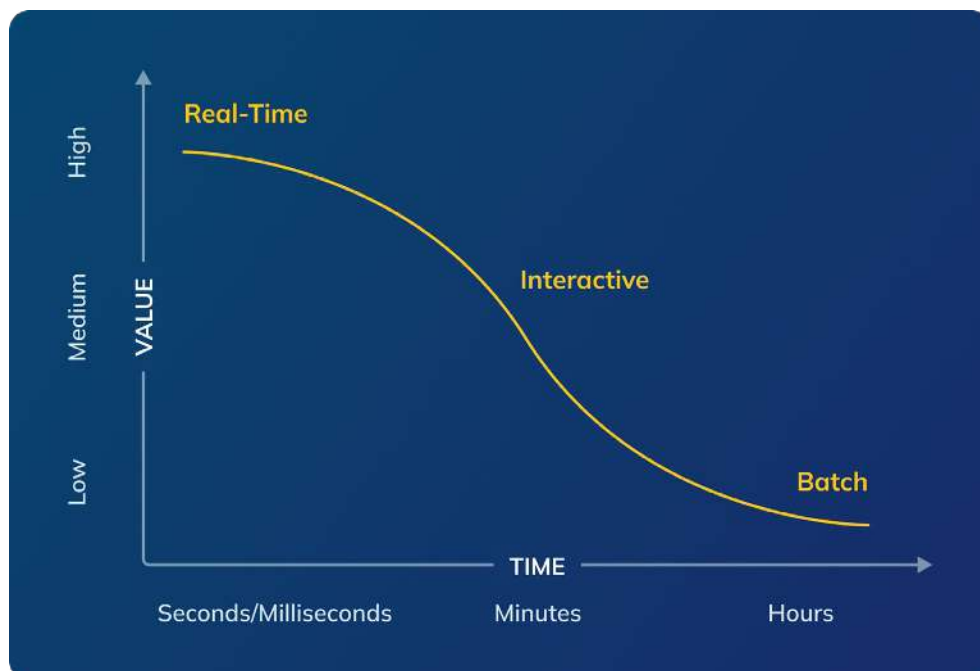
# The Real-Time Imperative

Imagine a world where businesses can predict customer needs before they arise, where supply chains adjust in real-time to prevent disruptions, and where fraud is stopped in its tracks before a transaction is complete. This isn't the stuff of science fiction – it's the reality of today's data-driven enterprises. Welcome to the age of real-time data, where the power of now is reshaping the business landscape at an unprecedented pace.

It is no surprise that over 120 zettabytes of data is created each year globally but here's the kicker – 73% of this data goes unused in analytics.<sup>1</sup> Why? Because by the time most businesses can analyze it, the moment to act has already passed.

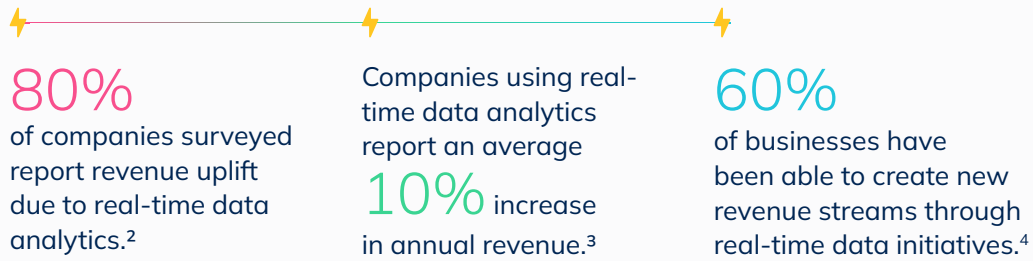
In this lightning-fast digital economy, the ability to make split-second decisions based on up-to-the-minute information isn't just an advantage – it's a necessity for survival. A mere 5-second delay in page load time can increase bounce rates by 20%, potentially costing e-commerce sites millions in lost sales. In the financial sector, high-frequency trading algorithms execute trades in microseconds, turning tiny time advantages into billions in profits.

But what exactly is real-time data? It's not just fast data – it's immediate, actionable information that reflects the current state of your business, customers, and market. It's the difference between knowing what happened yesterday and understanding what's happening right now.



