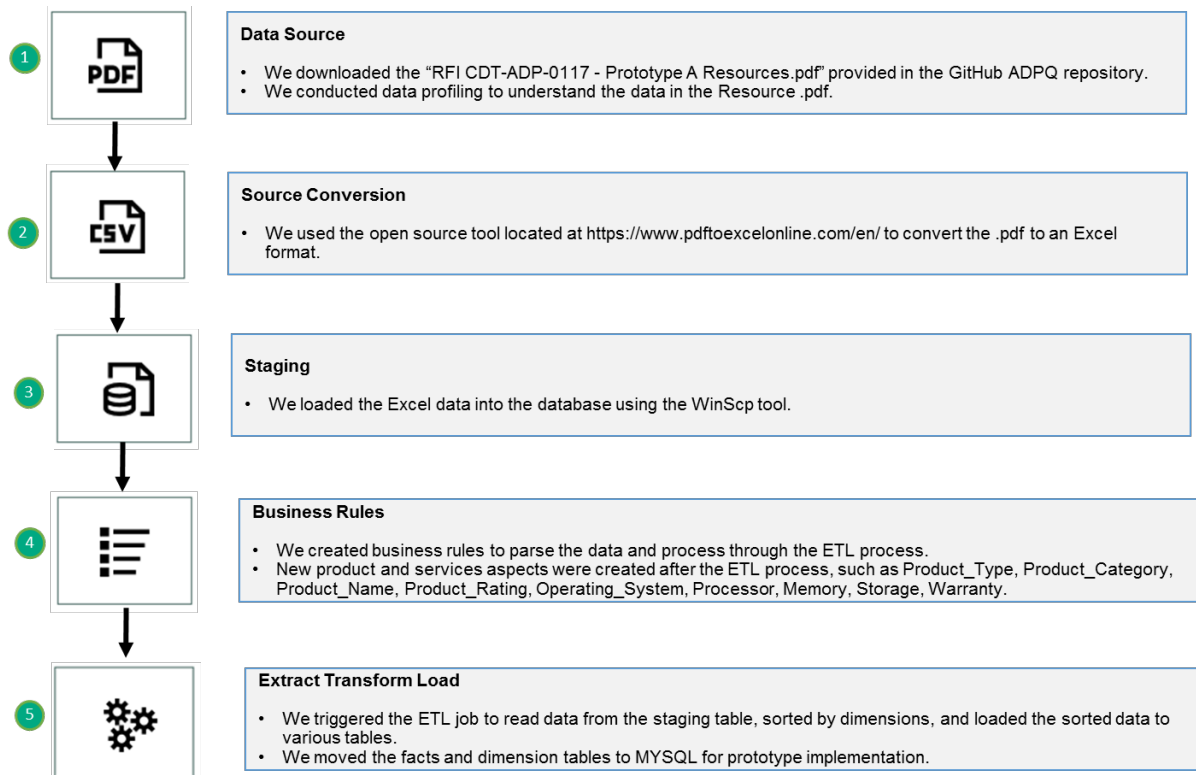


ETL Process for Cal eStore Prototype

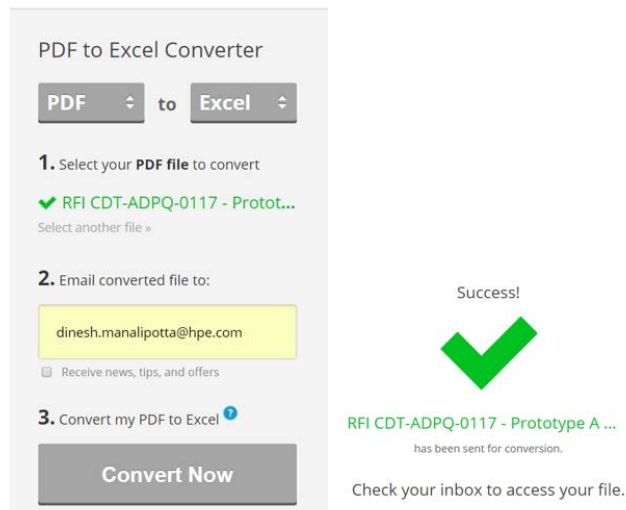
The following graphic depicts the Extract Transform Load (ETL) process the HPE team utilized in developing the Cal eStore prototype.



