

אנליטיקה רב-מנועית על פלטפורמה מנוהלת עם Amazon SageMaker

TLV-ANT305 — סיכום סשן, AWS Summit Tel Aviv 2026

Avichay Marciano & Vova Nevski · Sr. Analytics Specialist Solutions Architects, AWS

10 בספטמבר 2026 | משך: כ-40 דקות | הופק מהקלטת הסשן

מסלול: AWS For Data and Analytics | רמה: 300

על המסמך הזה

המסמך מסכם את הסשן TLV-ANT305 מתוך AWS Summit Tel Aviv 2026, בהתבסס על ההקלטה המלאה שפורסמה באתר הסאמיט. הוא נבנה מתמלול אוטומטי של ההקלטה ומהסליידים שחולצו מהווידאו עצמו, כך שאפשר לחזור לתוכן בלי לצפות שוב בארבעים הדקות. חותמות הזמן בגוף המסמך מפנות לנקודה המדויקת בהקלטה.

איך לקרוא את זה

- **פרק 1 — תקציר מנהלים:** חמש דקות קריאה. מה הוצג ומה רלוונטי לארגון שמנהל פלטפורמת דאטה.
- **פרקים 2-4 — הרקע והמבנה:** האתגרים שהוצגו, ארבע שכבות הקונטקסט, ומבנה השכבות של Amazon SageMaker.
- **פרקים 5-8 — הממשל בפועל:** SageMaker Catalog, מולטי-אקאונט, פרודיוסרים וקונסיומרים, אורקסטריציה ו-lineage.
- **פרקים 9-10 — שכבת האחסון והמנועים:** S3, Apache Iceberg ופדרציה לקטלוגים חיצוניים, ו-SageMaker Notebooks עם ה-Data Agent.
- **פרק 11 — הדמו:** חברת AnyInvesting וארבע הפרסונות שעברו על הבמה מקצה לקצה.
- **פרק 12 — מה לוקחים מכאן:** מסקנות ונקודות פתוחות.

הערת מקור המסמך הופק מתמלול אוטומטי של ההקלטה. התמלול מדויק בקטעים המקצועיים, אך שמות מוצרים ומונחים לועזיים מופיעים בו בשיבוש פונטי (למשל "סייג'מייקר", "אטנה", "אייסברג") — הם נורמלו כאן למונח האנגלי מול הסליידים. הציטוטים מובאים כלשונם מן התמלול ואומתו מול ההקלטה בחותמת הזמן המצוינת. מספרים ושמות שמופיעים על הסליידים צוינו ככאלה.

1. תקציר מנהלים



שקופית הפתיחה עם שמות שני הדוברים ותפקידיהם

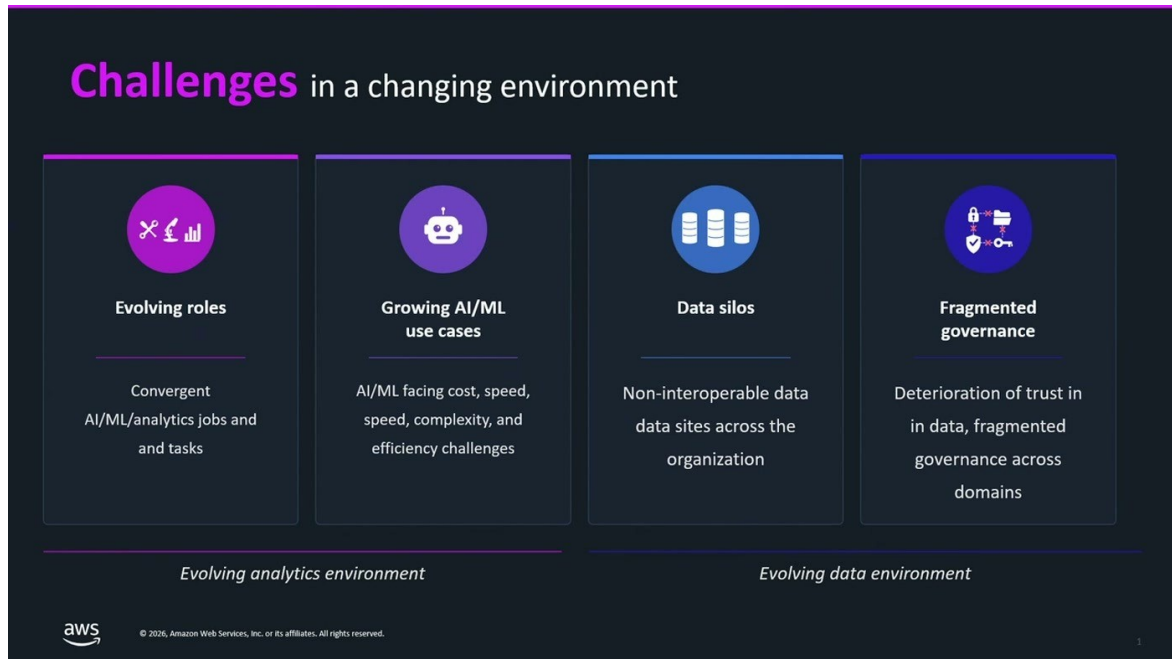
שני Specialist Solutions Architects מתחום האנליטיקה ב-AWS לקחו את Amazon SageMaker — בגלגולו החדש כפלטפורמת דאטה ו-AI מאוחדת — והראו אותו משתי זוויות: חצי ראשון תיאורטי על שכבת הממשל והקטלוג, וחצי שני דמו חי שבו ארבע פרסונות עבדו על אותו workspace אחד. הטענה המרכזית לא הייתה "יש לנו עוד כלי", אלא שכאשר הממשל יושב בתוך הפלטפורמה, הוא מפסיק להיות משימת day-two ונעשה חלק מתהליך הפיתוח.

- **הבעיה שהוצגה היא סיילואים בקצב חדש, לא סיילואים חדשים.** עידן ה-GenAI לא המציא את הבעיה אלא האיץ אותה: "וככה, תוך כמה פרומפטים פשוטים, נוצר עוד דאטה סיילו חדש בחברה, שכן אדם אחד אחראי עליו" [1:32].
- **השכבה החסרה היא Business Metadata.** מבין ארבע שכבות הקונטקסט שהוצגו, זו היחידה שאינה זמינה אוטומטית — והיא בדיוק זו שהמודל זקוק לה. "הוא לא מכיר את הארגון שלכם. הוא לא מכיר את הטבלאות שיש אצלכם" [3:06].
- **הממשל הוא ברזולוציית שורה ועמודה, ובאמת אטום.** מיסוך עמודות PII לא נראה גם ב-execution plan של Spark: "הם לא ידעו שהן קיימות בשום צורה" [16:51]. מאחורי הקלעים יושבים SageMaker Catalog ו-AWS Lake Formation.
- **ריבוי מנועים בלי ריבוי סביבות.** בתוך notebook אחד רצו תאים מעל Athena, Glue, EMR, EMR Serverless, DuckDB ו-Spark, וגם join ישיר בין טבלה פנימית לבין Snowflake [32:12].
- **פדרציה לקטלוגים של מתחרים, דרך Iceberg REST.** Glue Data Catalog, Databricks Unity Catalog, Snowflake Horizon או כל קטלוג Iceberg אחר — כולם מתחברים כ-connection ומופיעים כקטלוג נוסף. ב-AWS אנחנו גם "Iceberg First" [18:24].
- **המסר החותם: ממשל בגלל הסוכנים, לא למרותם.** "ולכן governance זה כבר לא [39:50] commodity — כי חלק גדל והולך מהקונסומרים של הדאטה הם agents ולא בני אדם."

