Written By

Eastern Analytics

COMMON PITFALLS WHEN BUILDING A CLOUD-BASED ANALYTICS PLATFORM

How to Avoid Them & Successfully Build Your Own

Free ★★★★★ Guide

EASTERN ANALYTICS

CONTENT INDEX

EASTERN ANALYTICS

More and more organizations are moving their analytics platforms to the cloud to take advantage of the flexibility, scalability, and cost savings. Not all cloud-based analytics platforms are created equal, however, and building a performance-driven platform is a complex process that requires a lot of time, effort, and skill.

It is essential to have a clear understanding of the purpose of the platform, the various functionalities it will include, and the required skillset so you can build it successfully while avoiding some of the most common pitfalls.

The Process of Building a Cloud-Based Analytics Platform

Part 1 -Designing the Platform

There are many pitfalls that organizations can face when building their platform. These pitfalls, however, can be avoided with proper planning.

While the process of building a cloud-based analytics platform is a little different for every organization in terms of the types of resources that are required, there are some commonalities. The following are common requirements to successfully build a cloud-based analytics platform regardless of organization:

- Define the platform goals
- Define your requirements
- Establish your data governance strategy
- Create a high level, comprehensive data model for the platform
- Define the platform's scalability
- Establish the platform's technical architecture

Defining Your Platform Goals

One of the most common pitfalls organizations face when building their platform is the failure to clearly define the business goals. Clearly defining your goals will allow you to break the project down into components and determine what clearly defines success.

What are the needs you are trying to fill? Do you have a specific use case where the deliverable is clear or are you trying to perform a 'lift & shift' without adding any new functionality for the business? Or are you trying to build a lake house (data lake & data warehouse in one) where all data streams are to flow and then build out data marts as requirements come in?

As you validate your understanding of your business drivers and challenges, identify your short-term needs, your long-term goals, and the data that you need to meet them. Start to build a roadmap with high-level objectives and an overall plan to meet them. Make sure that the roadmap has enough detail to provide a path, but make it clear that it will be refined based on lessons learned as you go.

EASTERN ANALYTICS

Defining Your Platform Requirements

After you define your platform goals, it is important that you have a good understanding of what you need and don't need. Determine the type of data you want to collect, the type of analysis you will be performing and a general approach to storage.

Next, you should decide what type of analytics software to use. There are many software packages available including a build-your-own approach where you select 'best-of-breed' depending upon your needs and integration points.

One of the components considered during the selection process should be cost, whether it be licensing costs or compute costs. Another component, commonly overlooked, is the cost of staffing & productivity. Selecting tools that are mature and easy to implement tends to keep this cost low.

There are many different analytics tools and technology stacks to choose from, each with its own set of features. At Eastern Analytics, we are partial to Microsoft Power BI and Microsoft Azure due to their tight integration with Azure Active Directory and Office365. There are multiple tools within the Azure stack and many of them are a progression from their on-site predecessors, so costs tend to be about average and Microsoft is a name you can depend on.

Establishing the Platform's Data Governance Strategy

As you identify your data requirements, you must understand your security and compliance requirements throughout the data lifecycle.

Your success in building a cloud-based analytics platform will depend on your ability to ensure accountability, control, and governance. Develop a cloud-governance framework that allows for data security and includes cost controls. Create policies that cover data access, storage and retention requirements and ensure that your employees are trained well enough to use the cloud safely and effectively.

Establishing a Comprehensive Data Model for Your Platform

Data visualizations often go wrong, making it hard to decipher the information they're trying to illustrate. Similarly, a BI dashboard or report is only valuable if it's easy to navigate and understand the data. Proper planning will ensure your visualizations are consistent and simplified so your users can't accidentally misrepresent that data to themselves.

When creating your data model, start with an ingestion layer to store data in its raw format. This will enable you to trace data back to its source and provide a thorough root cause analysis. In a second layer, enhance your data so that it conforms to your naming standards and transform it as required. In a third/top layer deliver consolidated/reporting quality data with your business user mind. There is no limit to the number of logical layers that you have in your repository. The primary rule is consistency. By consistently following the same approach for each data source the system will be easy to expand and easy to support.

Avoid common errors such as allowing different sources to follow their own source specific design, letting small sources fall outside the guidelines, and failing to account for how navigation through your data set could affect calculations within your model. Taking steps to design your platform first, then adjusting the design as needed will go a long way towards creating a flexible yet sound environment that is easy to control and support.

Defining the Platform's Scalability

Data grows exponentially, and it can quickly overload your system. A sudden change in data volume can cause a performance bottleneck that can lead to service delays or unknown issues. To run big data and analytics projects, compute architecture needs to be considered a primary component. Your platform design should be a combination of sound data architecture and a scalable compute that allows for volume fluctuations without impacting your delivery of service. Choosing a platform that is scalable – whether it be Databricks with auto-scaling compute clusters, Azure ML with its auto-scaling features, or by programmatically scaling your SQL database, your environment should cost as little as possible but be able to dynamically scale when the need arises.

Analytics that cannot scale does not give you the data access or data insights you need to make smart business decisions. Far too many data analytics projects fail to deliver on their promise because they don't provide the business actionable insights quickly enough. Your platform must provide a robust, secure scalable infrastructure that can affordably solve your immediate need for speedy, accurate data while having the flexibility to grow with your business.

