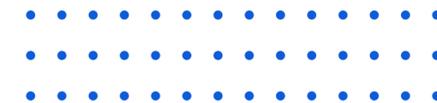
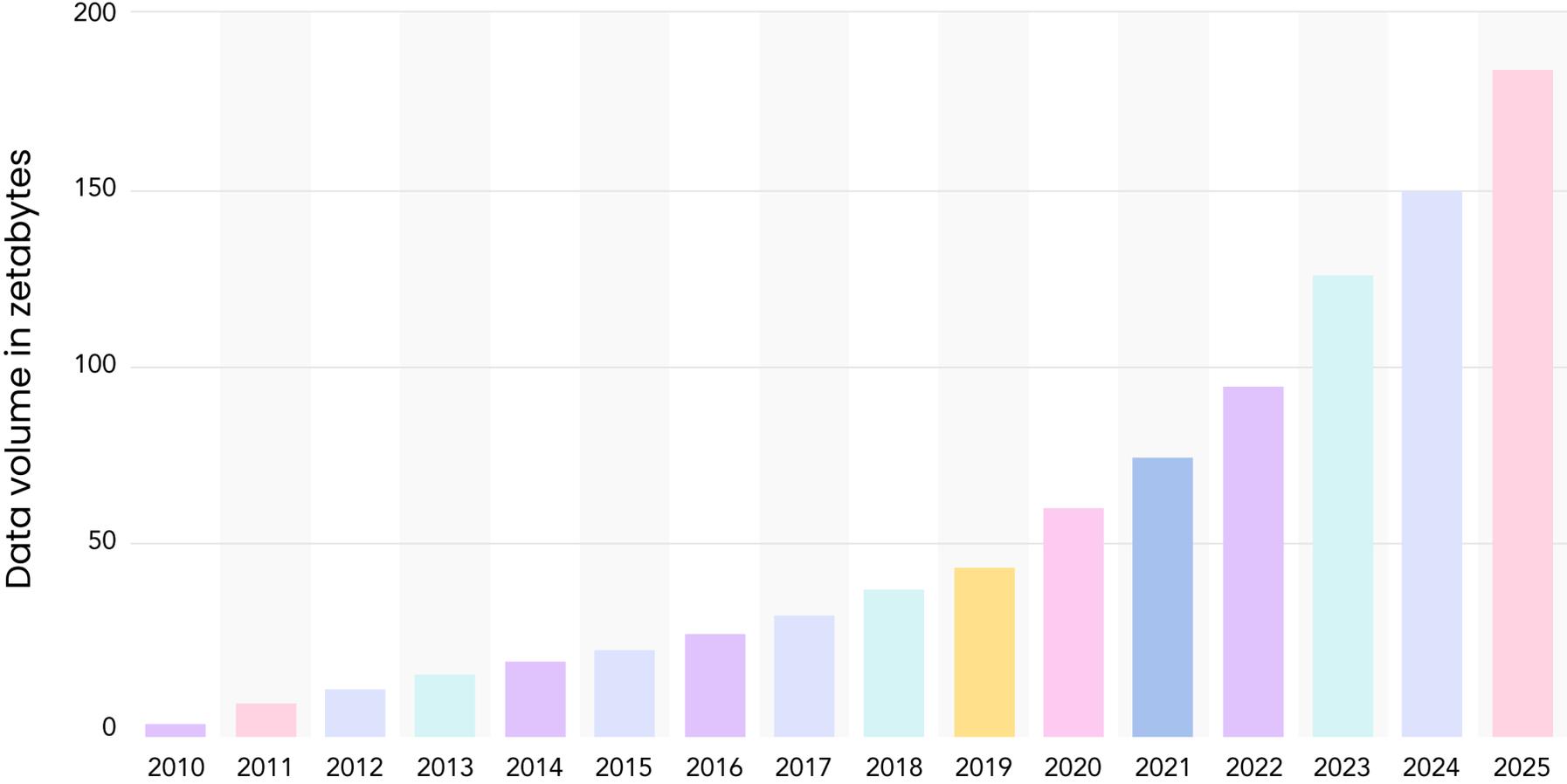


Informatica PowerCenter to Databricks Migration Guide



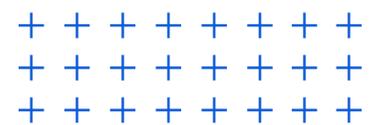


© Statista 2022

According to Statista, the total amount of data created, captured, copied, and consumed globally is increasing rapidly, reaching 64.2 zettabytes in 2022, and by 2025, it is projected to grow to more than 180 zettabytes.