למי זה רלוונטי

לארגונים שמריצים כבר היום כמה מנועי עיבוד במקביל ומשלמים על כך במטבע של ניהול הרשאות מפוצל: הרשאות ב-Lake Formation, הרשאות בכלי ה-BI, הרשאות בקלאסטר, הרשאות במחסן צד שלישי. הדמו שהוצג היה במובהק פיננסי-רגולטורי — ניהול תיקי השקעות, עמודות PII, דרישות בידוד בין צוותים — וזו גם הסיבה שהתרחיש של מולטי-אקאונט קיבל שקופית משלו.

2. מה השתנה בסביבת הדאטה



ארבעת האתגרים, מחולקים לשני צירים: סביבת האנליטיקה המשתנה מול סביבת הדאטה המשתנה

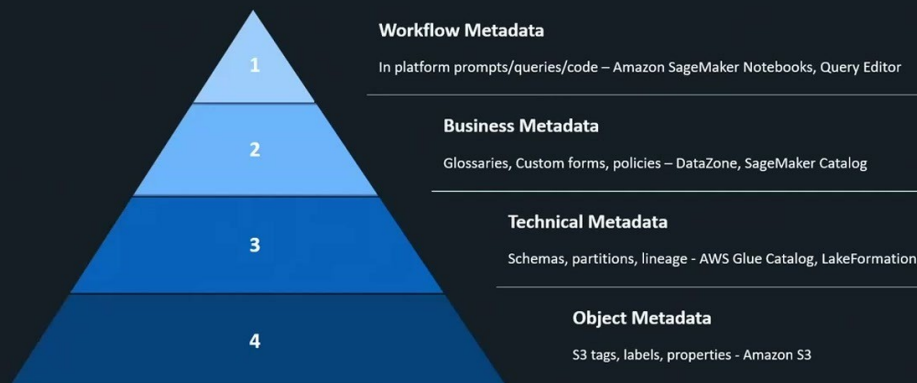
הדובר הראשון פתח בארבעה אתגרים. שניים מהם נוגעים לסביבת האנליטיקה ושניים לסביבת הדאטה עצמה.

- **גבולות התפקידים מיטשטשים.** "דאטה אנג'ניר יכול לקחת חלק מהמשימות של דאטה סיינטיסט והפוך" [0:00]. התוצאה המעשית: יותר אנשים צריכים גישה ליותר דאטה, ליותר כלים, ובעיקר צריכים להחליף בין כלים.
 - **ריבוי use cases של AI/ML.** עומס נוסף על צוותי הדאטה, אבל בעיקר על התשתית והדאטה עצמו — שברוב הארגונים אינם מוכנים לצריכה על ידי מודלים [0:00].
 - **דאטה סיילואים.** לא אתגר חדש, אבל הוא מחמיר: הדאטה מפוזר בין צוותים ומחלקות, וגם הממשל עליו מפוזר. כשצריך תמונה הוליסטית, צריך לעבוד מול ממשקים רבים.
 - **ממשל מפוצל.** על הסלייד: שחיקה באמון בדאטה עקב ממשל מקוטע בין דומיינים.
- הטוויסט של 2026 הוא הגישה האג'נטית: יותר אנשים מקבלים גישה לדאטה, "חלק מהזמן בצורה ישירה, אבל הרבה יותר פעמים דרך גישה אג'נטית" [1:32]. ובמקביל, יצירת סיילו חדש נעשתה זולה: משתמש שאינו טכני יכול לגזור נתונים, לעבד, להפיק דוח ולשתף — ובזה נולד סיילו חדש שכן אדם אחד אחראי עליו.

3. קונטקסט בארבע רמות

החלק התיאורטי נשען על שקופית אחת שחוזרת כחוט השני לאורך כל ההרצאה: פירמידת הקונטקסט של data lake טיפוס.

Context is important at every level



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

ארבע שכבות המטא-דאטה כפי שהוצגו, כולל השירות שמחזיק כל שכבה

שכבה	מה יש בה	היכן היא יושבת (מהסלייד)
Workflow Metadata	פרומפטים, שליפות וקוד שנוצרים תוך כדי עבודה שוטפת	SageMaker Notebooks, Query Editor
Business Metadata	משמעויות של טבלאות ועמודות, מטריקות, מילון מונחים, פוליסז	DataZone, SageMaker Catalog
Technical Metadata	סכמות, עמודות וטיפוסים, פרטישנים, lineage	AWS Glue Catalog, Lake Formation
Object Metadata	שם האובייקט, תיוגים ומאפיינים ברמת הקובץ	Amazon S3

הנקודה המעשית: מי שיש לו גישה לנתונים, בדרך כלל כבר יש לו גישה ל-Technical Metadata ול-Workflow Metadata. השכבה הסמנטית — Business Metadata — היא שחסרה כמעט תמיד, והיא בדיוק זו שמודל שאומן על מידע ציבורי אינו יכול להשלים בעצמו: "הוא לא מכיר את העמודות ומשמעויות שלהם" [3:06]. התוצאה, בלשון הדובר: "במקרה טוב תקבלו תשובה שלא מובין מה אתם רוצים ממנו" [3:06], ובמקרה פחות טוב — הזיה שתצטרכו להתמודד איתה.

המסקנה התפעולית ההשקעה בשכבת ה-Business Metadata אינה פרויקט תיעוד. היא תנאי סף להנגשת הדאטה לצריכה על ידי מודלים — וכפי שיודגם בהמשך, היא גם מה שמאפשר ל-Data Agent שבתוך ה-notebook לייצר קוד מדויק.

4. מהו Amazon SageMaker החדש

נקודה שדרשה הבהרה מהבמה: השם Amazon SageMaker מסמן היום מוצר אחר מזה שהיה לפני כשנתיים וחצי. המוצר הישן — זה של AI Machine Learning — קיים ונקרא היום Amazon SageMaker AI, גם כשירות נפרד וגם כרכיב בתוך SageMaker החדש [4:36]. SageMaker החדש נועד לאחד את עולמות ה-Data וה-AI תחת קורת גג אחת ולמנוע מעבר בין ממשקים.

המבנה שהוצג הוא מודל שכבות:

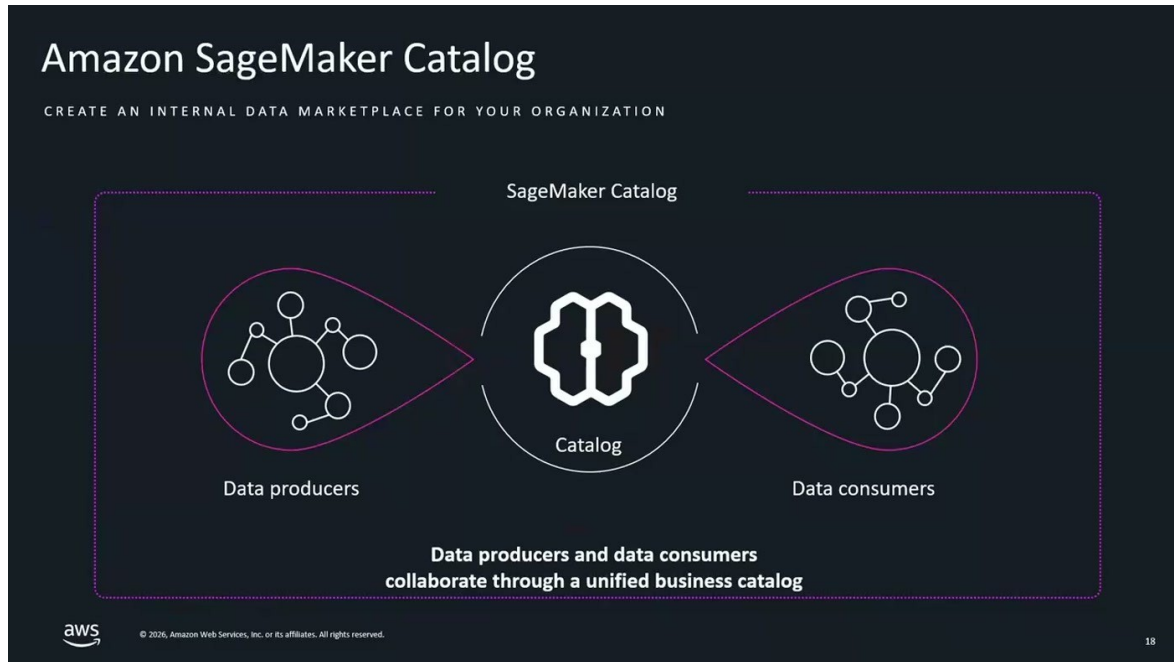
- **Data and AI Infrastructure**. שכבת הציפיים — Trainium, Inferentia ו-GPUs.
- **Lakehouse**. מתפצלת לשתיים: Storage (למשל Amazon S3 לאובייקטים, ו-S3 Tables לטבלאות) ו-Compute — המנועים שעובדים עם הנתונים: Glue, EMR, Athena, Redshift, וגם מנועי קוד פתוח וצד שלישי [4:36].
- **Catalog**, שכבה סמנטית וממשל. כולל את ה-Business Metadata, וממושת על ידי SageMaker Catalog.

• **Tools.** שכבת הכלים — SageMaker AI, notebooks וכיוצא באלה.

• **IDE.** הסביבה שבה המשתמש יושב בפועל.

לגבי השכבה האחרונה נאמר משהו שכדאי לשים לב אליו: אפשר לעבוד ב-IDE המובנה, אבל "הרבה המשתמשים שלנו מחפשים את החוויה של ה-IDE שלהם שרגילים לעבוד איתו" [6:06] — כולל המודל והפלאגנים שכבר מותקנים אצלם — ולכן אפשר לחבר IDE מקומי ל-SageMaker Amazon ולקבל את שתי הזוויות. בחלק הדמו הזכרנו במפורש Cursor ו-Kiro כסביבות שאפשר לחבר למוטבוק [36:48].

5. SageMaker Catalog: שלושה תפקידים



הקטלוג כמרקטפלייס פנימי: פרודיוסרים מצד אחד, קונסיומרים מצד שני, קטלוג עסקי אחד באמצע

הקטלוג הוצג כמענה לשלושה אזורים, ולא כמאגר מטא-דאטה בלבד.

Discovery

חיפוש טקסטואלי וגם חיפוש סמנטי מובנה מעל האסטים. מעבר לכך, יכולת לייצר את ה-Business Metadata אוטומטית — תיאורים ברמת הטבלה וברמת העמודה — כשאפשר גם למלא ידנית או לשלב בין השניים [6:06]. נוספים על כך Custom Glossaries (מילון מונחים פנים-ארגוני שאפשר לתייג בו אסטים), Metadata Forms, שכבת מטא-דאטה נוספת, ויכולות Data Quality ו-Data Lineage [7:40].

Governance

מי יכול לראות אילו נתונים, ושליטה על כך. הדגש שניתן מהבמה הוא הרזולוציה: המערכת "מספק את היכולות הזו עם גרונולריות מאוד גבוהה ברמה של שורה ועמודה" [7:40].

Collaboration

שתי רמות. הראשונה — אנשים תחת אותו צוות או פרויקט שעובדים עם אותם כלים ועל אותו דאטה. השנייה — שיתוף נתונים בין צוותים ופרויקטים שונים, בצורה קלה ואינטואיטיבית מצד אחד, ועם שליטה מלאה ובקרה על הפעולות מצד שני [7:40].

6. בידוד ומולטי-אקאונט

החלק הזה נועד לענות מראש על השאלה הרגולטורית. הארכיטקטורה שהוצגה: אקאונט ראשי אחד שמחזיק את הקטלוג עם כל המטא-דאטה ברמה הארגונית, וכל צוות יושב באקאונט נפרד לגמרי — עם הריסורסים שלו, הקומפיוט שלו והדאטה שלו — ועדיין משתמש בקטלוג המרכזי שיושב באקאונט אחר [7:40]. שיתוף דאטה בין הצוותים נשאר פשוט גם כשהם באקאונטים שונים.

מיד אחר כך הובהר הצד השני של אותה מטבע: אם אין לארגון דרישת בידוד כזו, אפשר להריץ את הכול תחת אקאונט אחד ולקבל את מלוא יכולות ההפרדה בין פרויקטים וצוותים [9:11]. הניסוח שנבחר היה ש-SageMaker בא לפגוש את הארגון במקום שבו הוא נמצא מבחינת דרישות בידוד — ולא כופה מבנה מולטי-אקאונט.

מה זה אומר בפועל הגבול הלוגי אינו האקאונט אלא הפרויקט. כפי שיתואר בפרק 7, פרויקט ב-SageMaker יכול לחיות באקאונט אחר ובריג'ן אחר, ולעבוד עם סט כלים משלו — כך שמעבר מארכיטקטורת אקאונט יחיד לארכיטקטורה מבודדת אינו משנה את מודל העבודה של המשתמשים.

7. פרודיוסורים, קונסיומרים ו-Data Products

הדובר השני לקח את הנושא והמשיג את הממשל דרך שני תפקידים. "הדאטה פרודוסרס למעשה מייצרים את הדאטה אסט שלי, את הדאטה פרודקט" [9:11] — והם גם אלה שמגדירים מה קורה סביבו.

מה פרודיוסור עושה

- **מייצר את הדאטה.** מקורות אפשריים: בסיסי נתונים רלציוניים (Aurora, RDS), אפליקציות SaaS כמו SAP ו-Salesforce, ודאטה בזמן אמת. שירותי Zero-ETL מושכים חלק מהמקורות האלה ישירות אל תוך הלייקהאוס [10:43].
- **מבצע טרנספורמציות.** ויזואלית בממשק drag & drop לפונקציות עסקיות, או בכלים המוכרים — AWS Glue ו-Amazon EMR — מתוך סביבת הנוטבוקים [12:16].
- **מייצר Business Metadata.** בלחיצת כפתור, תיאורים לכל העמודות ולכל הטבלאות. "במיוחד בעולם האיג'נטי שהדברים האלה מקבלים משנה תוקף נוסף" [12:16].
- **מגדיר Data Quality Rules.** חוקים שלומדים את התפלגות העמודות ומגדירים בייסליין תקין, ואפשר לאכוף אותם כתנאי לקידום הדאטה לקטלוג המרכזי [9:11].
- **מגדיר Fine-Grained Access Control.** הדוגמה שניתנה: דאטה סט עם עמודת PII, שבו data scientist רואה הכול ואנליסט רואה תת-קבוצה בלי ה-PII [9:11].
- **מפרסם.** פרסום לקטלוג המרכזי, כשמאחורי הקלעים עובדים SageMaker Catalog ו-AWS Lake Formation כמנגנון ניהול ההרשאות [12:16].

