Morgan Stanley

Using GenAI for Mainframe Operations

Alfred Loffredo, Executive Director

February 2025

Bio Page

Alfred Loffredo

Executive Director

Enterprise Z - Mainframe

Phone: +1 212 276-3630

Alfred.Loffredo@morganstanley.com

Years of Experience: Over 30 Years

Years at Morgan Stanley: 20 Years

Biography

I have spent much of my career in mainframe operations, supporting both infrastructure and batch processing. Over the years, I have progressed through roles ranging from overnight technician to leading a global team of 35 onshore and offshore professionals. I joined Morgan Stanley in 2006 as a day-shift technician and later advanced to shift manager overseeing overnight operations.

Over the past decade, my responsibilities have expanded from managing a single client's overnight support to directing the full mainframe operations function across two major businesses—Institutional Securities and Private Wealth Management—with teams based in the United States and Bangalore, India.

Introduced Artificial Intelligence into the production batch environment, driving automation, anomaly detection, and operational efficiency.

Led the Technology Award–winning z-Series Interchangeable Data Center Initiative for ISG and WM clients, enabling full production site swaps in under 30 minutes.

Unified mainframe infrastructure and batch operations into a single Command Center, delivering streamlined, end-to-end operational support for all Morgan Stanley mainframe environments.

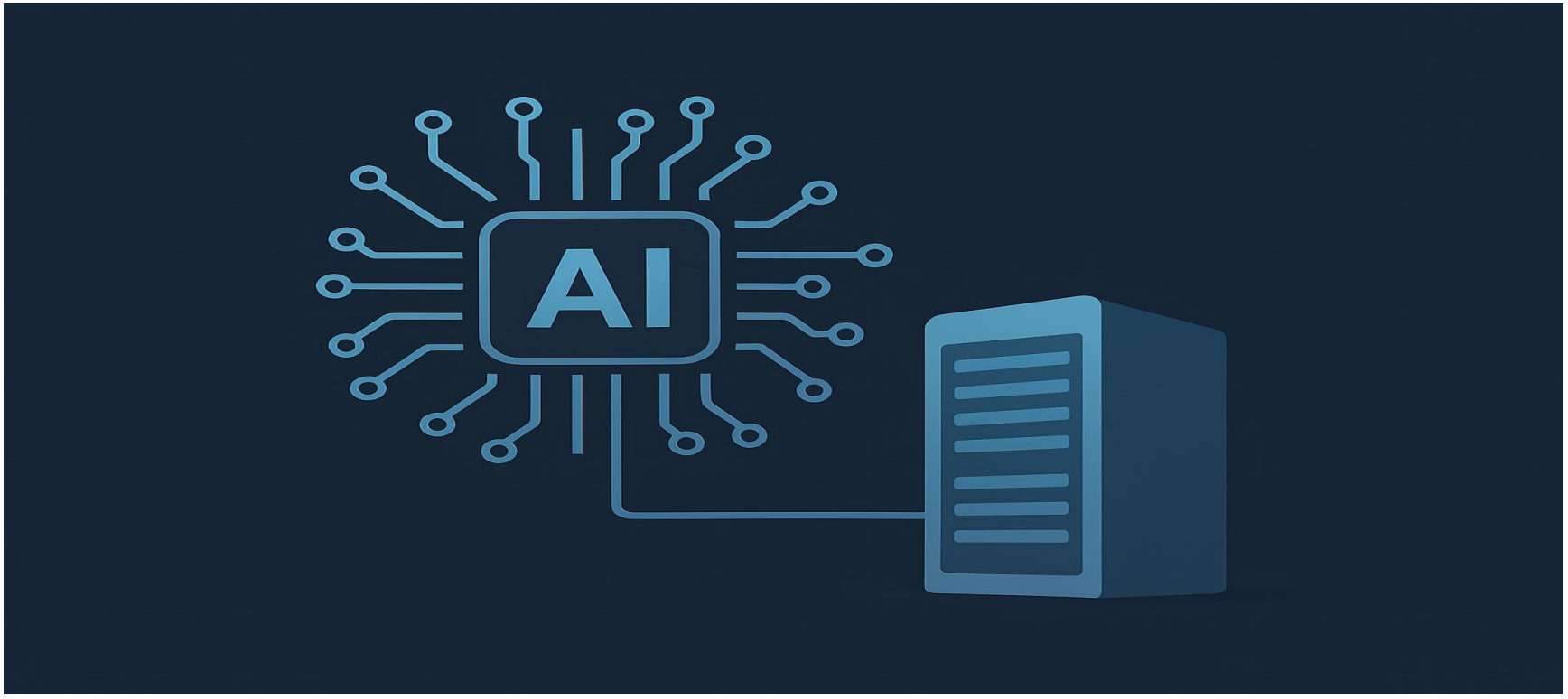
Managed four offshore teams since 2009 across Manila, Bangalore, and Shanghai, establishing resilient global follow-the-sun support models.

Directed the E*Trade mainframe integration, ensuring stability and operational continuity during platform consolidation.

Professional Objective

To consistently deliver award-winning operational excellence in technology that strengthen and accelerate Morgan Stanley's business outcomes.

Table of Contents



Why do GenAI for mainframe operations?

Mainframe batch job failures (abends) currently have an average Mean Time to Recovery (MTTR) of 58 minutes. This duration reflects the time from job failure to successful completion. A key driver of this extended MTTR is the fragmented and dispersed nature of diagnostic and technical data, which must be manually gathered from multiple systems and tools. This manual, multi-platform data collection process impacts SLA adherence, introduces downstream delays, and contributes to slower and inconsistent batch recovery.