In today's competitive world, there is an explosion of data from different sources. Enterprises are finding it extremely difficult to manage such vast quantities of data being generated from different silos systems in the organization. As the quantity and the complexity of the data are increasing with each passing day, the issue of managing it only worsens, widening the gap between the raw data and the valuable insights.

Getting a comprehensive cloud-based data platform for data analytics, scalable storage, data processing, and ML is crucial for businesses of all sizes. It is not enough to just have the right architecture to support your organization's data but having a solution to address and resolve all the data issues is imperative now.



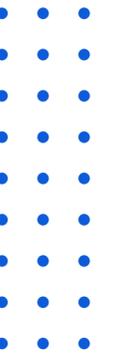
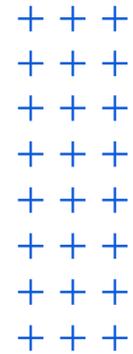
Looking into the future with data warehouse + data lakes – The Databricks Lakehouse

Databricks with its simple, open, and collaborative architecture has a lake-first approach. It stores all your data and helps you innovate faster and deliver ML-powered insights quickly. The following image will help you understand how customers use Databricks for their businesses.

When you are in Informatica PowerCenter, maintaining the data lakes, EDWs, BI, ML, Streaming, Security & compliance management, separately becomes a little complicated. However, migrating your data from Informatica to Databricks will pave way for staying ahead in the competitive edge as all of these can be done using the same platform.

Being a trusted Databricks partner, our team of data engineers ensures you a safe and smooth transition from Informatica to Databricks without any data leaks. We are concerned about the security of the data as much as you are, and we make sure that we migrate the data and workloads without compromising on security & compliance. We develop a clear-cut migration plan suitable for your business requirements. We love adding value to your business and thus we also work on error logging, data reconciliation, and data quality framework as add-ons to our services.





By migrating the workloads and data to Databricks you get,

- 2x faster data pipeline performance
- Reduced operational costs
- Simplified infrastructure management
- 10x more capacity to store business data
- Faster query time
- Matured DevOps services by integrating to GitHub and GitLab
- Improved team productivity
- Faster innovation with collaborative notebooks
- Centralized data analytics architecture