מה קונסיומר עושה

הקונסיומר ניגש למרקטפלייס הפנימי, מחפש — והחיפוש רץ גם על שמות הטבלאות וגם על ה-Business Metadata וה-Custom Glossaries — ומגיש בקשת רישום. הפרודיוסור יכול להגדיר מראש חוקים לאישור אוטומטי, או לאשר פרטנית [16:51]. אחר כך הוא עובד עם הדאטה בכלים שמוגדרים בפרויקט שלו. ההערה החשובה: "כל קונסומר יכול להיות פרודוסור בעצמו" [10:43].

המיסוך הוא אמיתי

זה הרגע החד ביותר בחלק התיאורטי. על מיסוך עמודות לאנליסטים נאמר: "יכול להיות שספציפית לאנליסטים, אנחנו נמסך עמודות PII מסוימות, הם לא ידעו שהן קיימות בשום צורה, גם כשהם יפתחו ספרק, אפילו אם הם יפתחו execution plans, לא יהיה שום בעצם אינדיקציה לכך שהן היו קיימות" [16:51]. כלומר, לא סינון בשכבת התצוגה אלא אכיפה למטה, בשכבת המנוע.

פרויקטים ו-Data Products

הגבול בין פרודיוסורים לקונסיומרים ב-SageMaker הוא הפרויקט: "כל פרויקט למעשה יכול לחיות באקאונט אחר, בריג'ן אחר, לעבוד עם כלים משלו" [15:19]. יחידת השינוף היא Data Product — אריזה של כמה Data Assets יחד עם ה-Business Metadata וה-Custom Glossaries שלהם, כך שאפשר להירשם לכולם בלחיצת כפתור אחת: "נניח כל הדאטה הכלכלי, כל הדאטה מדומיין מסוים" [16:51].

8. אורקסטריציה ו-Lineage

התהליך שתואר — דאטה נוחת, רץ עליו ETL אחד או כמה, נאכפים חוקי Data Quality ואז מתבצע פרסום לפרודקשן — דורש אוטומציה. הכלי שנבחר הוא Apache Airflow, בשתי תצורות מנוהלות ב-AWS ששתיהן מוטמעות בפלטפורמה [13:47].

תצורה	מה מקבלים	מודל תשלום
Airflow Domains (MWAAs מנוהל)	webserver, scheduler וכל החלקים המזכירים, בגרסאות תקניות תואמות אופן-סורס	על הקלאסטר שנמצא באוויר
Airflow Serverless	אין ניהול קלאסטר — כותבים DAG, מגישים לשירות, והשאר מנוהל	רק על זמן הריצה של הטסקים הבודדים

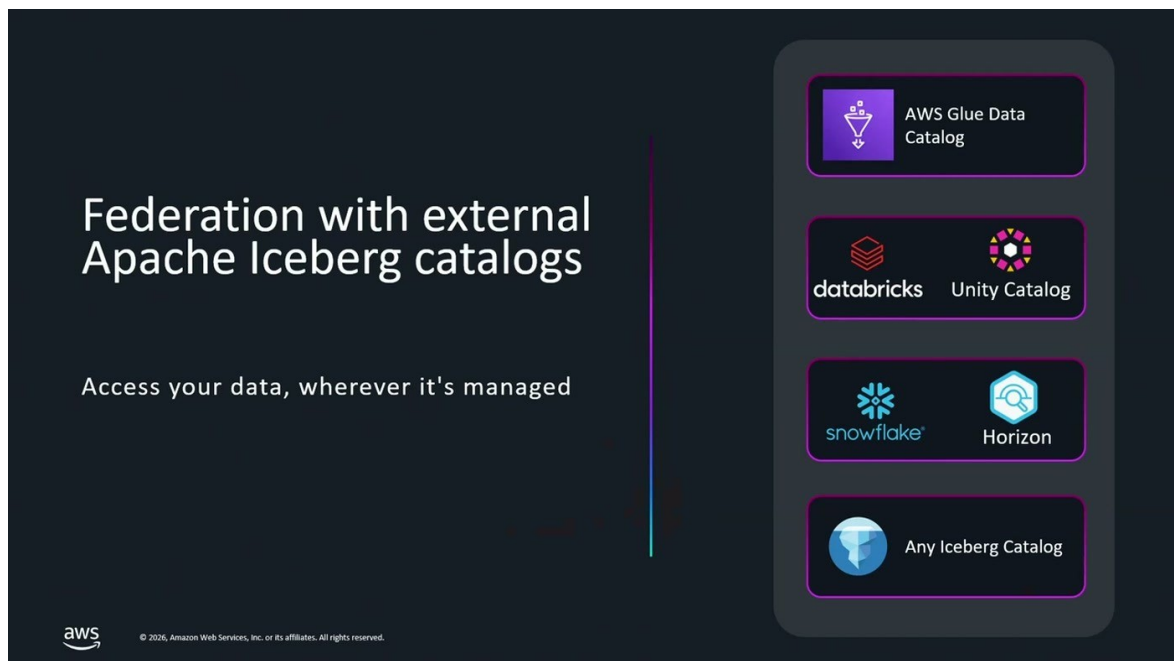
Airflow Serverless הוכרז, לפי הדובר, בתחילת השנה [13:47]. ההשוואה שנבחרה: "יש לנו מודל ריצה שהוא מאוד דומה לקונספט של למדה" [15:19], ולכן "אנחנו משלמים רק על זמן הריצה של הטסקים הבודדים שרצו" [15:19]. אפשר גם לעשות mix and match בין שתי התצורות. בשתיהן קיים ממשק ויזואלי לבניית DAG ב-drag & drop עם כל הקונקטורים של AWS.

Lineage שמגיע בחינם

הדרישה לעקוב אחרי מחזור החיים של אסט — מהדאטה הגולמי שנחת, דרך הטרנספורמציות, ועד האסטים החדשים — נענית בלי עבודה נוספת: "אנחנו מקבלים את ה-Traceability הזה בעצם בצורה מובנית מעצם העובדה שאנחנו עובדים מעל הכלים בפלטפורמה של [13:47] SageMaker. המימוש נשען על הפרוטוקול הפתוח OpenLineage.

9. איפה הדאטה יושב: S3, Iceberg ופדרציה

התשובה הטבעית היא Amazon S3, בזכות הדורביליות ושכבות האחסון. אבל השינוי שצוין מהבמה הוא עליית ה-Open Table Formats בכל התעשייה, ועמדת AWS בעניין נאמרה במפורש: "ב-AWS אנחנו גם Iceberg First" [18:24] — ומכאן האינטגרציות הרבות סביב Apache Iceberg.



