

Different Types Of Joins In Spark SQL

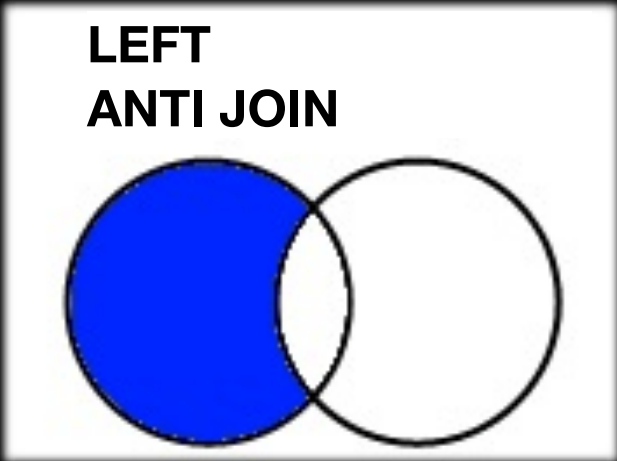
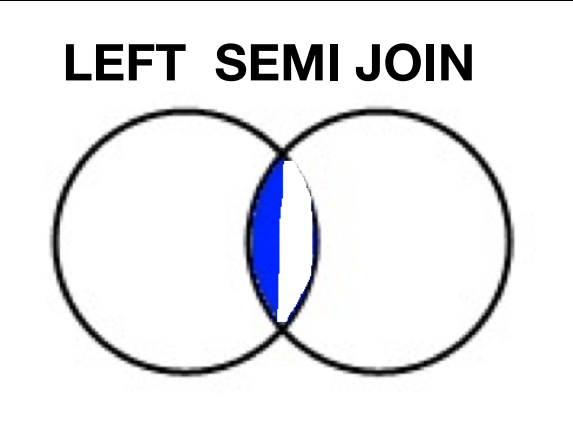
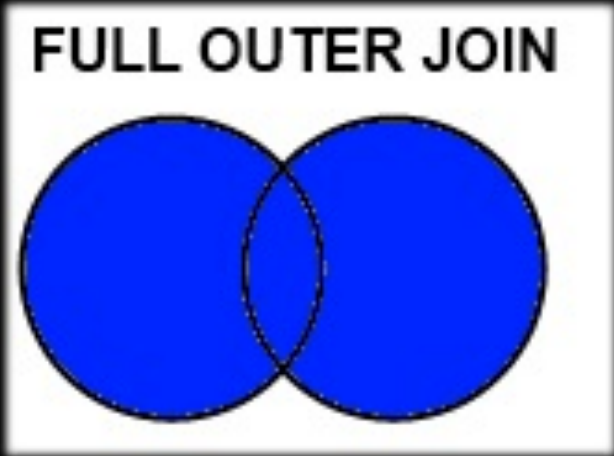
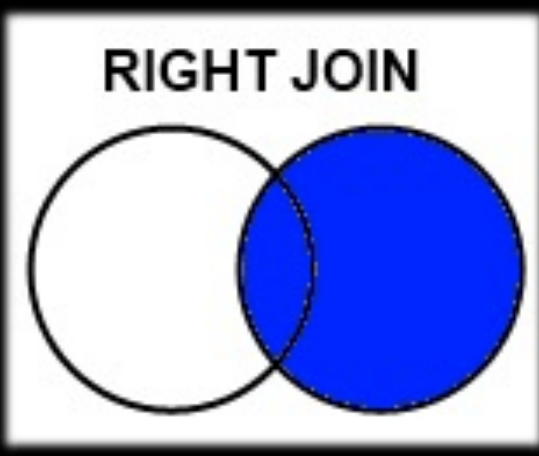
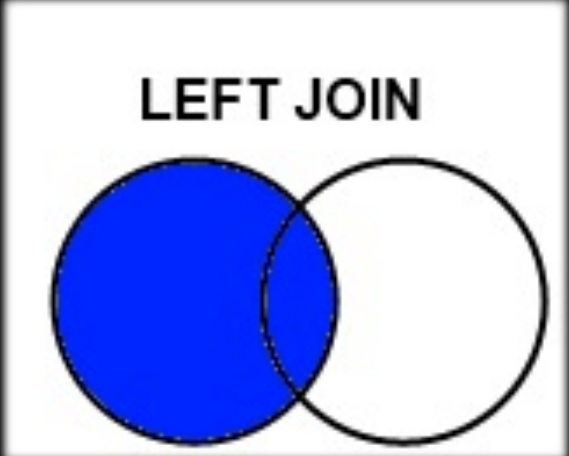
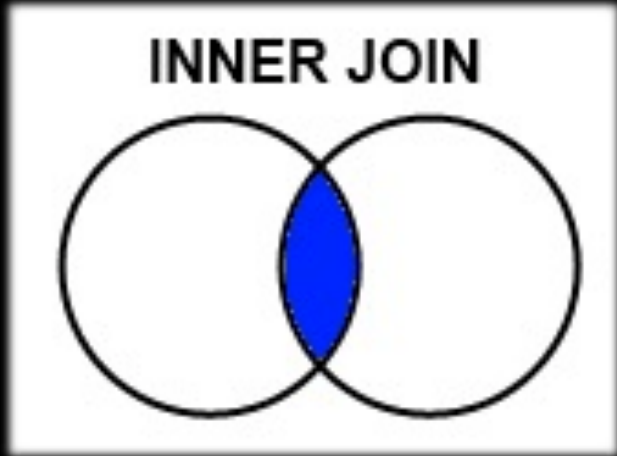


