

The Newest Products for Your Newest Designs.

M

MOUSER ELECTRONICS.

TDK

Featuring high-reliability without joints & magnetic flux cancellation realizing low EMI.

HPL Wound Ferrite Inductors

[Learn More](#)

MORNSUN® Power Solution for **Charging Station**

ISSI Long-term Support World Class Quality

DDR4 DRAM
Long lifecycle support and world-class quality in a 96-ball BGA package

ISSI **Big-Key** AUTHORIZED DISTRIBUTOR

[LEARN MORE](#)

M

MOUSER ELECTRONICS.

Methods

Keep up to date on what's important in technology, innovation and engineering design.

[Learn More](#)

eletimes.com

Big Data: A Guide to Choosing the Right Big Data Analytics Tools

December 29, 2018



In order to choose the right big data analysis tools, it's important to understand the transactional and analytical data processing requirements of your systems and choose accordingly.

Big data keeps getting bigger, but not every activity involving the use of this data is created equal. Sometimes, utilizing data is like running a small but critical errand at the corner store. At other times, it's about going for a leisurely stroll through a warehouse and taking a good hard look at the inventory. The objectives, and therefore the technology needed to handle transactional data, as opposed to the tools needed for analytical data processing are quite different. In order to choose the right big data analysis tools for the job, it's important to understand both the big differences and subtle nuances that differentiate operational data from data that is more analytical.

Operational workloads are about getting things done right, right now

Operational or transactional data handling has a focus on low latency for response times and handling many concurrent requests. Some real-time analytics may be involved, but they are typically limited to a small set of variables that are relevant to immediate decision-making processes for the end user. Such information might be displayed on a simple dashboard that allows business users to run standard or custom reports based on their own needs and experience level.

One of the most important features of a data transaction is reliability. “In a bank transfer, you need the from account and the to account to maintain transactional consistency so the money doesn’t fall on the floor if something breaks in the middle. You’re interested in updating a very small amount of records and making a rigid concept around transactionality,” Jacques Nadeau, vice president of Apache Arrow and Apache Drill, said.

Analytics and the right answers

In contrast, analytics typically involves the ability to process large volumes of data throughput using complex query structures. While streaming analytics may be a feature for specific use cases, analysis for many enterprises is still focused primarily on review of historical Big Data for longer range planning and prediction. As an example, a business might want to analyze sales in the last quarter or use machine learning operations to see what customers buy in a given situation. In the most challenging cases, businesses may not know exactly what they are looking for — or they may be intentionally experimenting with different ways to derive value from their existing data stores. Data scientists may be called upon to craft the right queries that deliver relevant business insights.

Julien Le Dem, principal architect at Dremio and Apache Parquet co-founder, offered a simple way of thinking about the difference: Moving data around is transactional, processing it is analytical. “You are working with a lot of records at the same time versus working with only one or a few records at one time. Analytics is about extracting the parts you are interested in very efficiently and producing results based on that data.”

Choosing the right solution for your data

Big data analysis tools have emerged for real-time, interactive workloads and retrospective, complex analysis of larger data sets. MongoDB and IBM, both major players in the big data analysis tools space, offer some key insights into the differences between the two. Here’s a brief overview.

According to IBM, NoSQL systems such as document databases and key-value stores are common solutions for fast and scalable operational databases. With an appropriate NoSQL database, transactions can be processed more quickly, and the system can handle many small transactions at the same time during periods of peak activity. Transactions per second are viewed as a more relevant indicator of performance than response time.

Massively Parallel Processing (MPP) databases and MapReduce — including variants like Hadoop — are key solutions in the analytical space. There are even emerging solutions that are designed to meet the needs of enterprises in analyzing data across both SQL and NoSQL, presenting Graph, R and MapReduce within a single analytics platform.

Distinguishing features for operational vs. analytical data processing systems

Experts at MongoDB offer additional detail about the technical distinctions between analytics and online transaction processing systems.

Transactional systems are optimized for short, atomic, repetitive, select-oriented operations and transactions — these systems can be very finely tuned for frequently used operations. They feature heavy reliance on caching, lots of resource sharing and prescribed code paths.

Analytical systems provide functional richness; processing speed, or fast response time; and ease of use. They typically feature lots of capacity within an MPP. Such systems have the ability to move data quickly when needed but are designed to reduce data movement overall. They rely on few shared structures. The functions may be built into the server and extensible to meet evolving end-user requirements.

Relying on a single database system to handle both types of activity is labor intensive for IT, since conventional database systems demonstrate a great deal of variability in performance when asked to handle analytic and transactional workloads. Of course, not all big data analysis tools suit every possible need, which means that at the enterprise level, most organizations end up using complementary systems to meet all their data workload needs.

Here are five steps to choosing the right data analytics tools for your organization.

