

Executive Summary of Hewlett-Packard Analysis

The following table summarizes the key results. Details are contained in the report.

| Questions | Result/Recommendation | | | | | | | | | | | | | | | | | | |
|---|---|-------|--------|-------|-----------------|-------|--------------|------|----------|------|-------|------|------------------|-----|--------------------|---|---------|--|--|
| Number of individuals retiring | 33,118 – approximately 10% of worldwide employee base | | | | | | | | | | | | | | | | | | |
| Number of individuals to be hired | Based on 1-1 Replacements. See additional recommendations in detailed report. <table><tr><td>count</td><td>titles</td></tr><tr><td>13651</td><td>Senior Engineer</td></tr><tr><td>12872</td><td>Senior Staff</td></tr><tr><td>2711</td><td>Engineer</td></tr><tr><td>2022</td><td>Staff</td></tr><tr><td>1609</td><td>Technique Leader</td></tr><tr><td>251</td><td>Assistant Engineer</td></tr><tr><td>2</td><td>Manager</td></tr></table> | count | titles | 13651 | Senior Engineer | 12872 | Senior Staff | 2711 | Engineer | 2022 | Staff | 1609 | Technique Leader | 251 | Assistant Engineer | 2 | Manager | | |
| count | titles | | | | | | | | | | | | | | | | | | |
| 13651 | Senior Engineer | | | | | | | | | | | | | | | | | | |
| 12872 | Senior Staff | | | | | | | | | | | | | | | | | | |
| 2711 | Engineer | | | | | | | | | | | | | | | | | | |
| 2022 | Staff | | | | | | | | | | | | | | | | | | |
| 1609 | Technique Leader | | | | | | | | | | | | | | | | | | |
| 251 | Assistant Engineer | | | | | | | | | | | | | | | | | | |
| 2 | Manager | | | | | | | | | | | | | | | | | | |
| Number of individuals available for mentorship role | 1549 – (based on criteria of “active employees born in 1965) | | | | | | | | | | | | | | | | | | |
| Additional Recommendation | Use a different criteria for selecting mentors which is better correlated to their success. | | | | | | | | | | | | | | | | | | |
| Additional Recommendation | Obtain updated salary data so salary impact can be accurately assessed. | | | | | | | | | | | | | | | | | | |

For questions concerning this analysis and its recommendations please contact:

Thomas Cottrell

tcottrell@pewlett-packard.com

mobile: 650-245-4033

Pewlett-Hackard Analysis

Overall Description

Pewlett-Hackard is a large established company with over 300,000 employees worldwide. They have a large number of people near retirement age, and the company wants to offer early retirement packages to move those employees on to retirement, do succession planning in the management ranks and refresh the workforce. Senior Managers and HR need reports to answer the following questions:

- 1) How many employees overall will be offered the retirement package?
- 2) How many retirees are there by each of the 9 functional departments?
- 3) Who are the managers of those departments who will need the lists for planning?
- 4) What is the impact on overall salary costs?
- 5) What is the impact by job title?
- 6) Who are potential mentors which might be able to help train the new hires?

Resources

- 6 CSV files from HR on employee and department data.
- PregresQL Database (free version)
- QuickDBD ERD Design Tool (free version)
- Visual Studio (capturing queries)

Exploratory Analysis

We were provided with the following CSV files to start the analysis:

- 1) departments.csv (2 X 10 array of data listing the 9 Departments and their names).
- 2) dept_emp.csv (4 X 331,604 listing historical assignment of employees to departments).
- 3) dept_manager.csv (4 X 25 – listing historical assignment of managers to the 9 Departments).
- 4) Employees.csv (6 X 300,025 – listing historical and active employees with name, hire date, birth date, gender and employee number).
- 5) Salaries.csv (4 X 300,025 – listing salary by employee and dates that salary went into effect).
- 6) Titles.csv (4 X 443,309 – listing historical titles for current and past employees).

Using excel, each file was quickly examined for data types missing records, etc. Given that some tables contained current data (i.e. departments) while others contained historical data (i.e. dept_manager) this suggested some additional attention would need to be given to duplicate rows, filtering etc.

