מבחינת כתיבת הדוח והגשתו, נראה שהחמצתם את מה שעושה דוח לדוח - את החלק של המילים שאמורות לעטוף את הגרפים ולשמור על הרצף הסיפורי והתיאורי של העבודה. את הסקת המסקנות מן הגרפים. את תיאור העבודה, את הנימוקים. את ה captions של הגרפים במקרים רבים.

בחלק האחרון חסרים התיאורים הבסיסיים של מה שעשיתם, והתוצאות אינן עקביות. חלק מהניתוח נעשה היטב, אבל חלקו לא מאפשר לראות את התוצאות. בנוסף, יש סתירות בין תוצאות שונות.

**שלב 2:**

בגרפים אנו מדברים על ציר אופקי וציר אנכי, לא ציר X וציר Y. תארו לעצמכם שאתם מציגים את התלות של X ב Y, ואז בכלל X בציר האנכי ו Y בציר האופקי...

בגרף 2-8 לא ניתן לראות FLURRIES או BURSTS, ניתן לראות כמה שיאים. לא מצוין בדוח מה הגרנולריות של יצירת ה PDF, לכן קשה להסיק מסקנות מן הגרף. לעומת זאת, ניתן לראות אכן שיש כמה RUNTIMES נפוצים מן העליות התלולות באיור 2-9.

אותה הערה לאיור 2-10. אלה לא FLURRIES או BURSTS. ליתר דיוק, יש דבר אחד שנין לקרוא לו BURST, וזה ה INTERARRIVAL TIMES הקטנים. ביצעתם את הניתוח של המידע כאן כאילו הוא מידע בדיד, נאבל זה מידע רציף שמעוגל לשניות. כלומר אין הרבה הבדל בין INTERARRIVAL TIME של שניה לשתיים. זה יכול להיות קוצו של יוד, לאן עיגלו את ההפרש בין הריצות.

-פרק 2 שאלה 2 / 3

באיור 2-7  -  זו שגיאה בנתונים או בניתוחם. נכון שקשה להצביע על הגורם הספציפי, אבל יש מספר קטן של ג'ובים שיכולים להוות את הבעיה - רק כאלה שהיו פעילים בזמן זה. כאן ציפיתי שתאתרו את המשימות הספציפיות שפעלו בזמן הבעייתי, בפרט משימות שהתחילו כאשר הבעיה התחילה או הסתיימו כאשר הבעיה הסתיימה, ואולי אפילו תתאימו את מספר המעבדים החורגים למספר המעבדים במשימות. לחליפין, אם לא ניתן לראות התאמה כזו, הייתי מצפה ל ZOOM IN על הקטע הבעייתי שמבהיר מדוע לא ניתן לעשות כן.

במתלב ביצעתם את ההמרה, אבל לא את הנקיון שהיה דרוש. למשל, כל הג'ובים שנקטעו בגלל שהסתיימה הקלטת הלוג (כם מי שירד בדיוק בסוף הטרייס) הוא בעצם ג'וב שזמן הריצה שלו לא נכון, ולצורך ECDF של זמני ריצה למשל צריך היה לנקות אותו. זה censored data. אופציה אחרת הייתה להפעיל כלים של censored data על הנתונים הללו, כדי לתקן את ה ECDF.

**שלב 3:**

לא "ניתן לראות" מהגרפים כמעט שום דבר מהדברים שנטען ש"ניתן לראות". ודאי שלא "ניתן לראות" התפלגות (לוג) נורמלית ב PDF ב 2-8. וודאי שלא "ניתן לראות" ארלנגים, בייחוד לא כשבשלב  2 לא ציירתם את ה CCDF בסקלת לוג לוג.

זה גם לא מה שהייתם אמורים לעשות בשלב זה. בשלב הזה הייתי מצפה או לדברים שבאמת אפשר לראות מהגרף, כמו חוק חזקה שנראה כמו קו ישר בלוג לוג (ב CCDF), או להצדקות תיאורטיות, כמו למשל - נבדוק עבור RUNTIME פונקציות שידוע מפרק 3.2 שהן HEAVYTAIL או LONGTAIL, למשל Pareto.