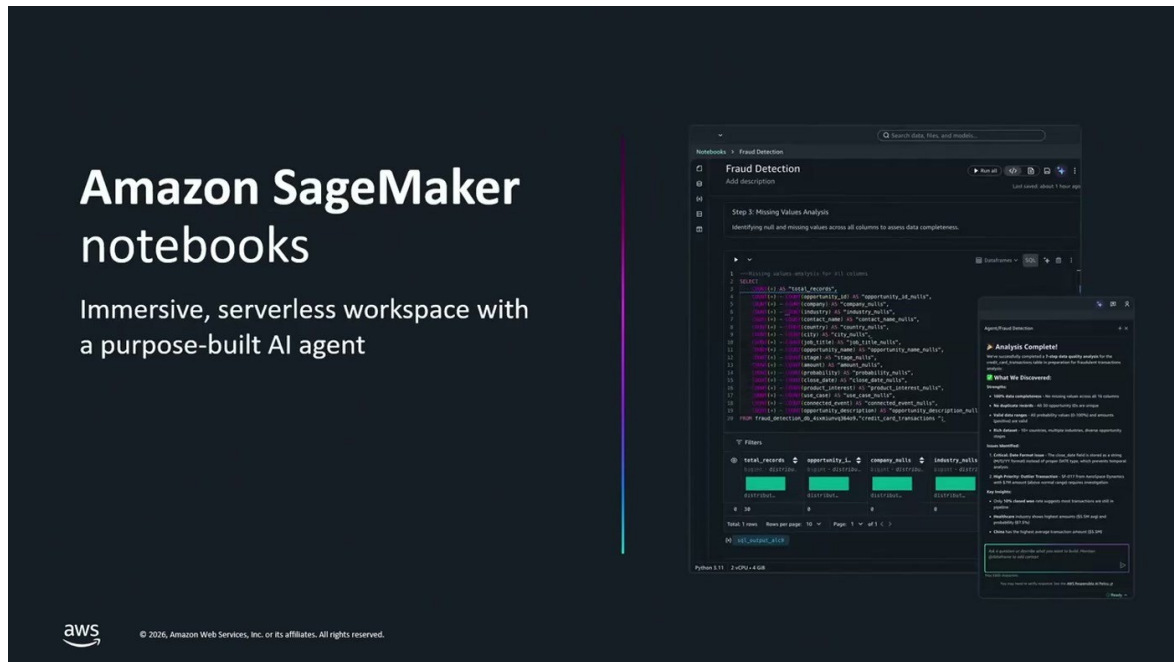
ארבע האפשרויות לקטלוג Iceberg שאפשר לחבר, כפי שהופיעו על הסלייד

האינטגרציה שקיבלה את הדגש היא הפדרציה. כשעובדים עם Iceberg יש קטלוג שמנהל את הגישה לטבלאות; מעל זה AWS יכול להיות Glue Data Catalog — אבל דרך פרוטוקול Iceberg REST אפשר לבחור לעבוד גם מול דאטה שמנוהל על ידי Unity Catalog של Databricks, מול Horizon של Snowflake, ואפילו מעל קטלוג שהארגון מרים בעצמו, כמו Gravitino [18:24]. בפועל: יוצרים connection מתוך SageMaker, "ובעצם מאותו רגע זה מופיע לנו כקטלוג שאנחנו יכולים לצרוך את הדאטה מתוכו" [18:24].

למה זה משנה הפדרציה היא מה שהופך את הטענה "פלטפורמה אחת" לשימה בארגון שכבר השקיע במחסן נתונים של צד שלישי. אין צורך להעביר את הדאטה כדי לשלב אותו — הוא נצרך במקום שבו הוא מנוהל. בדמו זה הודגם בפועל, כש-join אחד חיבר טבלה פנימית לטבלה שיושבת ב-Snowflake.

10. SageMaker Notebooks וה-Data Agent

הצורך שתואר מצד הלקוחות: להמשיך לעבוד עם ה-Athena, EMR, Glue, Redshift — best in class tools — אבל להפסיק לנהל הרשאות לכל אחד מהם בנפרד [19:56]. המענה: "SageMaker Notebooks למעשה מאפשרים לנו לעבוד עם כל הכלים של AWS בצורה קוליבורטיבית" [19:56].



הסלייד שהציג את הנוטבוקים כסביבה סרברלסית עם סוכן AI ייעודי

הדוגמה הקונקרטית שניתנה היא ברמת התא: תא אחד מריץ קוד מעל EMR on EKS, תא אחר מעל EMR Serverless, אחר מעל Glue, מעל Athena, ומעל DuckDB — שמשולב כחלק מהסביבה הפייתונית של SageMaker [19:56]. זו בפועל המשמעות של "multi-engine" בכותרת הסשן: לא בחירת מנוע ברמת הפרויקט, אלא ברמת התא.

הסוכן שיושב בצד

ה-Data Agent הוצג כסוכן שחי לצד המפתח ועוזר בכתיבת ETL-ים ומודלים. ההבדל ממנו לסוכן קוד גנרי הוא הקונטקסט: "יש לו גישה ישירה לביזנס מתה דאטה, כך שהוא גם מבין את ההקשרים" [19:56], ולכן הקוד שהוא מייצר מדויק יותר. ההדגמה הקצרה שקדמה לדמו הראתה אותו מתאר טבלאות בקטלוג לא מוכר, מציע שאלות SQL, מייצר ויזואליזציות בלחיצת כפתור (bar chart, scatter plot), וכשהבקשה מורכבת יותר — מייצר קוד פייתון עם matplotlib [21:27].

11. הדמו: AnyInvesting

החצי השני של הסשן היה דמו מוקלט שמקשר את כל הקונספטים דרך עיניהם של משתמשים שונים. התרחיש: חברה בדיונית בשם AnyInvesting, שמנהלת תיקי השקעות עבור לקוחותיה, מייעלת אותם באמצעות אנליזה ומודלים של Machine Learning, ופועלת בשוק רגולטורי עם מגבלות שצריך לעמוד בהן [21:27].