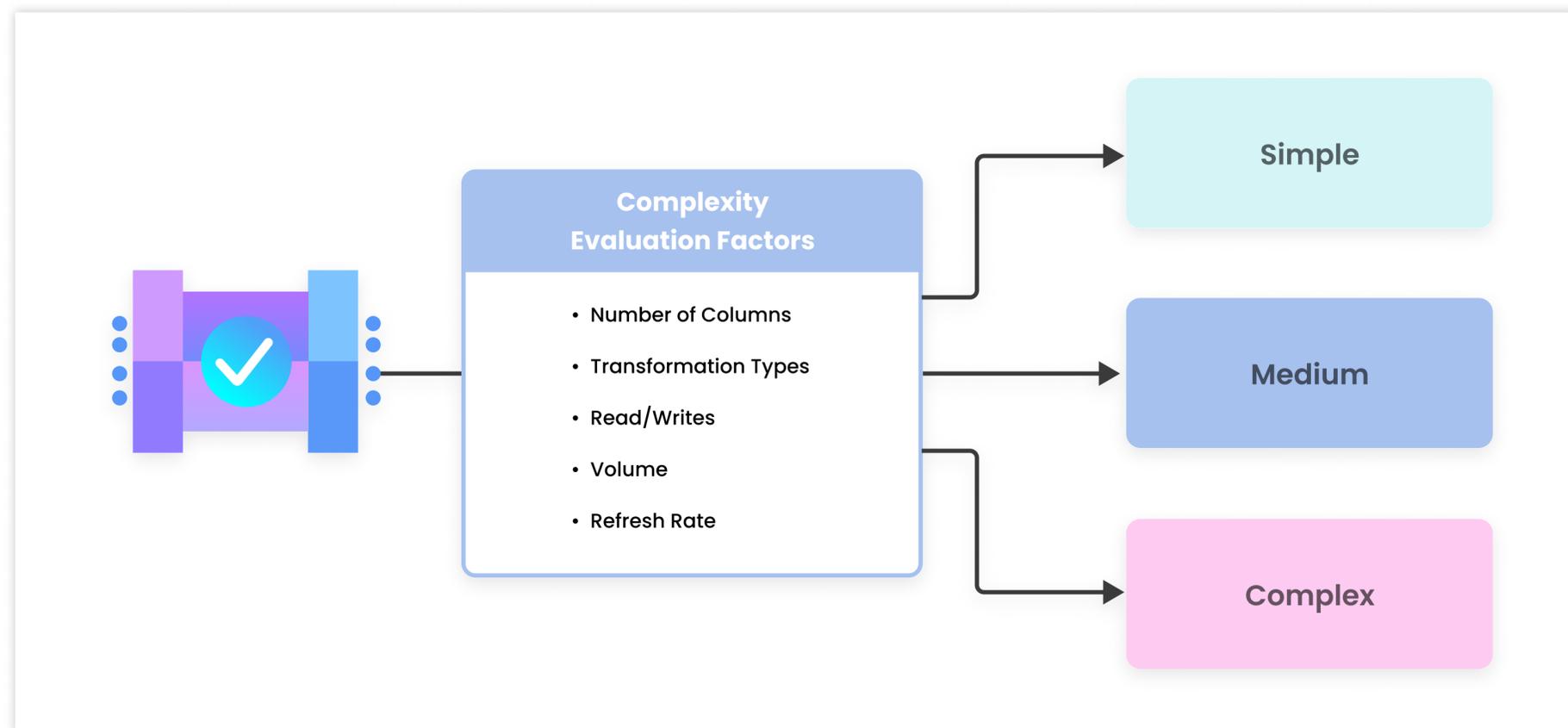
Let us now explore the step-by-step guide for the Informatica PowerCenter to Databricks migration. Here’s how we at AVASOFT, perform the migration and this is by far the best migration approach you will encounter.

Analysis	Rationalization	Migrate	Remediate	Cut Over
Assess your existing architecture	Optimize the data model and determine the approach for the initial load of data	Build the migration framework	Identify the gaps and issues	Go Live with complete Sync
Take Inventory of the existing pipelines, Orchestration and tables	Explore the data sync process, remove inconsistencies and nail down data sync process	Online and Offline historical data load transfer	Resolve them and fix the gaps	Ensure the applications are working perfectly in the target
Identify the pain points and determine the limitations of the existing architecture	Determine the Migration framework and curate the list of datasets to be migrated	Incremental data load transfer for keeping both existing and new architecture in sync	Identify the unsupported functionalities	Retire the previous database and sync
Classify the complexity of data pipelines and identify critical workloads	Develop separate lists for database programs for as-is migration and a list for rebuilding	Implement data quality framework using notebooks to validate generic edge cases and specific business logics	Ensure the target database is ready to use for the application	Complete Cutover to Target database
Propose future architecture with purpose-built database recommendations	Identify transformation patterns in the existing ETL pipeline and determine the future transformation pipeline	Parameterize the notebooks using metadata-based implementation		Post-production Support

Step 1: Analysis and Recommend

Define your requirements, explore your existing architecture, and initiate the migration process!

Migrating the workloads and data from one platform to another demands a lot of work. You cannot simply migrate whatever's there in your legacy environment. Vetting the legacy platform, exploring the in and out of the architecture, and assessing the datasets is very important.



Drafting the architectural outline:

Before we begin the migration, we assess your existing environment. Identifying the limitations in the current architecture and determining the drawbacks helps us understand the roadblocks and prepare for them better. We build the database inventory to ensure that we have covered everything from your existing database.

Our team works on analyzing all the source systems, and different formats of data generated from each source, and list them. We analyze the complexities of the data pipeline in the existing environment and then map the pipelines based on their complexity.

No matter what complexity your existing environment has, we can easily assess your legacy environment, inspect the development and deployment practices, and jot down the complexities. With our in-depth expertise, we develop an architectural diagram and include details that require reengineering work.