Leveraging AI to unify diagnostic and technical data from key operational sources enables us to:

- **Accelerate recovery and lower MTTR** by delivering rapid, high-confidence analysis.
- **Improve SLA adherence** by reducing downstream delays and ensuring more predictable batch execution.
- **Enhance decision-making speed and operational efficiency** through automated insight generation.
- **Strengthen platform resilience** with faster, more consistent issue resolution across batch processing.
- **Empower less-experienced technicians** with clear, actionable guidance to resolve abends quickly.
- **Free skilled resources** to focus on higher-value initiatives rather than manual triage.
- **Scale institutional knowledge** through AI-assisted pattern recognition and standardized workflows.
- **Lay the foundation for safe, progressive auto-remediation** of recurring batch failures.

Our approach to using GenAI

Our overall approach integrates **GenAI**, **AI for Operations**, and **MTTR optimization** into a single, end-to-end strategy focused on accelerating decision-making, reducing operational toil, and strengthening resiliency across Enterprise Z operations.

AI-Driven Consolidation & Contextualization

We use GenAI as an AI-powered investigation and decision-support assistant that consolidates data across TOM, SDSF, logs, JCL, historical resolutions, and incident systems into a single, unified context. This reduces cognitive load, eliminates multi-system pivoting, and accelerates the signal-to-decision cycle.

Rapid, Consistent Analysis

GenAI delivers concise, under-one-minute analysis—probable cause, impacted components, and restart instructions—replacing manual investigation delays. This establishes a consistent, repeatable triage process that reduces operator variance and error.

Intelligent Automation of Operational Workflows

We integrate GenAI with operational tooling (e.g., ServiceNow, SmartIS, batch systems) to automate or semi-automate parts of incident workflows, beginning with low-risk, high-repeatability tasks and progressing toward safe, closed-loop remediation as confidence grows.

Standardization, Governance & Knowledge Reuse

The approach relies on firm-approved prompt patterns, centralized prompt libraries, and reusable playbooks that encode operational knowledge. This governance ensures accuracy, compliance, and scalable reuse across teams and regions.

End-to-End MTTR Reduction

By aggregating abend data and automating the analysis workflow, GenAI accelerates time-to-first-action, driving faster recovery. Guided steps and pattern-based recommendations standardize responses, improving SLA adherence and reducing human-driven variability.

Continuous Improvement & Coverage Expansion

We iterate on edge cases (e.g., protected jobs, large outputs, missing JCL), expand supported data sources, and enhance observability (health checks, analytics dashboards). Over time, this broadens operational coverage and enhances system resilience.

Our Journey

Established a cross-platform framework by partnering across Mainframe Computing, Technology Services, and application teams to align on objectives and required capabilities.

Mapped the end-to-end recovery workflow, mirroring current manual investigation steps to ensure the AI-driven process matched real operational behavior.

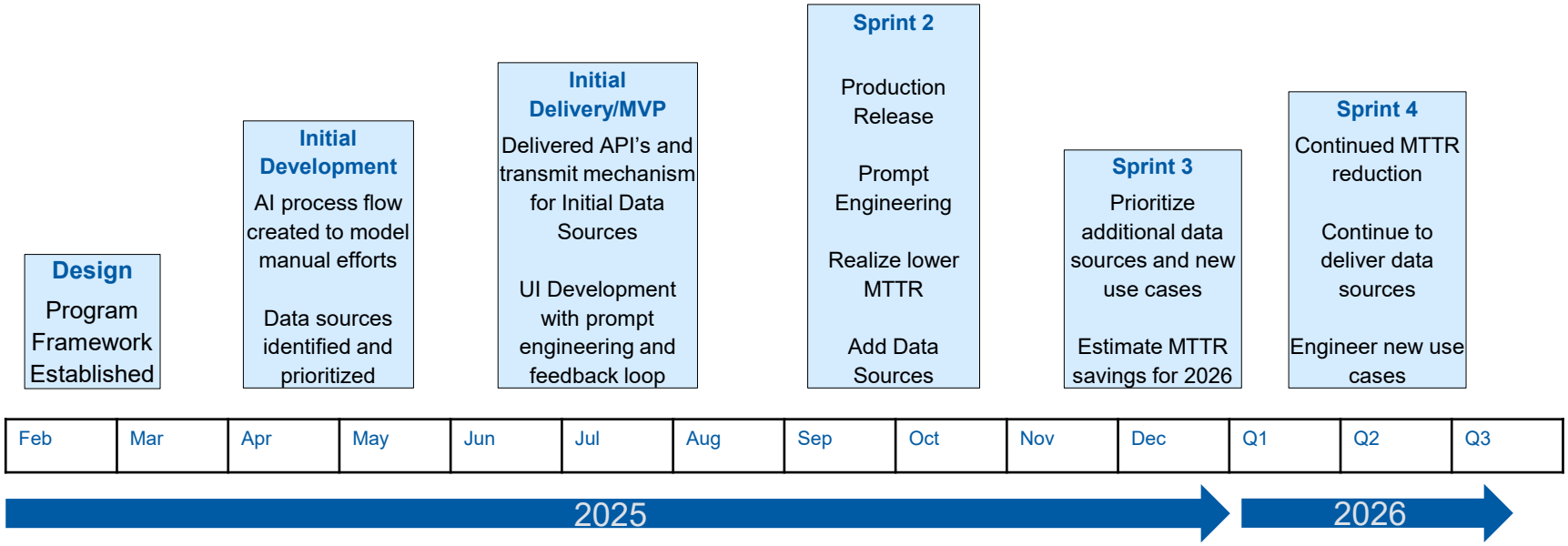
Identified and prioritized critical data sources (abend trigger, abend logs, and abend history) selecting only those required to accelerate our initial production release.