החברה הבדיונית שסביבה נבנה הדמו, כפי שהוצגה על הסלייד

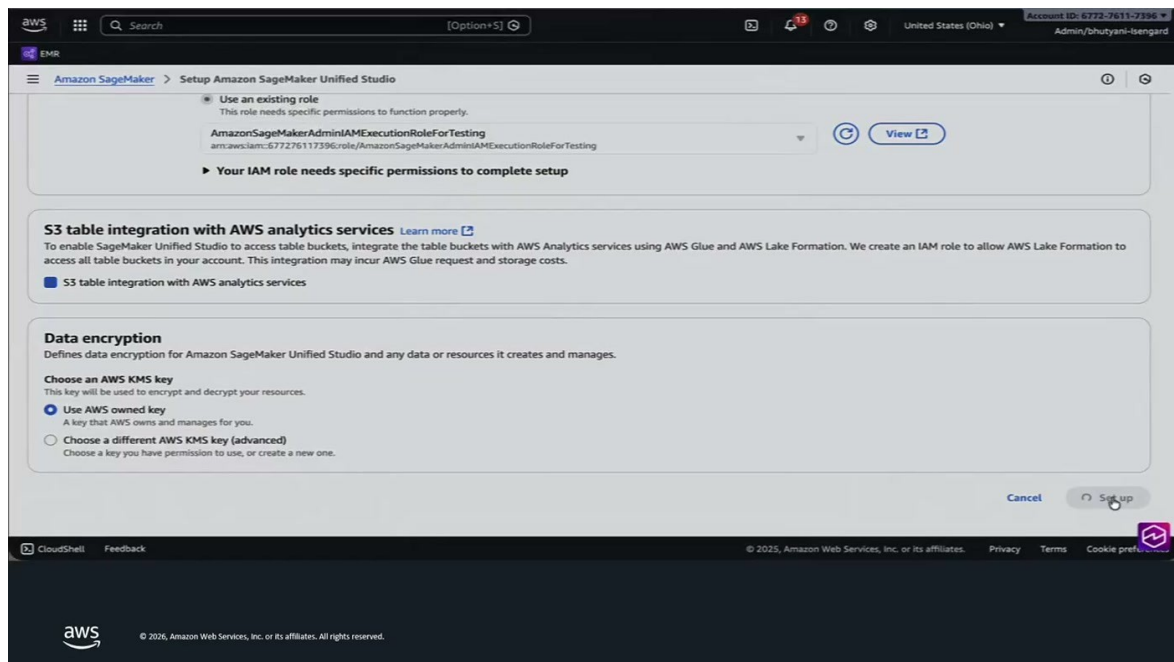
פרסונות וכלים

פרסונה	המשימה בדמו
אדמין	יצירת הדומיין של SageMaker מהקונסול
אנליסטית	אנליזה של ניירות ערך עם פוטנציאל רווח והפסד הגבוה ביותר
Data Engineer	פיתוח ETL להנגשת ממשק דאטה לאפליקציה פנים-ארגונית, ואורקסטרציה שלו
Data Scientist	איתור פורטפוליו לקוחות בסיכון, אימון מודל ו-deploy לפרודקשן
GenAI Builder	deploy של LLM פנימי שישרת את הצ'טבוט הארגוני

הכלים שהוצגו בדמו: Amazon Athena, AWS Glue, SageMaker AI, MWAA, בנוסף — חיבור ל-Snowflake כמנוע צד שלישי, Glue Data Catalog כקטלוג, ו-Lake Formation לניהול ההרשאות מאחורי הקלעים. הדאטה יושב גם ב-S3 סטנדרטי וגם ב-S3 Tables [23:00]. הודגש שאלו ארבעה כלים מתוך רבים הזמינים בפלטפורמה.

אדמין: הקמת הדומיין

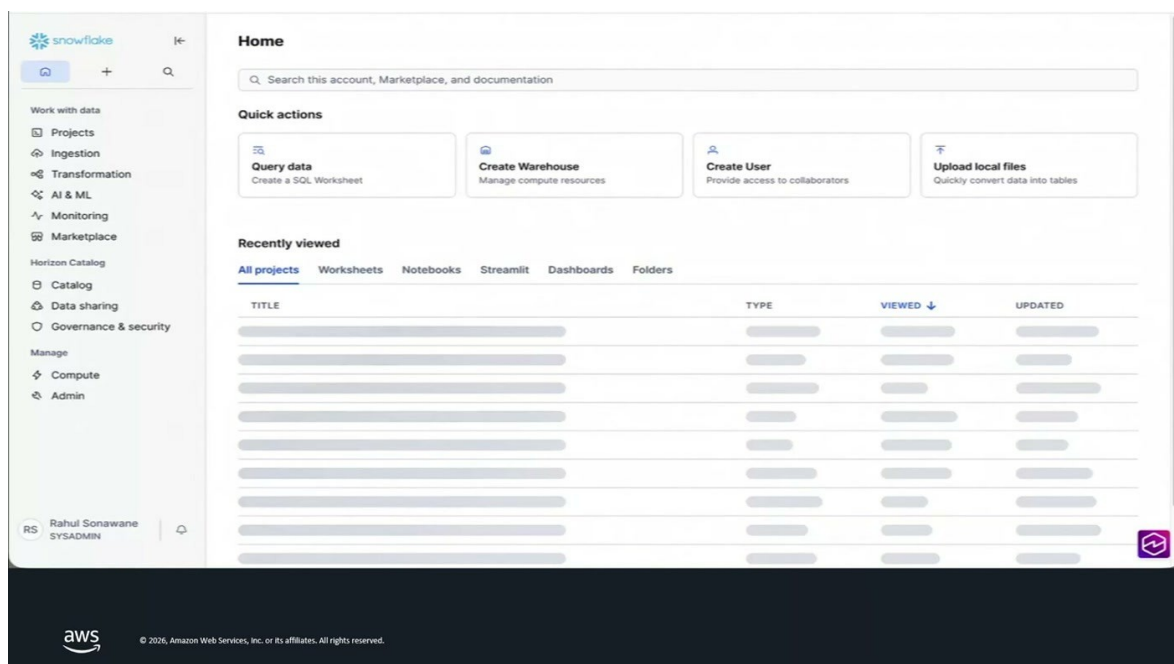
מתחילים בקונסול של AWS, נכנסים ל-SageMaker, ובהיעדר דומיין קיים לוחצים Get Started. אפשר לתת לשירות ליצור IAM Role בעצמו או לבחור רול קיים — שצריך להרשות גישה לכל השירותים ולכל הדאטה הנדרשים. האינטגרציה עם S3 Tables מופעלת כברירת מחדל וניתן לכבות אותה. לאחר Setup: "בעולם אמיתי זה לוקח משהו כמו שתי דקות, זה לא מייד" [23:00].



מסך הקמת הדומיין: אינטגרציה עם S3 Tables והגדרות הצפנת דאטה

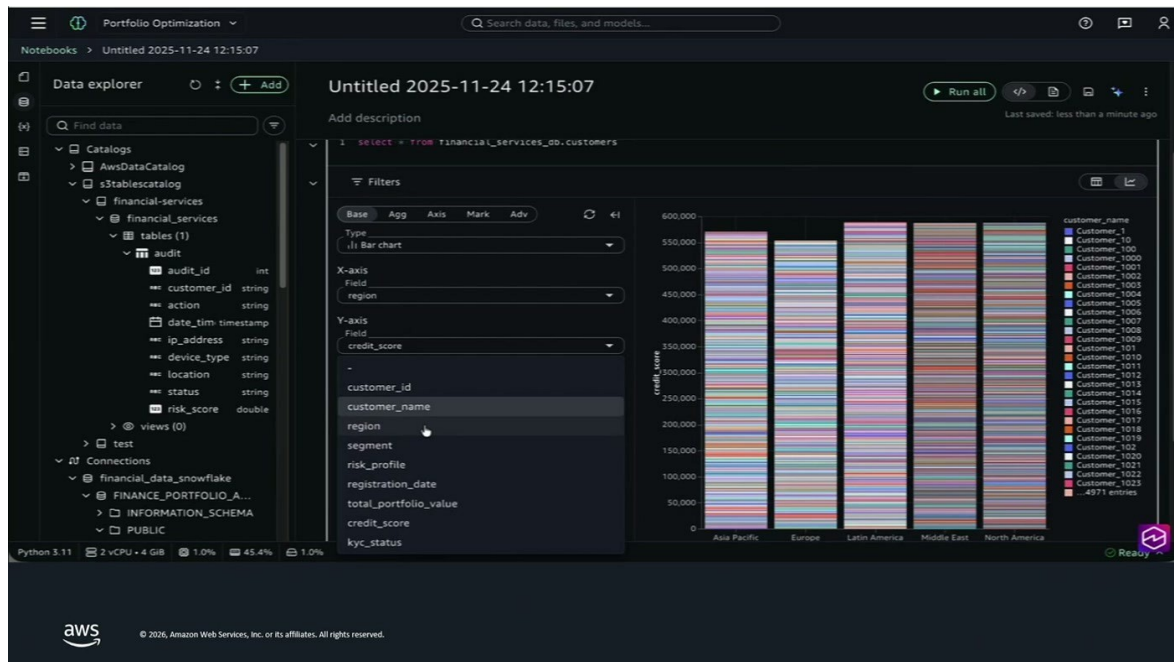
אנליסטית: מהחיפוש עד הגרף

היא זוכרת שהטבלה שלה נמצאת תחת סכימת financial, מחפשת, ובוחרת בטבלת Customers. בקטלוג היא רואה מידע על הטבלה, את העמודות, Preview של הנתונים, ותחת Details גם את מיקום הטבלה ב-S3. גם טבלאות S3 Tables זמינות באותו ממשק [24:33]. מכאן היא עוברת ישירות ל-Query Editor, מריצה SQL, ומורידה תוצאות כ-JSON או CSV.



