|  |  |
| --- | --- |
| **CREATED BY** | **UST Global** |
| **CREATED ON** | **11/01/2017** |
| **VERIFIED BY** |  |

CMS Fusion

***EQUIFAX***

**DOCUMENTATION**

**Revision History**

| **Revision** | **Author** | **Date** | **Status and Description** |
| --- | --- | --- | --- |
| 1.0 | UST Global | 11/01/2017 | Initial Draft |
| 1.1 | Equifax | 11/15/2017 | Review |
| 2.0 | UST Global | 11/16/2017 | Review Comments Incorporated |

#### EQUIFAX

Equifax Inc. is a [consumer credit reporting agency](https://en.wikipedia.org/wiki/Consumer_credit_reporting_agency). Equifax collects and aggregates information on over 800 million individual consumers and more than 88 million businesses worldwide. Founded in 1899 and based in [Atlanta](https://en.wikipedia.org/wiki/Atlanta,_Georgia), [Georgia](https://en.wikipedia.org/wiki/Georgia_(U.S._state)), it is one of the three largest credit agencies along with [Experian](https://en.wikipedia.org/wiki/Experian) and [TransUnion](https://en.wikipedia.org/wiki/TransUnion).

#### BUSINESS OVERVIEW

Equifax has operated primarily in the business-to-business sector, selling consumer credit and insurance reports and related analytics to businesses in a range of industries.

Business customers include retailers, insurance firms, healthcare providers, utilities, government agencies, as well as banks, credit unions, personal and specialty finance companies and other financial institutions.

Equifax sells businesses credit reports, analytics, demographic data, and software. Credit reports provide detailed information on the personal credit and payment history of individuals, indicating how they have honored financial obligations such as paying bills or repaying a loan. Credit grantors use this information to decide what sort of products or services to offer their customers, and on what terms. Equifax also provides commercial credit reports containing financial and non-financial data on businesses of all sizes.

#### DATA ASSETS

**ACRO** (Automated Credit Reporting Online) is the default credit bureau accessed in retrieving credit application for customers. The data from data providers are forwarded to the *QA team*. Here the collected data gets cleansed and consolidated, and then passed to ACRO (Automated Credit Reporting Online). So when there is a credit score request/ enquiry, it’s checked in the ACRO. The ACRO data are used in almost every Equifax products. The ACROFILE report comprises of identification information (PII), fraud/verification alert, model result, inquiry history, public records, collections, trades, consumer narrative statements and other ancillary products and features.

The Inquiries can be of two types – **Hard inquiry** and **Soft inquiry**. The inquiries initiated by the consumer is known as a hard inquiry whereas the internal inquiries (Ex. The bank forecasts one’s chances to default on the loan), are Soft inquiries. All these inquiries are updated in the credit file of the corresponding consumer.

NCTUE (National Consumer Telecommunication & Utility Exchange), or simply NC+ is not owned by Equifax. It mainly provides utility information. IXI data source is owned by Equifax which provides wealth and investments data. It is based on zip code and not on consumers. The Workforce Solution contains the income information. Other data sources include GEO, PDA, TWN, MLA etc.

Some of the data sources and their corresponding description are shown below:

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Source** | **Description** | **Internal/ External DS** | **Key Data for Reference** |
| ACRO | Credit data | Internal | CID, CONNEXUS |
| IXI | Wealth data | Internal | CONNEXUS |
| PDA | Property and address information | Internal | CONNEXUS |
| NCTUE | Utility data | Internal | CONNEXUS |
| TWN | Employment and work data | Internal | CONNEXUS |
| AUTO | Automobile Data | Internal | CONNEXUS |
| Commercial | Financial information and transactions data | Internal | CONNEXUS |
| Demographics | Demographic data. The data comes from a company called Epsilon. | External | CONNEXUS |
| Department of Defense | Military data | External | SSN,DOB and Class Name |
| Workforce | Income, tax and employment data | Internal | CONNEXUS |

The **Connexus** is a key, used for linking a dataset to another data set which doesn’t have any keys in common. Within Equifax, there is a home grown key processing system which helps us to associate different data sets.

#### INTERCONNECT

Interconnect is a decision management platform that provides authorized users the ability to generate and view real-time reports through the UI. When an applicant entry is received by the Interconnect platform, the customer solution shall utilize the Equifax’s ACRO credit file product.

