**מגישים: בת חן אברהם 316476985, ודויד חלפון 206745762**

1. **רקע**

אחסון בענן הינו כרוך באחסון נתונים בשרתים מרוחקים אליהם ניגש דרך האינטרנט, במקום לאחסן בהתקני אחסון מקומיים כגון כוננים קשיחים או שרתים.

אחסון בענן מציע מספר יתרונות משמעותיים, כולל אפשרות לאחסן נתונים מרחוק, גישה נוחה, גמישות בהתאמה לצרכים, אמינות ושירות קבוע, אבטחה מתקדמת, יכולת שיתוף פעולה, אפשרות לביצוע גיבוי ושחזור. כל אחד מהיתרונות הללו משקפים את החשיבות העקרונית של אחסון בענן, ומסייעים להבין את היתרונות שהוא מציע למשתמשים.

אחסון נתונים מרחוק: בעזרת אחסון בענן משתמשים וארגונים יכולים לאחסן את נתוניהם מרחוק בשרתים המתוחזקים על ידי ספקי שירותי ענן. נתונים אלה יכולים לכלול מגוון רחב של פריטים, כולל קבצים, מסמכים, תמונות, סרטונים, מסדי נתונים ועוד.

נגישות: אחסון בענן מספק גישה נוחה ונמצאת בכל מקום לנתונים של משתמשים, באמצעות קישוריות לאינטרנט. משתמשים יכולים לגשת לקבצים ולאפליקציות שלהם ממכשירים שונים, כולל מחשבים, סמארטפונים וטאבלטים.

מדרגיות: אחסון בענן מציע מדרגיות כמעט בלתי מוגבלת, ומאפשר למשתמשים להגדיל או להקטין בקלות את קיבולת האחסון שלהם לפי צרכיהם. ספקי ענן בדרך כלל מציעים מודלים של תמחור לפי תשלום, המאפשרים למשתמשים לשלם רק עבור האחסון שהם משתמשים בו.

אמינות ויתירות: שירותי אחסון בענן מספקים לרוב רמות גבוהות של אמינות ויתירות. נתונים משוכפלים על פני מספר שרתים ומרכזי נתונים, מה שמפחית את הסיכון לאובדן נתונים עקב כשלי חומרה או אסונות.

אבטחה: ספקי אחסון בענן מיישמים אמצעי אבטחה חזקים כדי להגן על נתוני המשתמשים. זה כולל הצפנה של נתונים במעבר ובמנוחה, בקרות גישה, מנגנוני אימות ועמידה בתקנים ובתקנות בתעשייה.

שיתוף פעולה: אחסון בענן מקל על שיתוף פעולה ושיתוף קבצים בין משתמשים. הוא מאפשר למספר משתמשים לגשת ולערוך מסמכים בו זמנית, לעקוב אחר שינויים ולשתף קבצים בצורה מאובטחת עם אחרים, הן בתוך הארגון והן מחוצה לו.

גיבוי ושחזור: אחסון בענן יכול לשמש כפתרון גיבוי ושחזור, דבר המספק למשתמשים דרך מאובטחת ואמינה לגבות את הנתונים שלהם ולשחזר אותם במקרה של אובדן נתונים או אסון.

אחסון בענן מציע יתרונות רבים, כולל גמישות, מדרגיות, נגישות, אמינות, אבטחה וחסכוניות. זה הפך למרכיב חיוני בתשתית IT מודרנית, המאפשר לאנשים וארגונים לאחסן, לנהל ולגשת לנתונים שלהם בצורה יעילה ואפקטיבית יותר.

בפרוייקט שלנו נשתמש בשיטה של Deduplication שיטה שבה לא נשמר קובץ פעמיים.  
מניעת כפילויות באחסון בענן מתייחסת לתהליך של זיהוי וביטול עותקים כפולים של נתונים בתוך מערכת אחסון. טכניקה זו משמשת כדי לייעל את קיבולת האחסון ולהפחית את עלויות האחסון על ידי אחסון רק מופעים ייחודיים של נתונים, תוך הבטחת שלמות נתונים ונגישות.

מניעת כפילויות משחקת תפקיד מכריע באופטימיזציה של יעילות האחסון והפחתת עלויות בסביבות אחסון בענן. על ידי זיהוי וביטול עותקים כפולים של נתונים, מניעת כפילויות עוזרת לארגונים למקסם את משאבי האחסון שלהם ולשפר את שיטות ניהול הנתונים הכוללות.

