DATA ANALYSIS AND DESIGN

ASSIGNMENT

SEPTEMBER 2017

Sydney Twigg
M8C3XRSN8

CONTENTS

QUESTION 1

Staff (staffNo, name, dept, skillCode)
Skill (skillCode, description, chargeOutRate)
Project (projectNo, startDate, endDate, budget, projectManagerStaffNo)
Booking (staffNo, projectNo, dateWorkedOn, timeWorkedOn)

QUESTION 1.1.A.

List all skills with a charge out rate greater than 60 per hour, in alphabetical order of description.

```
1.  SELECT *
2.  FROM skill
3.  WHERE chargeOutRate > 60
4.  ORDER BY description ASC;
```

Screenshot showing SQL code

References used: (Connolly & Begg, 2004).

QUESTION 1.1.B.

List all staff with the skill description 'Programmer' who work in the 'Special Projects' department.

Assuming skillCode '1' is 'Programmer'

```
1.  SELECT staffNo, name
2.  FROM staff
3.  INNER JOIN skill
4.  ON staff.skillCode = skill.skillCode
5.  WHERE skill.skillCode = 1
6.  AND staff.dept = 'Special Projects';
```

Screenshot showing SQL code

References used: (W3 Schools, 2017) (Connolly & Begg, 2004).

QUESTION 1.1.C.

For all projects that were active in July 1995, list the staff name, project number and the date and number of hours worked on the project, ordered by staff name, within staff name by the project number and within project number by date.

```
1.  SELECT name, proj.projectNo, dateWorkedOn, timeWorkedOn
2.  FROM Staff staff, Project proj, Booking book
3.  WHERE staff.staffNo = book.staffNo
4.  AND book.projectNo = proj.projectNo
5.  AND endDate >= DATE '1995-07-01'
6.  ORDER BY name, proj.projectNo, dateWorkedOn;
```

Screenshot showing SQL code

References used: (W3 Schools, 2017).

QUESTION 1.1.D.

List all projects that have at least two staff booking.

```
1.  SELECT projectNo, COUNT(*)
2.  FROM booking
3.  GROUP BY projectNo
4.  HAVING COUNT(*) >= 2;
```

Screenshot showing SQL code

References used: (W3 Schools, 2017).

QUESTION 1.1.E.

How many staff have the skill 'Programmer'?

```
1.  SELECT COUNT(*)
2.  FROM staff s, skill sk
3.  WHERE s.skillCode = sk.skillCode
4.  AND description = 'Programmer';
```

Screenshot showing SQL code

References used: (Connolly & Begg, 2004)

QUESTION 1.1.F.

List all staff with a charge out rate greater than the average charge out rate.

```
1.  SELECT *
2.  FROM staff s, skill sk
3.  WHERE s.skillCode = sk.skillCode
4.  AND chargeOutRate > (SELECT AVG(chargeOutRate) FROM skill);
```

Screenshot showing SQL code

References used: (Connolly & Begg, 2004) (W3 Schools, 2017).

QUESTION 1.1.G.

Create a view of staff details, giving the staff number, staff name, skill description, and department, but excluding the skill number and charge out rate.