Delivered data-transmission mechanisms and APIs for the prioritized sources while developing a front-end Operations UI for abend analysis.

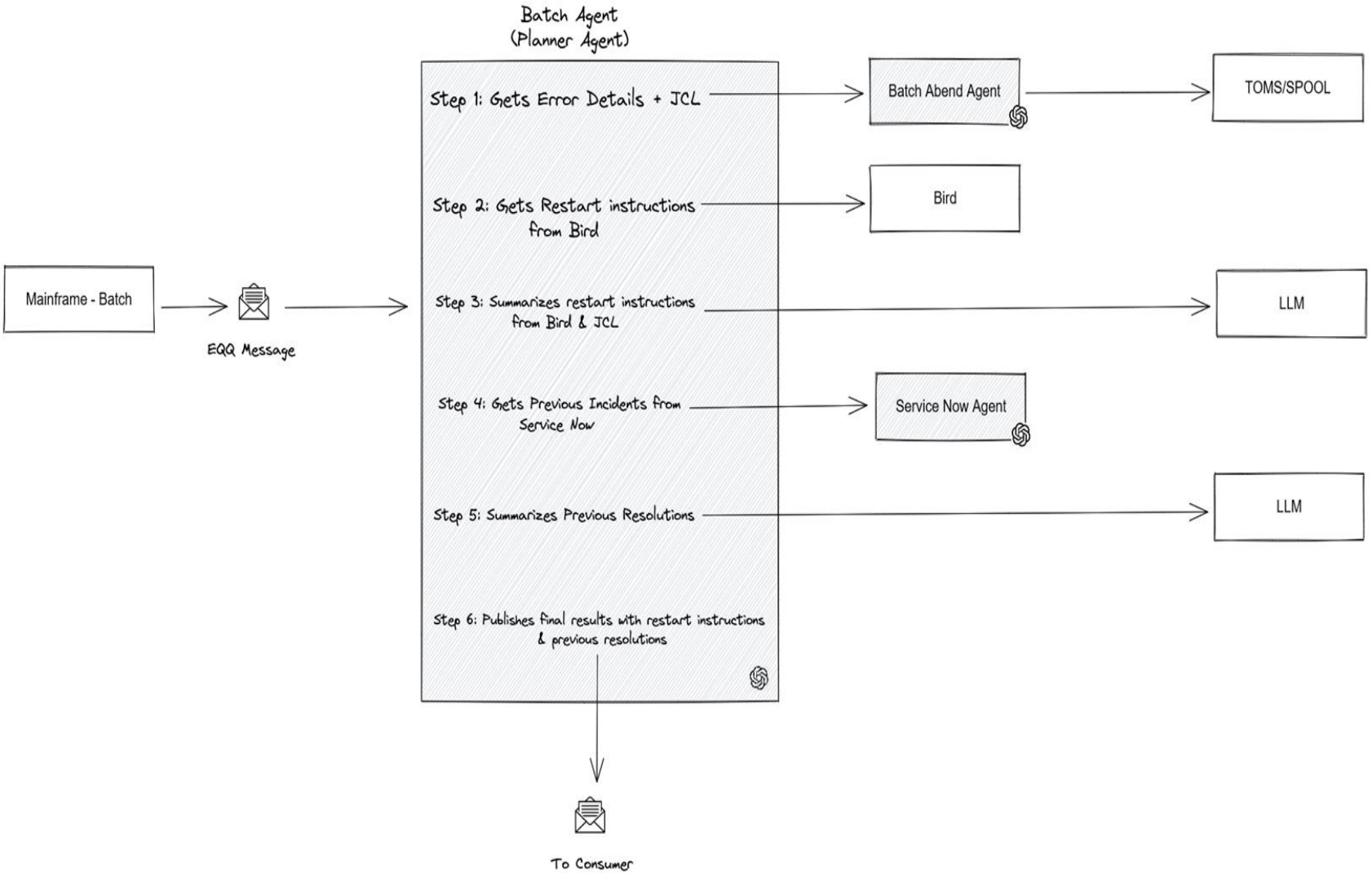
Built a feedback loop directly into the UI to capture accuracy, completeness, and improvement opportunities from Operations users.

Released the prioritized data sources into production, enabling immediate value and real-time learning.

Initiated prompt engineering and next-phase data onboarding immediately post-launch to expand coverage and enhance analysis quality