In addition, the “salaries” table of data looked odd or incomplete as a quick sort revealed that no salary increases had occurred since 2001 with over half of the employees not receiving an

increase since 1993. Although companies do freeze wages on occasion when experiencing financial difficulty, the length of this freeze suggests we do not have the latest data on “salaries” and we will have to do further investigation with HR and the Finance Dept.

Entity Relationship Diagram (ERD)

Using QuickDBD we designed an ERD shown below, then implemented the design using SQL queries.



Creating the Physical Database in PostgreSQL

Using PostgreSQL the SQL code was then written to define each table in the database. Some sample code is shown below with the full code shown in file `queries_original_6`:

Sample SQL to produce tables:

```
CREATE TABLE dept_manager (  
  dept_no VARCHAR(4) NOT NULL,  
  emp_no INT NOT NULL,  
  from_date DATE NOT NULL,  
  to_date DATE NOT NULL,  
  FOREIGN KEY (emp_no) REFERENCES employees (emp_no),  
  FOREIGN KEY (dept_no) REFERENCES departments (dept_no),  
  PRIMARY KEY (emp_no, dept_no)  
);
```


```
CREATE TABLE departments (  
  dept_no VARCHAR(4) NOT NULL,  
  dept_name VARCHAR(40) NOT NULL,  
  PRIMARY KEY (dept_no),  
  UNIQUE (dept_name)  
);
```

Error Checking and Data Loads

As the tables were built out, code was debugged for syntax and other errors using the SELECT statement.

```
SELECT * FROM dept_emp;  
SELECT * FROM departments;  
SELECT * FROM dept_manager;  
SELECT * FROM employees;  
SELECT * FROM salaries;  
SELECT * FROM title;
```

Prior to loading data, the SELECT statement would produce the following output if no errors were found. We were then ready to load the data from the CSV files:

| Data Output | Explain | Messages | Notifications |
|--|---------|-------------------------------------|---------------|
|  dept_no character varying (4) | | dept_name character varying (40) | |
| | | | |

Loading Data into Tables

Using the import function, we iteratively loaded the data into the tables, debugging as we went along. Errors were encountered based on:

- Order in which we executed loads
- Duplicate records caused by historical data.

These problems were solved by changing the loading order and removing some restrictions on table definitions such as UNIQUE. Tables were DROPPED and then recreated solving the loading problems.

Data Integrity Thoughts

Although, the loading problems were solved – to preserve data integrity and future needs of the HR Dept and Managers, the historical data (duplicate records) were loaded into the Database tables. This data was filtered out using specific queries for Challenge 7 but not during the exercises and requested outputs for Section 7.3.5 described below.

Section 7.3.5 Intermediary Queries

Additional queries and files were produced to answer the questions by the Sales Manager, Development Manager and other managers. The queries for each – and their related CSV files are located in the repository under queries_section7_3_5_intermediate_results and the data folder, respectively.

Specific queries for:

- employees_by_dept
- emp_info
- manager_info
- dept_info
- sales_info
- sales_dev_info

These were produced before duplicate records were filtered. Samples are shown below some of the CSV files.

| count | dept_no |
|-------|---------|
| 2199 | d001 |
| 1908 | d002 |
| 1953 | d003 |
| 8174 | d004 |
| 9281 | d005 |
| 2234 | d006 |
| 5860 | d007 |
| 2413 | d008 |
| 2597 | d009 |

| emp_no | first_name | last_name | gender | salary | to_date |
|--------|------------|------------|--------|--------|----------|
| 10001 | Georgi | Facello | M | 60117 | 1/1/9999 |
| 10004 | Chirstian | Koblick | M | 40054 | 1/1/9999 |
| 10009 | Sumant | Peac | F | 60929 | 1/1/9999 |
| 10018 | Kazuhide | Peha | F | 55881 | 1/1/9999 |
| 10035 | Alain | Chappelet | M | 41538 | 1/1/9999 |
| 10053 | Sanjiv | Zschoche | F | 67854 | 1/1/9999 |
| 10058 | Berhard | McFarlin | M | 52787 | 1/1/9999 |
| 10066 | Kwee | Schusler | M | 69736 | 1/1/9999 |
| 10067 | Claudi | Stavenow | M | 44642 | 1/1/9999 |
| 10070 | Reuven | Garigliano | M | 55999 | 1/1/9999 |
| 10072 | Hironoby | Sidou | F | 40000 | 1/1/9999 |
| 10076 | Erez | Ritzmann | F | 47319 | 1/1/9999 |
| 10088 | Jungsoon | Syrzycki | F | 65957 | 1/1/9999 |
| 10100 | Hironoby | Haraldson | F | 64308 | 1/1/9999 |

