GM

GM EXTRACT, TRANSFORM AND LOAD PLATFORM AS A SERVICE

How General Motors automated provisioning and increased productivity by 11 times

Extract, Transform and Load (ETL) platform is widely used at General Motors (GM) to perform data integration and data transformation. Currently, GM Information Technology (IT) is supporting over one hundred ETL servers. The ETL platform provisioning process and upgrade were manual, time consuming and error prone resulting in lost productivity and slow market delivery. To improve delivery speed, increase productivity and ensure quality, GM IT designed an ETL Platform as a Service (PaaS) solution.

The scope of this paper focuses on the ETL Platform as a Service (PaaS) solution to deliver the new platform 21 times faster, and increase productivity by 11 times. In addition, the solution targets to increase Platform Maintenance productivity by more than 7 times.

Scenario 1:
PLATFORM PROVISIONING

The current ETL platform provisioning process is time consuming due to the cross-team engagement to build out the infrastructure, configure the network, create the databases, install the middleware components, setup monitoring and configure security. In addition, the process is error prone due to manual installation and configuration of the ETL software.

Today, the GM IT ETL platform provisioning process, for a new platform installation and configuration, takes many hours from the time a request is initiated to the delivery of a working platform. Figure 1 illustrates the current ETL Platform Provisioning Process.
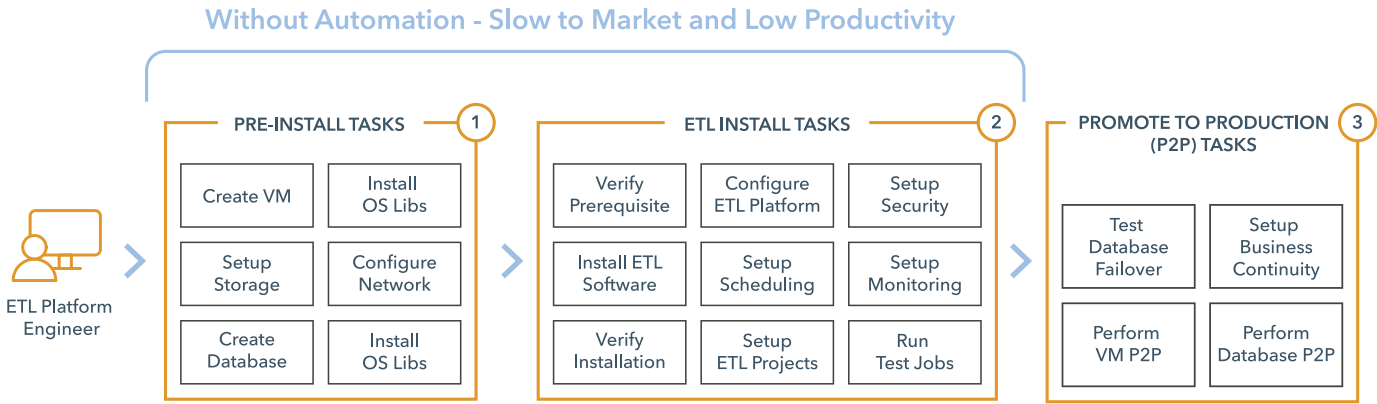

Without Automation - Slow to Market and Low Productivity

PRE-INSTALL TASKS (1)	
Create VM	Install OS Libs
Setup Storage	Configure Network
Create Database	Install OS Libs

ETL INSTALL TASKS (2)		
Verify Prerequisite	Configure ETL Platform	Setup Security
Install ETL Software	Setup Scheduling	Setup Monitoring
Verify Installation	Setup ETL Projects	Run Test Jobs

PROMOTE TO PRODUCTION (P2P) TASKS (3)	
Test Database Failover	Setup Business Continuity
Perform VM P2P	Perform Database P2P

ETL Platform Engineer

Figure 1 – Current ETL Platform Provisioning Process

Pre-install Tasks (1) include the infrastructure provisioning, setup OS with required OS libraries for ETL software installation; setup storage, configure network, create ETL meta database and install prerequisite middleware software. ETL engineers manually enter the requests in an online request system to get the required components from each team.