1. **מוטיבציה**

אנו נתקלים בתופעת כפילויות קבצים, שבה מספר רב של משתמשים מעלים קבצים דומים או זהים לשרת. כתוצאה מכך, השרת צריך לשמור מספר פעמים קבצים זהים, מה שיכול לגרום לתפיסת מקום מיותר, עיכובים זמניים בריצה, שימוש נוסף בזיכרון וכדומה.

1. **מטרות**

מטרת האחסון ענן שלנו הוא לפתור את בעיית ה Deduplication ובכך לייעל את שיטת האחסון ולהפחית עלויות בסביבות אחסון בענן. כמו כן, אנו שמים דגש על אבטחה ואמינות כלפי המשתמשים המשתמשים בשירותי הענן.  
אנו מבצעים זאת באמצעות עיבוד הצאנקים, כאשר אנו מחלקים כל קובץ ליחידות קטנות יותר הנקראות "צאנקים". חלוקה לצאנקים מאפשרת פיזור של הנתונים למספר שרתים, מה שמגביל את הסיכונים ומקנה אבטחה ותפוקה גבוהה יותר, תוך חיסכון במקום.  
לאחר החלוקה לצאנקים מעלים אותם לשרת. כאשר מעלים קובץ חדש לשרת, אנו מבצעים בדיקה האם חלק מהקובץ (צאנק) כבר קיים במאגר הנתונים. במקרה שכן, אין צורך לשמור את הצאנק הזה שוב, ולכן הוא לא ישמר מחדש.

1. **המצב הקיים**

מניעת כפילויות, טכניקת ה Deduplication היא פתרון נפוץ בתחום האחסון בענן, ומיועדת להפחית את העלויות ולשפר את היעילות של שירותי הענן. ספקי שירותי ענן משתמשים בשיטה זו בסוגים שונים של שירותים:

שירותי גיבוי ושחזור מאסון בענן: ממנפים מניעת כפילויות כדי להפחית את עלויות האחסון ולמטב את העברת הנתונים ברשת. בשירות זה מתבצעת ביטול כפילויות ברמת הבלוק או הנתח, כך שנתונים כפולים לא מאוחסנים על פני ערכות הגיבוי.

שירותי אחסון בענן: ספקי אחסון בענן, כמו Amazon Web Services (AWS) S3,   
Google-Cloud Storage, Microsoft Azure שבהם משתמשים בטכניקות של Deduplication ליעילות האחסון והפחתת העלויות. מניעת כפילויות מסייעת למזער נתונים מיותרים המאוחסנים באחסון אובייקטים, וכתוצאה מכך לחסכון משמעותי עבור המשתמשים.

שירותי הגנת נתונים וארכיון: שירותי הגנת נתונים וארכיון המוצעים על ידי ספקי ענן משתמשים לרוב במניעת כפילויות כדי להפחית את כמות הנתונים המאוחסנים ולשפר את ביצועי הגיבוי והשחזור. מניעת כפילויות מאפשרת לשירותים אלה לאחסן רק בלוקים או נתחים ייחודיים של נתונים, ובכך לייעל את קיבולת האחסון ואת השימוש ברוחב הפס של הרשת.

מכשירי גיבוי בענן: חלק ממכשירי גיבוי בענן ופתרונות משלבים פונקציונליות של מניעת כפילות כדי למנוע נתונים מיותרים לפני העברתם לאחסון בענן. מכשירים אלה עשויים לבצע ביטול כפילות במקור, ביעד או בשני הקצוות של תהליך העברת הנתונים, בהתאם ליישום.

שירותי ניהול נתונים ב-Cloud Native: שירותי ניהול נתונים מקוריים בענן, כגון פלטפורמות גיבוי ושחזור מבוססות ענן, פתרונות סנכרון ושיתוף קבצים וכלי העברת נתונים בענן, עשויים לשלב יכולות מניעת כפילויות כדי לשפר את יעילות הנתונים ולהפחית עלויות.

רשתות אספקת תוכן (CDNs): רשתות אספקת תוכן (CDNs) המפיצות ומספקות תוכן על פני שרתי קצה מפוזרים גיאוגרפית עשויות להשתמש בטכניקות של מניעת כפילויות כדי לייעל את האחסון במטמון ולהפחית את השימוש ברוחב הפס. מניעת כפילויות עוזרת למזער תוכן מיותר המאוחסן על פני מטמון קצה של CDN, ומשפרת את הביצועים והיעילות.

אלו הן רק כמה דוגמאות לשירותי ענן הממנפים שיטות מניעת כפילויות כדי לייעל את יעילות האחסון, להפחית עלויות ולשפר את יכולות ניהול הנתונים. מניעת כפילויות משחקת תפקיד קריטי במתן אפשרות לספקי שירותי ענן להציע פתרונות אחסון ניתנים להרחבה, אמינים וחסכוניים ללקוחותיהם.

