

Izzivi pri obstoječem sistemu

- Nenehno naraščanje količine podatkov.
- ETL procesi se soočajo s performančnimi težavami.
- Preobremenitev produkcijskih baz vpliva na delovanje sistemov.
- Sprejeta odločitev: **posodobitev ETL procesov** za boljšo učinkovitost.



Priprava DWH podatkov v družbi Generali zavarovalnica – primer iz prakse

Miha Petrič

Generali zavarovalnica d.d.
Specialist za podatkovna skladišča

Boris Oblak

Abakus plus d.o.o.

ORACLE® | CERTIFIED
PROFESSIONAL



ORACLE® Gold
Partner

Specialized
Oracle Database

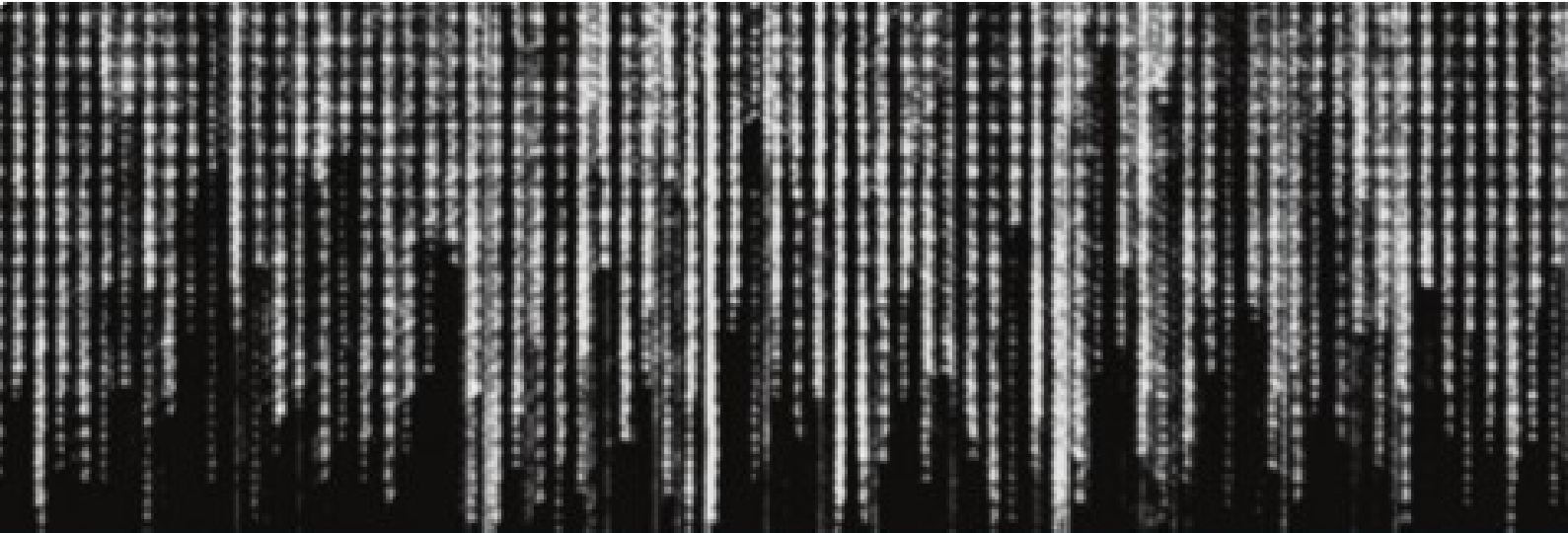
dejavu *Data at your service.*

Abakus Plus d.o.o.

- History
 - From 1992
 - ~20 employees
- DBA Applications
 - ♠ DeJaVu
 - ♠ ARBITER
 - ♠ APPM
(<https://appm.abakus.si/>)
(*free* edition available)
- Enterprise Applications
 - ♠ Document Management
 - ♠ Newspaper Distribution
 - ♠ Flight Information System
- Services
 - OS & Network admin
 - DBA, Programming
- Infrastructure
 - > 20 years of experience with
High Availability on GNU/Linux
- Hardware
 - Servers, SAN, Ceph,
 - Firewalls,
 - Backup Server

Abakus and Oracle

- Oracle Database on Linux
 - **Abakus: 1995**
(Oracle 7.1.5, Forms 3.0)
 - Oracle: 1997
- Parallel execution
 - **Abakus: 2004**
(SIOUG 2004: Vzporedno Izvajanje operacij s PL/SQL – Boris Oblak)
 - Oracle: 2007
dbms_parallel_execute



Deja Vu virtual databases

Data at your service.