Interconnect customers can choose from the set predefined of standard reports depending on business need. There are three broad categories of reports. They are listed as follows:

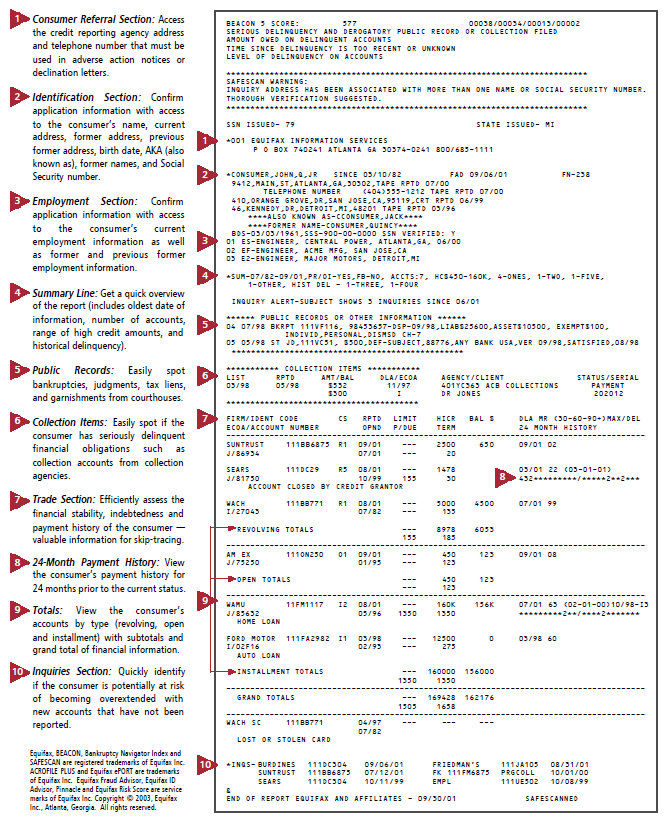
* Credit risk reports
* Operational reports
* Work Item reports

All standard reports enabled for a customer are available on the reports menu. Report preferences link enables user to set score cut range for score distribution report.

#### ACROFILE

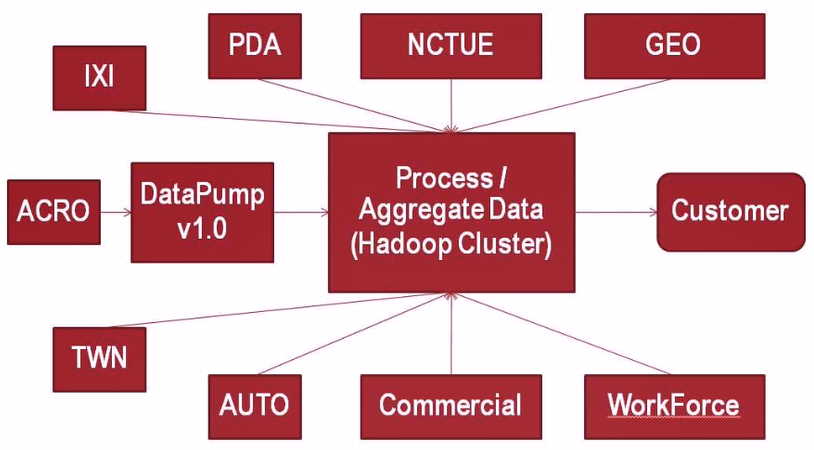
It refers to the ACRO report that comprises of multiple occurrences of various information – Identification information, fraud/verification alert, model results, inquiry history, public records, collections, trade, consumer narrative statements and other ancillary products and features. Currently we use ACRO 5 and 6.

A sample ACROFILE is shown below (This is an On-Line ACRO report and therefore differs from the “Sample File” report that CMS uses):



#### D360

D360 delivers consumer data based on a standard or a custom requirement from the consumer. It obtains data from different data sources and processes it by linking, merging and applying aggregation rules based on the customer needs. D360 provides various types of data including credit, commercial, property, income, employment, auto, demographics, utility etc. as obtained from various data sources. The illustration of D360 consumer products is shown below:



In this batch environment, data from ACRO is ingested into the **Datapump**. The data from ACRO (online DB), is split into 300 different parts (mdb files). Each copy of the data is referred to as **dataset**. The datasets are stored in a file system accessible by the datapump. Thus we can say that, the datapump maintains an offline copy of ACRO. It’s done on a daily basis.

