

**alteryx**  
COOKBOOK SERIES

# 6 Steps to Faster Data Blending Using Your Data Warehouse



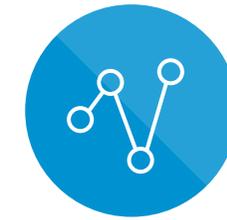
# Self-Service Data Blending and Analytics

Dynamic market conditions require companies to be agile and decision making to be quick—meaning the days of waiting for a centralized IT staff or data scientists to prepare data for insights are far gone.

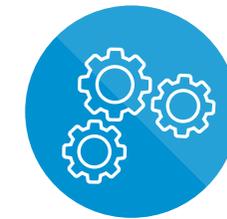
Self-service data blending allows line-of-business analysts to access, cleanse, and join data quickly and easily, and deliver insights faster.

## With self-service data blending and analytics, line-of-business analysts can:

- Connect to all data where insight may reside regardless of location—spreadsheets, local databases, corporate databases, cloud applications, social media, and more
- Prepare data by removing redundant or unnecessary data, and apply rules to fill in data that is missing or incorrect
- Blend multiple data sources quickly and easily without coding or IT involvement
- Deliver faster and deeper insights such as predictive and spatial analysis
- Share the results easily as static reports, or for data discovery in visualization software such as Tableau or Qlik



Connect to all the data types and locations where insight may reside



Blend your data without having to rely on IT, SQL coding or data scientists



Perform advanced analytics and share the results with business decision makers

# What is in-database data blending?

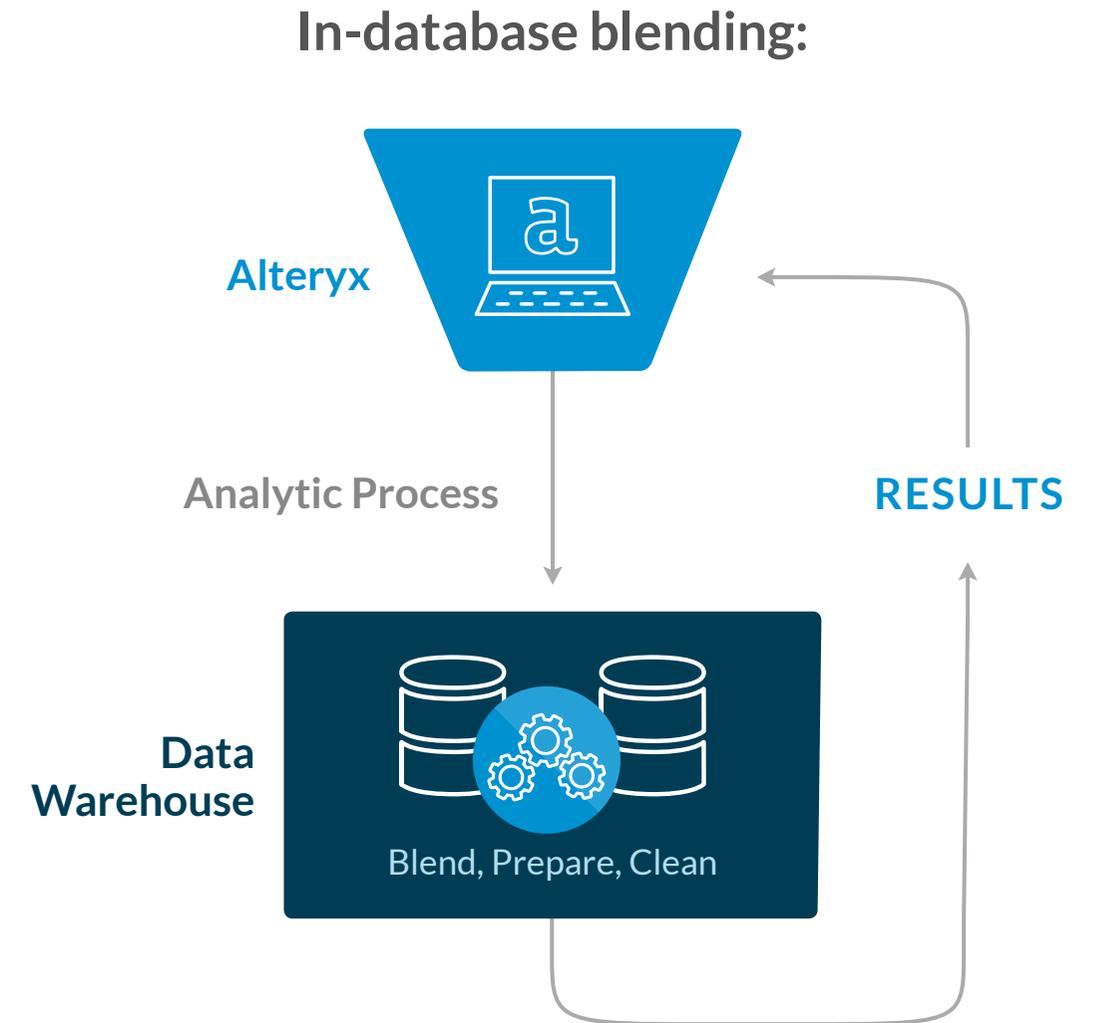
With the emergence of data from social media, cloud applications and sensors, analysts have a higher volume of data to deal with than ever before, making Big Data even more unwieldy.