<sup>1</sup> Forrester. (2020). Unveiling Data Challenges Afflicting Businesses Around The World

## The impact of real-time data is staggering:



Yet, despite these compelling statistics, only 33% of executives say their organizations are truly data-driven.<sup>5</sup> This gap represents both a challenge and an enormous opportunity.

As we stand on the brink of the fifth industrial revolution, where 5G networks, IoT, edge computing, and AI are converging to create an always-on, hyper-connected world, the question isn't whether your business needs real-time data capabilities – it's how quickly you can implement them to stay ahead of the curve.



<sup>2</sup> The Speed to Business Value report, KX and CEBR

<sup>3</sup> McKinsey & Company. (2021). The data-driven enterprise of 2025

<sup>4</sup> Harvard Business Review. (2020). Real-Time Analytics: The Key to Unlocking Customer Insights & Driving the Customer Experience

<sup>5</sup> IDC research sponsored by Tableau

# Understanding the Time-Value of Data

The concept of time-value is well established in finance, where the value of money changes over time. Similarly, in the realm of data, we must recognize that the value of information is not static—it decays over time. This phenomenon, known as data decay, underscores the critical importance of timely data processing and analysis.

**Data decay** occurs because the relevance and actionability of information diminish as time passes. For instance, a stock price from an hour ago is far less valuable for trading decisions than the current price. Similarly, a customer's current location is more useful for providing personalized offers than their location from yesterday. Fresh data often holds the most significant potential for impactful insights and actions.

The rate of data decay varies depending on the nature of the data and its application. Some types of data, like historical records, may decay slowly or retain their value over long periods. Others, particularly operational and transactional data, can lose their value rapidly. Understanding the decay rate of different data types in your business context is crucial for prioritizing real-time processing efforts.

Closely related to data decay is the concept of **data latency**—the delay between data generation and its availability for analysis. In fast-moving markets and operational environments, even seconds of delay can mean missed opportunities, increased risks, or diminished customer satisfaction. For example:

- ⚡ In algorithmic trading, microseconds of latency can result in millions of dollars of missed profits.
- ⚡ In e-commerce, delays in updating inventory information can lead to poor customer experiences and lost sales.
- ⚡ In manufacturing, late detection of equipment malfunctions can result in costly production downtime.

**Data Immediacy** is the antidote to data decay and latency. It refers to the ability to collect, process, and analyze data in real-time, enabling immediate action. Achieving data immediacy involves minimizing the time between an event occurring and the ability to respond to it effectively. This capability can provide a significant competitive advantage, allowing businesses to be proactive rather than reactive.

Implementing data immediacy requires a holistic approach, encompassing:

1. High-speed data ingestion mechanisms
2. Real-time processing and analytics capabilities
3. Low-latency data storage solutions
4. Rapid data distribution and visualization tools
5. Automated decision-making systems or streamlined human decision processes

By embracing the concept of data immediacy, organizations can maximize the value extracted from their data, leading to more agile operations, improved customer experiences, and ultimately, business supremacy.

# Real-Time Data Across Industries

The power of real-time data is transforming businesses across a wide spectrum of industries. Let's explore how various sectors are leveraging data immediacy to drive innovation, improve customer experiences, and optimize operations.

- In **retail and e-commerce**, real-time data is revolutionizing the shopping experience. Online retailers use real-time analytics to personalize product recommendations, dynamically adjust pricing, and manage inventory levels. Brick-and-mortar stores leverage real-time foot traffic data and sales information to optimize staffing and layout. The result? Enhanced customer satisfaction, increased sales, and improved operational efficiency.
- The **finance and banking** sector has long been at the forefront of real-time data adoption. High-frequency trading algorithms process market data in microseconds to make split-second investment decisions. Banks use real-time analytics for fraud detection, risk management, and personalized customer service. Real-time data enables these institutions to safeguard assets, comply with regulations, and provide cutting-edge services to their clients.
- **Healthcare and life sciences** are harnessing real-time data to save lives and improve patient outcomes. Hospitals use real-time patient monitoring systems to detect critical changes in patient conditions instantly. In research, real-time data analysis accelerates drug discovery and helps track the spread of diseases. The ability to process and analyze health data in real-time is paving the way for more personalized and effective treatments.
- In **manufacturing and supply chain management**, real-time data is driving the fourth industrial revolution. Smart factories use real-time sensor data to optimize production processes, predict equipment failures, and ensure quality control. Supply chain managers leverage real-time tracking and analytics to optimize inventory levels, route shipments efficiently, and respond swiftly to disruptions.

