HOW TO BUILD A MODERN DATA ARCHITECTURE

WITH AZURE SYNAPSE

zirous

Analytics, machine learning, and AI continue to be buzzy topics appearing on organizations' future roadmaps. However, as many teams within those organizations begin to think through how to make those initiatives a reality, some are beginning to question, "Considering our data environment's current state and our daily struggles with business-led reporting, how will we be able to enable advanced analytics?".

The pace of innovation is swift, and many IT leaders are realizing a re-evaluation of their data environment is necessary to support the types of projects and initiatives the business is starting to demand.

Today, with how quickly technology and customer expectations change, a modern data architecture must allow organizations to be nimble and flexible - to adapt just as their businesses will need to. Hybrid architectures across data lakes and data warehouses, on-prem and in the cloud, facilitated by modern ETL/ELT tools, allow for companies to begin their journey towards analytics no matter their starting point. And when done with the correct approach, quick wins can be accomplished every step along the way.

An organization's data is unique, making it one of their most important and strategic assets. Treating it as such requires an investment in technology, staff skill sets, and ultimately employee and customer experience.

Data is spread out in every organization. Data consolidation, within the context of use cases that drive business value, is an achievable necessity for those that want to keep pace with their competition, streamline manual processes, and enable the discovery of new and differentiated insights from the data they already have.
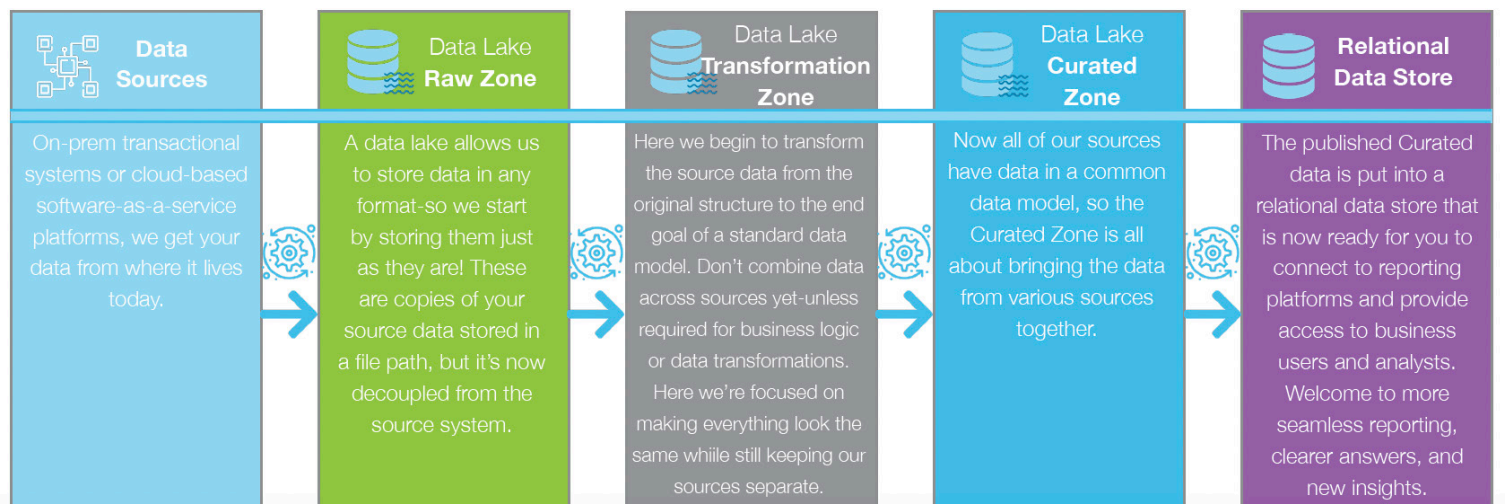
zirous®
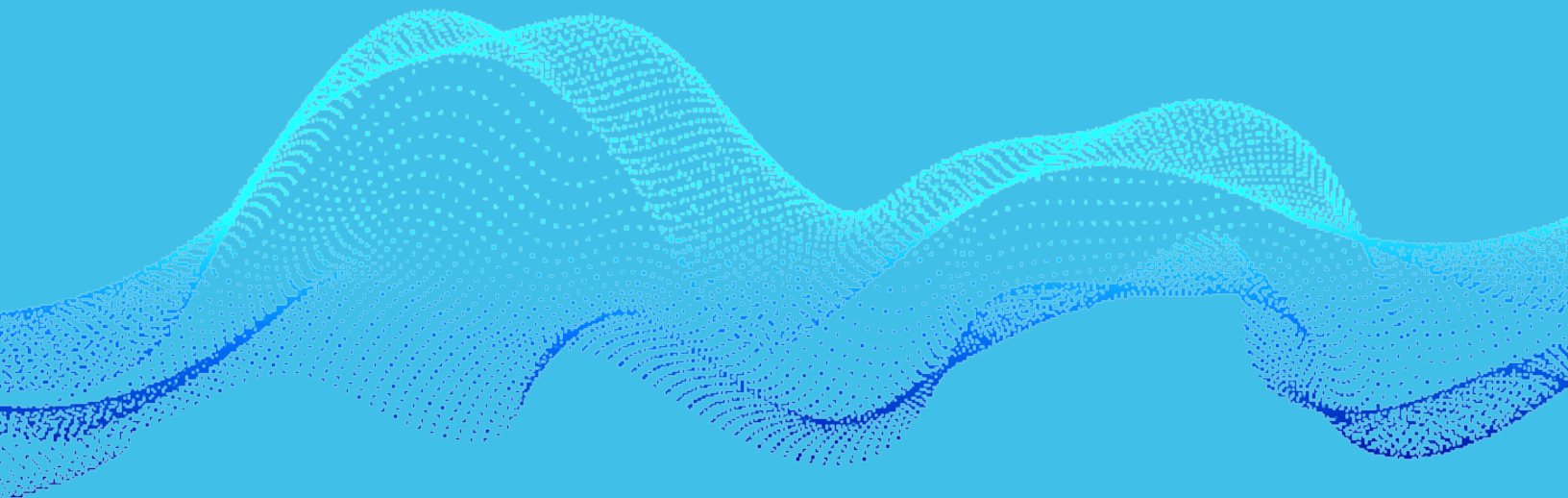
What is a Modern Data Architecture