Suggestions for future-state architecture

Based on the existing architecture issues, the historical data and workloads, we recommend and propose a future state architecture that perfectly fits the business requirement and allows flexibility for ML use cases. We also provide an estimation of the cost structure for the environment and ETL pipelines along with the project deadlines.



Our team provides you with an estimated cost structure to help you get clarity on the value of your investments.

Step 2: Reorganize and Rationalization

Streamlining the process and determining best-fit strategies for a seamless migration are essential!

As an organization, you will have unique business needs. Our team of experts will collaborate with your key stakeholders to identify the crucial datasets and develop a comprehensive migration approach for the initial load, tailor-made for your business needs.

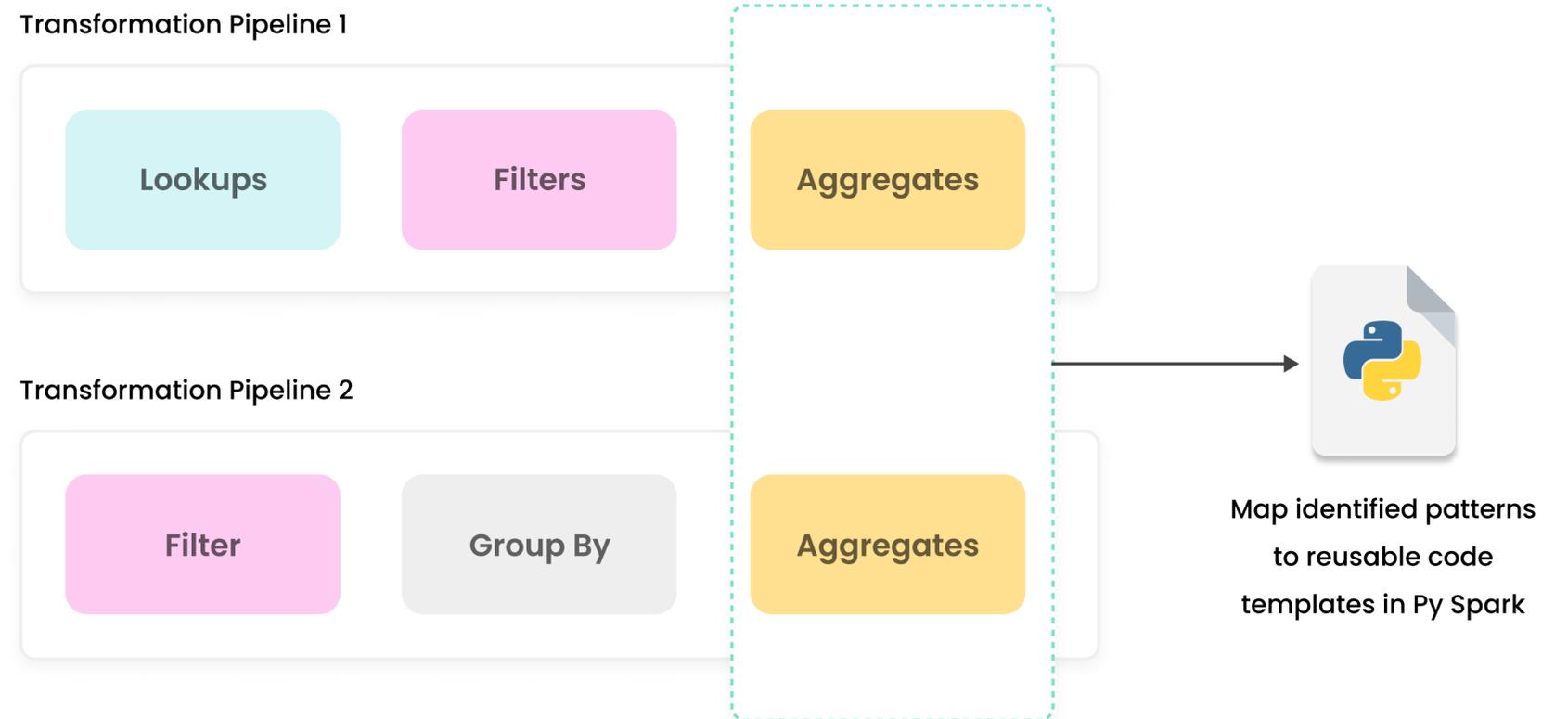
Optimize the data model to gain cohesion

We analyze the data models and remove the redundant and inconsistent ones. Then, we normalize the data tables without losing any information and eliminate inaccuracies. To ensure quality, our team cross-examines the business-critical metrics and identifies the gaps.