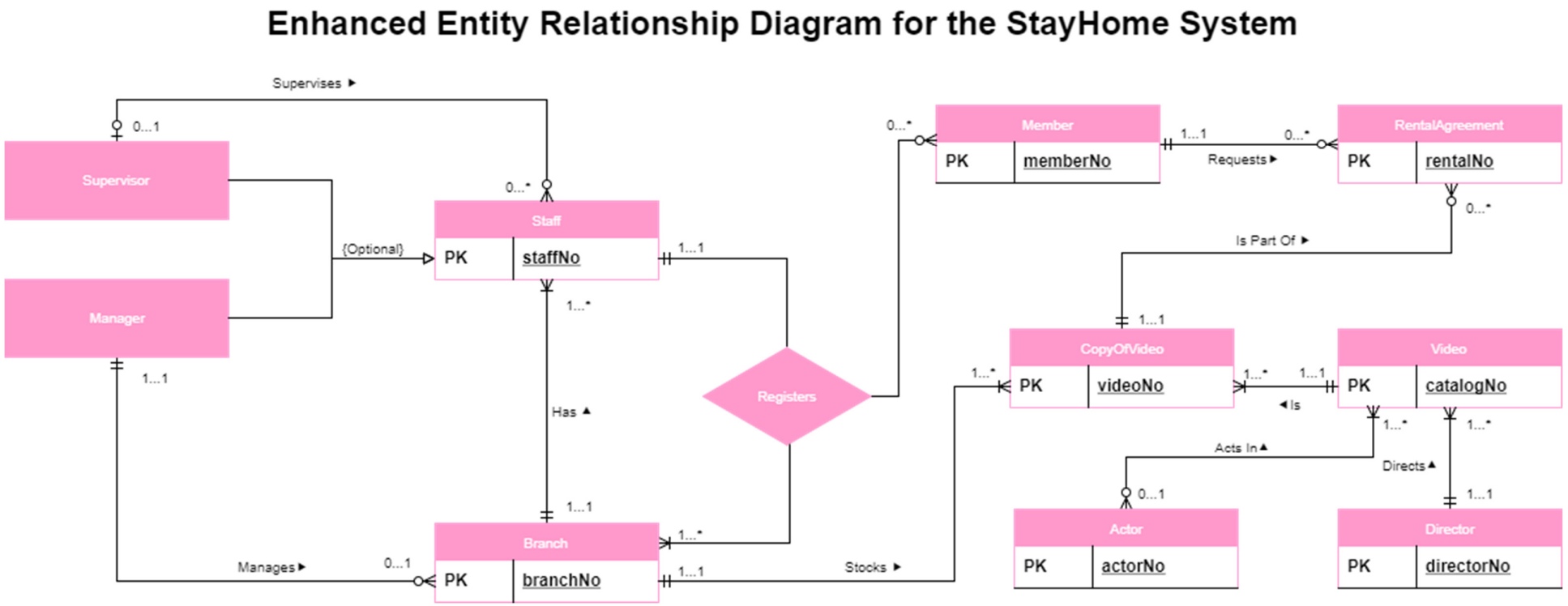
```
1.  CREATE VIEW staffDetails(staffNo, name, description, dept)
2.  AS SELECT staffNo, name, description, dept
3.  FROM staff s, skill sk
4.  WHERE s.skillCode = sk.skillCode;
```

Screenshot showing SQL code

References used: (W3 Schools, 2017).

QUESTION 2

QUESTION 2.1.A.



Enhanced Entity Relationship Diagram for the StayHome System

References used: (Connolly & Begg, 2004) (Connoly & Begg, 2015).

QUESTION 2.1.B.

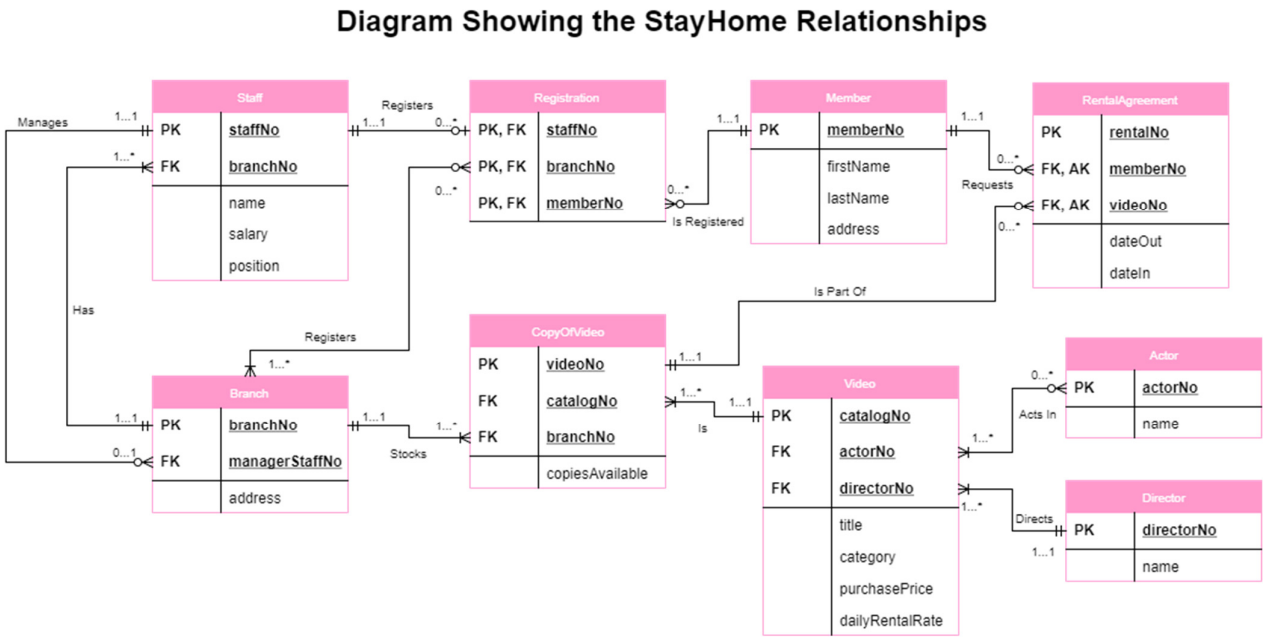


Diagram showing StayHome Relationships, Entities and Primary Keys.