מסך הבית של הסביבה המאוחדת, עם חיפוש רוחבי על דאטה, קבצים ומודלים

שאר היכולות שהודגמו באותו ממשק, בלי לצאת ממנו: מעבר לתצוגה גרפית, slice and dice במקום, הוספת תאי SQL או Markdown להסבר, החלפת הסכימה שמולה רצות השליפות, שמירה, ותזמון המחברת כך שה-SQL ירוץ באופן מתוזמן [26:08].



בניית ויזואליזציה מעל תוצאת שלפיה; בעץ משמאל נראים s3tables/catalog לצד connection ל-Snowflake

Data Engineer: Visual ETL ותקלה שנפתרה בכפתור

הוא בוחר בממשק Visual ETL. אפשר לתאר בטקסט מה רוצים ולקבל ETL מיוצר, אבל בדמו נבחר drag & drop: בחירת טבלה מתוך ה-Glue Data Catalog (רק מה שיש עליו הרשאות), הוספת טרנספורמציה — בדוגמה Drop Nulls — וקונפיגורציה ויזואלית. אפשר להריץ מתוך אותו ממשק, לתזמן משם, ולראות את הסקריפט של Glue שנוצר מאחורי הקלעים ולהמשיך ממנו בקוד [26:08, 27:39].

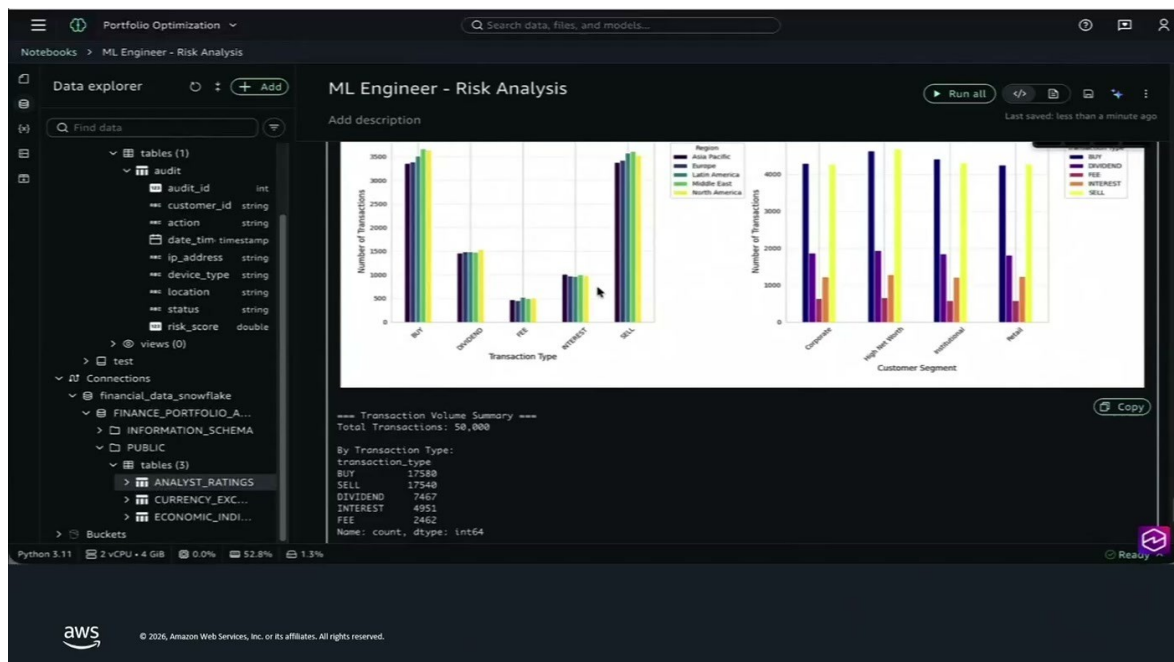
בהרצה של ETL מורכב יותר, עם כמה טבלאות ו-join-ים ביניהן, הריצה נכשלה — בכוונה. הלוגים זמינים בזמן אמת, ולצדם: "יש גם כפתור של Troubleshoot with AI לוחצים עליו ומקבלים הסבר מלא על מה נכשל ואיך להתקן את זה" [27:39]. התקלה בפועל: שתי עמודות באותו שם שנוצרו כתוצאה מה-join, והתיקון הוא rename לאחת מהן.

לאורקסטרציה הוא עובר לממשק Workflows, שהוא ממשק ל-SageMaker MWAA. תומך גם ב-MWAA Serverless וגם ב-MWAA מנוהל [29:11]. בקנבס הוויזואלי מופיעים האופרטורים המוכרים של Airflow, גוררים אותם, מקנפגים ומגדירים תלויות. בדוגמה: טאסק Glue שנבחר מתוך רשימת ה-Glue jobs שפותחו קודם, ותהליך שתלוי בשניים אחרים. אפשר לראות את קוד ה-YAML שנוצר מאחורי הקלעים ולהעתיק אותו החוצה. התהליך תוזמן לעשר בבוקר — והנקודה שהודגשה היא שכל האורקסטרציה נבנתה בלי כתיבת שורת קוד אחת ובלי ידע מוקדם ב-Airflow [29:11].

Data Scientist: join בין שתי טכנולוגיות

המשימה: לאתר פורטפוליו לקוחות בסיכון, לשלב מידע מכמה מקורות, לאמן מודל ולעשות לו [30:41] deploy. הוא מתחיל דווקא ב-UI של Snowflake, שם יושב חלק מהדאטה — סכימת Portfolio Analysis עם טבלאות של דירוגי אנליסטים ושערי המרת מטבע. הגישה מתקבלת על ידי יצירת connection ל-Snowflake מתוך SageMaker וה-credentials של ה-warehouse. ברמת הפרויקט קיים גם S3 bucket ייעודי להעלאה והורדה של קבצים.

בנוטבוק עצמו: שלפית טבלת הלקוחות הפנימית, הוספת פילטרים ומיון בלחיצת כפתור, ויזואליזציה שמתווספת אל תוך הנוטבוק. אחר כך שלפית טבלת שערי המטבע מ-Snowflake — בתא שמסומן בסמל של Snowflake ומשתמש בconnection אחר — ואז: "שוב, אני עושה join בין שתי טכנולוגיות שונות" [32:12].



נוטבוק ה-Data Scientist: בעץ משמאל טבלאות S3 לצד טבלאות Snowflake, ומימין הפלט של הניתוח

כאן נכנס ה-Data Agent פעמיים. בפעם הראשונה, אחרי תקלה: הסוכן מבין שבוצע join בין טכנולוגיות שונות ושנבחר המנוע הלא מתאים, מסתכל על ה-SQL ועל המנוע, ו"הוא מציע לי בעצם איך לשנות בתצורה שמאוד דומה לדיף בגיט" [33:42] — לוחצים accept, מריצים מחדש, עובד. בפעם השנייה, בהמשך, אחרי typo בשם ה-tracking server של [35:12] MLflow. אגב, נאמר שגם הדמו עצמו נוצר באמצעות ה-[33:42] Data Agent.