AI Ops Program: Flowchart



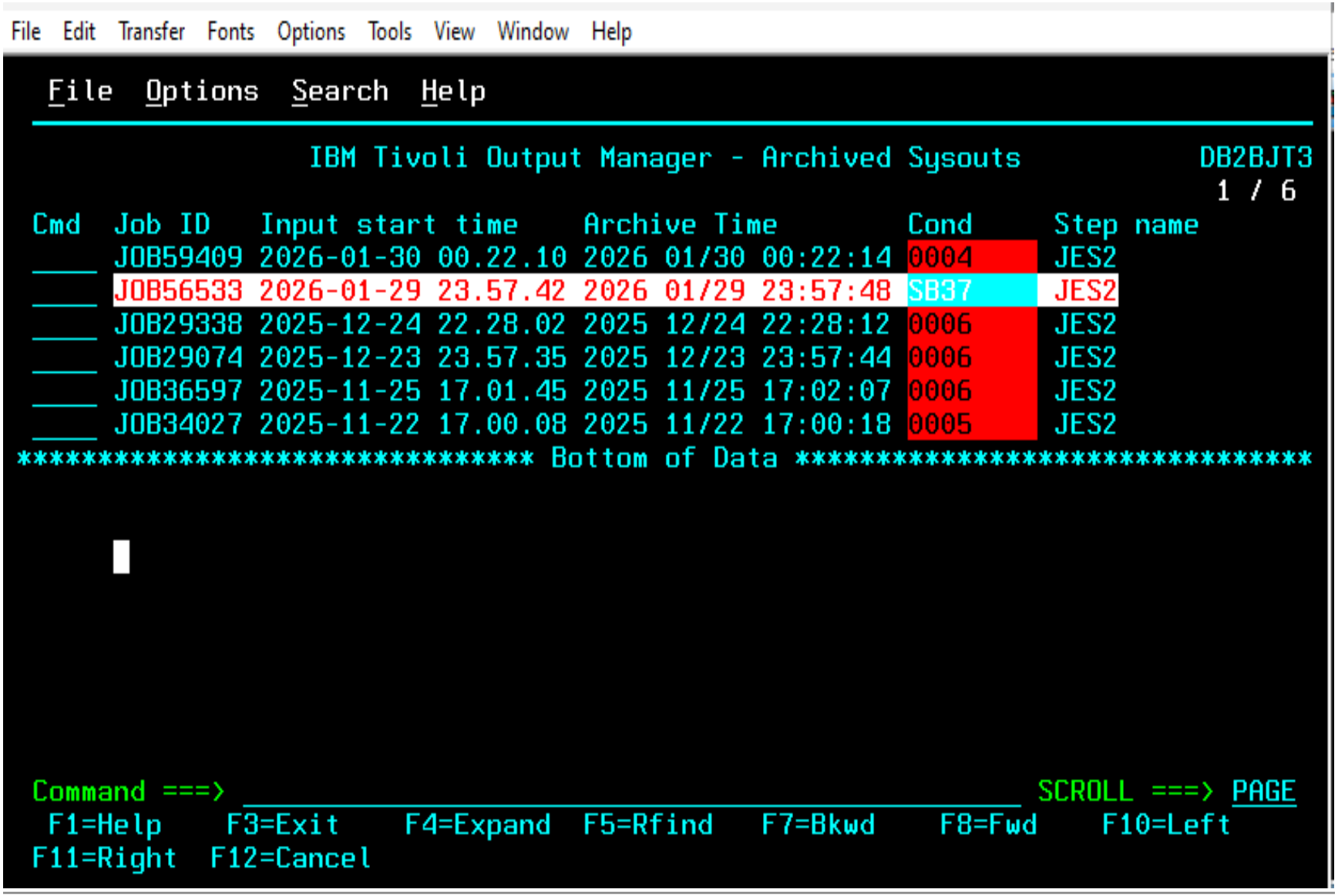
Our Journey Cont'd…

File Edit Transfer Fonts Options Tools View Window Help

```
H152NM (11.50.18)--------- Alert Monitor : History ----------Filter: QDCALERTW
Command ===>                                                  Scroll ===> CSR

          S/B=Browse T=Track N=Notes A=Analyze TT=TroubleTicket C=Close ?=More
    Date         Time     Description                              Resource
    29-JAN-2026 17.45.09 EQQ.63909 JOB26536 POPSDETE Has abend EQQ.63909
    29-JAN-2026 17.49.30 WTORMON Please check these pending re WTORMON
    29-JAN-2026 17.50.30 GRSALERT  CON: PQARDABP ON DSYS has b GRSALERT
    29-JAN-2026 19.00.00 GRSALERT  CON: PTEFD302 ON DSYS has b GRSALERT
    29-JAN-2026 20.13.02 EQQ.72782 JOB35782 PPSODPRE Has abend EQQ.72782
    29-JAN-2026 20.14.00 GRSALERT  CON: PALSDE71 ON DSYS has b GRSALERT
    29-JAN-2026 20.16.01 LATEALRT Job PMGSDXRF have not comple LATEALRT
    29-JAN-2026 20.25.06 EQQ.73506 JOB36607 PAECD126 Has abend EQQ.73506
    29-JAN-2026 22.03.32 EQQ.79412 JOB43267 PPLMRSWP Has abend EQQ.79412
    29-JAN-2026 22.03.33 EQQ.79413 JOB43266 PPLMRSWO Has abend EQQ.79413
    29-JAN-2026 22.30.07 EQQ.81007 JOB44950 PPLMRSWO Has abend EQQ.81007
    29-JAN-2026 23.01.20 EQQ.82880 JOB48939 PPLXRSWT Has abend EQQ.82880
    29-JAN-2026 23.26.00 WTORMON Please check these pending re WTORMON
    29-JAN-2026 23.44.20 EQQ.85460 JOB53859 PPLMRSWI Has abend EQQ.85460
    29-JAN-2026 23.57.44 EQQ.86264 JOB56532 PPLXRSWO Has abend EQQ.86264
    29-JAN-2026 23.57.45 EQQ.86265 JOB56533 PPLXRSWP Has abend EQQ.86265
    **END**
 F1=Help      F2=Split     F3=Exit      F4=Return    F5=Find      F6=Refresh
 F7=Backward  F8=Forward   F9=Swap                  F11=Right
```