ETL Install Tasks (2) include pre-installation verification, ETL software installation and post installation verifications, ETL platform configuration and ETL development project setup; along with security setup, ETL job scheduling, and platform monitoring. The install ends with a test job which is executed to verify that the ETL platform is built correctly and is ready to be handed over to ETL developers. Currently the above tasks are only partially automated.

Promote To Production (P2P) Tasks (3) are performed for production platforms only. The ETL production ready platform needs to pass a database failover and business continuity (BC) test. Business Continuity is required for disaster recovery.

To improve quality and reduce the ETL platform provisioning timeliness, the GM IT team designed an end-to-end automation solution to provide ETL platform as a service. Figure 2 illustrates the processes and tools involved in the automated ETL platform provisioning solution.

With Automation - Deliver Platform 21 Times Faster and Increase Productivity by 11 Times

GALILEO			CHEF					
PRE-INSTALL TASKS 1			**ETL INSTALL TASKS** 2			**PROMOTE TO PRODUCTION (P2P) TASKS** 3		
Create VM	Install OS Libs		Verify Prerequisite	Configure ETL Platform	Setup Security	Test Database Failover	Setup Business Continuity	
Setup Storage	Configure Network		Install ETL Software	Setup Scheduling	Setup Monitoring	Perform VM P2P	Perform Database P2P	
Create Database	Install OS Libs		Verify Installation	Setup ETL Projects	Run Test Jobs			
Partial Automation			Full Automation			Manual Requests		

ETL Platform Engineer
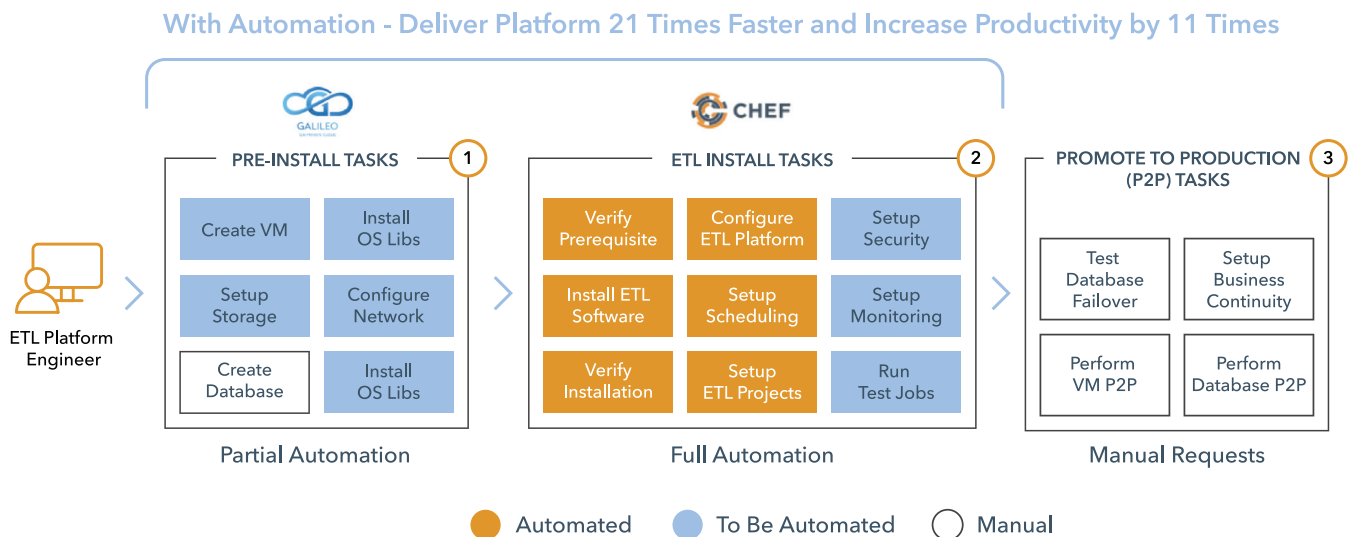
🟠 Automated   🔵 To Be Automated   ⚪ Manual