The **transportation and logistics** industry relies heavily on real-time data to keep the world moving. From optimizing delivery routes based on real-time traffic data to managing fleet performance through IoT sensors, real-time analytics help companies improve efficiency, reduce costs, and enhance customer service. Ride-sharing platforms use real-time data to match drivers with passengers and dynamically adjust prices based on demand.

As we can see, across these diverse industries, real-time data is not just a nice-to-have – it's becoming a critical component of business success. In the next section, we'll explore the technologies that are making this real-time revolution possible.

# Enabling Technologies for Real-Time Data

The real-time data revolution is powered by a suite of cutting-edge technologies that work in concert to collect, process, store, and analyze data at unprecedented speeds. Understanding these technologies is crucial for any organization looking to harness the power of real-time analytics.

At the foundation of real-time data processing are advanced data pipelines and stream processing systems. These technologies enable the continuous ingestion and processing of data as it's generated. Platforms like Apache Kafka and Apache Flink allow businesses to handle massive volumes of data in motion, transforming raw data streams into actionable insights in real-time.

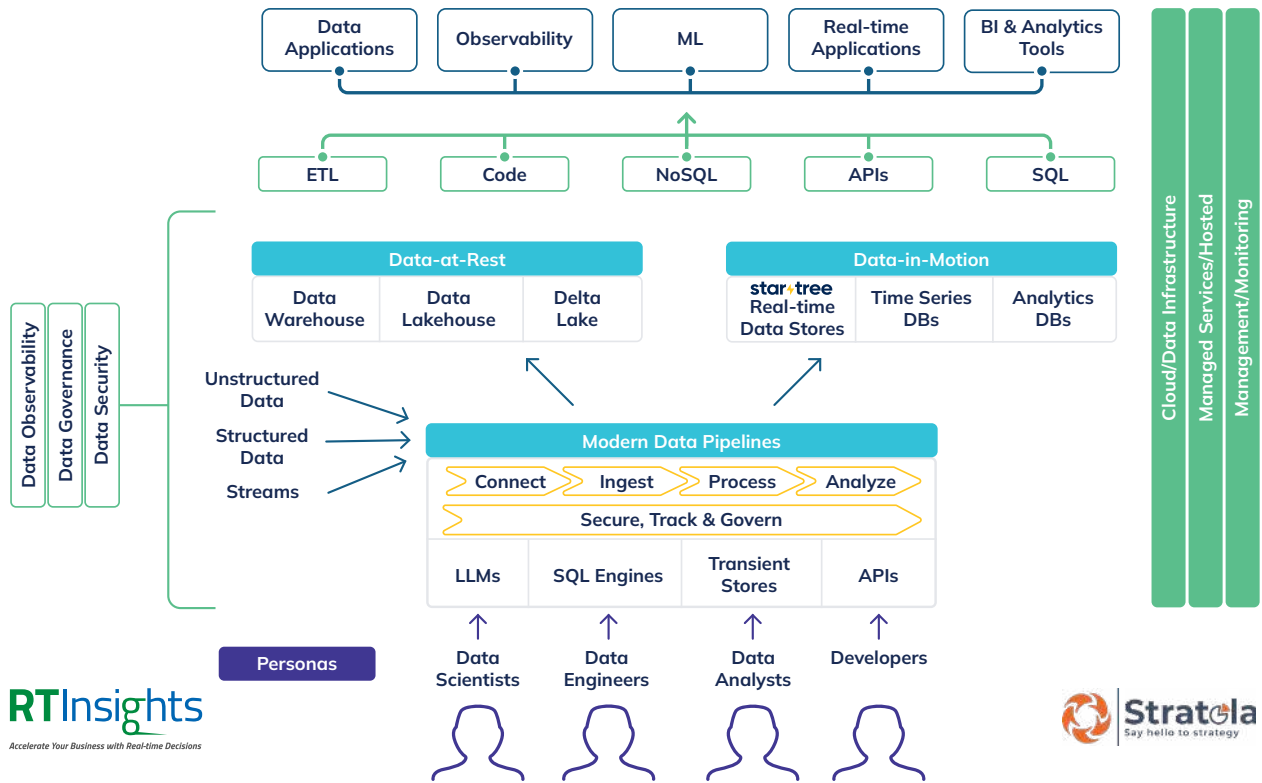
Real-time databases form another critical component of the real-time data ecosystem. Unlike traditional databases optimized for batch processing, real-time databases are designed to handle high-velocity data with low latency. Technologies like Apache Pinot enable businesses to query vast amounts of data in milliseconds, supporting real-time analytics and decision-making.

