



Delivering Value. Scientifically.

Instructor Lead Training

Brief: This was a customized ILT session, supported by handouts and user guides, to equip both instructors and learners.

Tools & Technologies:



Acrobat PDF



Microsoft PPT

Alteryx ETL for Amazon Redshift

This presentation will showcase the Alteryx ETL tool utilize AWS Redshift to acquire, extract, transform and build reporting tables which will be consumed by Tableau.



<https://www.youtube.com/watch?v=57qwPqubW78>

<https://help.alteryx.com/current/index.htm>



https://www.youtube.com/watch?v=UhQjSzdIO_g

<http://docs.aws.amazon.com/redshift/latest/mgmt/welcome.html>

Demonstration Context and Objectives

This demonstration intends to showcase:



Data acquisition from custom and third party API based sources.



Various ETL tools to create, manage, and execute different data management flows.



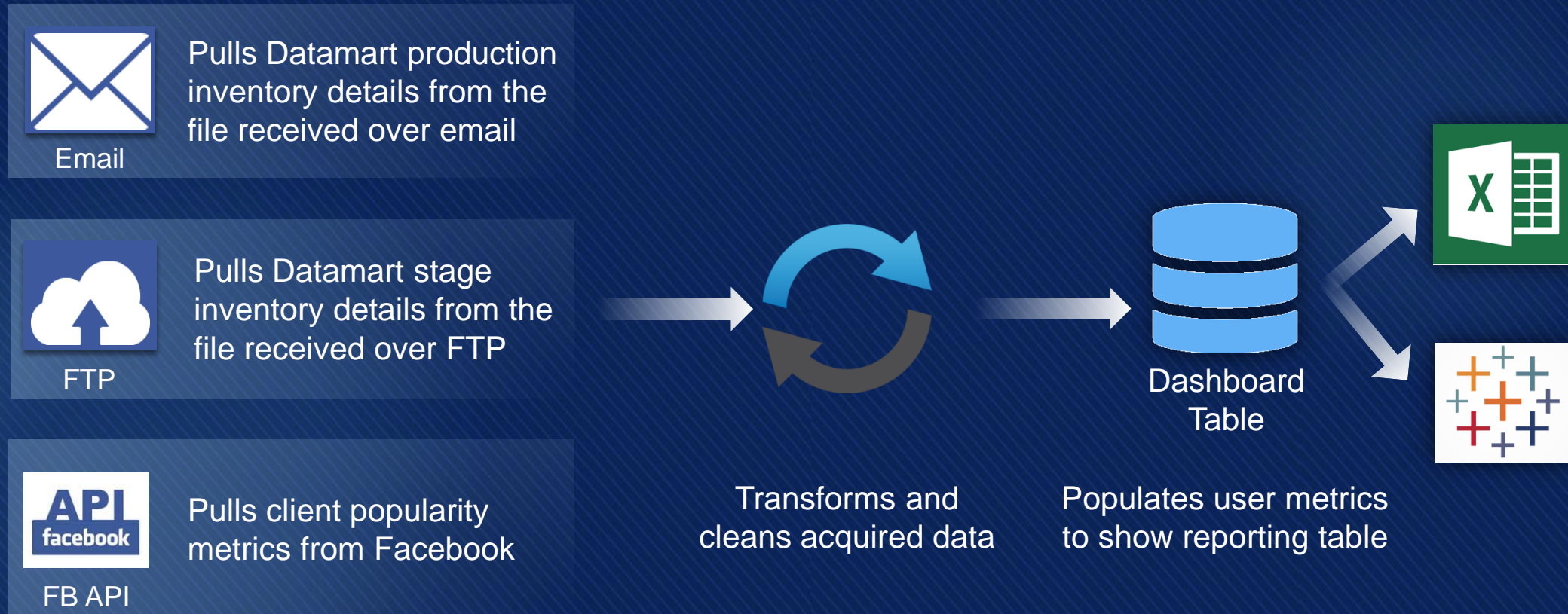
Storage repositories on the AWS infrastructure.



