TECHNICAL DOCUMENT

BEST PRACTICES FOR CI/CD INTEGRATION WITH ALTAIR® AI HUB

Enterprises can use the methods outlined in this document to set up a democratized data science development pipeline (CI/CD). This allows them to scale up their artificial intelligence (AI) initiatives using a structured approach including robust safeguards that ensure production integrity.

ALTAIR AI HUB ENVIRONMENT STRUCTURE

Separating development and production environments is best practice for any software development and/or deployment process. Since many Altair® RapidMiner® users have no background in software development, we recommend creating three or four isolated, version-consistent deployments of Altair AI Hub. This supports an agile and secure architecture:

Env 1: Development (DEV)

Use this environment to build new processes, try out ideas, or change the data pipeline safely and without affecting any other systems. This is your "sandbox" for experimentation and testing. The resources required for the DEV environment depends primarily on the number of active users, but, as a rule, each group of 10 to 20 active users requires a license to utilize Altair Units for 64 gigabytes and 8 cores.

The DEV environment is not typically configured to guarantee uptime reliability, although we do recommend the inclusion of backup systems, even though rollbacks are uncommon. DEV environment users should understand these limitations as normal in sandbox environments. Configure permissions generously in the DEV environment to encourage experimentation.

Env 2: Testing (TEST)

New projects should undergo internal testing before release and many enterprises prefer to divide this process into alpha and beta stages. It's best practice to test programs and models in a TEST or QA environment that can't break the DEV or PROD environments. This is the integration-test phase of software engineering projects.

The TEST environment uses samples of real-world data and processes to find flaws. Configure permissions on TEST to be like the DEV environment.

In most enterprises, TEST environments are wiped clean on a regular cadence to ensure a consistent testing environment for all processes headed to PROD.

Env 3: Staging (STAGING)

STAGING is a clone of the PROD environment that enables end-to-end testing without affecting the PROD environment.

For instances of Altair AI Hub running in a cloud infrastructure like AWS, Azure, Oracle, or Google Cloud, it can be economical to spin the STAGING instance up and down as needed.

Development teams typically heavily restrict permissions in STAGING and update it regularly to ensure it's a 100% simulation of PROD.

Production (PROD)

This instance of Altair AI Hub is typically a read-only environment. It runs processes that generate business value and problems in this environment can generate significant costs to the organization. PROD should be set up so it's completely unaffected by any issues that may arise with the DEV, TEST, and/or STAGING environments.

Permissions in PROD should be extremely limited and the infrastructure should support a guaranteed uptime of 99.99%+.

Most enterprises containerize their PROD deployments with dedicated RAM and storage to ensure stability. Changes to PROD should be infrequent and carefully controlled.

TYPICAL ALTAIR RAPIDMINER CI/CD LIFECYCLE

Initial Deployment

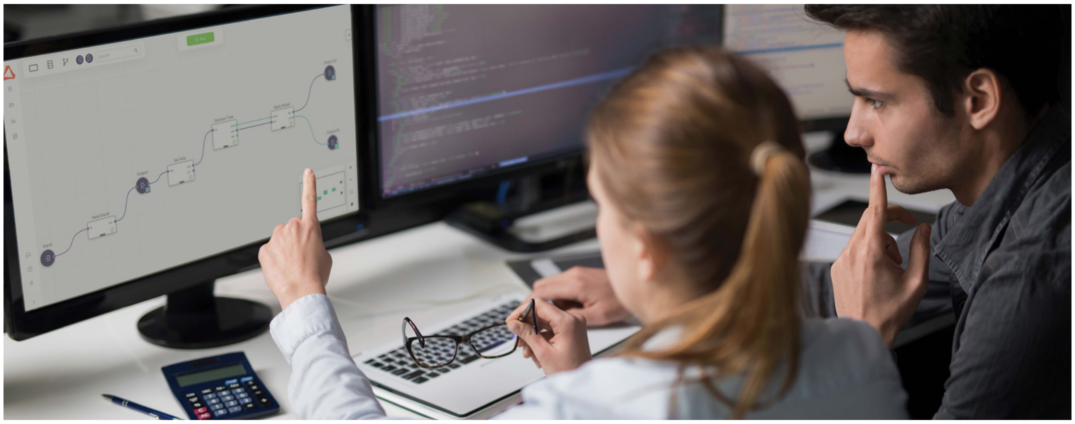
Step 1 - Engineer Eddie, an Altair RapidMiner-certified data engineer and machine learning professional, builds a Git-versioned Altair RapidMiner project on DEV. Examples of common projects include scoring models consumed by webservices, dashboards, email or SMS alerting workflows, scheduled ETL jobs, and more. Upon completion, the team considers the project to be in the "beta" phase and names it eddie-deployment-0.1.