Edge computing and Internet of Things (IoT) technologies are pushing the boundaries of real-time data processing. By moving computation closer to the data source – be it a smartphone, a factory sensor, or a connected vehicle – edge computing reduces latency and enables instant data processing. This is particularly crucial in scenarios where milliseconds matter, such as in autonomous vehicles or industrial safety systems.

Observability and IT monitoring tools have evolved to keep pace with the demands of real-time systems. These technologies provide real-time visibility into the health and performance of complex, distributed systems. Tools like Prometheus and Grafana allow teams to detect and respond to issues instantly, ensuring the reliability and performance of real-time data pipelines.

The synergy between these technologies – data pipelines, real-time databases, edge computing, and observability tools – creates a powerful ecosystem for real-time data processing and analytics. However, implementing these technologies effectively requires a well-thought-out strategy. In the next section, we'll discuss how to build a comprehensive real-time data strategy for your organization.

### The Modern Data-in-Motion Ecosystem Map





# Building a Real-Time Data Strategy

Transitioning to a real-time data environment is a journey that requires careful planning and execution. A well-crafted real-time data strategy can be the difference between merely collecting data and truly leveraging it for competitive advantage. Let's explore the key steps in building an effective real-time data strategy.

The first step is assessing your current data stores and their capabilities. This involves taking stock of your existing data infrastructure, identifying the types and volumes of data you're dealing with, and evaluating your current data processing speeds. Three key factors should be considered in this exercise - latency, throughput, and data freshness.

Next, focus on identifying high-impact use cases for real-time data in your organization. Look for areas where faster data processing could lead to significant improvements in customer experience, operational efficiency, or decision-making. This might involve reducing fraud in financial transactions, optimizing supply chain operations, or personalizing customer interactions in real-time.

Finally, map out the right use cases to the right type of data and analytics stores. You are probably not seeing the benefits of real-time insights for a use case where data freshness is critical and you are using a data warehouse instead of a real-time analytics database.

As you move forward, be prepared to overcome implementation challenges. These might include technical hurdles like integrating new systems with legacy infrastructure, or organizational challenges like changing data-driven decision-making processes. It's crucial to secure buy-in from stakeholders across the organization and to invest in training to ensure your team can effectively leverage new real-time data capabilities. But, also, remember that some solutions can simplify your data architecture by having in-built data ingestion or integration capabilities.

Consider starting with pilot projects to demonstrate the value of real-time data. These small-scale initiatives can help you refine your approach, build internal expertise, and generate quick wins to build momentum for larger real-time data initiatives.

Remember that building a real-time data strategy is not just about technology – it's about fostering a culture of data-driven decision-making. Encourage teams to think in terms of real-time insights and actions. Promote cross-functional collaboration to ensure that real-time data benefits the entire organization.

As you develop your strategy, keep scalability in mind. Your real-time data needs will definitely grow over time, so choose solutions and architectures that can scale with your business.

# Enabling Data Immediacy

Enabling data immediacy requires a thoughtful approach to data architecture and technology selection. The goal is to create a system that can handle high-velocity data streams, process them in real-time, and deliver insights at the speed your business demands.

One of the key challenges in real-time data architecture is striking the right balance between data velocity and scale. While you want to process data as quickly as possible, you also need to ensure your system can handle growing data volumes and user loads. This often involves distributed architectures that can scale horizontally to meet increasing demands.

The right data architecture for your real-time needs depends on your specific use cases and existing infrastructure. Some of the common elements of the data architecture include a robust data ingestion layer, a stream processing engine, and a highly-performant real-time analytics database.