| emp_no | first_name | last_name | dept_name |
|--------|---------------|-----------|-----------|
| 10053 | Sanjiv | Zschoche | Sales |
| 10088 | Jungsoon | Syrzycki | Sales |
| 10215 | Xiaobin | Duclos | Sales |
| 10355 | Satyanarayana | Cochrane | Sales |
| 10463 | Ung | Zaiane | Sales |
| 10504 | Xiong | Varker | Sales |
| 10548 | Ramalinga | Gundersen | Sales |
| 10767 | Monique | Doering | Sales |
| 10795 | Supot | Naudin | Sales |
| 10836 | Jianhua | Leivant | Sales |

Challenge 7

Due to the issues with duplicate data caused from historical records, it was decided that 2 new tables would be produced excluding these duplicate records and then all other related tables required for Challenge 7 to be redone. The alternative to this would be to use more complicated sql code to “clean up” joined tables. The former approach is cleaner, simpler and resolves the issue earlier in the data-pipeline resulting in less sql code.

Two new key tables produced were title1 and dept_emp1. Filters were applied to ensure they only contained “active employees”, and used the latest titles. From here we built 4 new queries contained in file *queries_for_challenge7* to answer the corresponding questions.

Number of Retirees

Called the “retirees” query, it produced a list as follows:

| 1 | emp_no | first_name | last_name | titles | from_date | salary | |
|---------|--------|------------|-----------|-----------------|------------|--------|--|
| 33109 | 499880 | Fen | Lenart | Senior Engineer | 10/24/1997 | 40000 | |
| 33110 | 499898 | Alagu | Azuma | Senior Staff | 9/11/1992 | 40000 | |
| 33111 | 499907 | Arnd | Garnham | Senior Engineer | 2/17/1994 | 49934 | |
| 33112 | 499908 | Toong | Coorg | Senior Engineer | 12/3/1995 | 40356 | |
| 33113 | 499920 | Christ | Murtagh | Senior Staff | 4/17/1995 | 73573 | |
| 33114 | 499922 | Mitsuyuki | Doering | Senior Staff | 7/3/1991 | 50525 | |
| 33115 | 499927 | Manohar | Heemskerk | Engineer | 11/19/1996 | 70546 | |
| 33116 | 499937 | Pantung | Litzler | Senior Engineer | 2/2/1993 | 40000 | |
| 33117 | 499940 | Yolla | Auria | Senior Staff | 4/25/1986 | 43417 | |
| 33118 | 499966 | Mihalis | Crabtree | Senior Staff | 6/13/1990 | 75579 | |
| → 33119 | 499986 | Nathan | Ranta | Senior Staff | 8/11/1992 | 91988 | |
| 33120 | | | | | | | |
| 33121 | | | | | | | |

Total Number of eligible retirees = 33118. (rows-1). The list contains all of the requested information on:

- Employee Number
- First and last name
- Current title
- The from_date on which that title became active
- Salary

The full file is contained in the repository under *retirees.csv*. It was filtered using the following conditions:

- Active employees only
- Employees whose birth_date is between 1952 – 1955
- Employees who were hired between 1985-1988

Retiring Titles

To understand how this large number of individuals will be replaced with new hires, its important to understand their titles and roles. A new query was created with its output called *title1_count*.

| count | titles |
|-------|--------------------|
| 13651 | Senior Engineer |
| 12872 | Senior Staff |
| 2711 | Engineer |
| 2022 | Staff |
| 1609 | Technique Leader |
| 251 | Assistant Engineer |
| 2 | Manager |

You can see that we are dealing with quite a large number (13,651) of Senior Engineers. This number of Engineers will be highly difficult to source so development of junior staff will need to be developed along with leaders such as “Technique Leader,” who control the processes within the company.

In addition, 2 Department Mgrs will be retired and must be replaced.

