Given the data below, we want to refactor this to meet 3NF requirements.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| PROJ\_NUM | PROJ\_NAME | EMP\_NUM | EMP\_NAME | JOB\_CLASS | CHG\_HOUR | HOURS |
| 15 | Evergreen | 103 | June E. Arbough | Elect. Engineer | $84.50 | 23.8 |
| 15 | Evergreen | 101 | John G. News | Database Designer | $105.00 | 19.4 |
| 15 | Evergreen | 105 | Alice K. Johnson | Database Designer | $105.00 | 35.7 |
| 15 | Evergreen | 106 | William Smithfield | Programmer | $35.75 | 12.6 |
| 15 | Evergreen | 102 | David H. Senior | Systems Analyst | $96.75 | 23.8 |
| 18 | Amber Wave | 114 | Annalise Jones | Applications Designer | $48.10 | 24.6 |
| 18 | Amber Wave | 118 | James J. Frommer | General Support | $18.36 | 45.3 |
| 18 | Amber Wave | 104 | Anne K. Ramoras | Systems Analyst | $96.75 | 32.4 |
| 18 | Amber Wave | 112 | Darlene M.Smithson | DSS Analyst | $45.95 | 44.0 |

To normalize this to 1NF, we must get rid of duplicate rows, make our fields atomic, and have each tuple identifiable by a primary key. There are no duplicate rows here, but we can atomize the employee name into first name, middle initial, and last name. We then see that PROJ\_NUM and EMP\_NUM make a good primary key.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| PROJ\_NUM | PROJ\_NAME | EMP\_NUM | FNAME | MINITIAL | LNAME | JOB\_CLASS | CHG\_HOUR | HOURS |
| 15 | Evergreen | 103 | June | E | Arbough | Elect. Engineer | $84.50 | 23.8 |
| 15 | Evergreen | 101 | John | G | News | Database Designer | $105.00 | 19.4 |
| 15 | Evergreen | 105 | Alice | K | Johnson | Database Designer | $105.00 | 35.7 |
| 15 | Evergreen | 106 | William | *Null* | Smithfield | Programmer | $35.75 | 12.6 |
| 15 | Evergreen | 102 | David | H | Senior | Systems Analyst | $96.75 | 23.8 |
| 18 | Amber Wave | 114 | Annalise | *Null* | Jones | Applications Designer | $48.10 | 24.6 |
| 18 | Amber Wave | 118 | James | J | Frommer | General Support | $18.36 | 45.3 |
| 18 | Amber Wave | 104 | Anne | K | Ramoras | Systems Analyst | $96.75 | 32.4 |
| 18 | Amber Wave | 112 | Darlene | M | Smithson | DSS Analyst | $45.95 | 44.0 |

To normalize this into 2NF, each attribute must depend solely on the full primary key. There are several rows that only depend on either the PROJ\_NUM (PROJ\_NAME) or the EMP\_NUM (FNAME, MINITIAL, LNAME, JOB\_CLASS). I am assuming any given employee has a *single* job class. In contrast, I not assuming that the CHG\_HOUR is an employee specific rate, but could potentially be a rate negotiated in the contract for the project. Looking at this data, we see that the two “Database Designers” are making the same rate of $105/hour for the Evergreen Project. I assumer that rate could be different for different projects, however. Given this,the CHG\_HOUR is dependent on both JOB\_CLASS and PROJ\_NUM, but *not* EMP\_NUM. Lastly, HOURS is assumed to depend on the full primary key. With all this in mind, we can now break out our relation to remove duplicity and fix these dependencies.

|  |  |  |
| --- | --- | --- |
| PROJ\_NUM | EMP\_NUM | HOURS |
| 15 | 103 | 23.8 |
| 15 | 101 | 19.4 |
| 15 | 105 | 35.7 |
| 15 | 106 | 12.6 |
| 15 | 102 | 23.8 |
| 18 | 114 | 24.6 |
| 18 | 118 | 45.3 |
| 18 | 104 | 32.4 |
| 18 | 112 | 44.0 |

|  |  |
| --- | --- |
| PROJ\_NUM | PROJ\_NAME |
| 15 | Evergreen |
| 18 | Amber Wave |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| EMP\_NUM | FNAME | MINITIAL | LNAME | JOB\_CLASS |
| 103 | June | E | Arbough | Elect. Engineer |
| 101 | John | G | News | Database Designer |
| 105 | Alice | K | Johnson | Database Designer |
| 106 | William | *Null* | Smithfield | Programmer |
| 102 | David | H | Senior | Systems Analyst |
| 114 | Annalise | *Null* | Jones | Applications Designer |
| 118 | James | J | Frommer | General Support |
| 104 | Anne | K | Ramoras | Systems Analyst |
| 112 | Darlene | M | Smithson | DSS Analyst |

|  |  |  |
| --- | --- | --- |
| PROJ\_NUM | JOB\_CLASS | CHG\_HOUR |
| 15 | Elect. Engineer | $84.50 |
| 15 | Database Designer | $105.00 |
| 15 | Programmer | $35.75 |
| 15 | Systems Analyst | $96.75 |
| 18 | Applications Designer | $48.10 |
| 18 | General Support | $18.36 |
| 18 | Systems Analyst | $96.75 |
| 18 | DSS Analyst | $45.95 |

Finally, we can look to see if we can put this into 3NF. To meet 3NF, there must be no attributes in any relation that depend on a non-primary key field. Looking at our first table in the upper right, we note that HOURS is dependent both on the project and employee. In the second table (upper right), the PROJ\_NAME is unique to the PROJ\_NUM. In the third table, all the name fields and JOB\_CLASS depend only on the employee number. Finally, in the forth table, CHG\_HOUR is dependent only on the primary key fields.

Thus, we are already at 3NF and do not need to refactor our tables further.