Choosing the right real-time analytics store is crucial for enabling data immediacy. Traditional data warehouses, while powerful for batch processing, often fall short when it comes to real-time analytics. Data warehouses will be the wrong tool of choice for your real-time data immediacy needs. You will incur significantly huge data latencies and inherent operational costs that are not ideal for real-time use cases. More importantly, the data will not be as fresh anymore for your business decisions.

## Consider the following factors when selecting a real-time analytics store:

- Query Latency:** The ability to execute complex queries on large datasets with sub-second response times.
- Data Freshness:** The capability to ingest and make data available for querying in real-time.
- Scalability:** The ability to handle increasing data volumes and concurrent users without performance degradation.
- Flexibility:** Support for various data types and the ability to adapt to changing schema requirements.
- Integration:** Easy integration with your existing data ecosystem and analytics tools.
- Consistency:** The level of data consistency required for your use cases.
- Cost-Effectiveness:** The total cost of ownership, including hardware, software, and operational costs.

# Apache Pinot

Apache Pinot stands out as an excellent choice for a real-time analytics database. Originally developed at LinkedIn and now an open-source project, Pinot is designed to deliver low-latency analytics at scale. It excels at handling high-velocity data ingestion while simultaneously serving analytical queries with sub-second response times.

By leveraging technologies like Apache Pinot, organizations can build data architectures that truly enable data immediacy, allowing them to harness the full power of their real-time data.

LinkedIn	stripe	Uber
1M+ EVENT/SEC	1PB+ DATA SIZE	200TB+ DATA SIZE
200K+ QUERIES/SEC	1T ROWS	30K+ QUERIES/SEC
1M+ QUERY LATENCY	< 1S QUERY LATENCY	<100MS QUERY LATENCY

**Key features that make Apache Pinot a great choice include:**

- Real-time ingestion and indexing of streaming data
- Support for both real-time and batch data ingestion
- Ability to handle complex queries on large datasets with low latency
- High availability and horizontal scalability
- Integration with popular data ecosystems like Kafka and Hadoop



# Measuring the ROI of Real-Time Data Initiatives

Implementing real-time data solutions represents a significant investment for many organizations. As with any major business initiative, it's crucial to measure the return on this investment to justify the costs and guide future strategies. The cost of real-time analytics can be quite high if you choose the wrong stack. Let's explore how to quantify the impact of your real-time data projects.

Key Performance Indicators (KPIs) for real-time data projects often fall into several categories:

- 1. Operational Efficiency:** Measure improvements in process speed, resource utilization, and cost savings.
- 2. Customer Experience:** Track metrics like customer satisfaction scores, retention rates, and personalization effectiveness.
- 3. Decision-Making Speed:** Assess the reduction in time-to-decision for critical business processes.
- 4. Revenue Impact:** Monitor increases in sales, cross-selling success rates, or fraud prevention savings.
- 5. Data Quality and Accessibility:** Evaluate improvements in data freshness, accuracy, and ease of access.

When setting up these KPIs, it's important to establish baselines before implementing real-time solutions. This allows for clear before-and-after comparisons.

The cost of a typical analytics data platform can be broken down into three components - storage cost, compute cost, and operational cost.

**Storage Cost** - This refers to the cost of storing and managing the data ingested from sources such as Apache Kafka or Amazon S3. A real-time analytics store like Apache Pinot can lower this cost with its ability to store data in a columnar format, use efficient compression algorithms, and leverage tiered storage to store highly fresh data in high I/O, SSD-based server nodes and less frequently used data in cheaper server nodes.

**Compute Cost** - This refers to the cost of processing queries as well as other tasks such as data ingestion, cluster maintenance, data governance, etc. Query optimization within Apache Pinot attempts to minimize this work, hence lowering overall resource consumption. You can also reduce query processing costs using various indexing strategies, query pruning, routing optimizations, and scaling the compute nodes as appropriate.

**Operational Cost** - This refers to the human cost of deploying, managing, and administering the datasets and clusters. Using a managed service provider (MSP) or the right Database-as-a-Service solution can help with reducing this cost. Such providers or solutions offer out-of-the-box automated cluster management, no-code ingestion utilities, easy data management operations, performance management, and troubleshooting.