Implementation and Hiring Strategy

It is suggested to:

- Divide this list down into 4 years to correspond to ages of the target retirees born between 1952 - 1955, resulting in about 7-8K new hires per year.
- Further division can be done across the 9 departments putting the number somewhere at about 800 employees per department per year for retiring and corresponding hiring.
- Additional information on geographic location of employees would allow HR and Managers to understand “where” openings need to be filled.
- In addition, First Level Managers will need to evaluate if a replacement is needed, or if efficiencies/changes in the business conditions require a replacement – with the opportunity to do things differently at a lower cost structure.

Additional Query

An additional query was built managers_by_dept1 showing all 9 Department Mgrs.

| dept_name | dept_no | first_name | last_name | emp_no | titles |
|--------------------|---------|------------|------------|--------|---------|
| Marketing | d001 | Vishwani | Minakawa | 110039 | Manager |
| Finance | d002 | Isamu | Legleitner | 110114 | Manager |
| Human Resources | d003 | Karsten | Sigstam | 110228 | Manager |
| Production | d004 | Oscar | Ghazalie | 110420 | Manager |
| Development | d005 | Leon | DasSarma | 110567 | Manager |
| Quality Management | d006 | Dung | Pesch | 110854 | Manager |
| Sales | d007 | Hauke | Zhang | 111133 | Manager |
| Research | d008 | Hilary | Kambil | 111534 | Manager |
| Customer Service | d009 | Yuchang | Weedman | 111939 | Manager |

Use of Mentors

To assist with this massive training and development effort, “Mentors” must be identified. It was suggested by HR that a list be created filtered by age (those born in 1965). The filtered list can be found in the *mentors.csv* file and a sample appears below.

| emp_no | first_name | last_name | titles | from_date | to_date |
|--------|------------|-------------|------------------|------------|----------|
| 10095 | Hilari | Morton | Senior Staff | 3/9/2000 | 1/1/9999 |
| 10122 | Ohad | Esposito | Technique Leader | 8/6/1998 | 1/1/9999 |
| 10291 | Dipayan | Seghrouchni | Senior Staff | 3/30/1994 | 1/1/9999 |
| 10476 | Kokou | Iisaka | Senior Staff | 9/20/1994 | 1/1/9999 |
| 10663 | Teunis | Noriega | Technique Leader | 1/23/1993 | 1/1/9999 |
| 10762 | Lech | Himler | Senior Staff | 1/21/1999 | 1/1/9999 |
| 10933 | Juyoung | Seghrouchni | Senior Engineer | 8/2/1993 | 1/1/9999 |
| 12155 | Keiichiro | Glinert | Senior Engineer | 9/16/2001 | 1/1/9999 |
| 12408 | Rasiah | Sudkamp | Senior Engineer | 4/18/1995 | 1/1/9999 |
| 12643 | Morrie | Schurmann | Staff | 12/31/1998 | 1/1/9999 |
| 13499 | Kazuhiko | Sidou | Technique Leader | 9/1/1991 | 1/1/9999 |
| 14075 | Denny | Tchuente | Engineer | 5/30/1998 | 1/1/9999 |
| 14104 | Sudhanshu | Demian | Senior Staff | 7/23/1999 | 1/1/9999 |
| 14127 | Magdalena | Cannard | Senior Engineer | 4/22/2002 | 1/1/9999 |
| 14158 | Caolyn | Setiz | Senior Staff | 2/6/1996 | 1/1/9999 |

Number of Mentors

1549 potential mentors have been identified using the “born in 1965” criteria. However, given that this role may lead to promotions and higher visibility down the road, it may be a violation of the “discrimination” laws, or result in an “overweighted class” of individuals rather than a cross-section of the future employee base. At the very least, using age as a selection criteria may be a poor business decision as “age” may not correlate with the candidates ability to “mentor,” or their “job performance.”

Recommendation on Further Analysis on Data Set

It is recommended that mentors be selected on a different criteria such as:

- Candidates ranked X on performance reviews.
- Candidates nominated by their supervisors based on performance.
- Candidates in their current position X minimum amount of time.

Additional data would need to be added to the database and queries written for the new criteria. The current query is contained in file *mentors*.

Final Recommendation on Salary Impact

As mentioned previously, the salary data is highly questionable based on the lack of recent entries within the last 20 years. It is recommended that:

- Updated salary data be obtained from HR or Finance, rerunning the retiree list so managers can accurately assess impact on reduction of salary vs retirement package