Figure 2 – Proposed ETL Platform Provisioning Solution

Galileo is the cloud services platform and process automation hub for General Motors Information Technology teams. It is the initiation point for the ETL PaaS solution where a requestor submits a request to provision a new ETL platform.

Entry Request (1) A requestor enters the ETL application information, ETL server size, ETL software version, environment information and project specific requirements through the Galileo User Interface. Upon the submission of the request, the pre-defined workflow will validate user input, request approval and invoke infrastructure provisioning, setup OS with required OS libraries, storage setup, network configuration, create ETL meta database and install DB2 client automation jobs. The validation rules and approval information are predefined and stored in a database.

ETL Install Workflow (2) Then workflow kicks off ETL Install job which includes pre-installation verification, ETL software installation and post installation verifications, ETL platform configuration, security setup, ETL job scheduling setup, platform monitoring setup, ETL development project setup. Finally, the workflow runs a test job to verify ETL platform is built

correctly and is ready to be handled over to ETL developers. The ETL platform Install Tasks are fully automated and implemented by Chef recipes, which are invoked by ETL's new provisioning cloud service.

Promote to Production (3) For production platform build, ETL platform engineers need to enter a request manually to start the Promote to Production (P2P) process. The P2P work includes database High Availability failover test, setup of business continuity (BC) for disaster recovery and BC test for server and database. Currently this part of work is performed manually by multiple teams.

Upon the completion of the automation process, Galileo will send the notification to the requestor to report the completion or failure of requests.

The ETL platform standard will be pre-defined which includes platform infrastructure requirements, ETL software version, monitoring standards, scheduling information and ETL project template configuration. The automation jobs access the standard platform information during the automation. The workflow also saves the automation status in the database so the requestor can query the request status at any time.

By automating the ETL platform provisioning process, the end-to-end delivery is 21 times faster and productivity is increased by 11 times.

Scenario 2:
PLATFORM MAINTENANCE

In addition to new server builds, maintaining platform health and upgrades are common manual activities that are time consuming and error prone.

The ETL PaaS solution can also be applied to platform maintenance to increases productivity and quality. The following use cases describe the PaaS solution for patch upgrades and platform standard compliance.

Use Case 1:
PATCH AND SECURITY UPGRADE

ETL software vendors typically release patches after a major release to address bug fixes and/or vulnerabilities. ETL platform also needs to constantly comply with the latest security policy changes.