Remember, ROI measurement should be an ongoing process. As your real-time data capabilities mature, continue to refine your metrics and look for new areas where real-time insights can drive value.

# StarTree - Enabling Data Immediacy with Real-Time User-Facing Analytics

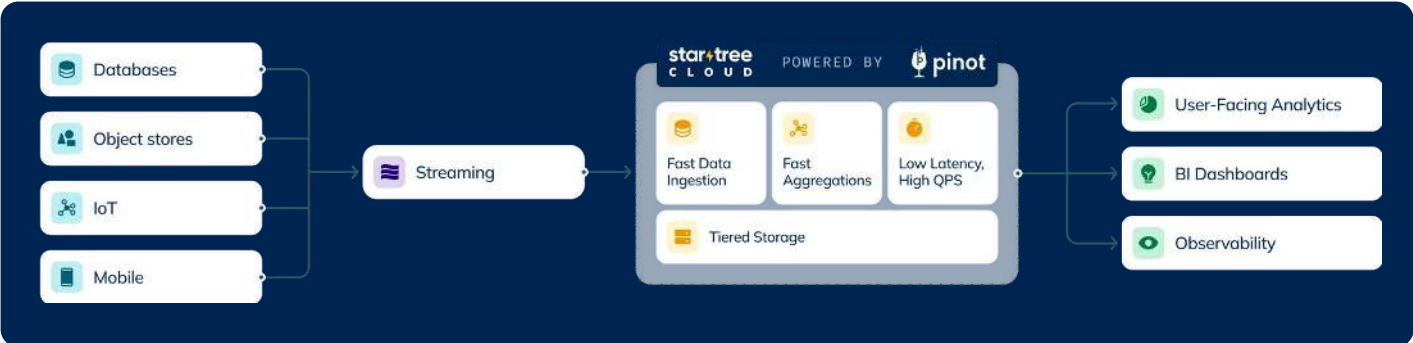
In the landscape of real-time analytics solutions, StarTree stands out as a powerful enabler of data immediacy. Built on the foundation of Apache Pinot, StarTree's platform is designed to deliver real-time, user-facing analytics at scale.

## Product Overview

StarTree's real-time analytics platform allows organizations to ingest, store, and query massive amounts of data with blazing-fast performance. It's specifically engineered to handle high-concurrency workloads, making it ideal for user-facing analytics applications where thousands of users might be querying the system simultaneously.