StayHome Relationships

Entity	Multiplicity	Relationship	Multiplicity	Entity
Staff	1…1	Manages	0…1	Branch
Staff	1…1	Registers	0…*	Member
Branch	1…1	Has	1…*	Staff
Branch	1…*	Registers	0…*	Member
Branch	1…1	Stocks	1…*	CopyOfVideo
CopyOfVideo	1…1	Is Part Of	0…*	RentalAgreement
Member	1…1	Requests	0…*	RentalAgreement
Video	1…1	Is	1…*	CopyOfVideo
Actor	0…*	Acts In	1…*	Video
Director	1…1	Directs	1…*	Video

References used: (Connolly & Begg, 2004) (Connoly & Begg, 2015).

QUESTION 3

QUESTION 3.1.

Data Warehousing is defined by (Connoly & Begg, p. 1225) as a subject-oriented, integrated, time-variant, and non-volatile collection of data in support of management's decision-making process. A data warehouse is a central repository for data collected by all of an enterprise's business systems, it emphasises the capture of data from a range of sources that is useful for analysis an access (Rouse, 2017). Data warehouses are run on enterprise mainframes, or more recently - the cloud, and allow for data from various sources to be extracted for use by analytical applications and user queries (Rouse, 2017).

Online Analytical Processing is defined by (Connoly & Begg, p. 1286) as the dynamic synthesis, analysis, and consolidation of large volumes of multidimensional data. OLAP allows a user to extract and view selective data from different points of view, with all data being stored in a multidimensional database - where each data attribute is considered a separate dimension; OLAP software locates the intersection of dimensions and displays them to show the stored data and its trends (Rouse, 2017).

The following table contrasts Online Analytical Processing (OLAP) and Data Warehousing Systems:

Characteristic	Data Warehousing System	Online Analytical Processing
Main Purpose	Support analytical processing (Connoly & Begg, 2015).	Support planning, problem solving and decision support (datawarehouse4u, 2009).
Data Age	Historic, however, it is trending towards including current data (Connoly & Begg, 2015).	Historical or Archival Data (datawarehouse4u, 2009).
Data Latency	Dependant on the length of cycle data for supplements to the warehouse, however, it is trending towards real time supplements (Connoly & Begg, 2015).	Dependant on the amount of data involved, batch data refreshes and complex queries can take a very long time - the speed can be improved through the use of indexes (datawarehouse4u, 2009).
Data Granularity	Detailed data - both lightly and highly summarised data (Connoly & Begg, 2015).	Provides summarized and consolidated data (TutorialsPoint, 2017).
Data Processing	Less predictable pattern of data queries; low-to-medium levels of transaction throughput (Connoly & Begg, 2015).	Complex queries involving aggregations (datawarehouse4u, 2009).
Reporting	Unpredictable, multidimensional, dynamic reporting (Connoly & Begg, 2015).	Provides summarized and multidimensional view of data (TutorialsPoint, 2017).
Users	Used by managerial users, however, it is trending towards supporting the analytical requirements for operational users (Connoly & Begg, 2015).	Used by knowledge workers such as executives, managers, and analysts (TutorialsPoint, 2017).

In data warehouse environments, different types of analysis form a platform to support business intelligence - OLAP is one of these types of analysis. OLAP supports business intelligence applications by allowing applications to support complex multidimensional calculations directly against the database without the need to replicate the large volumes of data needed - thus providing a single database that is manageable, scalable and accessible to business intelligence applications (Connoly & Begg, 2015).

QUESTION 3.2.

Data mining is defined by (Connoly & Begg, p. 1316) as the process of extracting valid, previously unknown, comprehensible, and actionable information from large databases and using it to make crucial business decisions. Data mining is the process of analysing data to find hidden and unexpected patterns and relationships in sets of data (Connoly & Begg, 2015).

Data mining, Online Analytical Processing (OLAP) and data warehousing are connected through business intelligence - data mining and OLAP form a platform within data warehousing to support business intelligence (Connoly & Begg, 2015). The data warehouse provides the foundation for the data - storing the repository of a business's collected data; and as discussed previously in Question 3.1., OLAP provides a single manageable, scalable and accessible database from the information stored in the data warehouse (Connoly & Begg, 2015). Data mining can be used to support business intelligence by analysing the OLAP produced data, as well as all of the stored data in a data warehouse to provide users with hidden and unexpected patterns and relationships from the data.

An example situation of how data mining, Online Analytical Processing (OLAP) and data warehousing could be used together is when analysing sales: the data could be data mined to determine which items are commonly purchased together, and from this data an analyst could compare various dimensions of shopping patterns to determine how to market these items together better (TechTarget, 2017).