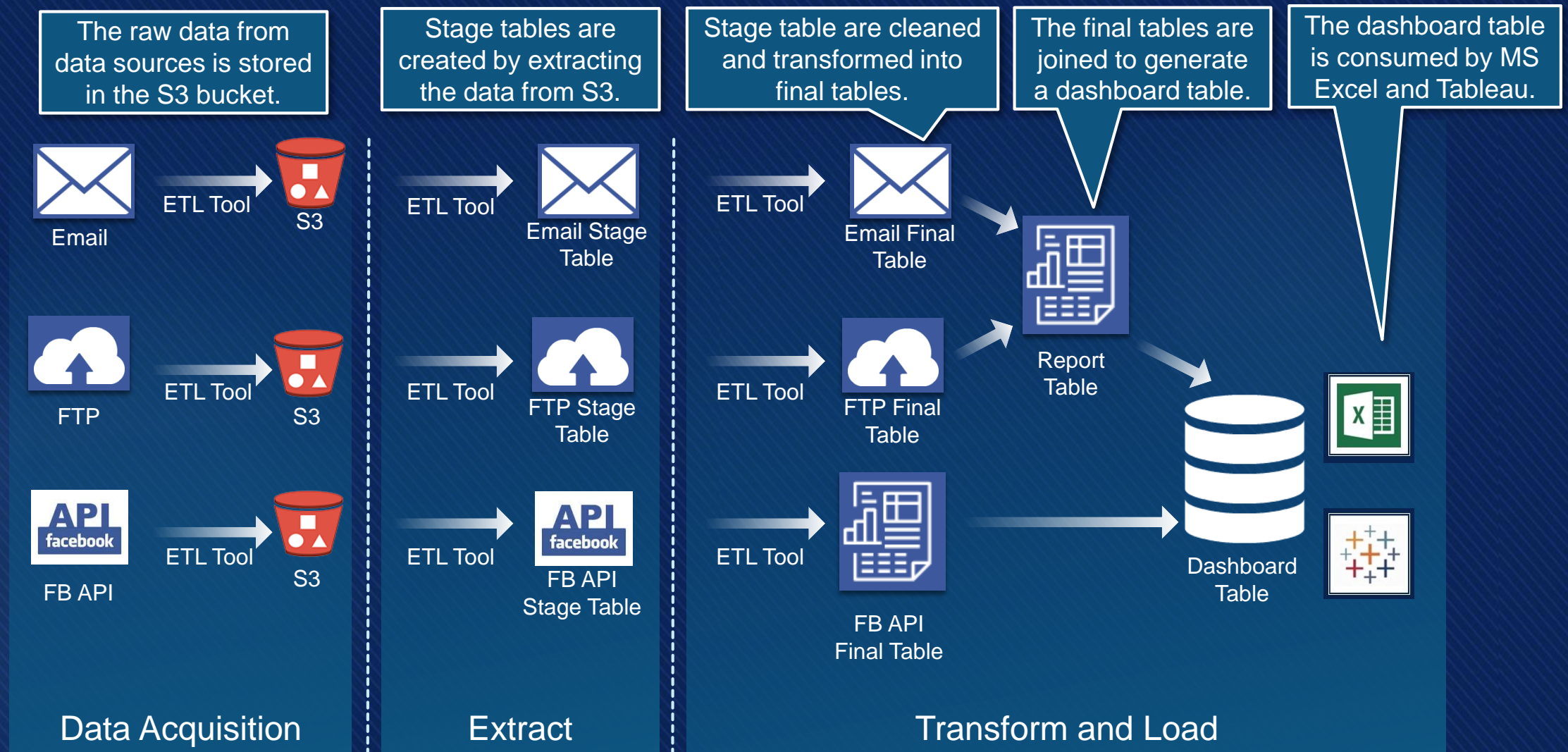
Ad-hoc reporting and data visualization.

Use Case

We choose a simple use case to demonstrate this solution:



Data Flow



Alteryx Product Lines

Alteryx is a licensed tool that is installed on Windows OS. It comprises of four major product lines:

- Alteryx Designer - helps you design your workflows
- Desktop Automation - helps you create and control scheduling from a desktop
- Alteryx Server - helps you schedule the workflows in a timely manner from a server
- Alteryx Analytics Gallery - helps you run and view the created workflows

Annual pricing is based on subscription model: 1 year and 3 year

Sr. No	Description	1 Year Price (\$)	3 Year Price (\$)
1	Alteryx Designer	5195	3995
2	Desktop Automation (add-on)	6500	5000
3	Alteryx Server *	58500	45000
4	Alteryx Analytics Gallery	1950	1500

* Alteryx Server is also available on AWS marketplace as an EC2 AMI -
Prices starting from \$8.02/hr. or \$58,500.00/yr. + AWS usage fees

P.S. Additional details about Alteryx is available in the Overview presentation.

Alteryx Designer Interface

The screenshot displays the Alteryx Designer interface. At the top is a menu bar (File, Edit, View, Options, Help) and a toolbar with icons for various actions. Below this is the **Tool Palette**, which is a horizontal bar containing icons for different tool categories: Favorites, In/Out, Preparation, Join, Parse, Transform, In-Database, Reporting, Documentation, Connectors, Developer, and CReW Macros. Each category has a sub-menu of specific tools, such as Browse, Input Data, Output Data, Text Input, Data Cleansing, Filter, Formula, Sample, Select, Sort, Join, Union, Text To Columns, Summarize, and Comment.