## Key Capabilities

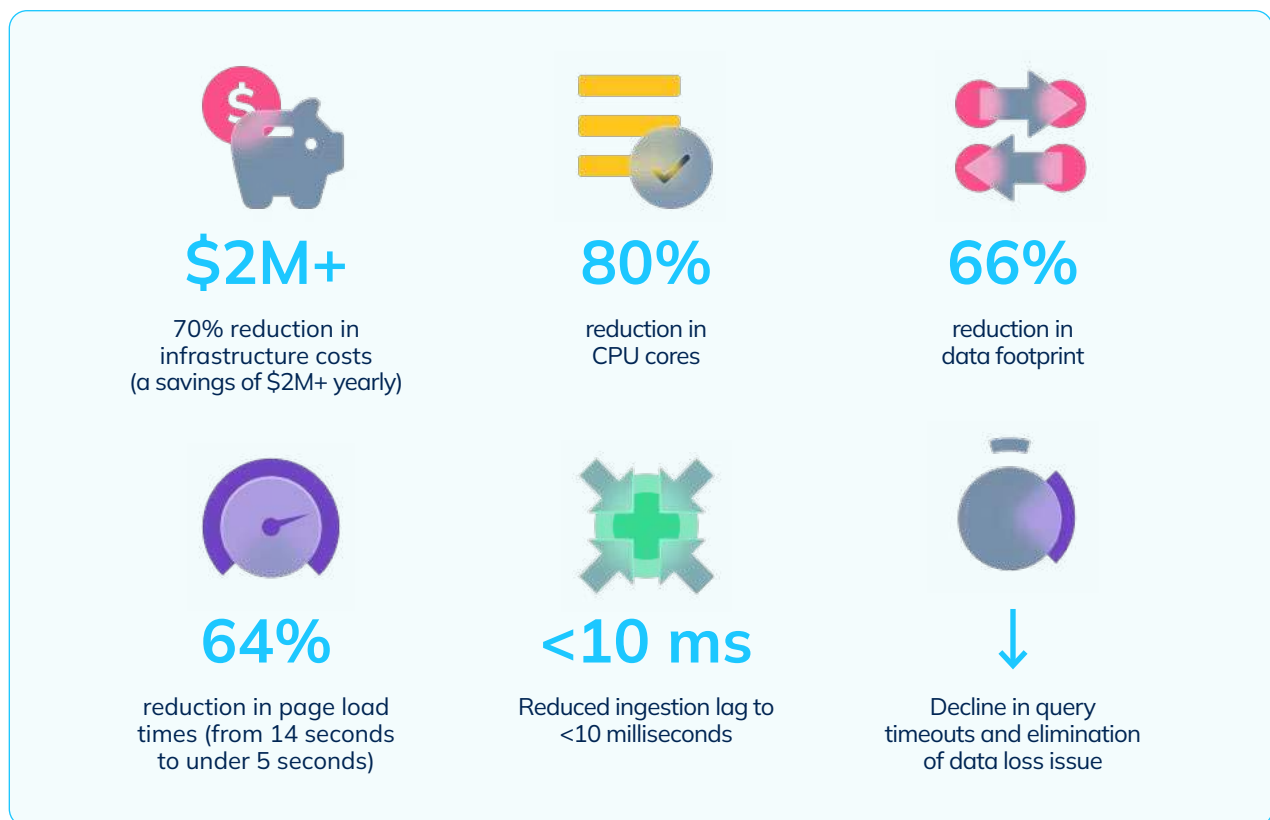
- ⚡ **Real-Time Ingestion:** StarTree can ingest millions of events per second, ensuring your analytics are always based on the freshest data. StarTree's real-time indexing allows for immediate availability of data.
- ⚡ **Low-Latency Queries:** Even with billions of records, StarTree delivers query responses in milliseconds. StarTree offers support for complex OLAP queries including filters, aggregations, and joins.
- ⚡ **High Concurrency:** The platform can handle thousands of queries per second without performance degradation.
- ⚡ **Scalability:** StarTree's architecture allows for seamless horizontal scaling to meet growing data and user demands.
- ⚡ **Flexible Deployment:** Available as a fully managed cloud service on major cloud platforms or for on-premises / hybrid deployments.
- ⚡ **Easy Integration:** StarTree supports popular data sources and sinks across the enterprise data spectrum such as Snowflake, AWS S3, Apache Kafka, Apache Flink etc.
- ⚡ **Advanced Analytics:** StarTree also supports time series analysis, anomaly detection, and predictive analytics.



## Why Choose StarTree

Built on the foundation of Apache Pinot, StarTree offers a robust, fully managed real-time analytics cloud solution designed for low-latency data processing. The platform is optimized for ingesting and querying data as it arrives, ensuring that users have access to the most current insights without delay. This capability is crucial for applications requiring immediate data responsiveness. With such advanced capabilities listed above, StarTree can outperform most of the competitive products in the market in terms of scale, speed and low cost. StarTree can truly deliver on the promise of low-cost and fast ROI on real-time data initiatives.

Uber, a company that relies heavily on real-time data to offer rides and food pickups to its customers, migrated to Apache Pinot to power their in-house tool that monitors mobile app crash analytics. Here are some of the benefits that Uber saw with Apache Pinot, compared to the previous analytics engine that they had -



StarTree's platform enables organizations to move beyond traditional batch-oriented analytics and embrace true data immediacy. Whether you're looking to enhance customer experiences, optimize operations in real-time, or make split-second decisions based on live data, StarTree provides the tools and performance to make it happen.

# Conclusion

The journey to real-time data mastery is not without its challenges. It requires a strategic approach, the right technologies, and often a cultural shift within the organization. But the rewards – increased agility, enhanced customer satisfaction, and improved bottom line – make it a journey worth undertaking.

As we've seen, industries across the board are reaping the benefits of real-time data. From retail to finance, healthcare to manufacturing, organizations that have embraced data immediacy are setting new standards in their fields. They're not just keeping up with the pace of business; they're actively shaping it.