היחודיות בפרויקט שלנו היא שאנו בוחרים לעשות את שיטת ה Deduplicationברמת הצאנקים, אחרי חלוקת הקובץ לצאנקים, אנו מבצעים HASH לכל צאנק. התוצאה היא רצף ייחודי של תווים, המייצג את התוכן של הצאנק. במקרה שהצאנק כבר קיים במערכת, מאגר הנתונים יתעדכן בהתאם וישייך את הצאנק לקובץ המתאים לו.

1. **ניתוח ואפיון דרישות המערכת:**

המערכת תוכל לבצע את כל הפיצ'רים הבאים:

הרשמה

התחברות

התנתקות

העלאת קבצים

הורדת קבצים

מחיקת קבצים

חיפוש בקבצים שהועלו

צפייה בקבצים שהועלו (פאנל)

פירוק הקובץ לצאנקים

הרכבת הקובץ מצאנקים

אי חזרה על צאנקים (דידופליקיישן)

חלוקת הצאנקים ל4 דאטה בייסים בישביל לייעל זמני ריצה

1. **תוכנית עבודה על בסיס ניתוח המטלות הקריטיות וניהול סיכונים ייחודיים לפרוייקט:**

**חשיבות:**

גבוהה

בינונית

נמוכה

לוועדה 1:

פירוק הקובץ לצאנקים

הרכבת הקובץ מצאנקים

העלאת קבצים

הורדת קבצים

מחיקת קבצים

צפייה בקבצים שהועלו (פאנל)

חיפוש בקבצים שהועלו

הרשמה

התחברות

התנתקות

ועדה 2-3:  
אי חזרה על צאנקים (דידופליקיישן)

חלוקת הצאנקים ל4 דאטה בייסים לייעול זמני ריצה

**סיכונים:**

בעיות בתיאום לוחות זמנים בין חברי הצוות.

בעיות עם טכנולוגיות כגון: בעיות בחיבור לDB , בעיות בפיתוח ה UI וכו.

כשל במניעת חזרה של צ'אנקים(דידופליקיישן).

כשל בחלוקה ל4DB שונים.

1. **תיאור הסביבה והפיתוח:**

פיתוח צד הלקוח יתבצע בFlask (Flask היא סיפרייה לפיתוח יישומי אינטרנט בפייתון)

פיתוח צד השרת יתבצע בPython

פיתוח בסיס הנתונים יתבצע בMariaDB (MariaDB היא מערכת ניהול מסדי נתונים המבוססת על MySQL)

יתבצע שימוש עזר בטכנולוגיות הבאות:

Visual Studio Code (עורך קוד לשימוש של כלל קבצי הפרוייקט)

HeidiSQL (כלי לצפייה במסד הנתונים בצורה נוחה)

XAMPP Control Panel (לוח בקרה להפעלה לשליטה בשירותים כגון: MariaDB)

1. **תיאור העבודה שבוצעה בסמסטר א':**
2. לימוד הרקע – מחקר על אופן הפעולה של עננים, פירוק והרכבה של צ'אנקים, שליטה בטכנולוגיות הנדרשות ותיכנון העבודה.
3. פיתוח המערכת – בניית בסיס הנתונים, בניית השרת (פיתוח פונקציונאליות לוועדה 1) ופיתוח UI תואם.
4. בדיקת כל הפיצ'רים הנדרשים לוועדה 1:

~~הרשמה~~

~~התחברות~~

~~התנתקות~~

~~העלאת קבצים~~

~~הורדת קבצים~~

~~מחיקת קבצים~~

~~צפייה בקבצים שהועלו (פאנל)~~

~~פירוק הקובץ לצאנקים~~

~~הרכבת הקובץ מצאנקים~~

~~חיפוש בקבצים שהועלו~~

**References:**  
  
J. Hai, "Network Cloud Storage Service Architecture Analysis and Research," 2016 Eighth International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Macau, China, 2016, pp. 413-416, doi: 10.1109/ICMTMA.2016.105.

P. Malathi and S. Suganthidevi, "Comparative Study and Secure Data Deduplication techniques for Cloud Computing storage," 2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2021, pp. 1-5, doi: 10.1109/ICSES52305.2021.9633960.

G. Ali, M. Ilyas Ahmad and A. Rafi, "Secure Block-level Data Deduplication approach for Cloud Data Centers," 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, 2020, pp. 1-6, doi: 10.1109/iCoMET48670.2020.9074109.