Determine the future transformation pipeline

We analyze the business and transformation logic in the existing ETL pipelines, and based on the requirement and your organizational needs, we suggest the future transformation pattern and finalize it. We put our effort to map the relevant PySpark code templates in the current ETL pipeline.

We develop a list of datasets to be migrated as such, and a list for modification and rebuilding of database programs.



Step 3: Evaluate and Migrate

Establishing the right migration plan and adopting a feasible approach is crucial for a successful migration!

Error logging and reusable code templates

We create Generic Databricks notebooks that support multiple use cases (Extraction, Transformation, Machine Learning, and Orchestration) to form the required table structure and perform error logging to detect and record the error logs. Further, these reusable code templates developed using the mapped tables will help your organization immensely to use the codes as and when needed.

Historical data load

With our data sync strategy, we validate and sync the data from the existing data warehouse, without any duplications or errors. Based on the amount of data and priorities in your existing environment, and the firewall environments, we select the approach for migrating the historical data.

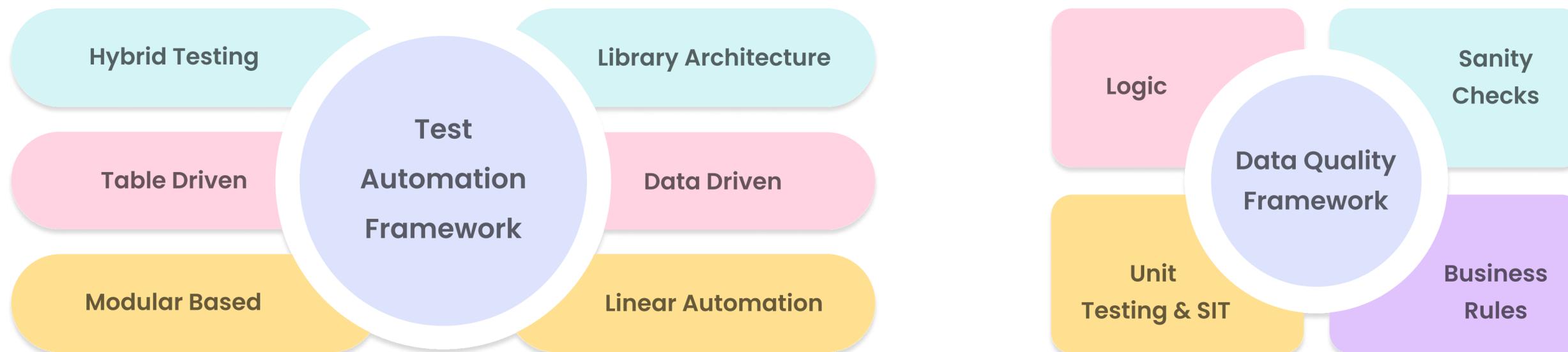
Push approach - Traditional approach for making changes or establishing the changes

Pull approach - For cloud-based scenarios to manage the workflows

Once the historical data load transfer is done, we move the incremental data and sync it.

Establish the data quality framework

We are dedicated towards ensuring a standard quality framework that will enable high-quality code development with consistency and accuracy throughout the process. Hence, we design a Data Flow diagram in line with your business requirement.



Further, we perform Unit testing and data validation test cases to validate the pipelines and identify the gaps. If there are any issues, we review them and fix them immediately.

Step 4: Validate and Remediate

Our team helps you sort and remediate all the issues in terms of data and orchestration. We also identify the issues pertaining to the different source systems and mitigate the differences and deviations in the migration. We create appropriate strategies to fix the migration issue if any.

Prepare the target database for the applications

To ensure everything falls in the right place, we ensure that the Target database is up and running and supports all the functionalities of the application.