The following narrative and graphical depictions show the steps in more detail.

Data Conversion

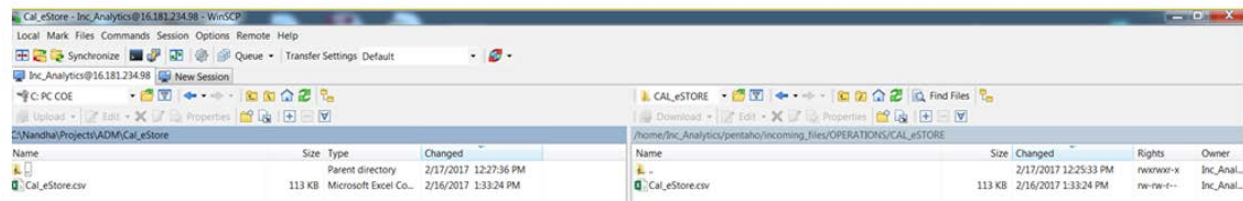
RFI CDT-ADPQ-0117 - Prototype A Resources.pdf provided in GitHub for prototype A was not a MYSQL DB (database) consumable format; therefore we used an open source PDF to Excel conversion tool. This tool can be found at <https://www.pdf-to-excel.com/en/> — to extract the data and convert it to a .csv file.



The screenshot shows the 'PDF to Excel Converter' interface. It includes a dropdown menu to select the input file type (PDF) and the output format (Excel). A list of files for selection is shown, with 'RFI CDT-ADPQ-0117 - Protot...' selected. Below this, there's a section for email notification with the email 'dinesh.manalipotta@hpe.com' entered. A 'Convert Now' button is at the bottom. To the right, a green checkmark and the text 'Success!' indicate the conversion was completed. Below that, it says 'RFI CDT-ADPQ-0117 - Prototype A ... has been sent for conversion.' and 'Check your inbox to access your file.'

Data Loading

We converted the input Excel file into .csv format, then moved the file to a server drop zone where it was loaded into the database using the WinSCP tool.



Business Rules

The Prototype A use case requires a comparison of products and services, as well as grouping into high-level categories like Hardware, Software and Services. Data in the *RFI CDT-ADPQ-0117 - Prototype A Resources.pdf* file is **not** conducive for item comparison and application layout. For example, hardware product aspects like product name, operating system, processor, memory, storage, warranty, connectivity type, HDD RPM, and RAM type, were merged together in a single column “Item description.” The same issue was there for all other products and services in the resource .pdf file.

To efficiently parse the item description text, we created an Extract, Transform and Load (ETL) process. The ETL process required development of business rules to ensure consistent parsing of the data.

NOTE: The following example is for illustration purposes and is not a complete list of business rules.

