

RND-2: PYSPARK ??? ??????????? ??????? AA-TEST ?? HYPEX

????????? ? ??????????

???? ?????????????? ? ??????, ??? ??????????? ?????????? AA-test ?? spark-????????? ????????, ???, ????? ??????? ?????? ??? ?????????? ?????? ??????.

????????? ? ????????

? H1: ??????? ?????? ? ?????????????? ?????????? ?? ?????? ?????????????? ??? ?????? ?????????????? ?????? ?????????????? ??????.

? H2: ??????? ? one-pass ?????????? ?????????? ?????? ? ?????? shuffle-like ???????.

? H3: ??? ??????? ?????? (mean, var, quantile proxy) ?????????? ??????????, ??????? ?????? ?????? ?????? ?????? ??????.

????????????? ? ??? ??????????????

? ?????????? ?????????? ?????????????? spark-like ????????: ?????? ?????????? ?? ????????, ?????? ??? ?????????? 3 ??????????:

1. ??????: ?????????? ?????? ?? ?????? ??????.

2. Batch-?????????: ??? ?????? ?? ?????? ??????????.

3. Incremental combine: ?????????? summary + merge.

????????: ?????? ?????? N ? ??? ?????? M. ?????????? wall-clock ? ??? ??????.

???????????

????????? ??? ?????????????? ? ??????? summary ? ????????. ?????????? ?????? ?? ??????????: batch/incremental ??? ??????? 1.5?3x ??? ????? N ? M.

????????? ??????? ? ?????????? ?. ? ??????? "????????????? 1" ? "????????????? 2" ???????.

????? ? ????????????

? ??? AA-test ? pyspark-????????? ?????? ?????????????? API ?????? ?????????????? sufficient statistics.

? ??????? ? ??? ?????? ?????? ?? ?????? ? ?????????????? ?????? ?????????? ??????????.

? ??????? ?????????? 'groupBy/agg' ?????? ?????????? ?????? ??? ??????.

?????

? ?. RnD-1: '01_bonferroni_aa_matching/report.md'.

? ??????? ??????????: 'docs/CONTEXT_TZ.md'.