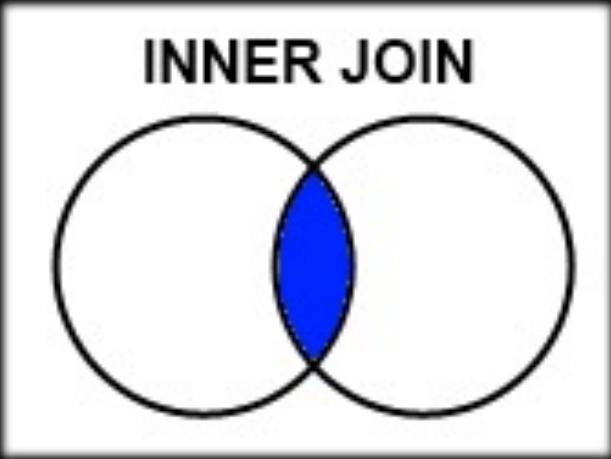
EMP Table Data

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO
7369	SMITH	CLERK	7902	1980-10-12	800	null	20
7499	ALLEN	SALESMAN	7698	1981-10-02	1600	300	30
7521	WARD	SALESMAN	7698	1981-12-02	1250	500	30
7566	JONES	MANAGER	7839	1981-02-04	2975	null	20
7654	MARTIN	SALESMAN	7698	1981-08-09	1250	1400	30
7698	SGR	MANAGER	7839	1981-01-05	2850	null	30
7782	RAVI	MANAGER	7839	1981-09-06	2450	null	10
7788	SCOTT	ANALYST	7566	1987-09-04	3000	null	20
7839	KING	PRESIDENT	null	1981-07-11	5000	null	10
7844	TURNER	SALESMAN	7698	1981-08-09	1500	0	30
7876	ADAMS	CLERK	7788	1987-03-05	1100	null	20
7900	JAMES	CLERK	7698	1981-03-12	950	null	30
7902	FORD	ANALYST	7566	1981-03-12	3000	null	20
7934	MILLER	CLERK	7782	1982-03-01	1300	null	10
1234	SEKHAR	doctor	7777	null	667	78	50

DEPT Table Data

Deptno	Dname	Loc
10	ACCOUNTING	NEW YORK
20	RESEARCH	DALLAS
30	SALES	CHICAGO
40	OPERATIONS	BOSTON

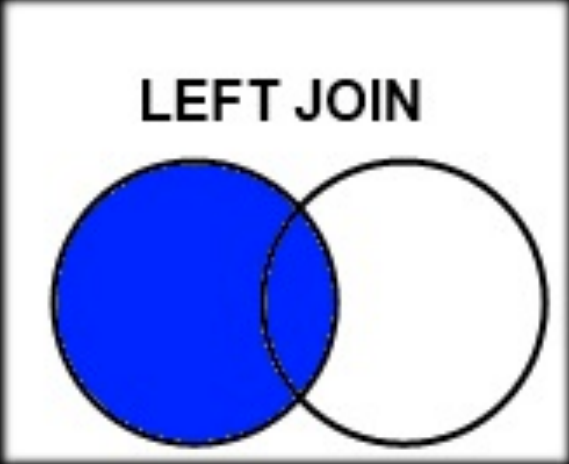




```
1 --techlake
2 --Spark SQL Joins -- inner join
3 select * from emp as e join dept as d on e.deptno = d.deptno
```

✓ inner join: Matching data from both tables/datasets based on join condition.