לסקיילים גבוהים הוא מעביר תא להרצה מעל Spark, ומשם אפשר לפתוח בלחיצת כפתור את ה-Spark UI כדי לצלול לתוך הטרנספורמציות, וגם את ה-Spark Logs — שם, כדבריו, לעיתים יש מידע רלוונטי כמו out of memory. [33:42]

הניסוי ל-endpoint

שלב ה-ML מתחיל ביצירת MLflow Tracking Server, שאליו ידווחו המטריקות של כל הניסויים. משימת AutoML מריצה מספר training jobs על פרמוטציות של פרמטרים, וב-MLflow אפשר לבחור את הניסוי ולראות את המטריקות — F1, accuracy ואחרות [35:12]. המודל שענה על התוצאות הרצויות מקבל inference endpoint ישירות מתוך הנוטבוק, והסוכן מייצר גם את הקוד לבדיקת ה-endpoint, כולל הצורה שבה ה-API מצפה לקבל את ה-input [35:12].

The screenshot displays the MLflow Experiments page for an experiment named 'portfolio-optimization-automi'. On the left, a sidebar lists experiments: 'Default', 'fraud-detection-financial...', 'financial-risk-classification', and 'portfolio-optimization-aut...' (selected). The main area shows a table of runs under the 'Experimental' tab. The table has columns for Run Name, Created, Dataset, Duration, Source, and Models. Three runs are listed, all with a duration of 46.2min. The first run was created 28 seconds ago, the second 5 days ago, and the third 7 days ago. The interface includes search filters, sorting options (Sort: Created), and a '+ New run' button. The footer shows the AWS logo and copyright information for Amazon Web Services.

Run Name	Created	Dataset	Duration	Source	Models
automi-portfolio-opt-2...	28 seconds ago	-	46.2min	main...	-
automi-portfolio-opt-2...	5 days ago	-	46.2min	main...	-
automi-portfolio-opt-2...	7 days ago	-	46.2min	main...	-

שרת ה-MLflow עם רשימת הניסויים; הניסוי הנבחר מציג שלוש ריצות AutoML

GenAI Builder: deploy ל-LLM פנימי

המשימה האחרונה: deploy ל-LLM פנימי שישרת את הצ'טבוט הארגוני. "אני פותח את אזור המודלים שלי בתוך הפלטפורמה של [36:48] SageMaker — המחקר על לידרבורדים כבר נעשה, והבחירה היא מודל מסוג Mistral. דרך JumpStart מתקבל sample notebook שכולל את כל הקוד ההתחלתי ל-deploy ולבחירת ה-endpoint. ה-endpoint הוקם על מכונה מסוג [36:48] ml.g5.2xlarge. משם — הרצת קלטים שונים, בדיקה מול הדרישות העסקיות, ואפשרות למחוק את ה-endpoint ולהרים מודל אחר בהרצה של תא בודד [38:18].

12. מה לוקחים מכאן

הסיכום מהבמה נעשה מהסוף להתחלה, ושלוש הנקודות שנבחרו מתמצות היטב את הטענה של הסשן.

- **Workspace אחד, בלי פרגמנטציה של הרשאות.** "כולם עבדו בתוך workspace אחד. לא היה לנו בעצם איזושהי פרגמנטציה של ניהול הרשאות" [38:18] — אדמין, data engineer בתפקיד פרודיוסר, data scientist בתפקיד קונסיומר, ו-GenAI specialist, כולם באותה סביבה.
- **מטא-דאטה ו-lineage אינם day-two.** "חלק מתהליך הפיתוח הוא יצירה של הביזנס מטה דאטה" [39:50], וה-lineage מתקבל built-in. זה שינוי סדר פעולות, לא תוספת כלי.
- **ownership מגיע ממודל ה-Data Products.** עצם העבודה במודל של data products היא שמייצרת בעלות ברורה על האסט, במקום טבלה יתומה בלייק.
- **הממשל הוא בגלל הסוכנים.** "ולכן governance זה כבר לא [39:50] commodity" — להגן על הדאטה ועל ה-PII ולקבוע מי ניגש למה, בין אם מדובר ב-workforce האנושי ובין אם ב-workforce האג'נטי.

נקודות שונות פתוחות

נושא	מה לא נאמר	לבדיקה
עלות	לא הוצגו מספרי תמחור לאף אחד מהרכיבים, למעט מודל ה-per-task של Airflow Serverless	15:19
איכות ה-Business Metadata המיוצר	הודגם ייצור אוטומטי, לא הודגמה בקרה על נכונותו	12:16, 6:06
הגירה מסביבה קיימת	הדמו התחיל מדומיין חדש; לא נגעו במעבר מפלטפורמה קיימת	23:00
Data Quality Rules בפועל	הוזכרו כמה פעמים אך לא הודגמו על המסך	12:16, 9:11
ביצועי הפדרציה	לא נאמר מה ה-latency של שאילתה שחוצה קטלוגים	18:24

הדוברים הפנו את הקהל להרצאה שבאה אחריהם, בנושא Apache Iceberg ו-S3 Tables [39:50].

13. מונחון

מונח	משמעות
Amazon SageMaker	בגלגול הנוכחי — פלטפורמה מאוחדת לדאטה, אנליטיקה ו-AI; המוצר הקודם בשם זה נקרא היום SageMaker AI
SageMaker Catalog	שכבת הקטלוג, המטא-דאטה העסקי והממשלי; משמשת כמרקטפלייס דאטה פנים-ארגוני
Business Metadata	השכבה הסמנטית: משמעויות טבלאות ועמודות, מטריקות, מילון מונחים ופוליסיז
Data Product	אריזה של כמה Data Assets יחד עם המטא-דאטה שלהם, שאפשר להירשם אליה כיחידה אחת
Project	גבול העבודה ב-SageMaker; יכול להיות באקאונט ובריג'ן נפרדים ולהחזיק סט כלים משלו
Fine-Grained Access Control	הרשאות ברזולוציית שורה ועמודה, נאכפות ברמת המנוע
AWS Lake Formation	מנגנון ניהול ההרשאות שמתחת ל-SageMaker Catalog
Apache Iceberg	פורמט טבלאות פתוח; AWS מציגה עמדת Iceberg First
Iceberg REST Catalog	הפרוטוקול שדרכו SageMaker מתחבר לקטלוגים חיצוניים כמו Unity Catalog או Horizon
S3 Tables	אחסון טבלאי מנוהל ב-Amazon S3, מעל Iceberg
Zero-ETL	אינטגרציות שמושכות דאטה ממקורות כמו SAP, Salesforce ובסיסי נתונים ישירות אל הלייקהאוס
OpenLineage	פרוטוקול פתוח לתיעוד lineage, שעליו נשענת יכולת ה-Traceability בפלטפורמה
MWAA	קיים בתצורת domains ובתצורת Amazon Managed Workflows for Apache Airflow serverless
Data Agent	סוכן AI בתוך הנוטבוק, עם גישה ל-Business Metadata של הארגון
JumpStart	ספריית מודלים מוכנים ב-SageMaker, עם sample notebooks ל-deploy
DuckDB	מנוע אנליטי מוטמע, זמין כחלק מהסביבה הפייתונית של SageMaker



Thank you



Please complete the session survey

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

שקופית הסיום, עם ה-QR לסקר הששן