A modern data architecture allows you to access all your data from across your organization's ecosystem in one place, while still decoupling it from the transactional source systems. This ensures data that is optimized for analytics (OLAP), removes manual steps, and delivers holistic insights, all while not impacting the performance of operational applications.

Our recommended modern data environment includes components in both a data lake and a more traditional relational data store, like a data warehouse or a data mart. It will also include a modern ETL/ELT tool to move and transform data between all zones of the environment.

Organizing these types of environments is a learned skill, and can be as much of an art as a science. Proper organization of your data lake hierarchical file structures is what keeps it from turning into the feared "data swamp", and the proper data model within a data warehouse requires knowledge of both the business and the intricacies of the data feeding it to blend and join data appropriately.

Data Sources	Data Lake Raw Zone	Data Lake Transformation Zone	Data Lake Curated Zone	Relational Data Store
On-prem transactional systems or cloud-based software-as-a-service platforms, we get your data from where it lives today.	A data lake allows us to store data in any format-so we start by storing them just as they are! These are copies of your source data stored in a file path, but it's now decoupled from the source system.	Here we begin to transform the source data from the original structure to the end goal of a standard data model. Don't combine data across sources yet-unless required for business logic or data transformations. Here we're focused on making everything look the same whiile still keeping our sources separate.	Now all of our sources have data in a common data model, so the Curated Zone is all about bringing the data from various sources together.	The published Curated data is put into a relational data store that is now ready for you to connect to reporting platforms and provide access to business users and analysts. Welcome to more seamless reporting, clearer answers, and new insights.