Business Rule Objective	Description	Applicable Column	Rank
Divide each records in to three Product Types (Hardware, Software and Services) and Product Category	BR1. When the Horizontal Category equal to Configuration (Hardware) and Category equal to (Standard Desktop Hardware or Workstation Hardware or Thin Client Hardware or Power Desktop Hardware or All in One Hardware) then Product type equal to "Hardware" and Product Category equal to "Desktop"	Product_Type Product_Category	1
	When the item description contains "DDPE" then Product type equal to "Software" and Product Category equal to "Data Encryption"		
	When the item description contains "Computrace" and then Product type equal to "Software" and Product Category equal to "Device & Data Security"		
Divide each records in to three Product Types (Hardware, Software and Services) and Product Category	When the item description contains "Intel Standard Manageability" or "INTEL VPRO" and then Product type equal to "Software" and Product Category equal to "Others"	Product_Type Product_Category	2
Divide each records in to three Product Types (Hardware, Software and Services)	BR2. When the Item Description contains = (FHD and LCD or LED and HD or Widescreen or Touchscreen or EliteDisplay or LED and FHD or 4K and LCD) and Sub_Category_Detail (Upgrades or Upgrade or MONITOR) then Product type equal to "Hardware" and Product Category equal to "Monitor"	Product_Type Product_Category	3
Create the product name for each record	BR3. Product name equal to "Manufacturer" + first part of the "Item Description"	Product_Name	4
Divide each records in to three Product Types (Hardware, Software and Services) and Product Category	BR4. When the Horizontal Category equal to Configuration (Hardware) and Category equal to (Integrated Microphone or External Speakers or Business Speakers or Wireless Headset or Stereo Speakers or LCD Speakers) then Product type equal to "Hardware" and Product Category equal to "Audio Devices"	Product_Type Product_Category	5
Divide each records in to three Product Types (Hardware, Software and Services) and Product Category	BR5. When the Horizontal Category equal to Configuration (Hardware) and Category equal to (DVD ROM or External Drive or Optical Disk Drive or USB DVD RW Drive) then Product type equal to "Hardware" and Product Category equal to "DVD Drive"	Product_Type Product_Category	6
Divide each records in to three Product Types (Hardware, Software and Services) and Product Category	BR6. When the Horizontal Category equal to Configuration (Hardware) and Category equal to (NVIDIA Quadro or AMD Radeon or AMD Graphics card or AMD R5) then Product type equal to "Hardware" and Product Category equal to "Graphics Card"	Product_Type Product_Category	7
Divide each records in to three Product Types (Hardware, Software and Services) and Product Category	BR7. When the Horizontal Category equal to Configuration (Hardware) and Category equal to (Solid State Drive or Hard disk Drive or Optical Disk Drive) then Product type equal to "Hardware" and Product Category equal to "Hard Drive"	Product_Type Product_Category	8
Create the Display Attributes for comparison	BR8. When the Product type equal to "Hardware" and Product Category equal to "Monitor" then move the numerical (14" or 15" or 15.6" or 17.3" or 19" or 19.5" or 21" or 22" or 23" or 24" or 28" or 30") in the Item Description to a new attribute Display_Size	Display_Size	9
Create the Display Attributes for comparison	BR9. When the Product type equal to "Hardware" and Product Category equal to "Monitor" then move the (LCD or LED) in the Item Description to a new attribute Display_Type	Display_Type	10
Create the Display Attributes for comparison	BR10. When the Product type equal to "Hardware" and Product Category equal to "Monitor" then move the (4K or FHD or HD) in the Item Description to a new attribute Display_Definition_Type	Display_Definition_Type	11
Create the Warranty Attributes for comparison	BR10. When the in the Item Description contain the word Warranty then move the ((Numeric and year description) and (Lifetime)) to a new attribute "Warranty"	Warranty	12

ETL Job Steps

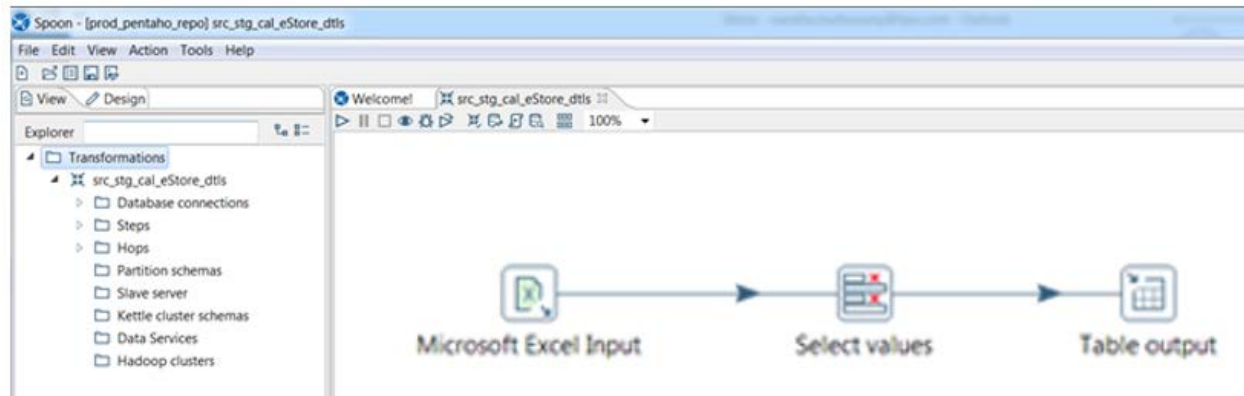
We triggered the shell script 'job_cal_eStore_load.sh' to load the data through the ETL job.

```
Inc_Analytics@IAS-RHEL:~/pentaho/Scripts/Cal_eStore
[Inc_Analytics@IAS-RHEL Cal_eStore]$ pwd
/home/Inc_Analytics/pentaho/Scripts/Cal_eStore
[Inc_Analytics@IAS-RHEL Cal_eStore]$ ls -ltr
total 4
-rw-rw-r--. 1 Inc_Analytics Inc_Analytics 1349 Nov  8 13:59 job_cal_eStore_load.sh
[Inc_Analytics@IAS-RHEL Cal_eStore]$ sh job_cal_eStore_load.sh
```