When dealing with extremely large datasets, it's best to limit movement of the data to mitigate network bottlenecks and processing latency. After all, why move a massive data file from a data warehouse to your desktop when all you want are a few pieces of data from it?

In-database data blending allows you to push the processing steps into the database and retrieve **only** the data you need, rather than pulling the entire dataset to the processing location (typically, your desktop).

## With in-database data blending and analytics you can:

- Avoid the need for IT or specialized analytics staff to write SQL code or other query languages for blending and preparing data
- Leverage existing Big Data platform investments and ecosystems
- Utilize the in-database processing power of the data warehouse to answer new questions faster



# Recipe for In-Database Data Blending

- 1 Identify all of your data sources
- 2 Determine the best place to work with your data
- 3 Connect to your data
- 4 Cleanse, filter and transform your data
- 5 Join data from multiple sources
- 6 Stream data from the data warehouse

## Ingredients you need

- A copy of Alteryx  
[alteryx.com/download](https://alteryx.com/download)
- A list of the data sources you want to blend
- Access to each database or source you are going to use
- A rough specification of the dataset you need
- A clear understanding of the analysis you want to deliver

See demo videos of in-database blending at [alteryx.com/solutions/in-database-processing](https://alteryx.com/solutions/in-database-processing)

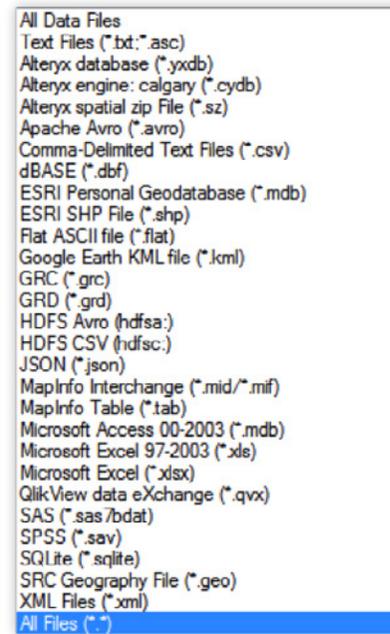
# 1

## Identify all of your data sources

An ever-increasing volume and variety of data is being stored in a range of locations—local hard drives, data warehouses, and the cloud—making it difficult to aggregate data in one location.

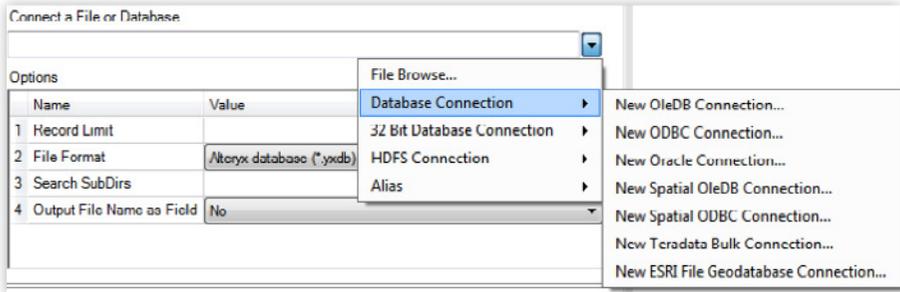
Alteryx gives you the ability to access and blend data types from any source, without size limitations, and all within a single view:

- Bring in data from spreadsheets, databases, and other common file formats
- Access data from corporate data warehouses, third-party data providers, and cloud-based storage such as Amazon S3 or Redshift
- Incorporate cloud-based data from Salesforce, Marketo, Google Analytics and more
- Connect to social media feeds such as Twitter and Foursquare to include customer sentiment



**Tip:**

- Work with IT to ensure you have the right credentials to access your data sources and databases.



Twitter tool



Salesforce Input tool



Marketo Input tool



Amazon S3 Download tool



Google Analytics tool

2

## Determine the best place to work with your data

When working with data stored in data warehouses such as Oracle, SQL Server, Amazon Redshift, Cloudera Impala, Spark and Teradata, you have two options:

1. Move an entire dataset from the data warehouse into Alteryx for blending with other data, or
2. Push the processing from Alteryx into the data warehouse to segregate just the data you need.