#### MODEL IMPLEMENTATION (MI)

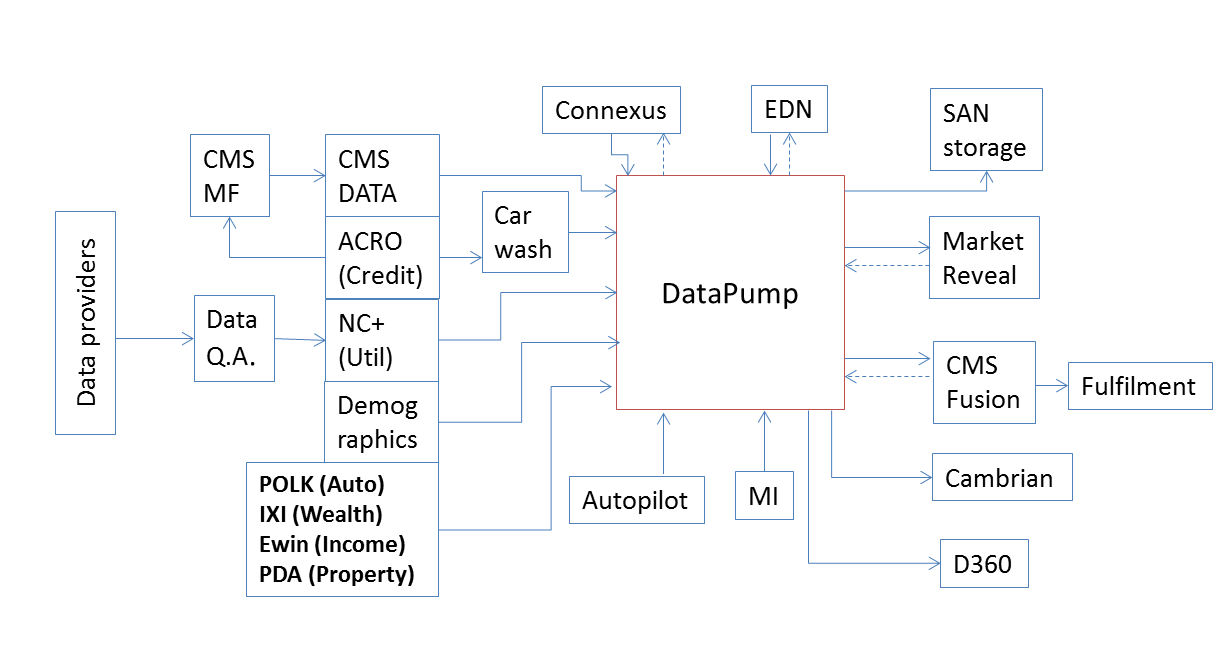
**Model** refers to a set of rules, procedures or strategies that defines the decision making process to establish a solution (Ex. whether a loan could be provided to a particular consumer or not). It calculates scores.

The MI is developed by programmers based on the datasets obtained from the datapump and other data sources. The **autopilot** module is an internal IDE for coding attributes. It calculates the **attributes.** The C++ code that’s generated by Autopilot is provided to MI to be compiled into the Model code. The attributes refers to the information or its combination, from the credit file. Ex: Credit file enquires bankruptcies, court decisions, average balance etc.

#### DATAPUMP

The datapump is a distributed platform application, which has a cluster of Linux nodes, which provides the framework for credit scoring. Datapump receives data from various data sources which includes the Credit, Utilities, Automobile, Demographics, Wealth, and Employment etc. It processes and archives the data and provides the infrastructure for attribution and score calculation using these data sources.

The Component Diagram of datapump platform is as follows:



Datapump receives data from various data sources, builds it and archives it. It dynamically parses the data at run time and provides the infrastructure to run Models and Attribution. It is a batch processing system and provides service to Market Reveal team, *CMS Fusion*, Cambrian and D360.

The Market Reveal and *CMS Fusion* teams cater to Customers (such as Banks, Financial institutions, Insurance companies, etc.). According to the customer requirements it may be necessary to run different score models and attributions on consumer data for a selected list of consumers or for a particular geographical location or for the entire consumer database (All records). Datapump receives these specifications from the teams in the form of an xml file known as **job.xml**. Datapump then parses the xml specifications, loads various data sources and dynamically loads the shared libraries of the models and attributes to be run as part of the batch job. The output of the jobs will be used to create reports for the customers.