Our Journey Cont'd…

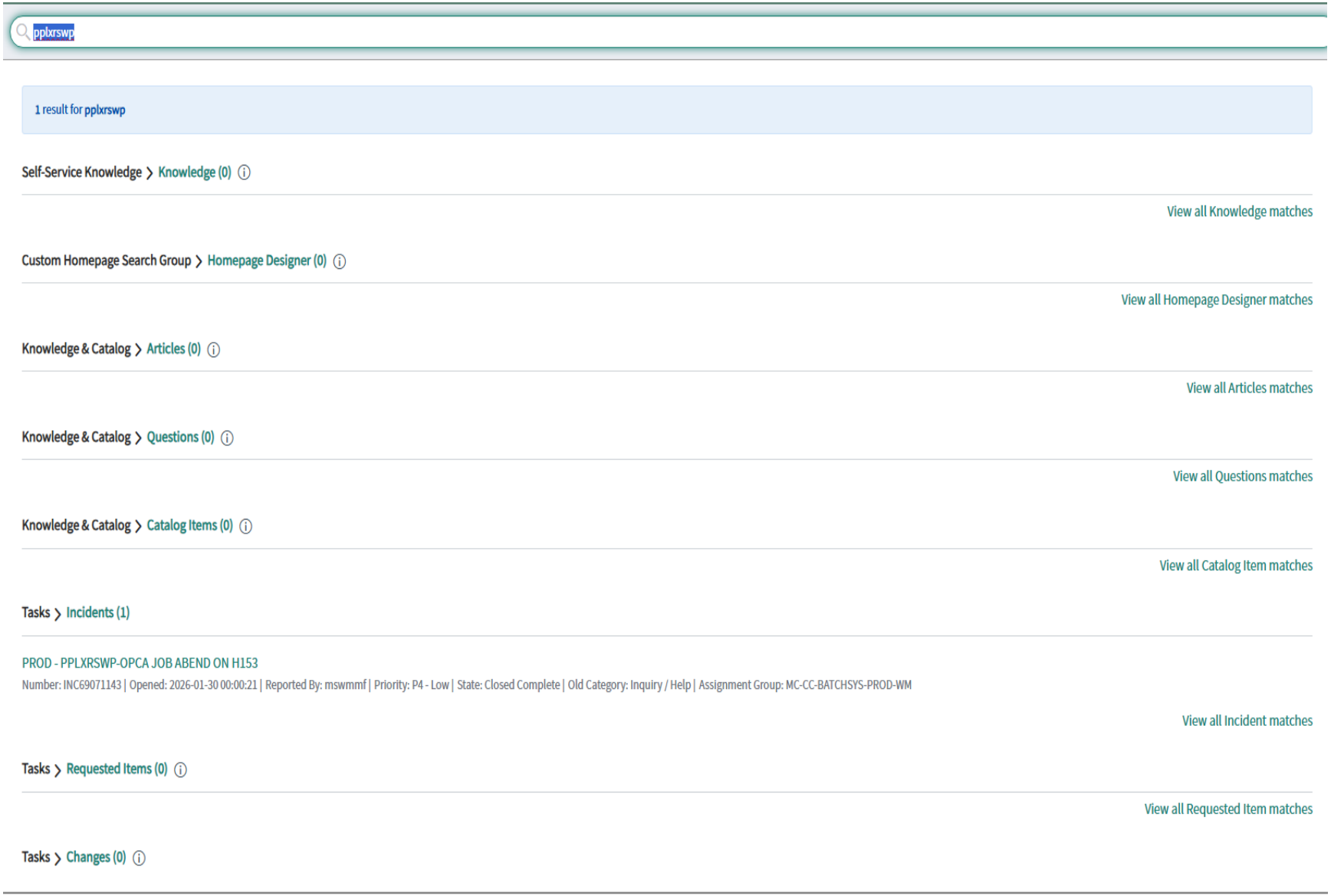


Our Journey Cont'd…

File Edit Transfer Fonts Options Tools View Window Help

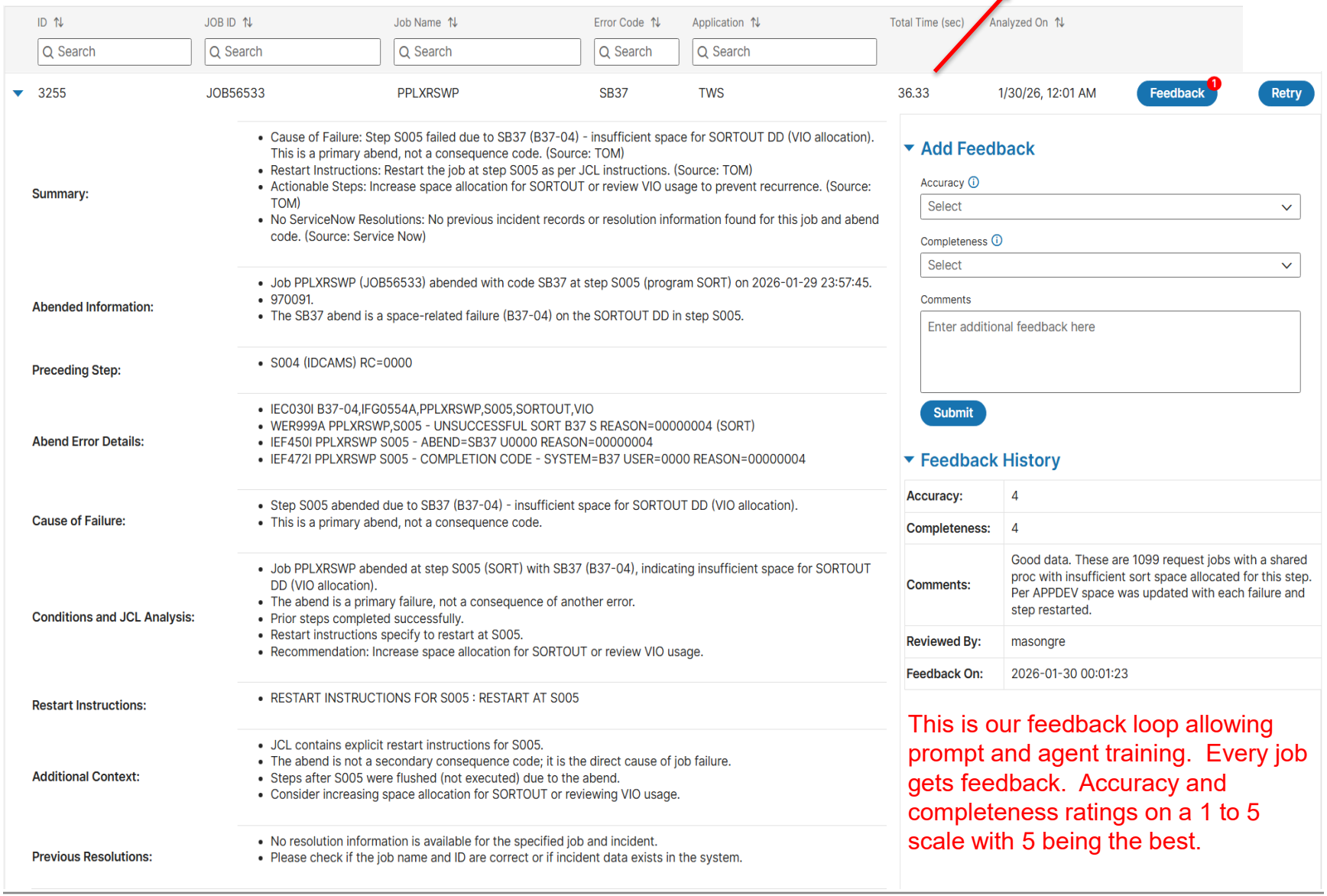
```
 File  Options  Help
BROWSE    PPLM1099SWHS#000 009 2601292203 > PPLXRSWP JOB56533 JES2 JESM Page 1
JOB56533  SURF920I STEP START: JOB=PPLXRSWP STEP=S005     PGM=SORT     STEP#=01
JOB56533  IEC030I B37-04,IFG0554A,PPLXRSWP,S005,SORTOUT,VIO ,      ,  824
          SYS26029.T235744.RA000.PPLXRSWP.RCSET2.H0A
JOB56533  WER999A PPLXRSWP,S005    ,         -  UNSUCCESSFUL SORT B37 S REASON=0
JOB56533  IDI0034I Fault analysis skipped due to: EXCLUDE option specification
JOB56533  IEA995I SYMPTOM DUMP OUTPUT  828
          SYSTEM COMPLETION CODE=B37  REASON CODE=00000004
           TIME=23.57.45  SEQ=31968  CPU=0000  ASID=0126
           PSW AT TIME OF ERROR  075C1000   80E2BAE6  ILC 2  INTC 0D
             NO ACTIVE MODULE FOUND
             NAME=UNKNOWN
             DATA AT PSW  00E2BAE0 - 41003038  0A0DB20A  00509808
             AR/GR 0: 009B0D20/00000000_00E2BD64   1: 00000000/00000000_A4B3700
                   2: 00000000/00000000_000211B8   3: 00000000/00000000_00E2BD2
                   4: 00000000/00000000_009AC4A8   5: 00000000/00000000_00E4400
                   6: 00000000/00000000_009AC74C   7: 00000000/00000000_009AC7A
                   8: 00000000/00000000_009AC76C   9: 00000000/FFFFFFFF_009AD9D
                   A: 00000000/00000000_009B23B0   B: 00000000/00000000_00E2B77
Command ===> _____________________________________________ Scroll ===> PAGE
 F1=Help    F3=Exit    F4=Mark    F5=Rfind   F6=Goto    F7=Up      F8=Down