**שלב 4:**

איור 4-1

גרף ה PDF של ה RUNTIME נתון בסקלה שבה כל הדיו על הצירים. לא ניתן להתרשם מן הגרף. ה CDF נתון בסקלה מנורמלת כלשהיא, עד 1. אבל לא ברור מה זה 1.

איור 4-2 נתון ללא צירים כלל, ללא מספרים.

הבעיה חוזרת גם בשאר הגרפים. באיור 4-10 ניתן לראות משהו בגרף ה PDF.

כל הגרפים נתונים ללא שום מלל! כפי שהיו בשלב הטיוטה. בדוח המסכם הציפיה שלי הייתה שתביאו דוח, עם מלל, שבו אתם מתייחסים לגרפים. זו הכוונה למשל במילים keep the story flowing בהסבר על הגשת הדוח בשלב 10. לצערי בשלב 4 הבאתם רק תוצאות בגרפים, ללא ניתוח מילולי, ללא עיבוד ראשוני. בשלב זה ציפיתי לקרוא משהו בסגנון "ניתן לראות מאיור 4-X שתיאור של ההתפלגות באמצעות פונקציית ... או ... הולם את הנתונים, כפי שציפינו בשלב 3 " או "בניגוד למה שציפינו בשלב 3, דווקא פונקציה ... שלא תוארה בספר תיארה את הנתונים באופן הטוב ביותר, על פי מדד ה ... (נניח שטח בין עקומות) שמימשנו.

בשלב זה נראה שלא התקדמתם מאז הטיוטה, ולהפך - נושאים ששוחחנו עליהם בפגישת ההנחיה בעל פה לא משוקפים בדוח המסכם.

**שלב 5**

עבודה טובה!

**שלב 6**

משהו לא בסדר בגרף 6-1. אני לא מאמינה שה think times כולם שליליים. זה ממש בלתי אפשרי.  יתרה מזאת, יש כאן 4 גרפים (6-1 עד 6-4), ללא CAPTIONS, ועם צירים אנכיים זהים בשני זוגות (interarrival ו think). כלומר, על פניו נראה שיש שני זוגות גרפים שאמורים להציג אותו דבר אבל נותנים תוצאות שונות. ממה שאני זוכרת ממה שדיברנו בפגישות ההנחיה, מדובר בשתי אינטרפרטיות שונות ל think time או ל interarrival time, אולי אחד מהם הוא user interarrival time. אבל אתם לא יכולים לסמוך על כך שהקורא ינחש מה המוצג בגרפים. גרפים חייבים להיות מתוארים כראוי. ומוסברים בטקסט.

**שלב 7**

גם כאן, הבאתם  גרפים, אבל לא הוספתם אף מילה על התהליך. מה עשיתם? באיזו שיטה השתמשתם? איך וידאתם שזו ההתאמה הטובה ביותר? מה הפרמטרים של ההתפלגות שמצאתם? היכן הקוד ששימש לביצוע השלב? חסרה השוואה כמותית, למשל במדד שדיברנו עליו של השטח בין העקומות. או לפי P P PLOT או Q Q PLOT. או כל מדד אחר שנלמד בקורס.

שלב 8

הייתי מצפה לראות תוצה של פעולה הCLUSTERING כאן. שתראו לי שאכן הדוגמאות שבחרתם נמצאות קרוב למרכזי קלאסטרים (אם זה אכן מה שעשיתם, כי דיברנו על כך, אבל לא אמרתם שעשיתם כך בדוח).

איך תכננתם שהקבצים השונים ייצרו עומס של 80% 120%? מה הבסיס לתכנון? מה הרעיון שגורם לכם לצפות לכזה שינוי בעומס? האם הכפלתם משהו ב 1.2 או ב .8? אם כן, איך התמודדתם עם שברים, כי מדובר כאן במספרים קטנים?