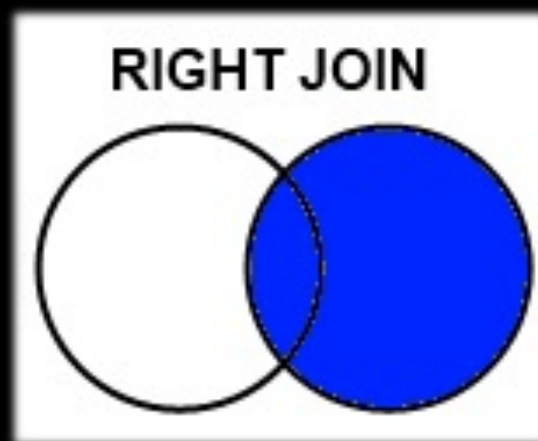
EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO	Deptno	Dname	Loc
7782	RAVI	MANAGER	7839	1981-09-06	2450	null	10	10	ACCOUNTING	NEW YORK
7839	KING	PRESIDENT	null	1981-07-11	5000	null	10	10	ACCOUNTING	NEW YORK
7934	MILLER	CLERK	7782	1982-03-01	1300	null	10	10	ACCOUNTING	NEW YORK
7369	SMITH	CLERK	7902	1980-10-12	800	null	20	20	RESEARCH	DALLAS
7566	JONES	MANAGER	7839	1981-02-04	2975	null	20	20	RESEARCH	DALLAS
7788	SCOTT	ANALYST	7566	1987-09-04	3000	null	20	20	RESEARCH	DALLAS
7876	ADAMS	CLERK	7788	1987-03-05	1100	null	20	20	RESEARCH	DALLAS
7902	FORD	ANALYST	7566	1981-03-12	3000	null	20	20	RESEARCH	DALLAS
7499	ALLEN	SALESMAN	7698	1981-10-02	1600	300	30	30	SALES	CHICAGO
7521	WARD	SALESMAN	7698	1981-12-02	1250	500	30	30	SALES	CHICAGO
7654	MARTIN	SALESMAN	7698	1981-08-09	1250	1400	30	30	SALES	CHICAGO
7698	SGR	MANAGER	7839	1981-01-05	2850	null	30	30	SALES	CHICAGO
7844	TURNER	SALESMAN	7698	1981-08-09	1500	0	30	30	SALES	CHICAGO
7900	JAMES	CLERK	7698	1981-03-12	950	null	30	30	SALES	CHICAGO



```
1 --techlake
2 --Spark SQL Joins -- left outer join
3 select * from emp as e left outer join dept as d on e.deptno = d.deptno
```

✓ left outer join : Matching data from both tables/datasets plus un-matched data from left table based on join condition.

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO	Deptno	Dname	Loc
7782	RAVI	MANAGER	7839	1981-09-06	2450	null	10	10	ACCOUNTING	NEW YORK
7839	KING	PRESIDENT	null	1981-07-11	5000	null	10	10	ACCOUNTING	NEW YORK
7934	MILLER	CLERK	7782	1982-03-01	1300	null	10	10	ACCOUNTING	NEW YORK
7369	SMITH	CLERK	7902	1980-10-12	800	null	20	20	RESEARCH	DALLAS
7566	JONES	MANAGER	7839	1981-02-04	2975	null	20	20	RESEARCH	DALLAS
7788	SCOTT	ANALYST	7566	1987-09-04	3000	null	20	20	RESEARCH	DALLAS
7876	ADAMS	CLERK	7788	1987-03-05	1100	null	20	20	RESEARCH	DALLAS
7902	FORD	ANALYST	7566	1981-03-12	3000	null	20	20	RESEARCH	DALLAS
7499	ALLEN	SALESMAN	7698	1981-10-02	1600	300	30	30	SALES	CHICAGO
7521	WARD	SALESMAN	7698	1981-12-02	1250	500	30	30	SALES	CHICAGO
7654	MARTIN	SALESMAN	7698	1981-08-09	1250	1400	30	30	SALES	CHICAGO
7698	SGR	MANAGER	7839	1981-01-05	2850	null	30	30	SALES	CHICAGO
7844	TURNER	SALESMAN	7698	1981-08-09	1500	0	30	30	SALES	CHICAGO
7900	JAMES	CLERK	7698	1981-03-12	950	null	30	30	SALES	CHICAGO
1234	SEKHAR	doctor	7777	null	667	78	50	null	null	null