ETL Tooling: Extract, Transform, and Load. Get the data from where it is, to where it needs to be, and looking like it should.

Data Lake: Allows you to store unstructured data, or data that doesn't all follow the same model or schema. Data is stored as files, and we help set up the environment, define what those file storage paths should be and what format they should be stored in based on your use cases and goals.

Relational Data Stores: Traditional data storage that most are very accustomed to. Even with advanced tools and techniques, it's still critical to eventually understand your data's purpose and define its schema in order to be used by your business.

zirous®

Benefits of an Organized Data Lake

The primary benefit of a data lake is its ability to consolidate the storage of data from any source, regardless of the sources' data formats. Structured, semi-structured, and unstructured data can all be stored in very large quantities in a single location, making data access and insights more easily accessible to a wide range of end users. The scale at which data can be stored and processed within a data lake has opened the door to many forms of advanced analytics that were not possible before cloud computing and the optimizations that are possible thanks to a data lake's storage structure became readily available.

However, turning this vast expanse of data into usable insights for the business takes a lot of continuous refinement. Few users are equipped with the skills to leverage data in its raw form, so a properly organized data lake will have zones that sequentially transform data based on use cases and the intended audience of the dataset. Even advanced use cases that incorporate complexities like machine learning require extensive cleaning, transformations, and refinement into trusted datasets. This can be done within the data lake, but most business users will not want to have to heavily manipulate data in order to do their reporting.

The Necessity of a Relational Data Store

When it comes to supporting repeatable reporting and more self-service insights and analytics to the business, a relational data warehouse will help ensure data is being used consistently, correctly, and without the need for extensive manual manipulation of the data.

A data warehouse will store data within a predefined relational data schema, which means it stores structured data that is still optimized for analytics. As previously mentioned, not all source data in its raw form will be in a structured format, so the combination of a data lake with a data warehouse will allow for the necessary structure to be applied before it can be added to the relational data model.

A data warehouse is an ideal option when connecting business intelligence and reporting platforms for self-service reporting and analytics. However, the amount of time it can take to structure the data and define and build the data model can be prohibitive. The combination of a data warehouse and a data lake means you can begin to store data in the data lake even before you know the specific way in which the business will need to model and use it in the future. This hybrid approach gives any organization the ability to begin down this path.

zirous®

Steps to Designing a Data Architecture

1. Define Use Cases

Begin by defining multiple use cases that, if accomplished, would deliver value to the business. Consider current pain points groups face within your organization when trying to accomplish it today. And ensure that there is clear value that the users of the ideal solution would appreciate, but also that leadership would see as beneficial to the overall health of the organization.

Prioritize the use cases based on the quantitative (ideally, financial) impact it would provide, and qualitative impacts as well, such as the visibility it would have within your organization based on breadth of use, or known executive priorities, impending regulatory or migration deadlines, etc.

2. Start Small & Achieve Quick Wins

Once the top use cases are identified and prioritized, scope the effort to ensure a valuable deliverable is achieved in a realistic but short amount of time. When you work with Zirous, we always look for that to fall within 8 - 24 weeks. (Does that seem surprisingly short to you? We hear that a lot.)