F10=Left   F11=Right  F12=Cancel
```

Our Journey Cont'd… Service Now Job History



Our Journey Cont'd… GenAI User Interface (UI)

This is the time it took AI to return the summary.



ID	JOB ID	Job Name	Error Code	Application	Total Time (sec)	Analyzed On
3255	JOB56533	PPLXRSWP	SB37	TWS	36.33	1/30/26, 12:01 AM

Feedback 1 | Retry

Summary:
- Cause of Failure: Step S005 failed due to SB37 (B37-04) - insufficient space for SORTOUT DD (VIO allocation). This is a primary abend, not a consequence code. (Source: TOM)
- Restart Instructions: Restart the job at step S005 as per JCL instructions. (Source: TOM)
- Actionable Steps: Increase space allocation for SORTOUT or review VIO usage to prevent recurrence. (Source: TOM)
- No ServiceNow Resolutions: No previous incident records or resolution information found for this job and abend code. (Source: Service Now)

Abended Information:
- Job PPLXRSWP (JOB56533) abended with code SB37 at step S005 (program SORT) on 2026-01-29 23:57:45.
- 970091.
- The SB37 abend is a space-related failure (B37-04) on the SORTOUT DD in step S005.

Preceding Step:
- S004 (IDCAMS) RC=0000

Abend Error Details:
- IEC030I B37-04,IFG0554A,PPLXRSWP,S005,SORTOUT,VIO
- WER999A PPLXRSWP,S005 - UNSUCCESSFUL SORT B37 S REASON=00000004 (SORT)
- IEF450I PPLXRSWP S005 - ABEND=SB37 U0000 REASON=00000004
- IEF472I PPLXRSWP S005 - COMPLETION CODE - SYSTEM=B37 USER=0000 REASON=00000004

Cause of Failure:
- Step S005 abended due to SB37 (B37-04) - insufficient space for SORTOUT DD (VIO allocation).
- This is a primary abend, not a consequence code.