The following depictions describe the individual steps in the ETL process.

Step 1 — ETL staging job

This job reads data from .csv file and loads it into a staging table.



Step 2 — Staging table

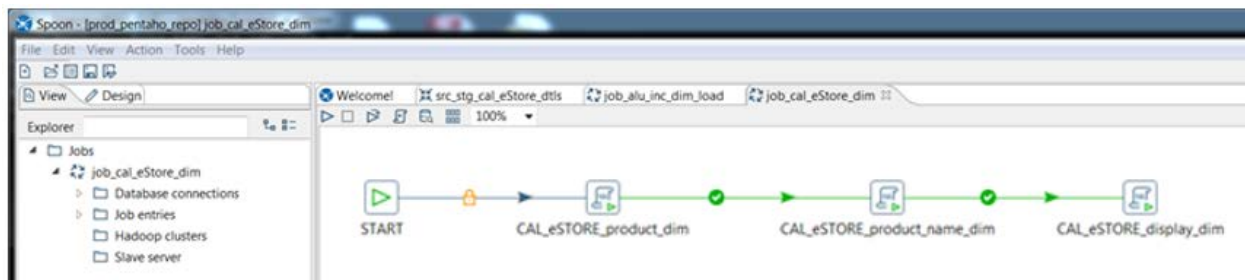
Data was successfully loaded into the staging table 'stg_Cal_eStore'.

The screenshot shows the DBeaver interface with the 'stg_cal_eSTORE' table selected. The table contains 200 rows of data. The columns are: Contract_Line_Item_Number, UNSPSC_Code, Manufacturer_Part_Number, Manufacturer, SKU_Item, Item_Description, Unit_of_Measure, and Quantity. The bottom status bar indicates '200 row(s) fetched - 109ms'.

Contract_Line_Item_Number	UNSPSC_Code	Manufacturer_Part_Number	Manufacturer	SKU_Item	Item_Description	Unit_of_Measure	Quantity
1	*Non Core	81110000	999-5033	NWN	999-5033	BASIC DEPLOYMENT & LOGISTICS	1
2	*Non Core	81110000	TIG-DLG-NBK	TIG	TIG-DLG-NBK	DEPLOYMENTS & LOGISTICS SERVICES	1
3	100-C	43211507	T6V32UP	HP	T6V32UP	HP ProDesk 600 G2 SFF Business PC SingleUn	1
4	1002c	NA	370-ACLY	DELL	370-ACLY	8GB (1x8GB) 1600MHz DDR3L Memory	1
5	1003c	NA	490-BCYL	DELL	490-BCYL	AMD Radeon R5 340X (2GB DR/DVI-I)	1
6	1004-A	43211902	L1G56AV	HP	L1G56AV	AMD Radeon R9 350 2GB DH PCIe x16 GFX	1
7	1005b	43211500	451-BBUV	DELL	451-BBUV	6 Cell (91 Whr) Long Life Cycle Lithium Polym	1
8	1005b	NA	400-AIOY	DELL	400-AIOY	3.5 TB 7200rpm HDD	1
9	1006	43211500	F2B56AA	HP	F2B56AA	External USB DVD+-RW Drive	1
10	1006b	NA	470-ABLQ 555-BCMT	DELL	470-ABLQ 555-BCMT	Intel Dual Band Wireless 8260 (802.11ac)	1
11	1007	43211500	313-7362	DELL	313-7362	USB Powered External Speakers	1
12	1007	43211500	E7U2SAA	HP	E7U2SAA	HP S803XL Notebook Battery	1
13	1007-A	43211500	N1T71&V	HP	N1T71&V	Intel R360 807 11.5in W/1.7in RT VDR	1

