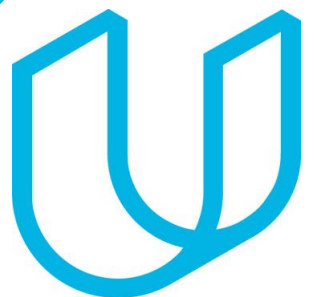# Tech ABC Corp - HR Database
Hock Chong - 12 April 2024

Business Scenario

Business requirement

Tech ABC Corp saw explosive growth with a sudden appearance onto the gaming scene with their new AI-powered video game console. As a result, they have gone from a small 10 person operation to 200 employees and 5 locations in under a year. HR is having trouble keeping up with the growth, since they are still maintaining employee information in a spreadsheet. While that worked for ten employees, it has becoming increasingly cumbersome to manage as the company expands.

As such, the HR department has tasked you, as the new data architect, to design and build a database capable of managing their employee information.

Dataset

The [HR dataset](#) you will be working with is an Excel workbook which consists of 206 records, with eleven columns. The data is in human readable format, and has not been normalized at all. The data lists the names of employees at Tech ABC Corp as well as information such as job title, department, manager's name, hire date, start date, end date, work location, and salary.

IT Department Best Practices

The IT Department has certain Best Practices policies for databases you should follow, as detailed in the [Best Practices document](#).

Step 1

Data Architecture

Foundations

Step 1: Data Architecture Foundations

Hi,

Welcome to Tech ABC Corp. We are excited to have some new talent onboard. As you may already know, Tech ABC Corp has recently experienced a lot of growth. Our AI powered video game console WOPR has been hugely successful and as a result, our company has grown from 10 employees to 200 in only 6 months (and we are projecting a 20% growth a year for the next 5 years). We have also grown from our Dallas, Texas office, to 4 other locations nationwide: New York City, NY, San Francisco, CA, Minneapolis, MN, and Nashville, TN.

While this growth is great, it is really starting to put a strain on our record keeping in HR. We currently maintain all employee information on a shared spreadsheet. When HR consisted of only myself, managing everyone on an Excel spreadsheet was simple, but now that it is a shared document I am having serious reservations about data integrity and data security. If the wrong person got their hands on the HR file, they would see the salaries of every employee in the company, all the way up to the president.

After speaking with Jacob Lauber, the manager of IT, he suggested I put in a request to have my HR Excel file converted into a database. He suggested I reach out to you as I am told you have experience in designing and building databases. When you are building this, please keep in mind that I want any employee with a domain login to be have read only access the database. I just don't want them having access to salary information. That needs to be restricted to HR and management level employees only. Management and HR employees should also be the only ones with write access. By our current estimates, 90% of users will be read only.

I also want to make sure you know that am looking to turn my spreadsheet into a live database, one I can input and edit information into. I am not really concerned with reporting capabilities at the moment. Since we are working with employee data we are required by federal regulations to maintain this data for at least 7 years; additionally, since this is considered business critical data, we need to make sure it gets backed up properly.

As a final consideration. We would like to be able to connect with the payroll department's system in the future. They maintain employee attendance and paid time off information. It would be nice if the two systems could interface in the future

I am looking forward to working with you and seeing what kind of database you design for us.

Thanks,
Sarah Collins
Head of HR

Data Architect Business Requirement

- **Purpose of the new database:**

  design and build a secure and efficient system for HR record keeping at Tech ABC Corp, taking into consideration the company's rapid growth and the need to manage employee information effectively. The new database will address

  - Data security: Ensure that sensitive information, such as salary details, is protected from unauthorized access. Implement role-based access controls to restrict access to salary information to HR and management level employees only.
  - Data Integrity: Ensure that the data stored in the database is accurate and reliable. Implement measures to prevent data duplication and ensure consistency across the database.
  - Scalability: Design the database to handle the company's projected growth, including the ability to accommodate a large number of employees and locations.
  - Compliance: Ensure that the database complies with federal regulations regarding the maintenance of employee data, including the requirement to retain data for at least 7 years.

- **Describe current data management solution:**

  The current data is stored in excel and it's a share document. If the wrong person have the access to the file, they would be able to access entire employee's salary including director.

- **Describe current data available:**

  The data lists the names of employees at Tech ABC Corp, as well as information such as job title, department, manager's name, hire date, start date, end date, work location, and salary.

- **Additional data requests:**

  The database able to connect with the HR department's payroll system in the future.

- **Who will own/manage data**

  The management and HR employees will be own and manage the data.

Data Architect Business Requirement

- **Who will have access to database**

    There will be two type of user who have the access to database:

    1. Employee have read access only to database but restricted the access to salary information. Estimate 90% of users will be read only.
    2. HR and Management level employees are the only one with write access.

- **Estimated size of database**

    The size of database consisting of 206 records, with eleven columns. With the projection of 20% growth per year, we estimate increase 41 new records with eleven columns per year.

- **Estimated annual growth**

    Current employees grow from 10 to 200 employees in only 6 months. Projection 20% growth per year within next 5 years.

- **Is any of the data sensitive/restricted**

    The only sensitive data is employee's salary. Only HR and Management Employee can access it.

Data Architect Technical Requirement

- **Justification for the new database**
    1. Improved Data Security: Moving from a shared spreadsheet to a database can significantly enhance data security. Databases offer features such as access control, encryption, and auditing, which can help protect sensitive information like employee salaries from unauthorized access.
    2. Scalability and Efficiency: With the company projected to grow by 20% annually for the next five years, a database provides a more scalable solution for managing employee information compared to a spreadsheet. Databases can handle large volumes of data more efficiently and can easily accommodate future growth and additional data needs.

- **Database objects**

    List the database objects (tables, views, special procedures) that will be created for the database.
    Hint - you may want to circle back to this answer after completing the logical ERD in step 2.

- **Data ingestion**

    Since excel file is flat file, ETL is the best approach based on IT Best Practice standards.

Data Architect Technical Requirement

- **Data governance (Ownership and User access)**

  Ownership: HR and management level employees

  User Access: HR and management level employees will have full access. Regular employee will have restricted access.

- **Scalability**

  Sharding should not be used as more te

  Should replication or sharding be used to ensure scalability based on user needs

- **Flexibility**

  Consider setup up automated ETL process if the flat file will regular updated

- **Storage & retention**

  Storage (disk or in-memory): Based on IT best practice, all databases are given a standard partition of 1 GB by the server group. We only need larger partition space if the data expected to exceed 10k row of data in next year.

  Retention: The database will store employee data for at least 7 years to comply with federal regulations.

- **Backup**

  Based on IT Best Practices, for critical business data, Backup schedule is full backup 1x per week, incremental backup daily.