Considering the inputs to the datapump, the consumer data is received from third party vendors and it undergoes cleaning and QA processes.

The following are the main data sources that are processed by Datapump:

* **ACRO Credit Data:** As mentioned previously, it is ~~the~~ Equifax’s first Automated Credit Reporting Online. All the credit details of a consumer are captured in the ACRO Credit record. This includes different segments such as Identification information (name, address, SSN, DOB etc.), Public records (Collections, Tax liens, Bankruptcy), Trade details, Inquiries etc.

The datapump receives data from ACRO on a daily basis in the form of MDB files in 300 packs. The data is split into 300 packs, each consisting of MDB records. The original ACRO data is in Binary EBCDIC format. This data is built, indexed and stored as datasets in datapump. The data is parsed and stored as a Credit Record object at runtime. The credit record is interrogated and used to calculate attributes and scores.

In datapump, the ACRO credit data for each pack is indexed on the CID (Consumer ID) which is the key for each consumer credit record in ACRO.

The archives of ACRO Credit data include *Golden data* - Copy of the entire credit database is frozen on the last Tuesday of each month. This copy is archived indefinitely. These copies are referred as a Golden dataset, *Daily data* - A copy of the On-Line ACRO database is written to a dataset (file) every single day. These datasets are normally retained for two weeks.

* **NC+:** The NCTUE team provides the NC+ data through Global Data Exchange. This is received on a weekly basis on every Tuesday. The NC+ data is indexed in Datapump based on Connexus key.
* **CMS segment:** CMS segment is a derived data source from ACRO. CMS mainframes (Credit Marketing Services) also receive the ACRO MDBs and it applies the logic within CMS mainframes to generate the CMS segment. CMS segment usually contains Y/N flags such as ‘Offensive Name’, ‘Valid Street Address’ etc. This CMS segment is also sent to datapump on a weekly basis every Tuesday. This data is also in Binary EBCDIC format.
* **Trended data:** Trended data is a derived source from ACRO and is provided by D360. The datapump provides the ACRO credit data to D360. The D360 extracts the trended history (The consumer’s payment history) for the past 24 months from the Golden snapshots archived by datapump and the trended data is sent back to the datapump on a monthly basis.

*CMS Fusion* uses its GUI to configure different process and put across the job xml to JET application. JET validates these job xmls and the jobs are queued and scheduled to be run in DP. DP then runs its multisource processor, which takes the job.xml as its input and creates the output. This output can be used by the *Fusion team*, which they extract and store in their green plum databases to create reports for the customers.

The *Fusion team* submits batch job requests through their GUI. Datapump receives the batch job specifications in xml format called the job xml. Different steps / plugins to be run as part of the batch job will be specified in the job xml along with the details of the datasources to be used. There can be two types of job requests - Source job and Direct extract job.

* The **Source job** runs only for a particular set of consumers specified by the customer. The list of consumers specified by the customer will undergo source match (which is a process that identifies the correct consumer record based on PII information) and the CIDs thus matched will be appended to the file. This appended file is used in datapump and the path for that file is stored.
* In **Direct extract job**, customers do not specify an explicit list of consumer ids for which this job is to be run. But there may be specific conditions to select the consumer records. Examples include the condition to run the job for consumers from a particular geographic location.

The **job xml** contains all the required details to run a particular job such as the data sources to be used, external files to be used, the models to run, The autopilot modules and other plugins to be run as part of this job and the order in which these steps are to be executed. Various tags in job xml are - <jobIdentifiers>, < jobRuns>, <parallel> etc.

#### AUTOPILOT

The Autopilot application is a GUI with point and click capability designed to improve coding speed and reduce programming errors while acting as the primary method for creating CMS criteria and LCR data attributes. It provides the facility to use the GUI instead of hand coding the attributes. Autopilot converts the Modules created in GUI into C++ shared libraries. Autopilot uses a complex expression analyzer that generates efficient and versatile C++ code that can run on virtually any processing platform including the mainframe, UNIX, and NT Windows.

The steps involved are:

* Fulfillment Analysts builds the Module through the GUI
* The Module code is converted into a metadata file.
* The metadata file is then converted into a CPP file by the autopilot compiler.
* The CPP file build takes place and the share object is created.