Currently the above tasks are performed by ETL Platform Engineers manually. On average, ETL platform applies patches twice a year which includes installation and verification. To comply with security, it is required to renew SSL certificates after they expire every two years. The ad-hoc security policy changes occur once a year by average. Figure 3 shows the manual ETL platform patch and security upgrade process
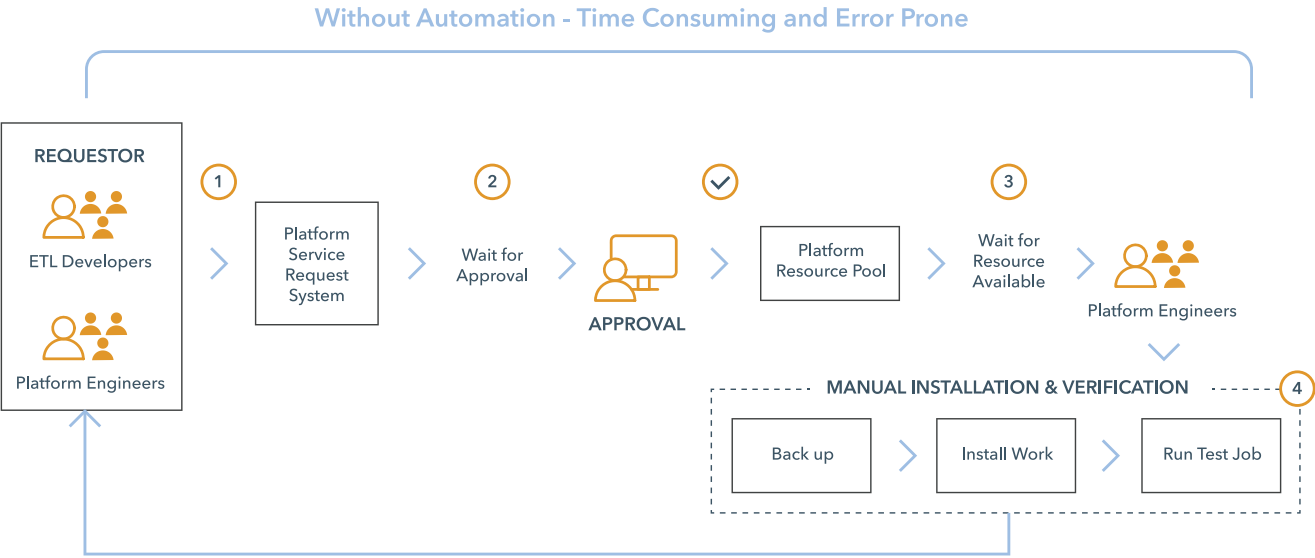
Figure 3 – Manual ETL Platform Patch and Security Upgrade Process

The manual steps involved in the ETL patch and security upgrade process are:

Entry Request (1) A requestor enters an ETL patch or a security upgrade request, enters patch version, ETL version, environment information, business justification and other information through ETL platform service request system.

Submission Approval (2) After the submission of request, the Platform Manager reviews and approves the request.

Request Queue (3) Once approved, the request goes to the queue for a platform resource to claim and begin work.

Manual Patch (4) The first available platform resource claims the request and performs the manual patch upgrade and verification

To improve quality and reduce the ETL platform patch and security upgrade timeliness, the GM IT team designed an end-to-end automation solution to provide ETL platform patch and security upgrade as a service. Figure 4 illustrates the processes and tools involved in the automated ETL platform patch and security upgrade solution.
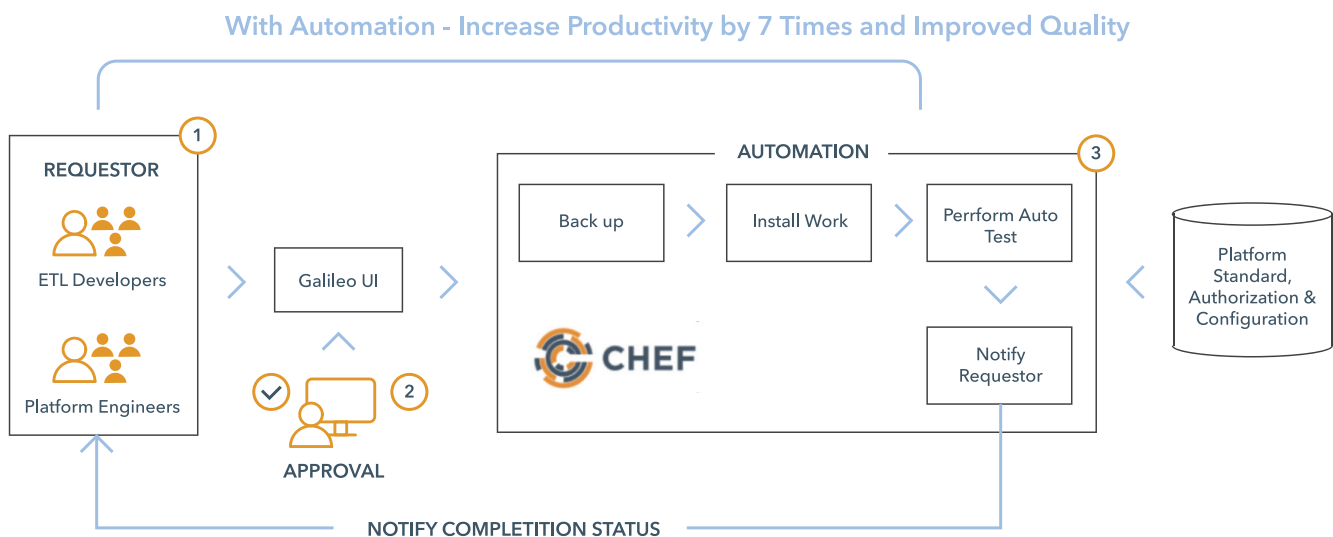

