Subjects:

* **SQL**
  + CASE
  + Delete / Truncate table
  + Delete from table
  + Join || Left || Right || Full
  + Union
  + Stored Procedures
  + Triggers
* **Relational Algebra**
* **Normalization**
* **XML XSD& JSON**
* **XPath & XQuery**
* **MongoDB**
* **MapReduce**
* **Neo4j**
* **TF-IDF**
* [**RDF & SPARQL**](https://moodlearn.ariel.ac.il/mod/resource/view.php?id=1396373)
* [**Java streams**](https://moodlearn.ariel.ac.il/mod/resource/view.php?id=1396373)
* **Cassandra**
* **ElasticSearch**
* **Basic Python**
* **Spark**
* **Redis**
* **Naive Bayes**
  + **Accuracy: how much did we actually get right / everything**
  + **Precision**
  + **Recall**
* **Linear Regression**
  + **NP Code**
  + **JavaStream Code**
* **Logistic Regression**
  + **NP Code**
* **Soft Max**
* **NoSQL – General**

**נורמליזציה דרך פעולה:**

* + 1. שדות שלא מופיעים בצד ימין בכלל, חייבים להיות מוכלים במפתח (ייתכן אותיות בודדות ולא כל הרצף)
    2. שדות שנמצאים רק בימין ולא בשמאל, חייבים להיות non-prime.
    3. NF1 תמיד מתקיים
    4. NF2 אם אין non-prime מימן לתת קבוצה ממש של candidates
    5. NF3 כל ה-nonprime נמצאים מימן ל-superKey| האם הוא תלוי רק בסופרקי

6. NF3 וחצי – כל הסופר קיס הם מצד שמאל

\*non-prime הוא כל מי שלא מופיע בKEYS

**NoSQL**

מאגר ערך מפתח

1. REDIS:

בסיס נתונים הפועל בזיכרון ומבוסס על "מפתחות" המאפשרות שליפת נתונים באופר מהיר מתוך מאגרי מידע ענקיים.

דוגמא: תפקידו של בסיס הנתונים הוא לאחסן מסמך לפי id ולאחר מכן לשלוף את כל תוכן המסמך. עם הזמן, ייתכן ונגדיר מסמכים ככאלה שלא צריך יותר, אך ליתר ביטחון נגדיר את משך החיים שלהם לחודש.

תכונות:

1. אחסון כמו ב-HASHMAP (ערך – מפתח)
2. ניתן לכתוב שאילתא לפי המפתח בלבד ואז כל התוכן ישלף
3. להגדיר ל-data להימחק אחרי x זמן

מאגר עמודה רחב

1. CASSANDRA: מסד נתונים המשתמש בטבלאות (שורות ועמודות) אך בניגוד למסד נתונים רלציוני, השמות והפורמט של העמודות יכולים להשתנות משורה לשורה באותה טבלה.

\* הערה : ניתן לראות כמאגר ערך, מפתח דו-מימדי. בקסנדרה אין מגבלה של נורמליזציה, שאילתות שחוזרות על עצמן המון, נעדיף לשמור המון מידע בצורה בזבזנית כדי לאפשר "מענה" מהיר יותר.

דוגמא: נרצה שהמידע יהיה מאורגן בדומה למערכת קבצים. כדי לגשת למידע, נהיה מעוניינים לתת את הערכים המדויקים של כל התכונות הקודמות. שפת שאילתא דומה לזו של relational data-base

תכונות:

1. כל מפתח מאפיין שורה בעלת אלמנטים
2. קריאת מידע מהירה יותר מאשר כתיבה
3. חלק מייעול הקריאות מבוסס על שכפול נתונים
4. לא ניתן לאחד טבלאות

מאגר מסמך

1. MongoDB: מסד נתונים המבוסס על מבנה של מסמך, כך שכל מפתח משוייך למסמך כלשהו.

דוגמא: מאגר מידע הבנוי ממסמכים בהם יש פרטים מספריים רבים. מסד נתונים שמאפשר לנו שאילתות עם חישובים מורכבים על פרטים מתוך המסמכים.

תכונות:

1. מסמכים יכולים להכיל זוגות רבים של מפתח-ערך או מפתח-מערך ואפילו מסמכים מקוננים
2. מבוססת על שימוש ב-JSON

\*הערה: בשונה ממאגר ערך-מפתח המסד הוא בעל "הבנה" מסויימת של המסמכים.

מאגר מנוע חיפוש:

1. Elastic Search: מסד מנוע חיפוש המאפשר לאחסן, לחפש ולנתח כמויות גדולות של מידע במהירות.  
   דוגמא: בסיס נתונים שמאחסן סיפורים ומאפשר גישה לסיפורים הרלוונטים ביותר לפי מילות מפתח.

תכונות:

1. ניקוד וקביעת רלוונטיות של מסמך
2. ניתן לכתוב ישר לקובץ בלי אינדקס או טיפוס

מאגר גרפי:

1. RDF | SPARQL: אחסון נתונים בגרף של שלשות (subject, predicate, object)

SELECT, ?var, Where

{?v1 attribute ?v2}

תכונות:

1. מאפשר מיזוג נתונים בין סכמות בסיסיות שונות.
2. תומכת באבולוציה של סכמות מבלי לשנות את כל המשתמשים בנתונים.

מבנה

מייצג גרף, כך שיש קשת בין 2 קודקודים. הקודקודים – נושא או אובייקט, הקשת – הצהרה, כלומר יש נושא שמחובר לאובייקט על ידי הצהרה.

מאגר גרפי כללי:

1. Neo4J: מערכת לניהול מסדי נתונים גרפיים. מתוארת כמאגר עסקאות כך שבין כל שני אובייקטים יש יחס מחבר.

דוגמא: בסיס נתונים אשר מאפשר לבטא קשרים בין ישויות שונות בסביבה בה יש קשרים מרובים. מעניין מי קשור למי גם אם בקשר עקיף .

תכונות:

1. תואם ל-ACID
2. מאפשר שאילתות מורכבות.