On the left side, the **Workflow - Configuration** window is open, showing the 'Canvas' tab. It contains 'Canvas Options' with settings for Layout Direction (Horizontal), Annotations (Show), and Connection Progress (Show Only When Running).

At the bottom right, the **Results - Workflow - Messages** window is open, showing a summary of workflow execution: 0 Errors, 0 Conv Errors, 0 Warnings, 0 Messages, and 0 Files. It also has buttons for 'Last Run' and 'Configuration'.

Three callout boxes provide additional context:

- Tool Palette:** Alteryx has an intuitive **Tool Palette**. You can select a tool to begin building a workflow.
- Configuration window:** The **Configuration** window displays the settings available for a selected workflow, tool, or connection.
- Results window:** The **Results** window displays the status, message, or errors that might occur while the workflow is running.

Master Workflow

The screenshot displays the Alteryx Master Workflow interface. The top toolbar includes various tools categorized under Favorites, In/Out, Preparation, Join, Parse, Transform, In-Database, Reporting, Documentation, Connectors, Developer, and CReW Macros. The main workspace shows a sequence of four sub-workflows: FTP Work Flow (FTP 1.1 to 1.4), FB API Work Flow (FB API 1.9 to 1.12), EMAIL Work Flow (Email 1.5 to 1.8), and Final Dashboard Work Flow (Temp DB Max 1.13 and Final Dashboard 1.14). A text box explains that CReW Macros are used for features not in-built in Alteryx. The bottom panel shows the Results - Workflow - Messages section with 0 Errors, 0 Conv Errors, 0 Warnings, and 0 Messages. A text box states that all workflows are sequentially executed using the Runner tool.

1.15 Master_Workflow.yxmd

CReW Macros is a set of tools provided by the Alteryx Community. In this POC, it is used, because a few required features are not in-built in Alteryx.

FTP Work Flow

FB API Work Flow

EMAIL Work Flow

Final Dashboard Work Flow

Results - Workflow - Messages

0 Errors 0 Conv Errors 0 Warnings 0 Messages

Last Run Configuration

In the **Master Workflow**, all the workflows (from acquiring data to generating a dashboard) are sequentially executed using the **Runner** tool.

Stage Inventory Extraction from FTP

Let's see how data is extracted by Alteryx

Data Extraction: FTP 1.1

The screenshot displays the Alteryx software interface with the workflow '1.1 FTP Data Source_Acquire.yxmd' open. The workflow canvas shows two tools connected in sequence: 'FTP File Details' and 'FTP Download Stage Datamart Inventory File to Local'. The 'FTP File Details' tool is highlighted with a red box and a callout stating: 'Text Input tool - holds FTP details, which are consumed by the next tool.' The 'FTP Download Stage Datamart Inventory File to Local' tool is also highlighted with a red box and a callout stating: 'Download tool - downloads data from FTP to a local folder or file.' A third callout at the bottom states: 'FTP data is acquired in the first FTP workflow.' The interface includes a top menu bar (File, Edit, View, Options, Help), a toolbar with various tool icons, and a 'Results - Workflow - Messages' panel at the bottom showing 0 Errors, 0 Conv Errors, 0 Warnings, and 0 Messages.

File Edit View Options Help

Search for tools, help, and resources

Workflow - Configuration

1.1 FTP Data Source_Acquire.yxmd

Canvas Workflow Runtime Events Meta Info

Canvas Options

Layout Direction: Horizontal

Annotations: Show

Connection Progress: Show C

Text Input tool - holds FTP details, which are consumed by the next tool.

1.1 Data Source FTP : Download FTP to Local

FTP File Details

FTP Download Stage Datamart Inventory File to Local

Download tool - downloads data from FTP to a local folder or file.

FTP data is acquired in the first FTP workflow.

Results - Workflow - Messages

0 Errors 0 Conv Errors 0 Warnings 0 Messages 0 Files All

Last Run Configuration

Data Extraction: FTP 1.2

The screenshot displays the Alteryx software interface with a workflow titled "1.2 Data Source FTP : Local to S3 Upload". The workflow consists of two main tools: "FTP Input Local Stg File Path" and "FTP Local to S3 Upload", connected by a data flow arrow. The interface includes a menu bar (File, Edit, View, Options, Help), a toolbar with various tool icons, and a sidebar with "Canvas Options" (Layout Direction: Horizontal, Annotations, Connection Pr). The bottom panel shows "Results - Workflow - Messages" with 0 Errors, 0 Conv Errors, 0 Warnings, 0 Messages, and 0 Files.