Step 3 — Facts and Dimension tables for Cal eStore

The ETL job in the following example reads data from the staging table, sorts by dimensions, and loads the sorted data to various dimension tables.



Records for Cal eStore were successfully loaded into dimension tables as illustrated in the following example.

Query

```
SELECT INTO 'eshop'..eshop ('STATUS_ID','STATUS_NAME','IS_CANCELLED_ALLOWED') VALUES (0,'Cancelled','Y');
```

2 Table Data

PRODUCT_ID	PRODUCT_NAME	PRODUCT_DESC	CATEGORY_ID	PRODUCT_TYPE	IMG_PATH
38	HP Care Pack Hardware Support - 3	HP 3y NextBusiness Exchange PC Only SVC	4 SE		HP/HP 3-Year Next Business
39	HP 3-Year Next Business Day On-Site Service	HP 3y 800 Onsite/24x7 Workstation Service	4 SE		HP/HP 3-Year Next Business
41	DELL Premier Backpack	Black/Gray, Nylon, Two Side Pocket, one external pocket and one mai	4 SE		Dell/Dell backpack.jpg
42	Dell - 5 Year Basic Hardware Service with WND on-s	5 Year Basic Hardware Service with WND on-site	4 SE		Dell/Dell VAS Services.
43	HP EliteOne 800 G2 23-in Touch All-in-One PC	800G2 All In One Desktop, HP EliteOne 800 G2 23-in Touch All-in-One	2 HW		HP/HP EliteOne 800 G2 2
44	Dell - All-in-One - Dell Optiplex 7440 AIO	All-in-One - Dell Optiplex 7440 AIO - Core i5-6500 3.20GHz, 4GB DDR4	2 HW		Dell/Dell - All-in-One
45	Dell - Workstation - Dell Precision T5810	Workstation - Dell Precision T5810 - Four Core E5-1607 v3 (4C, 3.1G	2 HW		Dell/Dell - All-in-One
46	Dell - Standard-Dell Optiplex 3040 Micro	Standard-Dell Optiplex 3040 Micro - Core i5-6500T, 2.50GHz, 4GB 1D3M	2 HW		Dell/Dell - All-in-One
47	HP EliteDesk 800 G2 DM 65W Business PC	HP EliteDesk FWR DT 800G2 DM i5, HP EliteDesk 800 G2 DM 65W Business	2 HW		HP/HP EliteOne 800 G2 2
48	HP - HP ProDesk 600 G2 MT Business PC	HP EliteDesk STANDARD DT 600G2 CMT, HP ProDesk 600 G2 MT Business P	2 HW		HP/HP EliteOne 800 G2 2
49	HP Business Top Load Case	Compact, Light Weight, Robust, Water Resistant.	4 SE		HP/HP Backpack.jpg
50	HP- NVIDIA Quadro K4200 4GB	NVIDIA Quadro K4200 4GB DL-DVI(1)+2xDP 1st No cables included Graph	8 HW		HP/HP NVIDIA Quadro K42
51	Dell - 2.5 inch 80008 7200rpm	2.5 inch 80008 7200rpm FIPS Certified Self-Encrypting Hard Drive,	8 HW		Dell/Dell - 2.5 inch 8C
52	HP- 1TB 2 Turbo Drive PCIe Solid State Drive	1TB 2 Turbo Drive PCIe Solid State Drive,	8 HW		HP/HP- 1TB 2 Turbo Driv