חסר לי כאן דיון בשיקולים שהביאו אתכם לבחירה בשיטה שבחרתם. כמו כן, חסר לי שלב של תיאום ציפיות - אילו מן השלבים של חלק 9 אתם מצפים לשחזר על סמך התכן, ואילו לא?

שלב 9:

1. בדיקת העומס - הגעתם ל 600 ואפילו 700 אחוז עומס. זה היה צריך להדליק אצלכם נורת אזהרה! זה אומר שלא ביצעתם את ה SCHEDULING כנראה בכלל, כלומר הכנסתם למערכת כל JOB שהגיע, גם אם לא היה מקום - והרי לא היה מקום, עברתם את ה 100%! הייתם אמורים להניח שכל המערכת מחוברת, אבל שלא ניתן לעבור את ה 100% של המערכת, כך שאם מגיע JOB ומבקש יותר משאבים ממה שיש במערכת, הוא צריך להמתין עד שיתפנו משאבים.

0.5

2. אני לא מצליחה לראות כמעט כלום בגרפים, כי הם גרפים של PDF במקום של CDF, או CCDF, בסקלת לוג לוג. בשלב זה הייתי מצפה מכם ליישם ולהפנים את מה שלמדנו בשלב 2 על אופן ההצגה הברור של התפלגויות.

0

3. משתמשים - הצגה טובה. אבל היה קל יותר להשוות בין הגרפים אילו היו כל הקווים נמצאים באותו איור, באותו גרף. ואז ניתן היה לראות מרחקים בין קווים. כך לא ניתן להתרשם מהשוואת המרחקים בין הגרפים השונים. למשל,, אני עדיין לא יודעת איך הגדלתם את העומס - האם הוספתם משתמשים או שיחקתם עם הפרמטרים של הג'ובים שלהם. אילו הגרפים היו אחד על השני, הייתי יכולה ללמוד את זה מהגרפים.0.5

קורלציה - הדגמה טובה.

1

4 SELF SIMILARITY - מה הצגתם כאן? זה חלק שלא למדנו בכיתה מדוע בחרתם בשיטה זו? מה היא מראה? איך עשיתם את החישוב?

0.5

5 LOCALITY OF SAMPLING - בחרתם בשיטה טובה, אבל הצגתם ציר אופקי בסקלה לוגריתמית. זה גורם לכך שהדבר העיקרי שהעין רואה זה הצטופפות הנקודות כתוצאה מהצטופפות הזמן לאורך הציר. כלומר לא ניתן להתרשם מן הנתונים באופן ויזואלi

0.5

6 SUBMISSION RATE - איך ה CDF וה ECDF הפוכים? מה אתם מתארים? אולי אחד מהם הוא CCDF?

אתם כותבים שאתם לא רואים הבדל ב SUBMISSION RATE בין הטרייסים השונים - האם זה מכיוון שניפחתם את גודל העבודות? כי אם ניפחתם מספר משתמשים, אני בהחלט מצפה להבדל ב SUBMISSION RATE. אתם לא מצפים להבדל ב SUBMISSION RATE בגלל ה THINK TIME כי את זה לא הכנסתם. מצד שני לא הגבלתם את המערכת ל 100%, אז לא השפעתם בכלל על ה THINK TIME.

0

7 טרנד ומחזורים  - קצת חיפפתם עם הצגת הטרנד אבל עשיתם עבודה יסודית שהוצגה יפה מאד עבור המחזוריות. הצגתם את המחזוריות גרף על גרף, כך שניתן ממש להתרשם מההבדלים ביניהם. יפה מאד!

1

8 WAIT TIMES - אני לא מבינה איך קיבלתם בכלל זמני המתנה עם אי הגבלת התפוסה. זה לא עקבי עם סעיפים קודמים. כשלעצמו הסעיף עשוי בסדר גמור, זה פשוט לא הגיוני.

0.5

9 סעיף חופשי - גם כאן עשיתם יפה, אבל גם כאן אני לא מבינה איך זה מסתדר עם ה600 700 % צריכה שהראיתם קודם.

0.5