Input Data tool - brings stage data in the workflow by connecting to the FTP file on the local machine.

Amazon S3 Upload tool - uploads data from the local drive to the S3 bucket.

Data from FTP is uploaded to the S3 bucket in the second FTP workflow.

Data Extraction: FTP 1.3

The screenshot displays the Alteryx software interface. On the left, the 'Stage File Download from S3 Bucket - Configuration' panel is open, showing fields for AWS Access Key, AWS Secret Key, Endpoint (set to Default), Bucket Name (dmrk.dev.us-east1.shai), and Object Name (Alteryx/WSR1_Datamart_Report_Stage.txt). Below these are options for authentication and file format (Comma-Delimited Text Files (*.csv)).

On the right, the '1.3 Data Source FTP : S3 to Redshift' workflow is shown. It consists of two tools connected by an arrow: 'Stage File Download from S3 Bucket' and 'Stage File Data moved from S3 to Redshift ftp_stage Table'. A red box highlights these two tools.

Three callout boxes provide additional context:

- Amazon S3 Download tool - retrieves the stage file from the S3 bucket.** (Points to the first tool in the workflow)
- Output Data tool – loads the stage data from S3 to a Redshift table.** (Points to the second tool in the workflow)
- Stage data is loaded to a Redshift table in the third FTP workflow.** (Points to the workflow title)

At the bottom, the 'Results - Stage File Download from S3 Bucket - Output' panel is visible, showing '0 of 0 Fields' and a message: 'No data available. (Use Ctrl+R to run the workflow.)'

Stage Inventory Transformation from FTP

Let's see how data is transformed by Alteryx

Data Transformation: FTP 1.4

The screenshot displays the Alteryx Data Transformation interface. On the left, the 'Input Table ftp_stage - Configuration' pane shows the connection to an ODBC data source named 'Alteryx_System' and the table 'dev"."etl2_alteryx_poc"."ftp_stage'. The main workspace contains a workflow diagram with several tools: 'Input Table ftp_stage', 'Convert Reportdate Column to Datetime', 'Convert Datamartdeleted on Column to Datetime', 'Convert datamartupdated on Column to Datetime', 'Select', 'Unique Component compulsory for Update Insert', and 'Ftp_stage Table Data moved to Ftp final Table'. Annotations in blue callout boxes provide details for specific tools:

- Input Table tool** - connects the input stage table from the database.
- DateTime tool** - converts Reportdate column from Varchar to Datetime.
- Select tool** - allows modifying the datatype, data lengths of columns. It also allows deselecting columns not required in the workflow.
- Unique tool** - distinguishes and sorts unique records based on one or more specified fields. The first record in each group is sent to a unique output stream while the remaining records are sent to a duplicate output stream.

The bottom of the interface shows a 'Results - Input' pane with '0 of 0 Fields' and a message: 'No data available. (Use Ctrl+R to run the workflow.)'

Dashboard Table Creation

Let's see how Dashboard Table is created in Alteryx

Dashboard Table Creation 1.14

The screenshot displays the Alteryx software interface. The top menu bar includes File, Edit, View, Options, and Help. Below it is a toolbar with various tools categorized by function: In/Out, Preparation, Join, Parse, Transform, In-Database, Reporting, Documentation, Connectors, Developer, and CREW Macros. The main workspace shows a workflow titled '1.14 Dashboard.yxmd'. The workflow canvas is titled '1.14 Creating Final Dashboard Table' and contains three tools connected in sequence: 'Query for creating Dashboard Table', 'In-DB workflow out', and 'Load Query Data to Dashboard Table'. A red box highlights this workflow. A text box above the 'Load Query Data to Dashboard Table' tool reads 'Select tool to modify the type and size of data'. The left sidebar shows the 'Workflow - Configuration' panel with tabs for Meta Info, Canvas, Workflow, Runtime, and Events. The 'Canvas' tab is active, showing 'Canvas Options' with settings for Layout Direction (Horizontal), Annotations (Show), and Connection Progress (Show Only). The bottom status bar shows 'Results - Workflow - Messages' with 0 Errors, 0 Conv Errors, 0 Warnings, 0 Messages, and 0 Files. The 'Last Run' and 'Configuration' buttons are visible.

Workflow - Configuration

Meta Info

Canvas Workflow Runtime Events

Canvas Options

Layout Direction: Horizontal

Annotations: Show

Connection Progress: Show Only

1.14 Creating Final Dashboard Table

Query for creating Dashboard Table

In-DB workflow out

Load Query Data to Dashboard Table

Select tool to modify the type and size of data

This workflow loads the final data to a dashboard table.