Conditions and JCL Analysis:
- Job PPLXRSWP abended at step S005 (SORT) with SB37 (B37-04), indicating insufficient space for SORTOUT DD (VIO allocation).
- The abend is a primary failure, not a consequence of another error.
- Prior steps completed successfully.
- Restart instructions specify to restart at S005.
- Recommendation: Increase space allocation for SORTOUT or review VIO usage.

Restart Instructions:
- RESTART INSTRUCTIONS FOR S005 : RESTART AT S005

Additional Context:
- JCL contains explicit restart instructions for S005.
- The abend is not a secondary consequence code; it is the direct cause of job failure.
- Steps after S005 were flushed (not executed) due to the abend.
- Consider increasing space allocation for SORTOUT or reviewing VIO usage.

Previous Resolutions:
- No resolution information is available for the specified job and incident.
- Please check if the job name and ID are correct or if incident data exists in the system.

Add Feedback

Accuracy ⓘ
Select

Completeness ⓘ
Select

Comments
Enter additional feedback here

Submit

Feedback History

Accuracy:	4
Completeness:	4
Comments:	Good data. These are 1099 request jobs with a shared proc with insufficient sort space allocated for this step. Per APPDEV space was updated with each failure and step restarted.
Reviewed By:	masongre
Feedback On:	2026-01-30 00:01:23

This is our feedback loop allowing prompt and agent training. Every job gets feedback. Accuracy and completeness ratings on a 1 to 5 scale with 5 being the best.

AI Improved this!

1.Initial Failure Alert
⏱ *1 min*
→ Triggered by system monitoring.

2.Logon to System
⏱ *4 mins*
→ Access the mainframe or monitoring platform.

3.Navigate to TOM
⏱ *2 mins*
→ Open the Tivoli Output Manager (TOM) interface.

4.Search Failed Job
⏱ *2 mins*
→ Use job ID or filters to locate the failed job.

5.Select Failed Job
⏱ *2 mins*
→ Click into the job details.

6.Navigate to Job Output
⏱ *2 mins*
→ Access the output logs or reports.

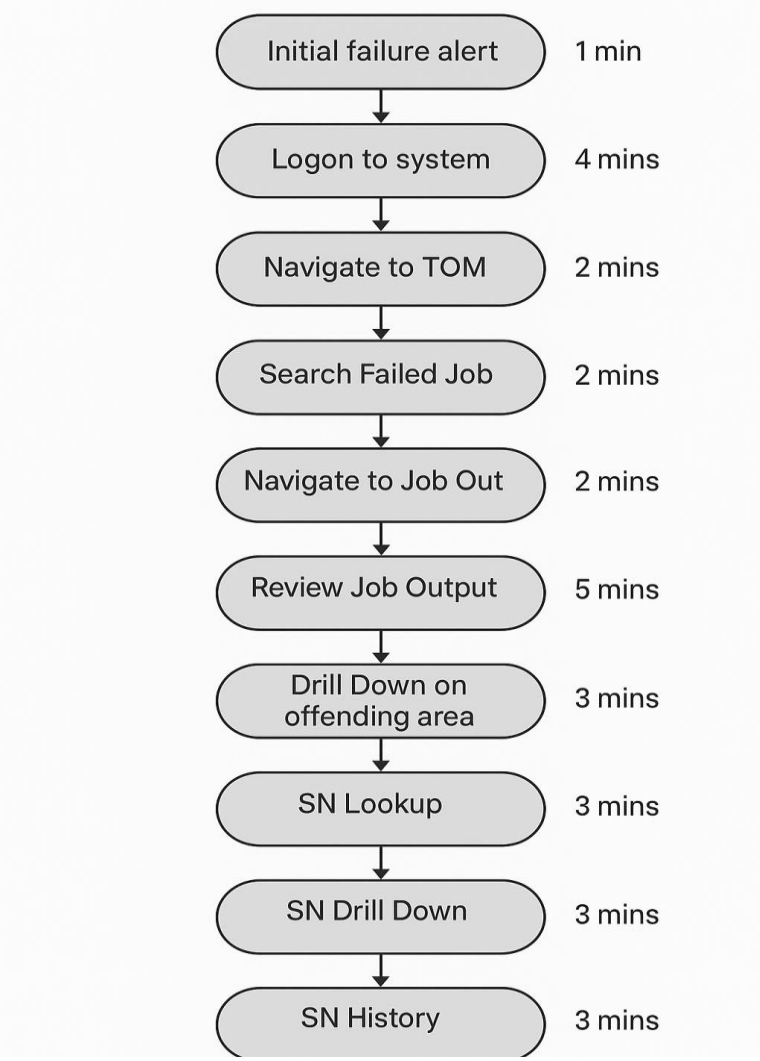
7. Review Job Output
⏱ *5 mins*
→ Analyze logs for error messages or anomalies.