Step 2 - Eddie asks a colleague, Engineer Elizabeth, an Altair RapidMiner-certified platform administrator, to review and test his deployment. Elizabeth provides feedback to Eddie and ensures that his project is safe, secure, well- documented, will consume reasonable resources, and is consistent in design with other deployments within the organization. After incorporating Elizabeth's ideas, they call the project eddie-deployment-0.2.

Step 3 - Eddie submits eddie-deployment-0.2 to System Administrator Sandra, who oversees the entire CI/CD pipeline and has full administrator access to both STAGING and PROD. Sandra loads Eddie's project into STAGING. She makes the final determination to deploy the project and decides what resources to allocate to it.

Step 4 - Sandra moves Eddie's project from STAGING to PROD and notifies Eddie that his deployment is now live as eddie-deployment-1.0. She shares relevant URLs and assigns permissions to Eddie so he can use the deployment as designed. Sandra adds this deployment to her list of other PROD deployments she monitors and sets up alerts as needed. Sandra also schedules an audit for 90 days in the future to evaluate the deployment.

Step 5 - The entire team begins using the deployment to generate business value.



Changes

Step 1 - At some point after deployment, Eddie wants to improve one of his processes. He works in DEV and commits eddie-deployment-1.1 to his project repository.

Step 2 - Eddie goes through the same stages as when he initially published his deployment. However, this process will consume less time since most of the project is unchanged. If successful, Sandra notifies all users that she will upgrade eddie-deployment-1.1 at a time that will minimize the operational impact of the upgrade. When it goes live, this is now called eddie-deployment-2.

Quarterly Review

Three months after going live with the most recent revision eddie-deployment-2.0, Sandra and her team meet with Eddie and other users to review performance and resource utilization. If the team recommends code changes, Eddie follows the procedure as shown in the Changes section above. If the review recommends deprecation, Eddie follows the procedure outlined in the Deprecation section below.

Deprecation

Step 1 - Eighteen months later during a quarterly review, Sandra and Eddie agree that his deployment is no longer generating business value and should be deprecated. Analyst Alyssa has been working a new solution to this business problem and is ready to start the process to move her deployment, alyssa-deployment-0.1, from DEV to TEST. Sandra alerts all existing users of eddie-deployment-2.0 that it will be deprecated on a specific date and time – far enough ahead so that adjustments can be made for time-critical applications.

Step 2 - At the designated date and time, Sandra removes eddie-deployment-2.0 from PROD and archives the process and all logs for audit purposes. She moves the resources to the general pool to be used by other deployments.