Figure 4 – Automated ETL Patch and Security Upgrade Solution

The automated ETL platform patch and security upgrade solution include the following steps:

Entry Request (1) A requestor enters an ETL patch upgrade request, enters patch version, ETL

version, environment information, business justification and other information through Galileo User Interface.

Review and Approval (2) After the submission of the request, platform managers review and approve the request.

Automated Patch Process (3) Automation process is invoked by Galileo workflow. The workflow first backups the current ETL installation, then applies the patch. After the patch installation, it invokes an ETL test job to verify the patch installation. Upon completion of installation and verification, it notifies the requestor the patch upgrade is complete. The ETL patch and security automation tasks are implemented by Chef recipes, which are invoked by ETL patch and security upgrade cloud service.

The automated ETL platform patch and security upgrade solution provides the following benefits:

- Shortens the wait time for manual approval processes by offering the workflow tracking to enforce Service Level Agreement

- Eliminates the wait time for available platform engineer resources by automating patch and security upgrade work

- Improves the quality by enforcing the standard and automated testing

> *By automating ETL platform patch and security upgrade process, productivity can be increased by 7 times with improved quality.*

Use Case 2:
PLATFORM STANDARD COMPLIANCE

After an initial build, the ETL platform can be changed unintentionally or by server maintenance processes which can cause deviations from the standard build. For production servers, to ensure business continuity (BC), active and passive servers need to be consistent. Deviations from a standard build are the major contributors to incidents that disrupt the normal business operations due to platform outrages. Deviations also cost both platform engineers and ETL developers extra hours on debugging issues, resulting in ineffective use of resources and low productivity. Figure 5 shows Current ETL platform process to handle deviations from standard build
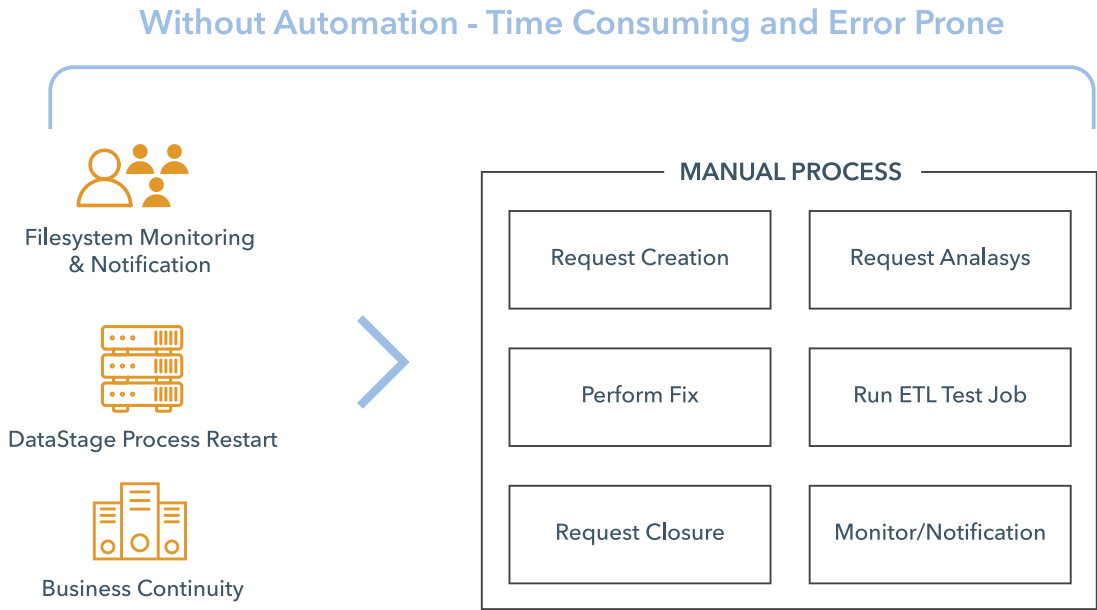
Figure 5 – Current ETL Platform Process to Handle Deviations from Standard Build