1. Research and discovery

First, business professionals must determine the current state of analytical tool implementation and analytical capabilities within their organization. To do so, they must conduct in-depth interviews with key stakeholders including business intelligence developers, administrators, and IT executives, said Levy. Essentially, you must interview the people that are going to both use and benefit from analytical tools, he added.

“That helps us to understand what the ins and outs are around who’s using these things, what they are using, what tools they use to do their work, what the tools can do, what they can’t do, and if they are being used correctly,” said Levy. “Are these tools being used to their fullest capabilities? Do they have the in-house knowledge required to make the fullest use of their portfolio of software?”

2. **Current state landscape**

The second step involves taking inventory of market’s current analytical tools and separating them into different classes. These tool classes include report writers, semantic layer reporting tools, MDX/Cube query tools, data discovery and visualization tools, embedded BI and reporting tools, data science and modeling tools, as well as AI and machine learning use case driven tools.

“Where is the next wave going? What’s the landscape like in terms of the various vendors and the tools that they offer?” Levy said. “You start to sift those into the different holes that you found in the first step.”

3. **Capability tree**

The third step uses a capability tree to compare the results from step one and step two, so you look at the classifications of your company’s current inventory against the overall market’s inventory, said Levy.

The capability tree is helpful because businesses can see areas they are doing well in, or are lacking in, based on the tools that are big in the market.

4. **Decision matrix**

“[The decision matrix] is where for each of these tool classes or sets, or if you’re doing a specific vendor selection for each of these vendors, you go in and actually go rate them in these various capabilities,” said Levy. The scoring will be based on the needs of the organization, providing more weight to the capabilities more important to the business.

“For instance, data science. We know from our experience that a Big Data science tool is really good at advanced algorithm creations, maybe not so good at displaying dashboards,” said Levy. “We can use our experience there to rate the various different classes on the capabilities that you defined.”






5. **Decision tool**




Finally, the organization uses a decision tool to match the best tool with each business capability. “[A decision tool] is a combination of the capability tree and the decision matrix, in the sense that you weigh each of the capabilities according to what’s most important to the organization, or for whatever particular project that they’re undertaking,” said Levy. “You weigh these various capabilities, and the decision tool should spit out the weighted score of all of these capabilities and tell you what the right candidate is.”

Regardless of the steps, however, business leaders need to spend a lot of time studying their own company and figuring out where the most help is needed, said Levy. None of the tools will be helpful if none of them are solving the actual gaps and problems within the organization.

1. Desktop Computer Deals	➤	3. Download Free Software Update	➤	5. Picasa Photo Editing	➤
2. PDF File to Word Document	➤	4. All-In-One PC Deals	➤	6. Desktop PC Deals	➤

- TAGS
- Analytics Tools
- Big Data
- Big Data Analytics Tools
- Machine Learning
- MapReduce
- MPP
- NoSQL





RELATED ARTICLES

MORE FROM AUTHOR



Photovoltaic Converters, Get More from the Sun with STM32 High-Resolution Timers



Improving the properties of diodes and transistors using spintronics



Unveiling new AI-based live sense SDK to keep drivers informed of unexpected road hazards



LEAVE A REPLY

Comment:

Name:

Email:

Website:

POST COMMENT



AC/DC Converters

RECOM

ELECTROLUBE
THE SOLUTIONS PEOPLE
**Poised To Lead The
Way In Thermal Management
Solutions:**
'Top '10 Most Promising Thermal
Management Solution Providers-2018'
FIND OUT MORE>
electrolube.in
Twitter LinkedIn Google+ Facebook

Experience our
handheld
spectrum analyzer.
[Click here!](#)



ROHDE & SCHWARZ

White Papers

How Public Safety Communications Can Close the Digital Gap

ELE Times - October 22, 2019

Introducing the DXOMARK Audio test protocol

ELE Times - October 22, 2019

How can we test smartphone speaker playback?

ELE Times - October 21, 2019

Simplify Connecting to Smarter Manufacturing

ELE Times - October 17, 2019

Why MQTT Is Important for Edge-to-cloud Connectivity?

ELE Times - October 14, 2019

How Time-sensitive Networking Is Revolutionizing Smart Manufacturing

ELE Times - October 11, 2019

Renewable Energy

India showing political will on renewable energy development: IRENA

ELE Times - November 1, 2019

30 gigawatts of renewable plants to be built along western border in India

ELE Times - October 19, 2019

The new thin film Solar Car will run for thirty days without pole-charging

ELE Times - September 21, 2019

Renewable energy to future-proof global trade

ELE Times - September 19, 2019

New design rule for generating excitons to enhance next-generation devices

ELE Times - September 13, 2019

India has already achieved half of its renewable energy target: Prakash Javadekar

ELE Times - September 11, 2019

India gifts solar panels to UN headquarters in New York to promote sustainable energy

ELE Times - September 9, 2019

<  >

Subscribe for Weekly E-newsletter

Your Email

Subscribe

