ANALYTICS SYNERGY IN THE MODERN SOFTWARE DEVELOPMENT ORGANIZATION

source allies

Analytics Synergy in the Modern Software Development Organization

EXECUTIVE SUMMARY

Decision makers in modern organizations need the ability to combine and analyze data from many systems in order to gain insights and drive business direction. Historically, this has been achieved by forming analytics teams responsible for transforming data from source systems into a central data store, on top of which they create reports and dashboards.

The problem is, with each additional data source coming online, and with each new question asked by the business, an additional burden is added to the analytics team. How does an organization embrace this growth? This executive white paper proposes a shared ownership strategy that allows analytics to scale with the organization. The idea is to combine development teams with analytics experts early on in the application development/configuration process and make the data analyzable from the get-go. This realignment makes increased data attributes, sources, and granularity an asset rather than a burden on the business. It also reduces the time and cost of managing the current infrastructure, and it positions an organization to use data to be competitive in their market.

THE CURRENT APPROACH

Many businesses are still using centralized data platforms, meaning the information is stored and maintained in one location (data warehouse or data lake) and often governed by a dedicated analytics team. These individuals are responsible for transforming data from source systems into a standardized model, then building dashboards and reports on top of this warehouse. The drawback to this approach is an organization-wide analytics project employing this strategy

might spend a year's worth of effort and resources before the first report is ever produced. During this time, a number of things could happen:

- Source systems add new data attributes

- New systems come online

- The current system reaches end-of-life

- Other companies, with their own data, are acquired or divested

- New regulations take effect

- New markets emerge (with new questions to answer)

- The initial results create follow-up analytics questions to answer

Analytics projects in the past, whether knowingly or not, have incorrectly assumed that the requirements gathered and executed on during the duration of the project remain true and relevant until the project is completed. There's also a high risk that a dedicated analytics team responsible for overseeing all the data will generate a bottleneck in business decision-making. As problems continue to arise, data analysts must balance a backlog of data integration work with existing systems and new requests to answer new questions.

Managing a centralized data infrastructure is costly and time-consuming. This is the reason many businesses are moving toward cloud computing platforms such as Amazon Web Services. They are aiming to reduce the dependency on data warehouses by delegating all infrastructure to free up their data analysts. That way, the analytics team can focus on solving business-related problems instead of problems related to infrastructure. However, moving to the cloud only tackles one half of the problem, because while that does outsource the infrastructure, there is still one central team that owns the cloud data lake and is responsible for the development of analytics on top of it.
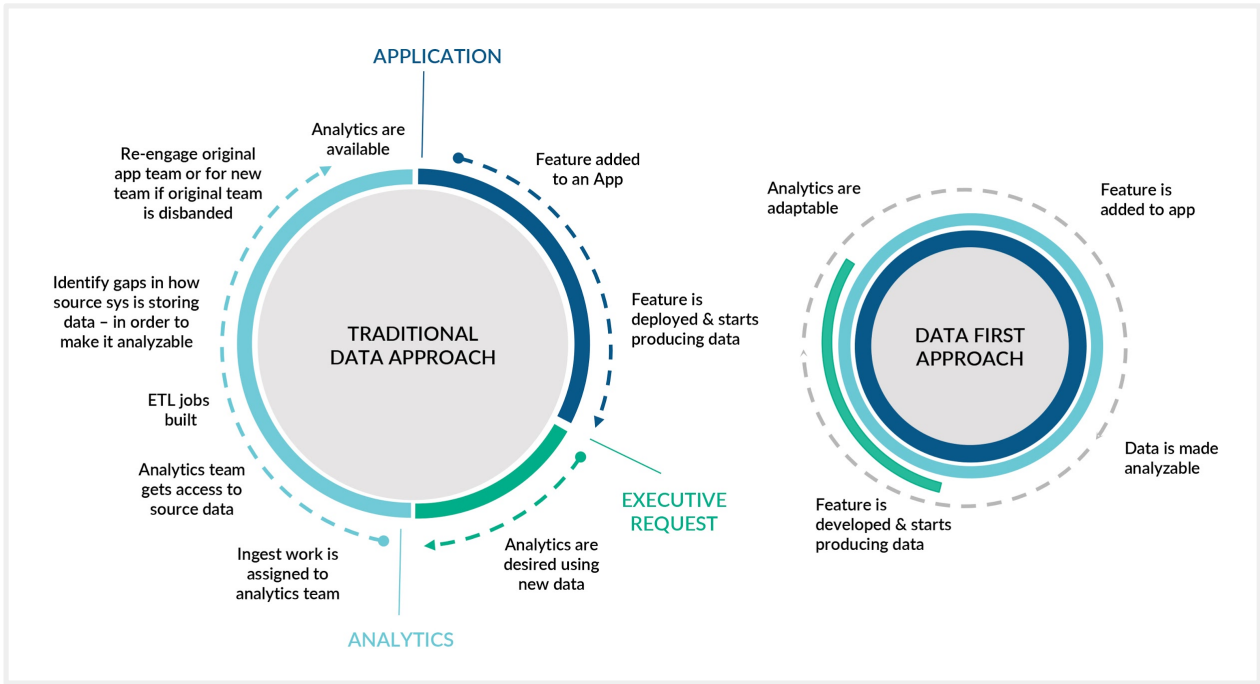
In a recent ThoughtWorks article, Zhamak Dehghani writes, "*We need to shift to a paradigm that draws from modern distributed architecture: considering domains as the first class concern, applying*