To reduce ETL platform incidents and effectively use resources, GM IT team designed a "self-healing" solution which evaluates the platform to identify gaps against "standards and policies" and then initiates the automation to correct the gaps. Figure 6 illustrates the processes and tools involved in ETL platform Self-Healing solution.

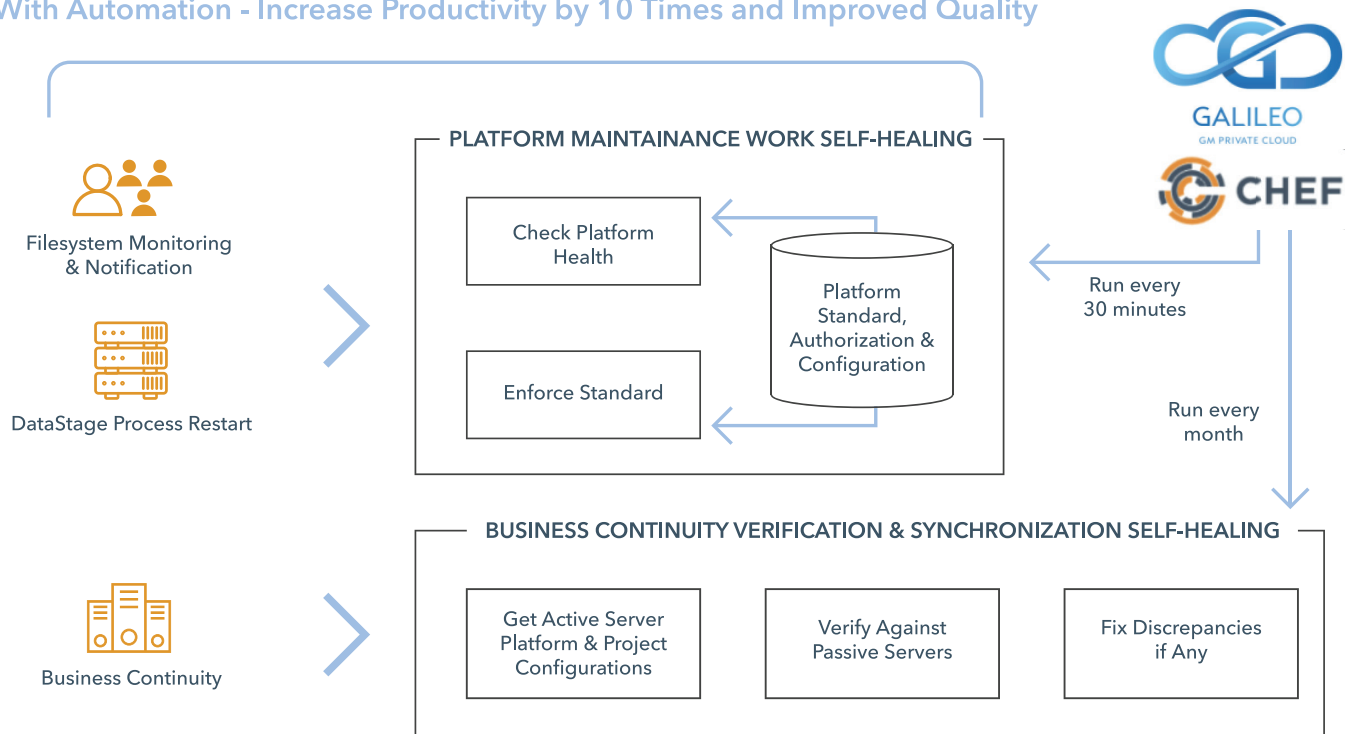## With Automation - Increase Productivity by 10 Times and Improved Quality