Your platform should provide:

- A hyper-scalable cloud data warehouse
- Fast, flexible data ingestion and transformation
- Data organized for consumption in the presentation layer
- Responsive dashboards and self-service analytics
- Machine learning tools that deliver at the speed of business

That's why Microsoft Azure is Eastern Analytics' tool of choice. It provides everything an organization could need including inexpensive blob storage, scalable architecture for almost of all of its services, and they take a look forward approach providing state of the art tools including Azure ML and Azure Cognitive Services (AI services that provide common, pre-defined AI to fulfill some of your common AI requirements - out of the box).

Establishing the Platform's Technical Architecture

A solid data analytics solution requires a robust architecture combining platforms, tools and services that serve as a framework for delivering analytics. A solid platform architecture will support state of the art approaches like Data Lakes and Lake Houses addressing all of your consumer needs.

Your architecture should include some/all of the following components:

- Inexpensive blob storage
- Database services – that support structured & non-structured data
- Industry standard compute capabilities such as Databricks and/or python
- AI/ML support – for training and using machine learning models
- Security features including built in firewalls, Denial of Service attack services
- Tight integration with your AIM/AD infrastructure – for user provisioning and access management

Keep in mind that your platform architecture will evolve over time. That's why flexibility is essential and it's one of the key elements in your overall success.

At Eastern Analytics we favor Microsoft Azure and its components. It includes best of breed services like Azure Databricks for processing, modeling, and storing large data sets and it can be combined with the Azure Data Factory for ETL/ELT. Additional functionality to support data streaming, IOT and AI/ML can be added later once the need arises. It even provides legacy supporting Azure SQL Database and Synapse so it is easy to migrate existing environments to the cloud for cost reduction and scalability as needed.

The Process of Building a Cloud-Based Analytics Platform

Part 2 -Establishing a Clear Timeline

All well-defined plans include a work breakdown structure and effort estimate. When establishing the project timeline, you will first need to scope out all the required tasks and subtasks, being as specific as possible and determine the project's milestones and target dates. Be sure to consider dependencies, resource allocations and durations and you build out the plan. A realistic timeline needs to consider not only your teams' allocations and deliverables, but also the availability of related parties as each phase is rolled out.

Projects can be long and involve many outside parties (parties other than your analytics team), so make sure to allocate the proper amount of time and have a contingency plan for when deviations and roadblocks are encountered. Make sure your timeline is flexible enough to work through unforeseen technical issues, scheduling conflicts and even gaps in requirements and/or data sources.

The Process of Building a Cloud-Based Analytics Platform

Part 3 -Establishing a Realistic Budget

The next step is to create a budget. Building a cloud-based analytics platform is a big-ticket venture, so be sure your funds will carry the commitment through. While developing your budget, you should account for upfront and recurring fees such as licensing and compute time. Some cloud related items include payments for data storage, processing/compute fees, security costs, maintenance, license and subscription fees, and data provider specific fees. Some of these are tough to estimate like compute fees or variable storage costs, but try your best. Many of these can be estimated based on a proof-of-concept project or one of the platform specific cost calculators. Proper budgeting & forecasting is the first step in determining the projects ROI. If your project plan is accurate, and your cost estimate is accurate then creating an accurate budget should fall within your standard IT budgeting approach.

The Process of Building a Cloud-Based Analytics Platform

Part 4 -Define Roles & Responsibilities

When creating a team to build your platform, it's important to create the right size team with the right skillsets and experience. You want to align your staffing with your deliverables and make sure each task can be met by your available resources. Without the right size team, or a team with all the necessary skillsets, deadlines will be missed and budgets will be exceeded. Make sure to have a contingency plan for unforeseen staffing issues. If necessary, align yourself with an external company such as Eastern Analytics, so when issues arise you have access to knowledgeable people, and they can meet any short-term staffing needs that you have. The earlier they are engaged the more pro-active they can be so that your timelines stay on track and under budget.

Preparing for Common Challenges

Once you have outlined your strategy, you'll want to prepare for some of the more common challenges. The best way to prepare for and overcome these challenges is to plan, plan and plan some more. With proper planning, you can mitigate, or even eliminate, some of the more common challenges such as:

- **Scope creep and increased costs** – Again, one of the best ways to keep costs on track is with proper planning. That being said, plan for the unexpected. If one part of the plan is thrown off, the whole timeline is affected resulting in higher costs. So, as with any project, be sure to pad your timeline and budget for the unexpected.

- **Security** – Cloud cybersecurity is a different approach to security than on-premises. Security measure should be naturally embedded in DevOps operations and must include the following:
  - Setting security configuration parameters in cloud instances
  - Automating security process
  - Building continuous monitoring systems

- **Data governance** - Make sure you understand the implications of regulations for cloud data. Data may need to be cleaned, or audited, or transformed in some way in order to be appropriately transferred.

- **Data quality issues** – Analytics platforms are only as accurate as the data they're built on. Utilizing high-quality data before beginning any analytics projects is crucial.

Conclusion

More and more organizations are moving their analytics platforms to the cloud. They are taking advantage of the flexibility, scalability, and cost savings that cloud computing offers. Done incorrectly, it can be costly. There is a lot to consider, but with proper planning and execution, the risk can be reduced or eliminated. That's why it's important to find the right analytics partner to assist you. With the right partner, building your cloud-based analytics platform can be smooth and easy, saving you time and money.

At Eastern Analytics, we have over 25 years' experience building analytics platforms on-premises and in the cloud. We have the technical and functional experience needed to help migrate or design a new environment from the ground up. Our dedicated team of consultants have the skills required to take you from concept to solution and support. We can help you design, develop, and deploy a scalable and cost-effective cloud-based analytics solution, and offer full support for all the platforms we build.

Call us today at 781-781-7610 to discover how we can help you maximize the value of your data.

Or visit us at: https://eastern-analytics.us to learn more.