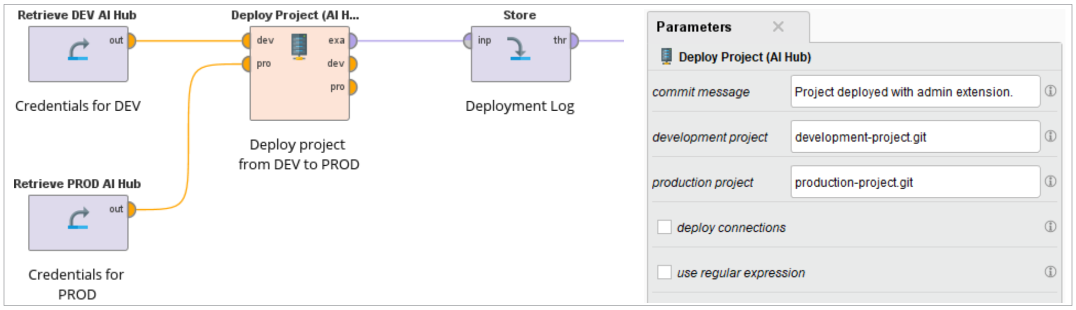
ADMINISTRATIVE TOOLS

All processes, ExampleSets, and associated objects to be deployed should reside in an Altair RapidMiner project. Technically, these projects are Git-backed, making deployment to different environments equivalent to a “clone - copy - push” operation in Git.

Teams can automate tasks, including project access, using Altair AI Hub’s remote access and control capabilities. This is realized through a REST-API. Click here to learn more.

The Admin Extension

The software’s Admin Extension provides further support for our code-optional approach to development and deployment. With a single operator, users can deploy projects from one instance of Altair AI Hub to another and stay within the same ecosystem.



Deploy Project (AI Hub)

commit message	Project deployed with admin extension.
development project	development-project.git
production project	production-project.git

□ deploy connections

□ use regular expression

Users will need connection credentials with appropriate privileges of both instances (for example, going from DEV to TEST) to create specialized Altair AI Hub connection objects. The user defines source and target projects in the software’s parameters panel. (Users can obtain access token and secrets by authenticating to the respective instances of Altair AI Hubs using this code: <URL>/get-token/ page.)

Since different environments usually have different data sources (typically databases), administrators should avoid deploying connections. Instead, they should create connections with the same name pointing to the data source(s) appropriate for the new environment.

ALTAIR'S UNIQUE UNITS LICENSING SYSTEM

Since different environments usually have different data sources (typically databases), administrators should avoid deploying connections. Instead, they should create connections with the same name pointing to the data source(s) appropriate for the new environment.

Altair Units, Altair's patented units-based licensing system, is scalable and shareable across an organization. It allows customers to maximize their software licenses by giving them the flexibility to run Altair software anywhere and the freedom to choose from a host of software tools with unparalleled value.

With Altair Units, customers have full access to Altair software instantly – more than 150 Altair and partner products – and can run these applications on-demand locally, in the cloud, or in a hybrid setup. Learn more about Altair Units.

Altair RapidMiner: A Comprehensive Range of Data Analytics Tools

The Altair RapidMiner platform gives teams the power to use data analytics and AI to gain competitive advantages and drive next-level business results:

AI and Machine Learning - Altair's industry-leading visual approach to analytics modeling minimizes repetitive tasks, shares knowledge across enterprises, and reuses steps within connected model workflows for faster analysis and shared insight. Altair tools support code written in the SAS language, Python, R, and SQL and offer a unique visual interface for creating machine learning models without requiring any written code.

Data Preparation - Access, cleanse, and format data from a wide variety of sources (including Microsoft® Excel®, CSV, PDF, TXT, JSON, XML, HTML, SQL databases, big data sources like Hadoop, and more) without any manual data entry or coding. Automate data transformation and report distribution tasks. Integrate data transformation workflows with enterprise content management (ECM) and robotic processing automation (RPA) platforms.

Business Intelligence and Streaming Data Visualization - Build, modify, and deploy sophisticated business intelligence, data visualization, and stream processing applications using a drag-and-drop interface. Connect to virtually any data source, including big data sources, SQL and NoSQL databases, message queues; develop complex event processing programs; and design visual user interfaces that provide the perspectives people need to make insightful, fully informed decisions based on massive amounts of data.

To learn more, please visit **altair.com/data-analytics**