Figure 6 – ETL Platform Self-Healing Solution

The ETL platform maintenance work Self-Healing solution includes the following tasks:

- Retrieve the ETL platform standards from the database. The standard includes the ETL software versions, infrastructure requirements and configurations, ETL project configurations, platform security setup. The standard is pre-defined and can be modified by ETL platform engineers at any time.

- Compare the standard with the values on the current ETL platform.

- If the automation process finds any deviations, it corrects the platform with the standard.

- After the correction, it runs an ETL test job to verify the platform healthy.

- If the test job is successful, it logs the audit information and the process is complete. If the test job fails, it logs the audit information and notifies the ETL Platform Engineers manual intervention is required.

The above self-healing tasks are implemented by Chef recipes. The Chef recipes are kicked off by the Galileo Chef pipeline process and run every 30 minutes, human resources are not required unless a test job fails, then manual intervention is required.

The ETL platform business continuity (BC) Self-Healing solution includes the following tasks:

- Get the platform and project configurations from each BC pair server

- Compare configurations from active server against passive server

- If the automation process finds any discrepancy, it corrects the passive server configuration with the values from its active server

- After the correction, it runs an ETL test job to verify the passive server healthy.

- If the test job is successful, it logs the audit information and the process is complete. If the test job fails, it logs the audit information and notifies the ETL Platform Engineers manual intervention is required.

The above self-healing tasks are implemented by chef recipes. The Chef recipes are kicked off by Galileo Chef pipeline process and runs every month, human resources are not required unless a test job fails, then manual intervention is required.

By implementing the self-healing solution, the productivity can be increased by 10 times with improved quality.

CONCLUSION

Global organizations are under pressure to provide high-quality products while simultaneously achieving lower costs, increase productivity and speed to market. The GM ETL Platform as a Service solution described in this article illustrates a practical implementation using automation approach for platform builds and maintenance work. GM Private Cloud service along with Chef infrastructure as code technology enabled the GM ETL PaaS solution.

> *In summary, the GM ETL Platform as a Service (PaaS) solution can deliver new ETL Platform 21 times faster and 11 times improved productivity. In addition, the solution can increase Platform Maintenance productivity by more than 7 times.*

GLOSSARY

Abbreviation	Description
Chef	Both the name of a company and the name of a configuration management tool
ETL	Extract, Transform and Load
Galileo	GM private cloud services platform and process automation hub for General Motors Information Technology teams. It provides self-service tools for developers to create and manage environments, project management teams with streamlined online workflows, and automates communication between IT Service Teams.
GM	General Motors, is an American multinational corporation that designs, manufactures, markets, and distributes vehicles and vehicle parts, and sells financial services
PaaS	Platform as a Service. The National Institute of Standards and Technology (NIST) defines PaaS as *"The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly application hosting environment configurations"*, cloud computing, https://cloudinfosec.wordpress.com/2013/05/04/nist-definition-for-saas-paas-iaas/
Self-healing	The "self-healing" evaluates the platform to identify gaps against "standards and policies" and initiates auto-correction to reach compliance during the operations.

AUTHOR

Jeanne Gu – Platform Architect, Enterprise Application Integration, General Motors

ACKNOWLEDGEMENT