platform thinking to create self-serve data infrastructure, and treating data as a product." This type of approach empowers development teams and promotes self-service.

THE ROLE OF DEVELOPERS AND DATA ANALYSTS

Development teams that are responsible for building and enhancing mission-critical source systems generally do not include analytics professionals in the initial scope of the planning. Therefore, development teams may not store data that could become valuable in the future simply because the information is not required now. And the data they do store may not be in a format best suited for analytics.

On the flip side, the analytics teams responsible for spearheading these kinds of projects are skilled in data modeling, warehousing, and analytics tools. They are not typically members of the product teams that build and maintain the source systems. Therefore, they lack the technical context needed to understand how those systems work, how their data is stored, and how those systems are changing.

APPLICATION

Analytics are available

Re-engage original app team or for new team if original team is disbanded

Feature added to an App

Identify gaps in how source sys is storing data – in order to make it analyzable

Feature is deployed & starts producing data

TRADITIONAL DATA APPROACH

ETL jobs built

Analytics team gets access to source data

EXECUTIVE REQUEST

Ingest work is assigned to analytics team

Analytics are desired using new data

ANALYTICS

Analytics are adaptable

Feature is added to app

DATA FIRST APPROACH

Data is made analyzable

Feature is developed & starts producing data

Because an analytics project typically executes after the source system gets built or purchased, there is little opportunity to change those systems to track new pieces of data that are found to be critical to decision-making. Business leaders cannot gain valuable insights from the source system data until analysts are nearly done interpreting it — when changes are the most expensive, time-consuming, and disruptive to the company. In addition, source systems often store data in non-obvious ways. So, when it comes time to analyze the data, it can often be ambiguous or mislabeled. Misinterpreting the source data can result in project delays or, worse, in faulty conclusions and poor business decisions.

TIGHTENING THE FEEDBACK LOOP

According to the <u>National Institute of Standards and Technology (NIST)</u>, the cost to meet an analytics requirement in the production stage of the Software Development Life Cycle (SDLC) is approximately 75–990 times greater than doing so during the requirements stage. To drive down the cost of answering analytics questions, companies should look to introduce analytics earlier in the SDLC and report on timely remediation.

Many businesses are also becoming more responsible for creating a culture of quality in order to avoid fixing costly problems down the line. This includes making sure analytics has been built in to an application from the get-go. According to an article in the *Harvard Business Review*: "*A company with a highly-developed culture of quality spends, on average, $350 million less annually fixing mistakes than a company with a poorly developed one.*"

The feedback loop of quality assurance and development has been shortened in many organizations by adding testers to the development team and making the team responsible for quality. A similar paradigm shift is occurring in the infrastructure space. As cloud technologies become more common and teams are better able to self-serve their infrastructure, teams are expected to be responsible for deployments and uptime. Major companies like Netflix have already started implementing this new approach and seen positive results. In one of their recent tech blogs, Netflix urges other businesses to "operate what you build":

> *"'Operate what you build' puts the DevOps principles in action by having the team that develops a system also be responsible for operating and supporting that system. Distributing this responsibility to each development team, rather than externalizing it, creates direct feedback loops and aligns incentives. Teams that feel operational pain are empowered to remediate the pain by changing their system design or code; they are responsible and accountable for both functions. Each development team owns deployment issues, performance bugs, capacity planning, alerting gaps, partner support, and so on."*

Data analytics can benefit from these same principles. Data analysts and development teams can opt to work together during the project's inception to ensure that both groups understand the source systems inside out, as well as making sure the data is analyzable and easily interpreted.

- Request that your application development teams expose their data as part of the initial app development (i.e., "shift left analytics").

- Empower application development teams to own the concept of making their data analyzable, thereby removing the bottleneck of a centralized organization.

- Serve the analytics needs of today and tomorrow.