**Attributes**, a main component of the autopilot module, are the fundamental building block in which we aggregate, manipulate, and translate raw credit data. The attributes can also form the output of the autopilot module. The different types of attributes includes aggregate attribute, collection attribute, value attribute, definition attribute etc.

Another component of the autopilot module is the output map. The output map allows you to define an output for your project.

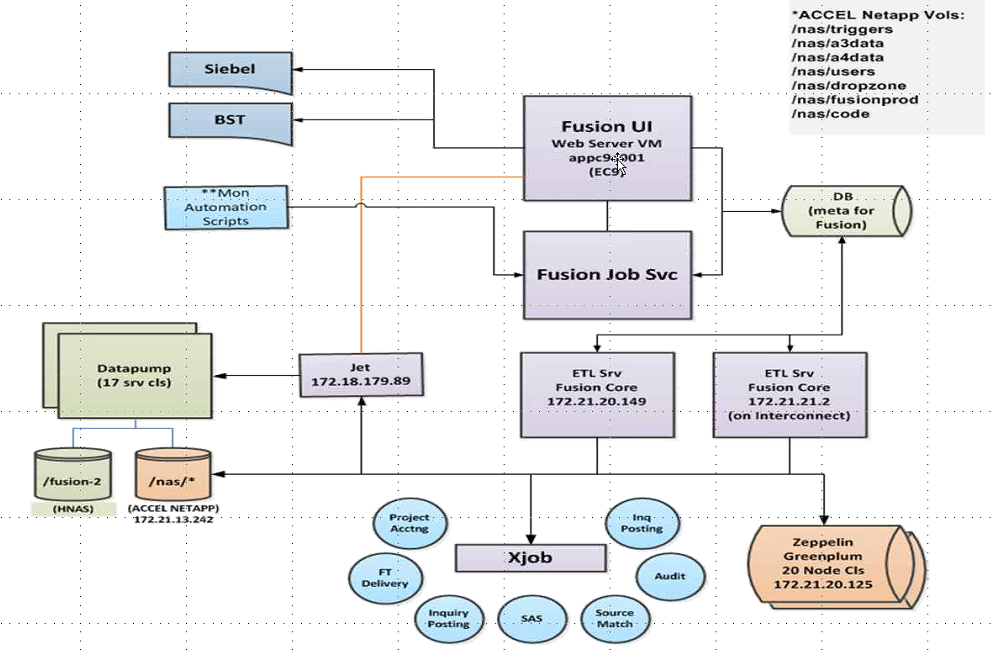
#### FUSION

The Fusion is the platform for processing batch requests and fulfilling various products – portfolio reviews, acquisitions, prescreen account monitoring etc. based on the specifications and requirements provided by the sales project manager to the user (fulfillment analyst), after consulting the customer. So the sales team creates the projects after working closely the customer and updates the necessary details in the **Siebel**. The fusion user utilizes this Siebel data. The *Professional Services Consultant Team (PSCT)* works with the customer and the sales team and acts as an intermediate between the fulfillment analyst and the customer. The fulfillment analysts seek the help of PSCT for more information, if required. The fusion comprises of the final output that must be given to the customer, the list of products that is needed to be run, if there is attributes that is need to be returned to the customer, criteria etc. The **criteria** refer to a set of selection rules that typically determines the credit worthiness of a consumer by analyzing the associated attribute values.

We can say that the fusion is a user interface to setup and configure various jobs and also a platform that involves the datapump and various data sources, process these data and feed it to the *Greenplum* *DB* to do all other ancillary processing and deliver the final output to the customer. The fulfilling organization is *CMS*.

Some of the key processes in CMS fusion are : Input, Search/Match, Pre-select, Datamenu, Sample file, Refinement, Householding, SAS, Filter, Match/Join, MLA, Rolling suppression, DNS, Custom Data Processing, Reporting, FFF, Output, Shipping, Billing, Enquiry posting etc.