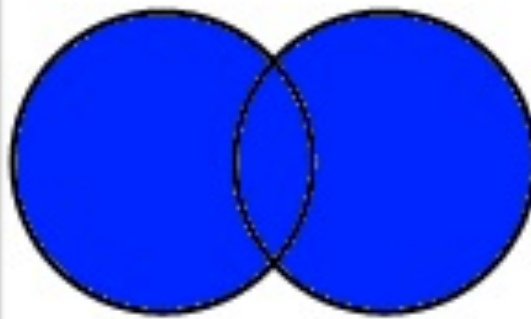


```
1 --techlake
2 --Spark SQL Joins -- right outer join
3 select * from emp as e right outer join dept as d on e.deptno = d.deptno
```

✓ Right Outer join: Matching data from both tables/datasets plus un-matched data from right table based on join condition.

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO	Deptno	Dname	Loc
7934	MILLER	CLERK	7782	1982-03-01	1300	null	10	10	ACCOUNTING	NEW YORK
7839	KING	PRESIDENT	null	1981-07-11	5000	null	10	10	ACCOUNTING	NEW YORK
7782	RAVI	MANAGER	7839	1981-09-06	2450	null	10	10	ACCOUNTING	NEW YORK
7902	FORD	ANALYST	7566	1981-03-12	3000	null	20	20	RESEARCH	DALLAS
7876	ADAMS	CLERK	7788	1987-03-05	1100	null	20	20	RESEARCH	DALLAS
7788	SCOTT	ANALYST	7566	1987-09-04	3000	null	20	20	RESEARCH	DALLAS
7566	JONES	MANAGER	7839	1981-02-04	2975	null	20	20	RESEARCH	DALLAS
7369	SMITH	CLERK	7902	1980-10-12	800	null	20	20	RESEARCH	DALLAS
7900	JAMES	CLERK	7698	1981-03-12	950	null	30	30	SALES	CHICAGO
7844	TURNER	SALESMAN	7698	1981-08-09	1500	0	30	30	SALES	CHICAGO
7698	SGR	MANAGER	7839	1981-01-05	2850	null	30	30	SALES	CHICAGO
7654	MARTIN	SALESMAN	7698	1981-08-09	1250	1400	30	30	SALES	CHICAGO
7521	WARD	SALESMAN	7698	1981-12-02	1250	500	30	30	SALES	CHICAGO
7499	ALLEN	SALESMAN	7698	1981-10-02	1600	300	30	30	SALES	CHICAGO
null	null	null	null	null	null	null	null	40	OPERATIONS	BOSTON

FULL OUTER JOIN

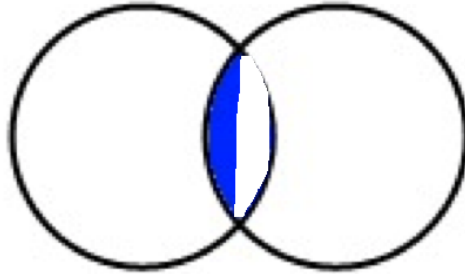


```
1 --techlake
2 --Spark SQL Joins -- full outer join
3 select * from emp as e full outer join dept as d on e.deptno = d.deptno
```

✓ Full Outer join : Matched and un-matched data from both tables based on join condition.