- Iteratively build analytics solutions starting with the first initial report in parallel with the application that sources the data.

- Have analytics specialists mentor and educate development teams in the use of data and tooling.

- Decentralize analytics to provide economies of scale, data permanence, etc.

- Include analytics professionals on product teams without changing who team members report to.
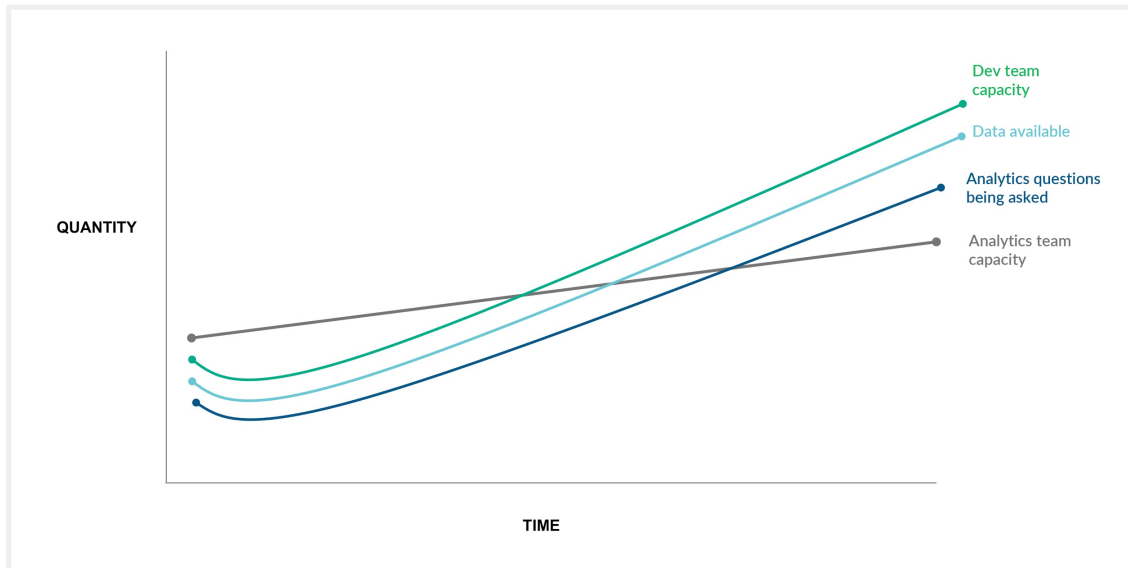
Business stakeholders overseeing the development of applications within the organization will need to expand their view to include making data analyzable, as they are the best equipped to understand the data their applications produce. Studies have shown this will help them in the long run, as it's much more cost-effective to make the data analyzable during development rather than tacking on an analytics solution after the fact.

CONCLUSION

As business needs continue to evolve alongside technological innovation, the more traditional data models and centralized analytics expertise can no longer effectively serve large corporations. There's a growing need for real-time data analytics, and businesses can't afford to make faulty decisions based on bad data. Sharing distributed data and analytics responsibilities is a transformative solution that allows your business to answer the questions it needs today without impeding your ability to answer tomorrow's questions.

ABOUT SOURCE ALLIES

Source Allies is a technology services company based in Des Moines, Iowa. Our multi-disciplinary teams are comprised of senior IT professionals whose areas of technical and process specialization include software development, cloud infrastructure, user-experience, and information security. We have extensive experience building distributed, test-driven, secure analytics solutions, whether it's legacy or greenfield, on-prem, or with any of the three major

cloud providers. Our company takes an iterative, cross-functional, team-focused approach to all development to ensure that we're building the right thing at the right time to solve the right problem. We work closely with our partners to solve mission-critical technical problems and deliver high-quality software solutions.

REFERENCES

Dehghani, Zhamak. "How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh." (2019). martinfowler.com. Accessed November 19, 2019. https://martinfowler.com/articles/data-monolith-to-mesh.html.

"The Economic Impacts of Inadequate Infrastructure for Software Testing." (2002). National Institute of Standards and Technology. 1-12, 1-13. https://www.nist.gov/system/files/documents/director/planning/report02-3.pdf.

Netflix Technology Blog. "Full Cycle Developers at Netflix—Operate What You Build." (2018). Medium. Netflix TechBlog. Accessed November 11, 2019. https://medium.com/netflix-techblog/full-cycle-developers-at-netflix-a08c31f83249.

Srinivasan, Ashwin and Bryan Kurey. "Creating a Culture of Quality." (2014). *Harvard Business Review*. Accessed November 19, 2019. https://hbr.org/2014/04/creating-a-culture-of-quality.