Prepare the downstream applications for the target database

To provide you with a smooth experience without any hurdles, we work on applications that have a dependency on the existing database. We optimize and configure it to make it easily adaptable in the Target database without any issues.



Step 5: Cutover

The final and the most crucial step of all!

We orchestrate this critical event by selecting the most suitable approach for your organization with adequate time to wrap up the process. Our team brainstorms with the executive team and plans for the cutover. We then finally perform the cutover either by using the big bang migration or trickle migration depending upon the business need.

Big Bang migration – Complete data migration with planned downtime for live systems

Trickle migration – The systems are synced in phases with parallel runtime of both the legacy and the modern cloud data platforms I.e. Informatica and Databricks respectively.

Before we Go Live and Sync the data, we check the application’s functionality and ensure that it’s fully operational. We Sync the data and verify the database for any issues or gaps. This way, you will have no issues in your Databricks platform.

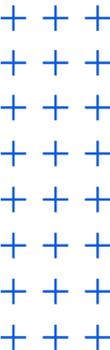
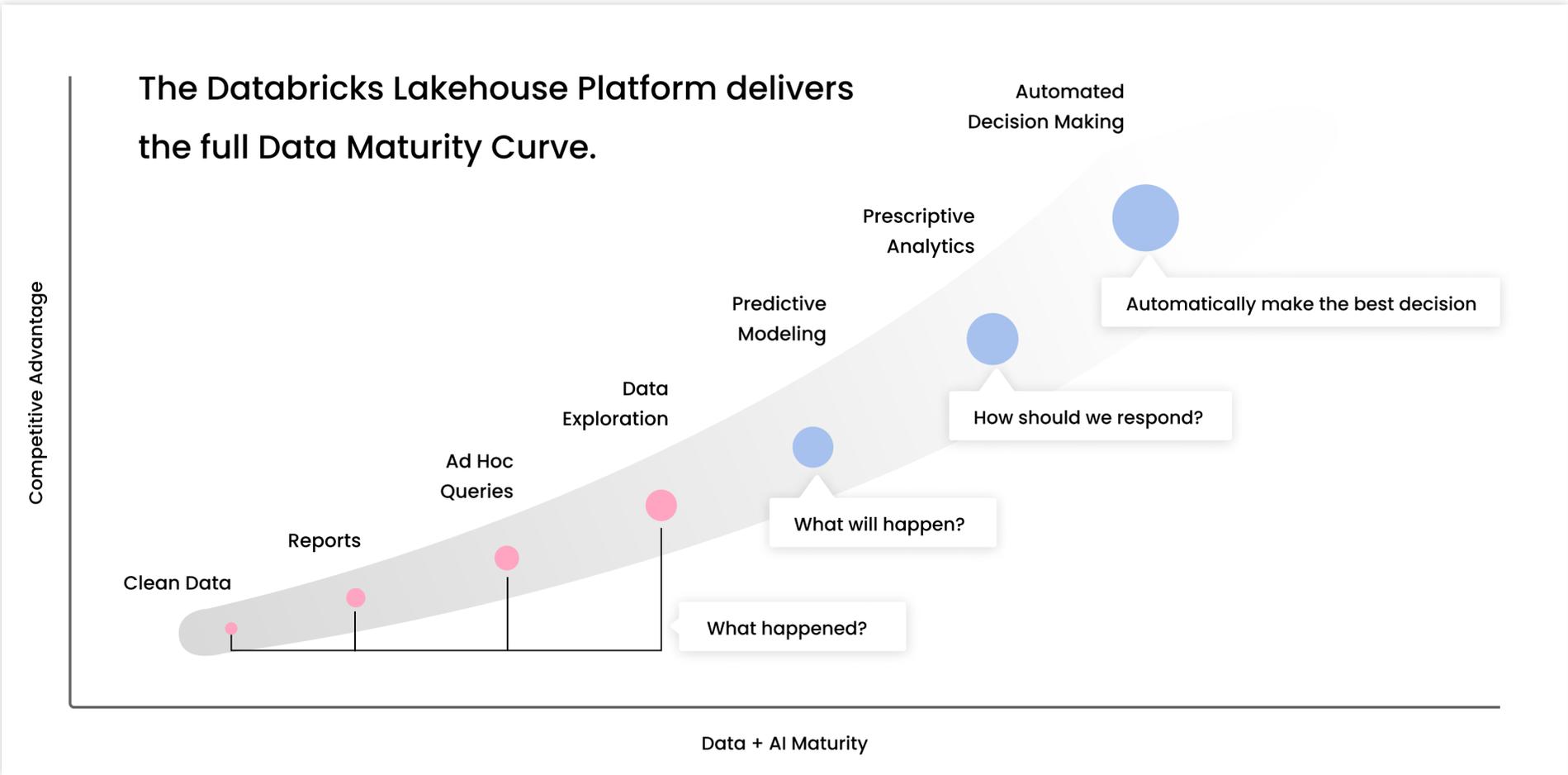
Post-production Support