The CMS fusion ecosystem is illustrated below:



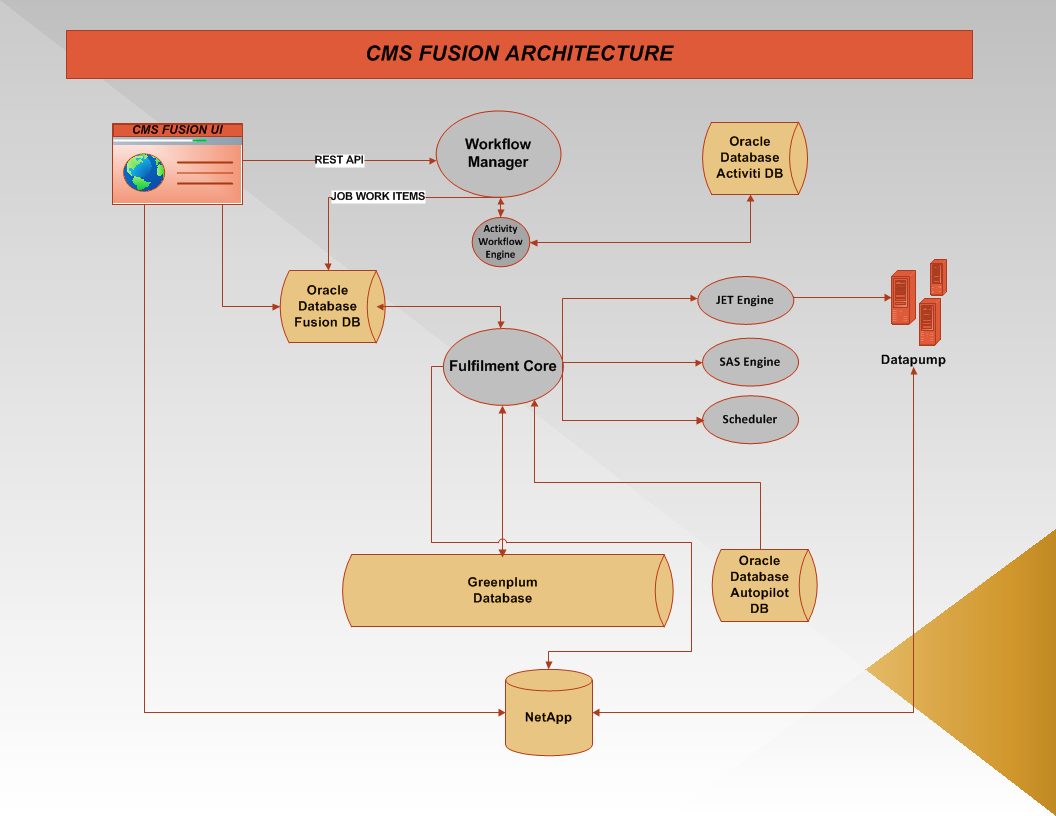
The main components of CMS fusion are listed as follows:

* UI Application
* Job Interface
* Core Application

The UI application includes individual process configurations, process stacking and submission. The Job interface deals with job submission, creation of dynamic workflow based on process configurations, workflow management of individual items, Leveraging Activiti BPM tool and Automation. The Core application manages the processors for the work items, Leverages the Greenplum DB for some pre/post processing and delivers scalable and extensive support to different work item executors.

The **Process** refers to the configuration whereas a *step* refers to the run instance of a process in a job. A **job** may contain one or more steps and a step is comprised of Workitems. A job can also be defined as a combination of steps – Input, Source Match and Datamenu. **Workitem** can be defined as the granular execution element. For example, Input process has two Workitems – INPREPARE and IN\_GPLOAD. The INPREPARE Workitem sequences the records in a file whereas the IN\_GPLOAD loads the data into Greenplum.

The CMS fusion architecture is illustrated below:

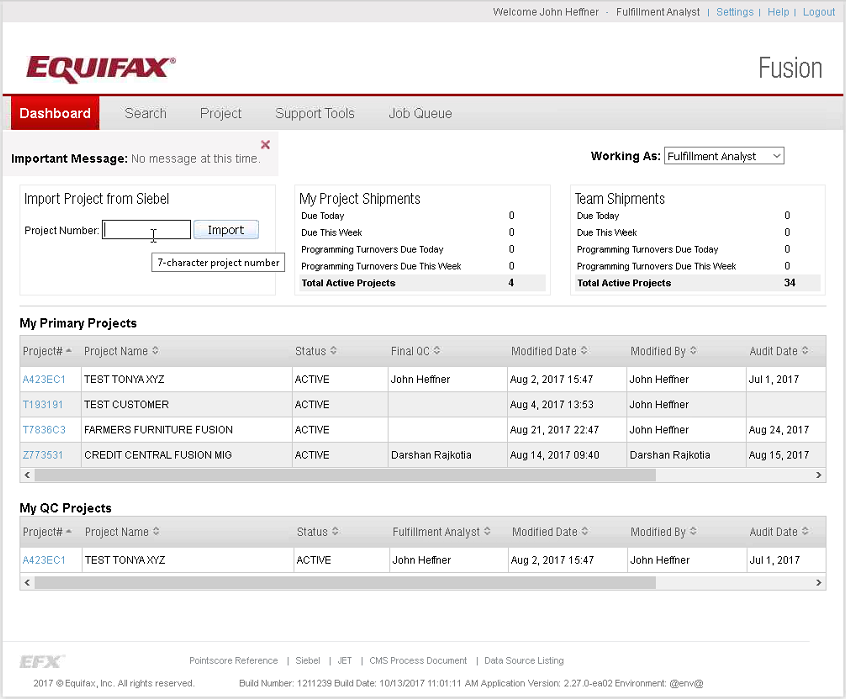


Fusion can be also defined as the delivery engine for CMS. It’s scalable and extensive architecture supports different work item processors. 24+ different processes are supported by fusion. Easy integration with external systems is made possible through fusion.

#### FUSION UI - DASHBOARD

The dashboard tab of the Fusion provides various information like the type of user, provision to import projects from Siebel, project shipment details, team shipment details, user’s primary projects, and QC projects.

The screenshot of a sample CMS fusion dashboard (fulfillment analyst user) is shown below:

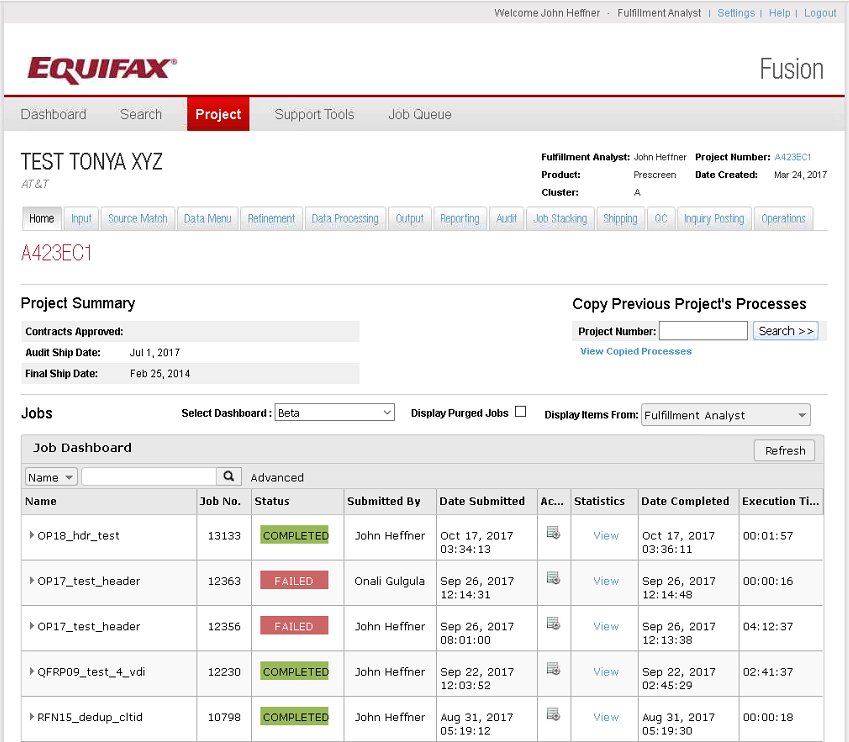


The type of fusion users includes *Fulfillment analysts, Quality analysts, Programmers and the System itself (automation)*. When the *PSCT* creates the customer orders or projects in Siebel, the fulfillment analyst gets assigned. This is how the fusion gets the input. To start a project in fusion the fulfillment analyst, first imports the project from Siebel.

So, once the project is imported from Siebel to fusion, open the corresponding project from the dashboard. The project will be opened in the Project tab of fusion.

#### FUSION UI - PROJECT > Home

A sample fusion Project > Home tab is shown below:



Here, TEST TONYS XYZ is the Project name and AT&T is the customer name. It includes various project details like project summary, name of the fulfillment analyst, type of product, project creation date in fusion, cluster and job dashboard.

The job dashboard provides the details about the jobs in each project. It contains the fields - Job name, Job number, Status, Submitted by, Date submitted, Actions, Statistics, Date completed and Execution time. Jobs can be further divided into steps, if necessary.