When working with extremely large datasets, you will find significantly improved performance with the second option because you're limiting movement of vast amounts of data.

The intuitive Alteryx interface allows you to quickly connect to your data at the source. In addition, the flexibility of Alteryx provides bi-directional functionality, enabling you to easily push data into the database, or pull data out.

### Tip:

- If you're not working with large datasets, you can easily stream data, or a data table, out of a database with the *Data Stream Out* tool.



# 3

## Connect to your data

The intuitive Alteryx interface allows you to quickly connect to your data regardless of location, size or format. When working with large datasets, the in-database tools in Alteryx make it easy to connect to the data you need, or easily pull a subset of that data.

- Use the **Connect In-DB** tool to establish a connection directly inside the data warehouse.
- If you are working with very large datasets and don't need all entries, use the **Sample In-DB** tool to limit the amount of data records, optimizing runtime and throughput.
- Use the **Input** tool to connect to other external datasets and use the **Data Stream In** tool to push the data into the database.



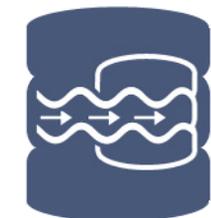
**Connect In-DB tool:**  
Connect to a variety of data directly in databases such as AWS, Cloudera, Oracle, and SQL Server



**Sample In-DB tool:**  
Work with a subset of data within your database if you don't need all of it



**Input tool**



**Data Stream In tool**

Push external datasets into your database to blend and analyze data in one location

# 4

## Cleanse, filter and transform your data

Alteryx makes it easy to cleanse and filter large volumes of data by pushing the data preparation steps to where large datasets reside.

- The **Filter In-DB** tool enables you to query records and return those records that meet the specified criteria, such as location, brand, or product SKU, or to filter out null values.
- The **Select In-DB** tool allows you to select the fields that contain the information you need, plus rename and re-sequence fields, modify data types, and add field descriptions.
- The **Formula In-DB** tool, a powerful processor of data, allows you to perform a broad variety of calculations and/or operations to create new data fields or update existing fields.
- The **Summarize In-DB** tool can group, sum, count distinct fields, and more.



**Filter In-DB tool:**  
Filter the data by a specific product, brand or service



**Select In-DB tool:**  
Eliminate unnecessary fields, or rename key fields



**Formula In-DB tool:**  
Understand the likelihood of a purchase



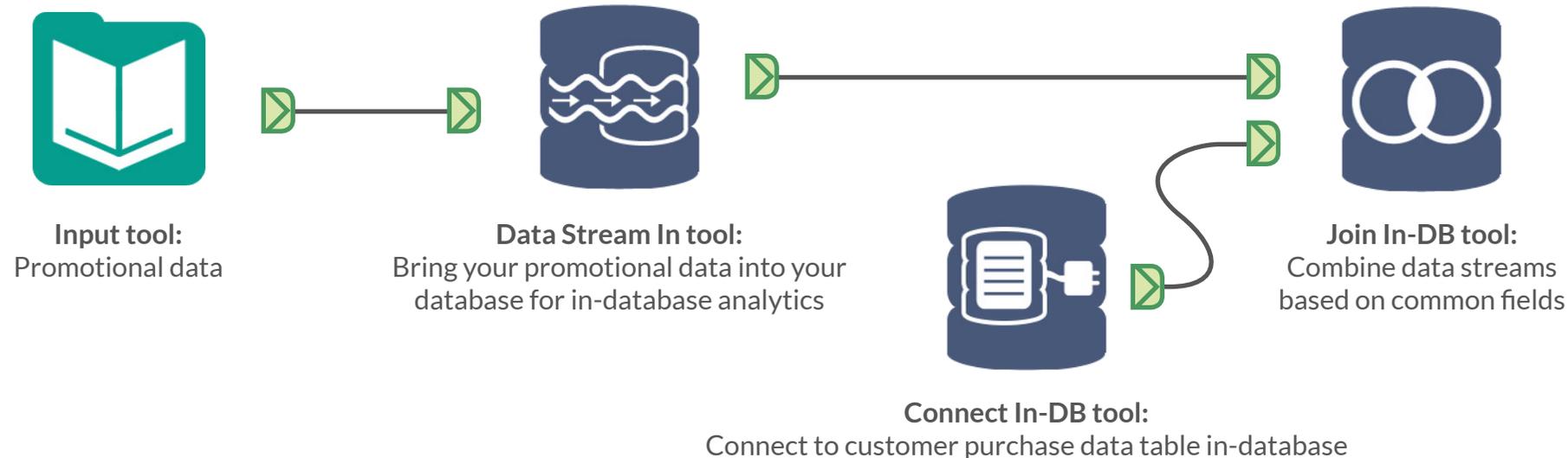
**Summarize In-DB tool:**  
Summarize the data by key groups such as “country”

# 5

## Join data from multiple sources

Alteryx provides analysts with an intuitive workflow for data blending that leads to deeper insights within hours, not the weeks typical of traditional approaches. This is all done in a single workflow, with no programming required.