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO	Deptno	Dname	Loc
7782	RAVI	MANAGER	7839	1981-09-06	2450	null	10	10	ACCOUNTING	NEW YORK
7839	KING	PRESIDENT	null	1981-07-11	5000	null	10	10	ACCOUNTING	NEW YORK
7934	MILLER	CLERK	7782	1982-03-01	1300	null	10	10	ACCOUNTING	NEW YORK
7369	SMITH	CLERK	7902	1980-10-12	800	null	20	20	RESEARCH	DALLAS
7566	JONES	MANAGER	7839	1981-02-04	2975	null	20	20	RESEARCH	DALLAS
7788	SCOTT	ANALYST	7566	1987-09-04	3000	null	20	20	RESEARCH	DALLAS
7876	ADAMS	CLERK	7788	1987-03-05	1100	null	20	20	RESEARCH	DALLAS
7902	FORD	ANALYST	7566	1981-03-12	3000	null	20	20	RESEARCH	DALLAS
7499	ALLEN	SALESMAN	7698	1981-10-02	1600	300	30	30	SALES	CHICAGO
7521	WARD	SALESMAN	7698	1981-12-02	1250	500	30	30	SALES	CHICAGO
7654	MARTIN	SALESMAN	7698	1981-08-09	1250	1400	30	30	SALES	CHICAGO
7698	SGR	MANAGER	7839	1981-01-05	2850	null	30	30	SALES	CHICAGO
7844	TURNER	SALESMAN	7698	1981-08-09	1500	0	30	30	SALES	CHICAGO
7900	JAMES	CLERK	7698	1981-03-12	950	null	30	30	SALES	CHICAGO
1234	SEKHAR	doctor	7777	null	667	78	50	null	null	null
null	null	null	null	null	null	null	null	40	OPERATIONS	BOSTON

LEFT SEMI JOIN

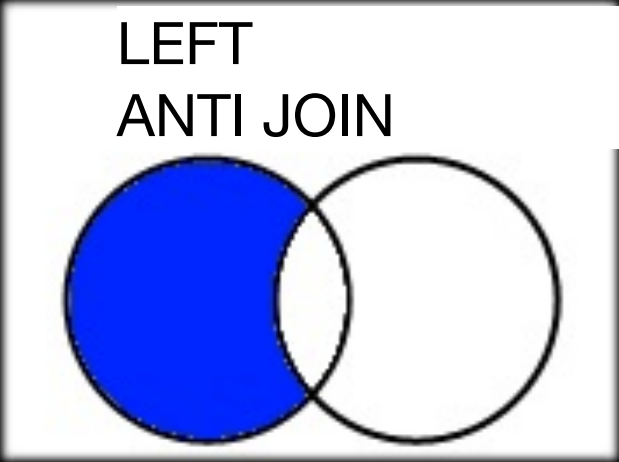


```
1 --techlake
2 --Spark SQL Joins -- left semi join
3 select * from emp as e left semi join dept as d on e.deptno = d.deptno
4 -- or using IN operator with where clause sub-query
5 select * from emp as e where deptno in (select deptno from dept as d)
6 --- using exists with correlated sub query
7 select * from emp as e where exists (select 'x' from dept as d where e.deptno = d.deptno)
```

✓ Left Semi Join : Only Matched data from left table based on join condition.

👉 Note: Its same as where clause subquery with IN or Correlated Sub-Query EXISTS

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO
7782	RAVI	MANAGER	7839	1981-09-06	2450	null	10
7839	KING	PRESIDENT	null	1981-07-11	5000	null	10
7934	MILLER	CLERK	7782	1982-03-01	1300	null	10
7369	SMITH	CLERK	7902	1980-10-12	800	null	20
7566	JONES	MANAGER	7839	1981-02-04	2975	null	20
7788	SCOTT	ANALYST	7566	1987-09-04	3000	null	20
7876	ADAMS	CLERK	7788	1987-03-05	1100	null	20
7902	FORD	ANALYST	7566	1981-03-12	3000	null	20
7499	ALLEN	SALESMAN	7698	1981-10-02	1600	300	30
7521	WARD	SALESMAN	7698	1981-12-02	1250	500	30
7654	MARTIN	SALESMAN	7698	1981-08-09	1250	1400	30
7698	SGR	MANAGER	7839	1981-01-05	2850	null	30
7844	TURNER	SALESMAN	7698	1981-08-09	1500	0	30
7900	JAMES	CLERK	7698	1981-03-12	950	null	30