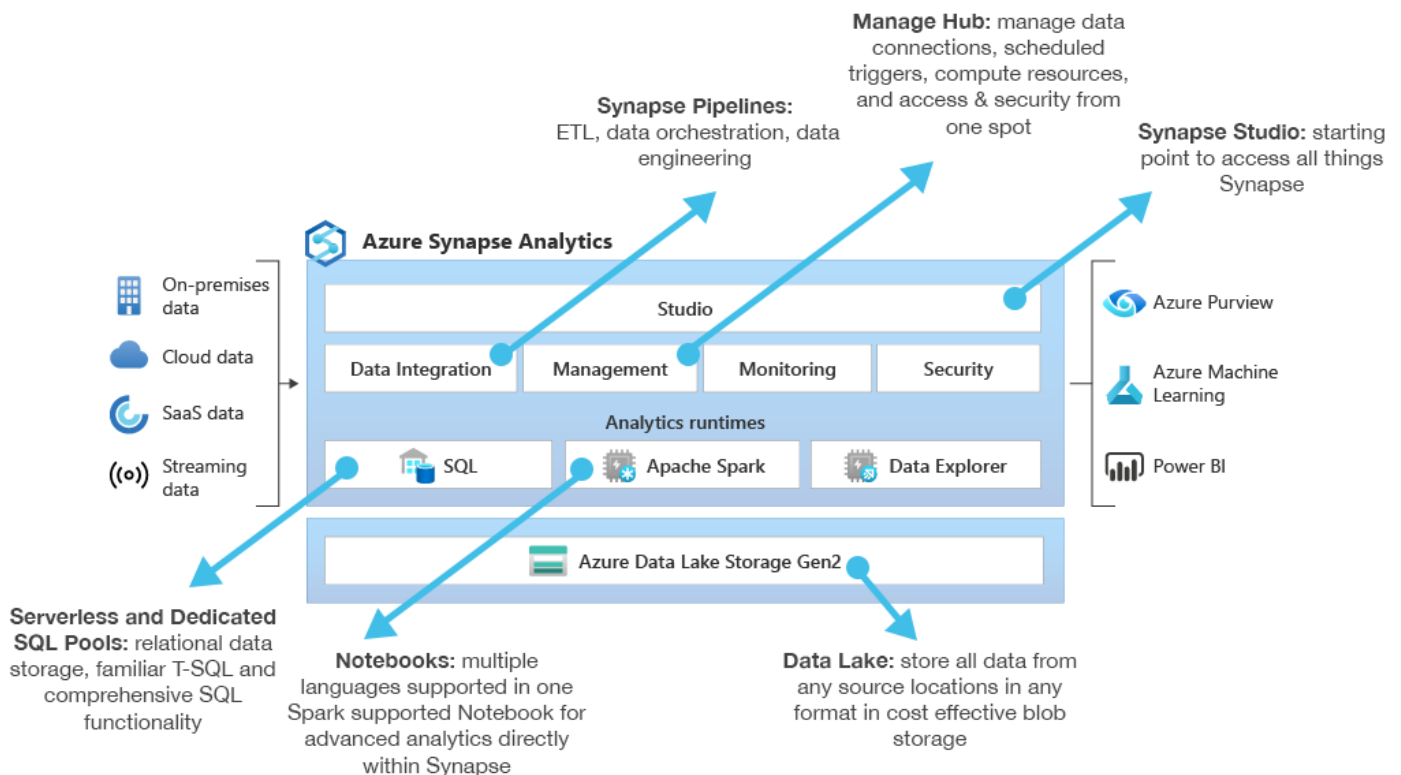## 3. Leverage an Experienced Partner

Working with a technology partner expands your team to include experts who have done this before. In addition to having a working understanding of best practices and how to avoid common roadblocks, the right technology partner will be able to perform certain steps through scripting and automation, saving you days or weeks on your timeline, which translates to monetary savings. A partner's guidance gives you the confidence that you're doing things the right way the first time. True partners will also provide substantial training, peer programming opportunities, and knowledge transfers to your team so you are able to maintain and expand the environment even after the partner is no longer directly involved.

zirous®

Selecting a Cloud Platform

When it comes to choosing a cloud platform to support you modern data architecture, it is important to first understand your organization's needs, goals, and priorities for what this environment is going to support and deliver. Some cloud platforms are more well-suited for enterprise, business-led adoption, while others are more tailored developer-driven teams. Organizations that have developer-led initiatives may prefer cloud platforms that provide a lot of tools that can be pieced together in a more customized manner to support their goals. Organizations that are looking for enterprise technologies should consider cloud platforms that offer more packaged solutions. Microsoft's Azure platform has packaged data lake, data warehouse, ETL/ELT integration, and more into one platform to allow businesses to more quickly adopt the technology.