8. Drill Down on Offending Area
⏱ *3 mins*
→ Focus on the specific section causing the failure.

9. SN Lookup
⏱ *3 mins*
→ Search for the ServiceNow (SN) ticket or reference.

10. SN Drill Down
⏱ *3 mins*
→ Investigate the SN ticket details.

11. SN History
⏱ *3 mins*
→ Review past incidents or resolutions related to the job



NOTE: 30-minute detection and analysis does not include resolution

Future Enhancements

1. Dynamic Agent-Based Automation - These enhancements standardize and accelerate recovery workflows. Next-gen dynamic AI agents will:

- Diagnose batch failures automatically
- Recommend or possibly execute fixes
- Reduce operator-dependent variability

2. Automated Compliance & Governance

- JCL standards
- Operational policy compliance
- Risk-flagging
- This reduces manual checks and removes friction in the recovery path.

3. Cross-Platform (Mainframe + Distributed) Scalability

- Automation capabilities will extend beyond mainframe into distributed batch ecosystems
- consistent MTTR improvements across all compute layers.
- Real-time dependency graphs between mainframe, distributed, Autosys/Control-M, network and storage
- Impact prediction - which business units and apps are affected?

4. AI-Driven Batch Scheduling Optimization - This leads to faster, more reliable batch completion.

- Detection of job dependency issues
- Cycle-time bottleneck prediction
- Visibility into cross-dependencies
- Hot-spot prediction zones
- Expected impact on upcoming batch and business SLAs
- Real-time guidance for Command Center staffing

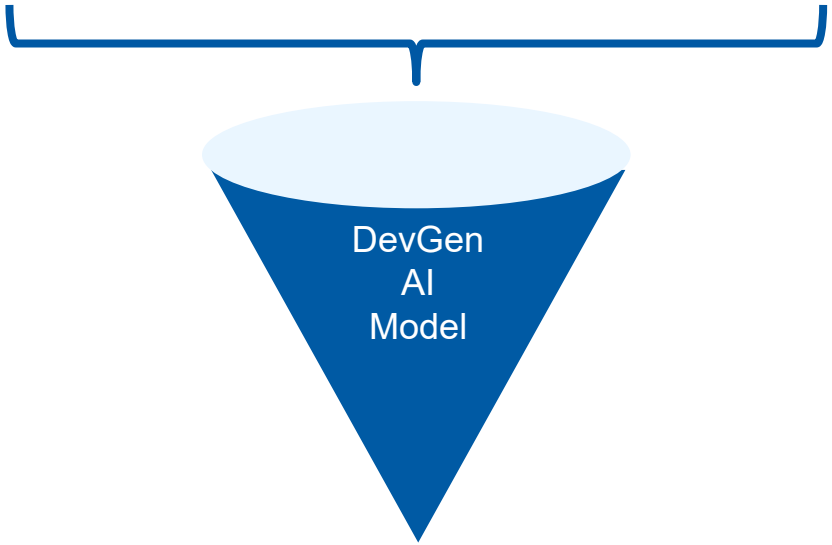
5. Standardized Recovery Playbooks - Driving consistent, predictable MTTR reduction across teams.

- Reusable AI playbooks
- Integration across L1/L2/L3
- Global reuse of institutional knowledge

These enhancements collectively form an AI-driven, proactive, standardized, cross-platform operating model designed to significantly reduce MTTR, improve reliability, and strengthen operational resilience across Wealth Management batch processing.

Future Enhancements Cont'd… Additional Batch Data Sources

- Dump data (Fault Analyzer)
- Raw Dump data & Syslog
- Job console output data (SDSF Spool)
- Historical file transfer logs (NDM log)
- Technical vendor Information (Whitepapers, Redpapers)
- Summarized technical vendor information (QuickRef)
- Agentic GenAI additional IBM information (WatsonX)
- Abend Output data (TOM)
- Incident History data (ServiceNow 1 & 2)
- Technical historical data (Splunk)
- Abend Restart Instructions data (Bird)
- JCL Abend Restart Instructions and Dependencies data (SmartIS)
- Ownership and TAI data (MRC)
- Program and JCL code data (GiT Code)



✓**Data Summary**
✓**Course of Action**

Future Enhancements Cont'd…

The next evolution is to extend this into infrastructure components (CICS, DB2, MQ, storage, network, etc…) by leveraging logs, metrics, and event streams to identify: This expands "job failure prediction" into true infra-level outage prevention.

1. AI-Driven Predictive Infrastructure Failure Detection

- Impending resource exhaustion
- Latency anomalies
- Configuration drifts
- Network degradation patterns

2. Intelligent Alert Noise Reduction & Severity Normalization - AI Ops can enhance alerting by:

- Clustering related alerts to eliminate noise
- Auto-classifying severity based on business impact and dependency analysis
- Combining signals from TOM, SDSF, NetMaster, ServiceNow, and monitoring tools
- Combine infra logs (CICS region dumps, MQ channel stats, DB2 performance traces, network logs)

5. Closed-Loop Infra Remediation via Dynamic AI Agents - Dynamic agent orchestration is already documented as part of DevGenAI's roadmap. All governed through policy-based approvals with human-in-the-loop.

- Health-check scripts triggered autonomously
- System health scorecards – Normalize comparison
- Automated restart/recovery for MQ channels, CICS regions, or DB2 threads when safe
- Pattern-based self-healing

7. Multi-Source Infra Data Unification (TOM + SDSF + NetMaster + Monitoring + SNOW) - future enhancements would broaden this to:

- SNMP traps
- Network telemetry
- Storage/HSM logs
- MQ/CICS performance logs

9. Intelligent Infra Alert Routing, Automated Ticketing & PagerDuty Escalation - specifically noted improved alert routing and automated ticketing as a key opportunity. Future enhancements:

- Auto-open ServiceNow incidents with enriched context
- Auto-assign to the correct resolver group
- Dynamic severity updates based on evolving conditions
- PagerDuty escalations for batch and infra

Questions ?