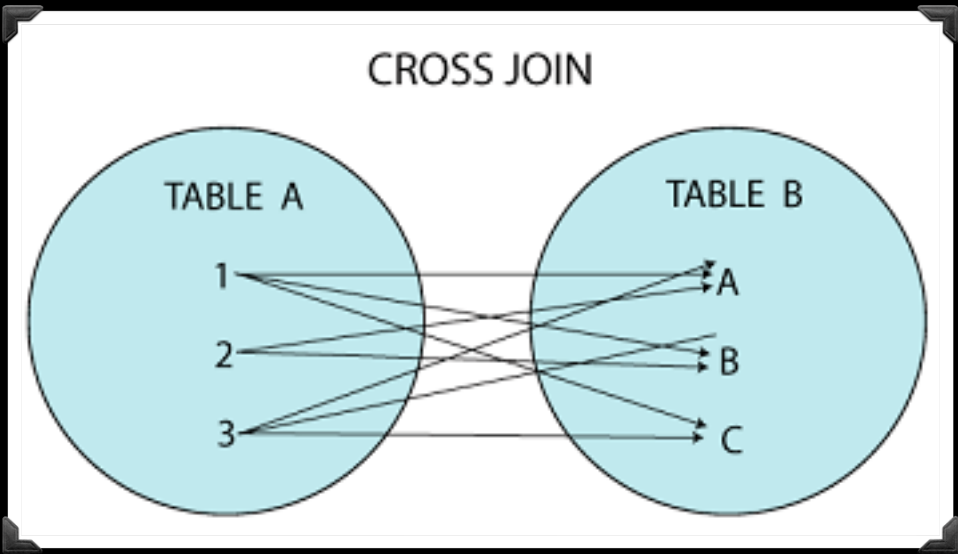


```
1 --techlake
2 --Spark SQL Joins -- left anti join
3 SELECT * FROM emp as e left anti JOIN dept as d ON e.deptno = d.deptno
4 -- or using NOT IN operator with where clause sub-query
5 SELECT * FROM emp as e WHERE deptno NOT IN (SELECT deptno FROM dept as d)
6 --- using NOT EXISTS with correlated sub query
7 SELECT * FROM emp as e where NOT EXISTS (SELECT 'x' FROM dept as d WHERE e.deptno = d.deptno)
```

✔ Left Anti Join : Only Un-Matched Data From Left table based on join condition.

👉 Note: Its same as where clause subquery with NOT IN or Correlated Sub-Query NOT EXISTS

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO
1234	SEKHAR	doctor	7777	null	667	78	50



```
1 --techlake
2 --Spark SQL Joins -- CROSS JOIN
3 SELECT * FROM emp as e CROSS JOIN dept as d
```

✓ Cross Join : if two tables does not have any common column for join condition then we can use cross join. it will return cross product of both tables.

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO	Deptno	Dname	Loc
7369	SMITH	CLERK	7902	1980-10-12	800	null	20	10	ACCOUNTING	NEW YORK
7369	SMITH	CLERK	7902	1980-10-12	800	null	20	20	RESEARCH	DALLAS
7369	SMITH	CLERK	7902	1980-10-12	800	null	20	30	SALES	CHICAGO
7369	SMITH	CLERK	7902	1980-10-12	800	null	20	40	OPERATIONS	BOSTON
7499	ALLEN	SALESMAN	7698	1981-10-02	1600	300	30	10	ACCOUNTING	NEW YORK
7499	ALLEN	SALESMAN	7698	1981-10-02	1600	300	30	20	RESEARCH	DALLAS
7499	ALLEN	SALESMAN	7698	1981-10-02	1600	300	30	30	SALES	CHICAGO
7499	ALLEN	SALESMAN	7698	1981-10-02	1600	300	30	40	OPERATIONS	BOSTON
7521	WARD	SALESMAN	7698	1981-12-02	1250	500	30	10	ACCOUNTING	NEW YORK
7521	WARD	SALESMAN	7698	1981-12-02	1250	500	30	20	RESEARCH	DALLAS
7521	WARD	SALESMAN	7698	1981-12-02	1250	500	30	30	SALES	CHICAGO
7521	WARD	SALESMAN	7698	1981-12-02	1250	500	30	40	OPERATIONS	BOSTON
7566	JONES	MANAGER	7839	1981-02-04	2975	null	20	10	ACCOUNTING	NEW YORK
7566	JONES	MANAGER	7839	1981-02-04	2975	null	20	20	RESEARCH	DALLAS
7566	JONES	MANAGER	7839	1981-02-04	2975	null	20	30	SALES	CHICAGO
7566	JONES	MANAGER	7839	1981-02-04	2975	null	20	40	OPERATIONS	BOSTON
7654	MARTIN	SALESMAN	7698	1981-08-09	1250	1400	30	10	ACCOUNTING	NEW YORK
7654	MARTIN	SALESMAN	7698	1981-08-09	1250	1400	30	20	RESEARCH	DALLAS



Learn & Lead



Like



Celebrate



Support



Funny



Love



Insightful



Curious