Data mining and data warehousing are complementary as data mining requires a single, separate, clean, integrated, self-consistent source of data - all of which can be provided by data warehousing (Connoly & Begg, 2015). Data warehouses contain data from various business enterprise systems, providing multiple sources to mine data from allowing the discovery of more patterns and interrelationships (Connoly & Begg, 2015).

The following diagram illustrates the data flow relationship between data mining, Online Analytical Processing (OLAP) and data warehousing:

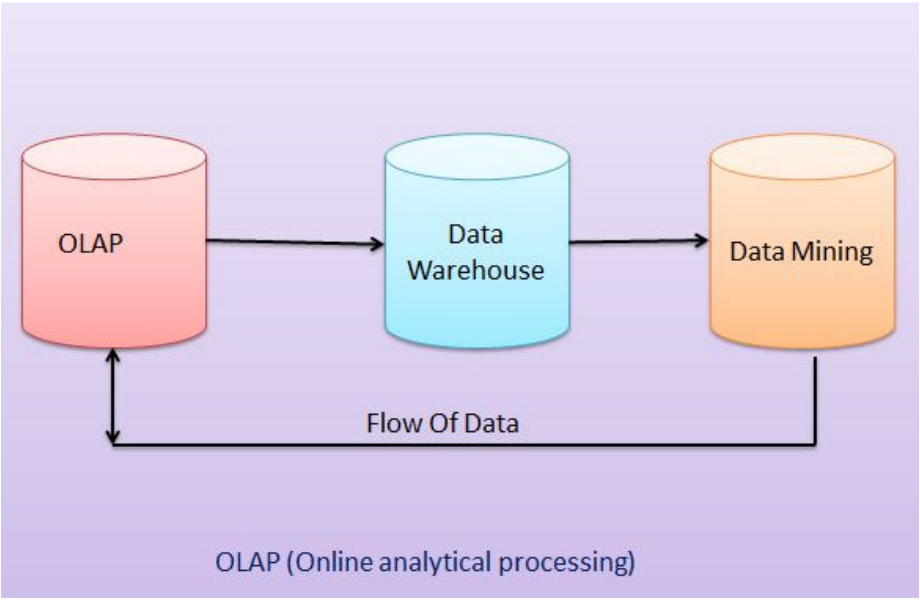


Illustration showing the relationship between data mining, Online Analytical Processing (OLAP) and data warehousing. Source: (Tiwari, 2014).

REFERENCES

Connolly, T. M. & Begg, C. E., 2004. *Database Solutions A Step-by-Step Guide To Building Databases.* 2nd ed. Harlow: Pearson Education.

Connoly, T. & Begg, C., 2015. *Database Systems: A Practical Approach to Design, Implementation, and Management.* 6th ed. Harlow: Pearson.

datawarehouse4u, 2009. *OLTP vs. OLAP.* [Online]
Available at: http://datawarehouse4u.info/OLTP-vs-OLAP.html
[Accessed September 2017].

Rouse, M., 2017. *Data Warehouse.* [Online]
Available at: http://searchsqlserver.techtarget.com/definition/data-warehouse
[Accessed September 2017].

Rouse, M., 2017. *OLAP (online analytical processing).* [Online]
Available at: http://searchdatamanagement.techtarget.com/definition/OLAP
[Accessed September 2017].

TechTarget, 2017. *The roles of data warehousing, data mining and OLAP in knowledge discovery.* [Online]
Available at: http://searchcrm.techtarget.com/answer/The-roles-of-data-warehousing-data-mining-and-OLAP-in-knowledge-discovery
[Accessed September 2017].

Tiwari, P., 2014. *Tools of Business Intelligence in QlikView.* [Online]
Available at: http://www.c-sharpcorner.com/UploadFile/35fef0/tools-of-business-intelligence-in-qlikview/
[Accessed September 2017].

TutorialsPoint, 2017. *Data Warehousing OLAP.* [Online]
Available at: https://www.tutorialspoint.com/dwh/dwh_olap.htm
[Accessed September 2017].

W3 Schools, 2017. *https://www.w3schools.com/sql/sql_view.asp.* [Online]
Available at: https://www.w3schools.com/sql/sql_view.asp
[Accessed September 2017].

W3 Schools, 2017. *SQL COUNT(), AVG() and SUM() Functions.* [Online]
Available at: https://www.w3schools.com/sql/sql_count_avg_sum.asp
[Accessed September 2017].

W3 Schools, 2017. *SQL HAVING Clause.* [Online]
Available at: https://www.w3schools.com/sql/sql_having.asp
[Accessed September 2017].

W3 Schools, 2017. *SQL INNER JOIN Keyword.* [Online]
Available at: https://www.w3schools.com/sql/sql_join_inner.asp
[Accessed September 2017].

W3 Schools, 2017. *SQL Self JOIN.* [Online]
Available at: https://www.w3schools.com/sql/sql_join_self.asp
[Accessed September 2017].