Results - Workflow - Messages

0 Errors 0 Conv Errors 0 Warnings 0 Messages 0 Files All

Last Run Configuration

Dashboard Generation 1.15

The screenshot displays the Dashboard Generation 1.15 software interface. The top menu bar includes File, Edit, View, Options, and Help. Below the menu is a toolbar with various icons for file operations and workflow steps. A search bar is located in the top right corner. The main workspace shows a workflow diagram titled "1.15 Master_Workflow.yxmd". The diagram consists of four sub-workflows: "FTP Work Flow" (containing steps FTP 1.1, FTP 1.2, FTP 1.3, and FTP 1.4), "EMAIL Work Flow" (containing steps Email 1.5, Email 1.6, Email 1.7, and Email 1.8), "FB ABI Work Flow" (containing steps FB API 1.9, FB API 1.10, FB API 1.11, and FB API 1.12), and "Final Dashboard Work Flow" (containing steps Temp DB Max 1.13 and Final Dashboard 1.14). The workflow is connected by arrows indicating the flow of data. At the bottom, a status bar shows "Results - Workflow - Messages" with 0 Errors, 0 Conv Errors, 0 Warnings, and 0 Messages. A blue box with white text states: "The workflows are now ready and can be scheduled." The status bar also includes buttons for "Last Run" and "Configuration".

File Edit View Options Help

Search for tools, help, and resources

Favorites In/Out Preparation Join Parse Transform In-Database Reporting Documentation Connectors Developer CREW Macros

Browse Input Data Output Data Text Input Data Cleansing Filter Formula Sample Select Sort Join Union Text To Columns Summarize Comment

1.15 Master_Workflow.yxmd

FTP Work Flow

FTP 1.1 FTP 1.2 FTP 1.3 FTP 1.4

EMAIL Work Flow

Email 1.5 Email 1.6 Email 1.7 Email 1.8

FB ABI Work Flow

FB API 1.9 FB API 1.10 FB API 1.11 FB API 1.12

Final Dashboard Work Flow

Temp DB Max 1.13 Final Dashboard 1.14

Results - Workflow - Messages

0 Errors 0 Conv Errors 0 Warnings 0 Messages

The workflows are now ready and can be scheduled.

Last Run Configuration

Scheduler

Let's see how workflows are scheduled in Alteryx.

Managing Schedules

The screenshot shows the 'View Schedules' window in a workflow management application. The window is titled 'View Schedules' and has a 'Controller' dropdown set to 'My Computer'. It features tabs for 'Workflows', 'Schedules', 'Queue', and 'Results'. The 'Schedules' tab is active, displaying a table of scheduled workflows. The table has columns for 'Workflow Name', 'Frequency', 'Next/Last Run', 'State', and 'Scheduled By'. A single schedule is listed: '1.15 Master Data Source_Workflow.yxml (In Scheduler)' with a frequency of 'Days/Weeks', a next run of 'November 14, 2017 9:00:00 AM', and a state of 'Active'. Below the table, detailed information for the selected schedule is shown, including its frequency, next and last run times, run count, status, and comment. A text box overlay explains the purpose of the screen. The bottom of the window shows pagination information and a 'Done' button.

Controller: My Computer

Workflows Schedules Queue Results

Filter: Workflow name

Workflow Name	Frequency	Next/Last Run	State	Scheduled By
1.15 Master Data Source_Workflow.yxml (In Scheduler)	Days/Weeks	November 14, 2017 9:00:00 AM	Active	Administrator

Frequency: Days/Weeks: at 9:00 AM on every day of the week
Next Run: Tuesday, November 14, 2017 9:00 AM
Last Run: Monday, November 13, 2017 9:00 AM
Run Until:
Run Location: Copy saved in Scheduler DB

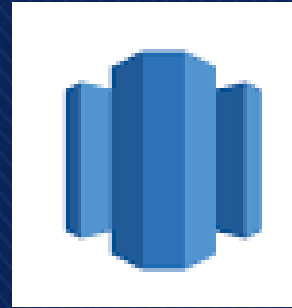
Created: Thursday, November 9, 2017 4:55:35 PM
Run Count: 4
Status:
Comment: None

1 schedules
Queried at 10:38:31 AM
Page 1 of 1
Display: 25

Done

The **View Schedules** screen allows you to view, edit or create a new schedule for a selected workflow.

Integration of BI Tools with Redshift



AWS Redshift can be accessed by BI and other advanced analytic tools.



Dashboard Table on Tableau