- Use the **Data Stream In** tool to import external files into your data warehouse for in-database blending and analysis.
- Use the **Join In-DB** tool to combine datasets based on common fields or record position. In the joined output, each row contains the data from both inputs.



### Tip:

The *Join in-DB* and *Union In-DB* tools merge datasets differently.

- The *Join In-DB* tool combines data streams based on common fields.
- The *Union In-DB* tool combines each stream of data into a single stream and can be used when bringing in multiple data streams.

# 6

## Stream data from the data warehouse

Once you've joined and prepared your data, Alteryx makes it easy to stream it from the data warehouse to feed your downstream analytic workflow.

Once it is in Alteryx it acts just like any other data source, and allows you to perform typical data blending and advanced analytics functions in Alteryx, including predictive analysis, spatial analysis or exporting for visualization in Qlik or Tableau.

- Stream data from your data warehouse into Alteryx using the **Data Stream Out** tool.
- Limit the replication of datasets by using the **Write In-DB** tool to create or update a table directly in the database for future use.



**Data Stream Out tool:**  
Use the Data Stream Out tool to feed downstream analytic process



**Write In-DB tool:**  
Use the In-DB stream tool to create or update a table directly in the database

### Tip:

- Use Alteryx Server to schedule updates to the files, ensuring analysts are working with up-to-date transactional information.
- Output data in a wide variety of flat file and relational database formats.

# Alteryx Supports Ambit Energy's Customer Strategy with Faster Data Blending and Predictive Insights



Ambit Energy uses Alteryx to quickly and easily blend large datasets within a database, and build predictive models to improve customer engagement, all without IT or data scientist involvement.

## Deeper Insights

Accessed, cleansed, and appended millions of customer records, product combinations, and all variable attributes to better understand existing customer profiles, and predict future behaviors and probability of attrition.

## Hours vs. Weeks

Reduced time to prepare customer behavior data and create predictive insights in hours instead of weeks by pushing data blending processes down into the database and leveraging the predictive tools in Alteryx, resulting in a flexible and nimble analytic team that delivers insights faster.

## Intuitive Workflow

Created a collaborative analytic workflow that allowed the entire analytic logic to be easily understood by the line-of-business users, removing the black box of analytics and ensuring that the delivered analytics provided actionable insights to improve customer engagement.

**“The simple drag-and-drop interface of Alteryx empowers my team to perform in-database data blending and build predictive models without requiring them to spend hours coding, enabling us to deliver faster and deeper insights that impact our bottom line.”**

**– Lloyd Tokerud  
Director of Analytics at Ambit Energy**

# Why should you use Alteryx for data blending and analytics?

An ever-increasing volume and variety of data is being stored by analysts in a range of locations—local hard drives, data warehouses, and the cloud—making it difficult to aggregate data in one location. But, it doesn't have to be.

## With Alteryx, you can:



Access and blend data from wherever insight resides: Excel spreadsheets, corporate databases, cloud-based applications, and much more



Blend multiple data sources quickly and easily using an intuitive workflow that doesn't require IT or coding



Deliver faster and deeper insights through advanced analytics such as predictive and spatial analysis

**“Our analysts are doing less low-value work and more high-value, satisfying work, which keeps them energized and produces a greater return on investment of time for the company.”**

**– Tom Sturgeon**

**Director of Business Analytics, IT Business,  
Schneider Electric US**

# Alteryx Delivers on the Three Things Analysts Need Most



Allows them to **access all the data they need**, when they need it, and analyze it in the optimal manner

Gives them a **single intuitive workflow** for a complete data blending & advanced analytics process

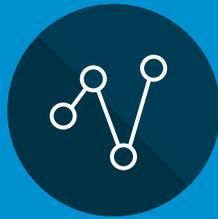


Delivers **deeper business insight** without relying on others for spatial or predictive analysis

# Next Steps



Learn more about in-database data blending and analytics in Alteryx  
[alteryx.com/solutions/in-database-processing](https://alteryx.com/solutions/in-database-processing)



Try in-database data blending and analytics in Alteryx  
[alteryx.com/download](https://alteryx.com/download)



View Customer videos  
[alteryx.com/customers](https://alteryx.com/customers)



# alteryx

## 6 Steps to Faster Data Blending Your Data Warehouse

---

Thousands of data analysts worldwide rely on Alteryx daily.  
[alteryx.com/solutions/in-database-processing](https://alteryx.com/solutions/in-database-processing)