Walk with you until the final walkthrough!

We provide multiple KT sessions and ensure that you have all the details at your fingertips. Our team walks with you throughout the process of the migration journey until the walkthrough. We understand that you'll have numerous concerns regarding the new platform after migration. Hence, we provide on-call support to address your ad hoc data and orchestration requests. We ensure that you have a safe and seamless migration experience with us.

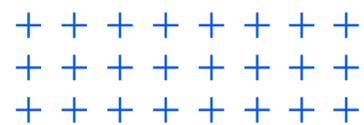
A walkthrough before release

After completing the process, we provide you with a walkthrough of the code, provide insights on the developments and approaches and finally get the confirmation before moving the release to UAT or PROD. Our team uses the most recommended blue/green deployment in a production environment.



Organizations relying on the legacy and expensive Informatica PowerCenter will only fail to leverage the future opportunities of big data and data analytics. Migrating from Informatica Power Center to Databricks is the most sensible decision that any organization can make right now.

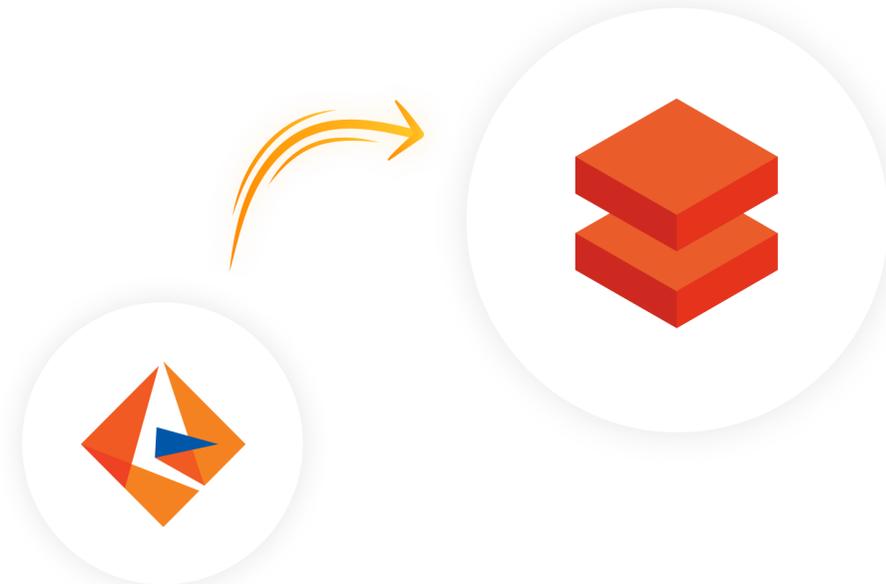
Data Migration is a time-consuming and challenging process, and our team of data engineers, data scientists, ML experts – data science specialists will make the migration process seamless for you. Being the official Databricks partner, we have gained extensive expertise in securely migrating the data to the data lakehouse without causing any business disruptions.



AVASOFT makes your Informatica PowerCenter to Databricks migration process seamless, and more efficient. Want to migrate and learn more about the process? What are you waiting for? Talk to us TODAY!



Databricks – with its fast, simple, scalable ways to build a data warehouse, eliminating the need to invest in the expensive ETL pipelines revolutionizes the way data sets are analyzed and stored. The interactive UI, secure data management, multi-user support, cluster sharing, and job scheduling make Databricks the most preferred of all.



Want to know more?

Talk to our Experts now!

Kumar: +1 732 737 9188

sales@avasoft.com

www.avasoft.com