607 Project3 Data Transformations For Storage

The data sources as provided were in denormalized layout, which was desirable for analysis but not for storage in a relational database. The following steps were taken to transform the data for storage.

1. ds\_user\_profiles\_skills
   1. Add a new column [SkillType] with values of {“General Skill”, “Software/Language Skill”}
   2. The file was split into two files which each will be loaded into the relational database
      1. ds\_user\_profiles\_general\_skills
      2. ds\_user\_profiles\_software\_skills
2. ds\_job\_postings\_general\_skills
   1. Each job posting site was previously its own column showing the count of postings having that skill. The data were unpivoted (gathered) to now have columns:
      1. [Site] which houses the name of the job site
      2. [N\_Records] which houses the number of postings on that jobsite mentioning that skill.
3. ds\_job\_postings\_software\_skills
   1. Each job posting site was previously its own column showing the count of postings having that software. Additionally there were separate columns for each site, showing the percent of postings mentioning that software. The data were unpivoted (gathered) to now have columns:
      1. [Site] which houses the name of the job site
      2. [N\_Records] which houses the number of postings on that jobsite mentioning that skill.
      3. [Pct\_Records] which houses the percent of postings on that jobsite mentioning that skill.