# 2023-03-29 PLA Sales, GMV: Offline > Realtime by 7%

|  |  |
| --- | --- |
| **Environment** | Production - Internal |
| **Description** | Brief Issue Description:  PLA Sales and GMV offline metrics greater than realtime by 7%.  Pre-existing query on prs\_t.pl\_org\_ads\_sales\_fact resulted in duplicate fields, artificially inflating the GMV.  This showed up as incorrect data in Clickhouse and Control Center (pulled from Clickhouse)  What was failing: Offline merge step for PLPS data was not deduping correctly.  Issue introduced with Halo release on 3/30. |
| **Reviewed by** |  |
| **Problem Report** | 3/30 Halo Attribution rollout- [prs\_t.pl\_org\_ads\_sales\_fact](https://wiki.corp.ebay.com/display/DataServicesandSolutions/PL_ORG_ADS_SALES_FACT)  Transaction status cps\_trans\_sts\_id for Halo rolled out with Halo Launch  4/17 Ads Infra team noticed offline and realtime data for PLA Sales was not consistent while manually investigating another issue and reported the issue to ADPP in case it was an issue with the clickhouse offline data processing.  [Tainov, Rinat](file:////display/~rtainov) began investigation.  4/18 Rinat provided sample queries in case saleType empty strings were not being accounted for.  [Wang, Yazhi](file:////display/~yazhwang) found the same query with eventType=PLPS and **sum(count)** as sales count returned different counts between offline and realtime, where offline returned some records with count=2.  4/20 Rinat reproduced the issue on offline data from Hadoop Apollo where a formerly validated query resulted in duplicates, which was causing gmv issues.  prs\_t.pl\_org\_ads\_sales\_fact, group by following fields: |  eventType, |  transactionId, |  adFeesCurrencyId, |  buyerId, |  campaignId, |  categoryId, |  itemId, |  placementId, |  saleCurrencyId, |  sellerId, |  transactionSite, |  listingSite, |  keyword, |  sellerKwId, |  sellerKeyword, |  adgroupId, |  orderId, |  checkoutStatus, |  wacko, |  firstPriceScale4, |  campaignType, |  exactMatch, |  priceType, |  plpRevenueInSiteCurrencyScale4, |  pdRevenueInSiteCurrencyScale4, |  duplicatePLPS, |  newForSeller, |  device, |  revenueInSiteCurrencyScale4, |  clickSrc, |  eventts, |  saleType  Rinat engaged [Bakkireddy, Veera](file:////display/~vbakkireddy) and [Venugopal, Shankar](file:////display/~shvenugopal) to check from DSS side  Veera identified the issue with one of the merge steps for PLPS which started with Halo attribution release, fixed the data, corrected the merge job code.  Accurate lower value Sales data was corrected by 4/20 in Apollo & picked up by Clickhouse.  4/21 - Data Issue resolved - Control Center corrected offline data was backfilled as part of [SHEPHERD-2942](https://jirap.corp.ebay.com/browse/SHEPHERD-2942) - Backfill corrected offline sales for PLA since March 29 Closed  [Roy, Puja](file:////display/~pnroy) confirmed that SellerHub reporting data uses latest version so there was no impact to seller and no seller communication was necessary. |
| **FCI** |  |
| **Revenue Impact** | 0.74% total CPC transactions were duplicated used in DSAM-23126.  Analytics GMV impact  No Revenue impact, no seller impact. |
| **Domain Impacted** |  |
| **Triage** | How issue was found and information supplied to indicate there was an issue |
| **Examine** | Steps taken to investigate issue |
| **Root Cause** | Issue with Hadoop merge job for eventType=PLPS for [prs\_t.pl\_org\_ads\_sales\_fact](https://wiki.corp.ebay.com/display/DataServicesandSolutions/PL_ORG_ADS_SALES_FACT) causing duplicates on some transactions which resulted in double counted PLA GMV related metric values for some transactionIds if the query did not check for the latest transaction eventts value. |
| **Cure** | Merge job corrected  Corrected data backfilled in Hadoop Apollo and populated to Clickhouse, then Control Center.  DSS correcting offline data DSAM-23126, base tables updated and Tableau dashboard will need to be refreshed. [Abzal, Wahid Khan](file:////display/~wabzal) |
| **Opportunities** | What can be improved   * Monitoring   + Enhance PLA Offline vs realtime | NuMonitor parity alerts - add PLA Sales and GMV parity check. PLA Revenue check exists. * Code   + Dual attribution testing for duplicate transactions coverage for PLS only.  Regression tests needed for PLA and combinations. [Bakkireddy, Veera](file:////display/~vbakkireddy)   + Duplicate transaction validations added for any downstream users. [Bakkireddy, Veera](file:////display/~vbakkireddy)   + Daily Check - 15% difference has been triggered by new products (Promoted Display) and should be split between program lines to isolate stable products from new emerging ones. Should Rinat/ADPP also add some additional checks before uploading.   + Transaction Count should be =1, added on Attribution and DSS Quality check * Automation * Process   + Current slack channel for ADPP & DSS to discuss data delay [Venugopal, Shankar](file:////display/~shvenugopal)   + DSS PagerDuty   + Indicator of where the data is coming from - DSS, Offline, etc. to contact correct team. * Staging   + Staging tests lack prod-like data due to firewall   + S3P can be queried to download prod data to staging tooling [Iyamperumal, Ganesh](file:////display/~gaiyamperumal)  and [Bakkireddy, Veera](file:////display/~vbakkireddy) to discuss |
| **Flow** | Data/Pool Dependencies |
| **Rollback** |  |
| **Related items** | [SHEPHERD-2942](https://jirap.corp.ebay.com/browse/SHEPHERD-2942) - Backfill corrected offline sales for PLA since March 29 Closed  [DSAM-23126](https://jirap.corp.ebay.com/browse/DSAM-23126) - CPC Attributed transaction - data issue Resolved |
| **Status** | Incident Status, RCA edit status |