The technologies enabling this real-time revolution – advanced data pipelines, real-time databases like Apache Pinot, edge computing, and sophisticated monitoring tools – are more accessible than ever. Platforms like StarTree are making it possible for organizations of all sizes to leverage the power of real-time analytics without the need for massive in-house data engineering teams.

As we look to the future, the importance of real-time data will only grow. The proliferation of IoT devices, the advent of 5G networks, and the increasing expectations of consumers for instant, personalized experiences will drive demand for even faster, more sophisticated real-time data solutions.

The question is no longer whether your organization should adopt real-time data strategies, but how quickly you can implement them. Those who act now to embrace data immediacy will be best positioned to lead in their industries, while those who delay risk falling behind.

By harnessing the power of now – the ability to turn data into action at the speed of thought – you're not just preparing your business for the future; you're actively shaping it. Welcome to the era of data immediacy, where every moment brings new opportunities for those ready to seize them.

# Next Steps

Now that you understand the importance of real-time data and the potential it holds for your business, it's time to take action. Here are some concrete steps you can take to begin or accelerate your journey towards data immediacy:

## 1. Assess Your Current State

- a. Evaluate your existing data infrastructure and processes.
- b. Identify areas where real-time insights could have the most significant impact.

## 2. Define Your Real-Time Data Strategy

- a. Set clear goals and KPIs for your real-time data initiatives.
- b. Prioritize use cases based on potential business value and feasibility.

## 3. Build Internal Expertise

- a. Invest in training for your team on real-time data technologies and best practices.
- b. Consider hiring or partnering with experts in real-time analytics.

## 4. Choose the Right Technologies

- a. Evaluate real-time data platforms like StarTree that align with your needs.
- b. Consider proof-of-concept projects to test different solutions.

## 5. Start Small, Think Big

- a. Begin with a pilot project to demonstrate value and gain buy-in.
- b. Plan for scalability from the start to support future growth.

## 6. Foster a Data-Driven Culture

- a. Encourage teams across the organization to think in terms of real-time insights.
- b. Celebrate early wins and share success stories to build momentum.

## 7. Continuously Optimize

- a. Regularly review and refine your real-time data processes.
- b. Stay informed about emerging technologies and best practices in the field.

## 8. Explore StarTree's Solutions

- a. Visit [startree.ai](https://startree.ai) to learn more about our real-time analytics platform.
- b. Contact our team for a personalized demo or consultation.

Remember, the path to data immediacy is a journey, not a destination. Each step you take brings you closer to unlocking the full potential of your data. Start your real-time data journey today, and position your organization at the forefront of the data-driven future.





Dinesh Chandrasekhar has been a data management veteran, thought leader, and data practitioner for 30+ years. As the chief analyst and founder of Stratola, he speaks and writes on the latest topics in data-in-motion, real-time analytics, IoT, Observability, GenAI, and more. Follow his work at [www.stratola.com](http://www.stratola.com).



**Stratola** is a business strategy consulting and full-stack marketing services firm for companies that need help with their product messaging, positioning, GTM strategies, sales enablement, pricing, and business development. With deep expertise in data management, analytics, data-in-motion, observability, and AI, Stratola successfully enables its clients with market penetration, category creation, thought leadership, product launches, customer acquisition, and overall growth. We are the business growth acceleration experts. Learn more at [www.stratola.com](http://www.stratola.com).

## RTInsights Research

**RTInsights Research** is the definitive voice in the evolving landscape of real-time insights and its transformative business impact of better decision making. We stand as a cornerstone for CxOs, operations leaders, and forward-thinking data and IT professionals, providing them with on-the-pulse industry trend analysis of emerging technologies that are leading towards business disruption.

The RTInsights Research Team has deep expertise across the real-time ecosystem, and how leading companies are implementing emerging technologies to create value. Our team has published reports and delivered webinars that inform business and technology strategy in areas such as Edge computing and IoT, Real-Time and Advanced Analytics, and the deployment of AI and Generative AI in enterprises. In addition, the RTInsights Research team has deep expertise on the supporting infrastructure that enables real-time insights, including Cloud and Hybrid Cloud Data Management, Machine Learning, DataOps and AIOps.

RTInsights Research provides business and technology leaders with actionable insights, helping to drive growth through emerging technologies.