Azure Synapse is an enterprise analytics service that accelerates time to insight across data warehouses and data lake systems. It is a unified analytics platform that brings together data integration, enterprise data warehousing, and big data analytics, and brings the divided worlds of data together with a unified experience to ingest, explore, prepare, transform, manage, and serve data for immediate BI and machine learning needs. It is designed to handle large-scale data volumes, all while separating the storage costs from the compute costs - so you can get the benefits of cost-effective cloud storage with the option to turn off compute instances when analytics is not being performed.

Azure Synapse is the best of SQL technologies used in enterprise data warehousing, Spark technologies used for big data, pipelines for data integration and ETL/ELT, and is deeply integrated with other Azure services such as Power BI, CosmosDB, Purview, and AzureML, meaning it will mature with you as your priorities and skill sets mature over time.

Manage Hub: manage data connections, scheduled triggers, compute resources, and access & security from one spot

Synapse Pipelines: ETL, data orchestration, data engineering

Synapse Studio: starting point to access all things Synapse

Azure Synapse Analytics

On-premises data

Cloud data

SaaS data

Streaming data

Studio

Data Integration | Management | Monitoring | Security

Analytics runtimes

SQL | Apache Spark | Data Explorer

Azure Data Lake Storage Gen2

Azure Purview

Azure Machine Learning

Power BI

Serverless and Dedicated SQL Pools: relational data storage, familiar T-SQL and comprehensive SQL functionality

Notebooks: multiple languages supported in one Spark supported Notebook for advanced analytics directly within Synapse

Data Lake: store all data from any source locations in any format in cost effective blob storage

Microsoft
Silver Partner

HOW TO GET STARTED

EVER FEEL LIKE THERE IS A HOLE IN YOUR DATA STRATEGY?

Modern data architectures leverage the strengths of data lakes and data warehouses to support more use cases and consolidate disparate data sources. We've turned our experience of setting customers up for data success into a free educational workshop so you can learn how to make a cohesive data environment incorporating data lakes, relational data stores, and modern ETL technologies. You'll learn:

- When and how to make a cohesive data environment incorporating data lakes, relational data stores, and modern ETL technologies
- How to effectively organize a data lake
- Data orchestration best practices with modern ELT/ETL technologies
- The benefits for reporting and support of advanced analytics

We know making any changes to your data environments is not an easy decision. That's why our experts will focus on equipping you with the knowledge to navigate the decisions you'll be faced with.

Learn more: www.zirous.com/donuts/

WHO IS ZIROUS?

For over 35 years, Zirous has served as an IT consulting firm working to improve our clients' business processes by leveraging cutting-edge technologies. We've been able to achieve this by engaging and retaining technologists who share our values of ingenuity, integrity, selflessness, and honesty.

Our experts become an extension of your team, and our clients tell us they appreciate how our staff matches their company cultures. With an average Zirous tenure of 9 years, we have the experience to lead you confidently through the projects we partner together on. And as a certified Microsoft partner, we have the know-how to design and deliver solutions, in Azure and on-prem, aligned with best practices and done right the first time.
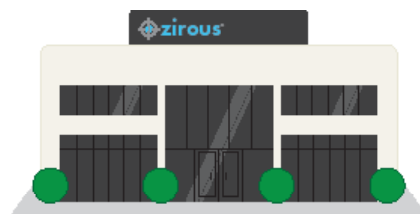
CONTACT US OR CONNECT

🌐 www.zirous.com/contact

📍 1503 42nd Street, West Des Moines, IA 50266

in www.linkedin.com/company/zirous/

📞 515-225-9015

◈ zirous®