

## 7. Duplicates loading the data

### Duplicates are solved in RDV load

Using hash diff to tell the difference and intra load-cycle changes activated in parameters

A duplicate BK that is exactly the same will generate the same HASHDIFF and will be loaded once in the SAT.

A duplicate BK with different values will have a different hashdiff and will be loaded to the SAT with both instances. Otherwise, you lose data. The load date of the second record is augmented by +2 microseconds. Order of the SATs is in this case not exact as we do not know it.

**EXTRA INFO**  
VaultSpeed also supports having an attribute that contains the sequence of the data (for example the logposition if you have a CDC system in place)

```

1 select * from w111ba1d.w111ba1d_0v_fl_sat_web_product where PRODUCTID in (20,21);
2
3 select PRODUCT_MKEY ,PRODUCTID, EQUAL, HASH_DIFF, PFLANZABSTAND, PFLANZORT, PREIS, TYP, UMFANG
4

```

PRODUCT_MKEY	PRODUCTID	EQUAL	HASH_DIFF	PFLANZABSTAND	PFLANZORT	PREIS	TYP	UMFANG
71CB392816AC4E218E852E9E2E18778DF61E2078	21	0	564912D36E43EECF90DE152917F3160969C141	35 x 20 cm	Freiland	5,00	3	7
71CB392816AC4E218E852E9E2E18778DF61E2078	21	0	1F6D851A88726522A4AD02B98951F65290BCF147	30 x 25 cm	Freiland	4,00	2	6
7F186E0FD9D7E45797A5C9269E3C26FF9755834	20	0	79E5E073F550791E019F7618018D22776FC4E4EE	70 x 50 cm	Freiland	8,00	2	6
7F186E0FD9D7E45797A5C9269E3C26FF9755834	20	1	79E5E073F550791E019F7618018D22776FC4E4EE	70 x 50 cm	Freiland	8,00	2	6

**STAGING: notice the EQUAL calculation and identical hashdiffs for productid 20**

```

1 select PRODUCT_MKEY ,PRODUCTID, LOAD_DATE, LOAD_CYCLE_ID, HASH_DIFF, PFLANZABSTAND,DELETE_FLAG ,PFLANZORT, PREIS, TYP, UMFANG from w111ba1d.w111ba1d_0v_fl_sat_web_product where PRODUCTID in (20,21);
2
3 select PRODUCT_MKEY ,PRODUCTID, EQUAL, HASH_DIFF, PFLANZABSTAND, PFLANZORT, PREIS, TYP, UMFANG from w111ba1d.webshop_stg_sat_web_product_top where PRODUCTID in (20,21);
4

```

PRODUCT_MKEY	PRODUCTID	LOAD_DATE	LOAD_CYCLE_ID	HASH_DIFF	PFLANZABSTAND	DELETE_FLAG
71CB392816AC4E218E852E9E2E18778DF61E2078	21	2023-06-17T16:09:06.405+0000	312	564912D36E43EECF90DE152917F3160969C141	35 x 20 cm	N
71CB392816AC4E218E852E9E2E18778DF61E2078	21	2023-06-17T16:09:06.405+0000	312	1F6D851A88726522A4AD02B98951F65290BCF147	30 x 25 cm	N
7F186E0FD9D7E45797A5C9269E3C26FF9755834	20	2023-06-17T16:09:06.409+0000	312	79E5E073F550791E019F7618018D22776FC4E4EE	70 x 50 cm	N

**RAW VAULT SATELLITE: the record that was equal was filtered out**

VaultSpeed used the hash diff to tell the difference between records. Intra load-cycle changes should be enabled in our parameters.  
 A duplicate BK that is exactly the same will generate the same HASHDIFF and will be loaded once in the SAT.  
 A duplicate BK with different values will have a different hashdiff and will be loaded to the SAT with both instances. Otherwise, you lose data. The load date of the second record is augmented by +2 microseconds. Order of the SATs is in this case not exact as we do not know it.