dejavu

Customers

Gorenjska Banka

GENERALI
Zavarovalnica

Ljubljana Airport

Iskra

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA OBRAMBO

NOVA
BANKA

MILENIJUM[®]
OSIGURANJE

DRINA
OSIGURANJE

MM
KARTON

KONTROLA
ZRAČNEGA
PROMETA
SLOVENIJE

Horia

Mestna občina
Ljubljana

skbbanka
otp group

PH

triglav

GOODYEAR

TRELLEBORG

UNIVERZITETNA PSIHIATRIČNA
KLINIKA LJUBLJANA
University Psychiatric Clinic Ljubljana

BANKA
SLOVENIJE

Blubit

SAVA
INFOND

PRVA

ELES

Фонд
здравственог осигурања
Републике Српске

NLB Vita
Življenjska zavarovalnica

G96

EKDIS

CENTROSINERGIJA

SAVA
HOTELS & RESORTS

terme čatež

NLB Skladi

Mercator

MAGNETIK

MERKUR

SPL d.d.

ZAVOD ZA
ŠPORT RS
PLANICA

editel

NFOTRANS

LASERLINE

studio item

PRONET
CHOOSE THE FUTURE

hit alpinea
Kranjska Gora

ORACLE

ROS d.o.o.

sij acroni

ANDRITZ

jata emona
LJUBLJANA

ADRIA ANKARAN
HOTEL & RESORT

AS
Abakus
As na disku.

dejavu Data at your service.

Uvod

- Generali je ena izmed 3 največjih zavarovalnic v Sloveniji s širokim naborom zavarovalnih produktov življenjskih in neživljenjskih zavarovanj.
- Imamo veliko potrebo po podatkih.
- Posebej pomembna nam je časovna komponenta.

Stanje pred prenovo

- Anchor DW na oracle bazi, kjer smo imeli vse stopnje ETL.
- Ogrodje in ideja arhiviranja je že bila postavljena – vsak dan prenašati podatke (spremembe po primarnem ključu) iz izvornih tabel in jih shraniti na bazi v kopijah teh izvornih tabel oz. tako imenovanih arhivskih tabelah.
- Na arhivskih tabelah smo potem gradili DW, kjer smo vedno imeli možnost poustvariti zgodovino.

Prenova

- Zaradi vse večjih količin podatkov smo imeli performančne težave.
- Na obzorju je bila združitev z Adriatic Slovenica zavarovalnico.
- Imeli smo nekaj napak v kodi kar je ogrožalo zanesljivost podatkov.

Izzivi pri obstoječem sistemu

- Preobremenitev, ETL procesi niso končali pravočasno.
- Obremenitev produkcijskih baz.
- Težave z referenčno integriteto (podatki so se prenašali iz živih baz).
- Omejitve glede popravkov podatkov za pretekli teden.
- Velikost baz: približno 40TB.

Zahteve za novi sistem

- Razbremenitev produkcijskih baz.
- Rešitev za referenčno integriteto.
- Zagotovitev, da ETL procesi končajo pravočasno.
- Možnost popravkov podatkov za en teden nazaj.
- Skalabilnost.
- Ohraniti strukturo in vsebino arhivskih podatkov.

Ideja za novo rešitev

- Podatke bi prepisovali iz read-only standby baze (razbremenitev produkcije).
- Active Data Guard za omogočanje naprednega dostopa.

Ideja za novo rešitev

- Podatke bi prepisovali iz read-only standby baze (razbremenitev produkcije).
- Active Data Guard za omogočanje naprednega dostopa.
- DejaVu rešitev omogoča read-only baze v različnih časih zgodovine.
- Več DejaVu in DWST baz omogoča skalabilnost.
- ETL procesi bi morali delovati vzporedno:
 - Oracle SE - dbms_parallel_execute

Implementacija nove arhitekture

- Postavitev strežnika za arhive (DWSTA).
- Standby baze na rešitvi Abakus DejaVu za dostop do podatkov iz preteklosti:
 - Virtualne baze podatkov.
- Skalabilnost: več DWSTN in DejaVu strežnikov, softver mora podpirati skalabilno postavitev.
- Omogočeno hitrejša in učinkovitejša izvajanje ETL procesov:
 - razdeliti ETL procese na korake, ki se lahko izvajajo vzporedno.

Kako deluje DejaVu

- navijanje standby baze;
- dvakrat dnevno se naredi snapshot;
- enkrat dnevno se kreira read-only baza (VDB);
- zgodovine je več kot 7 (zahtevanih) stanj (ločenih instanc baze odprte v poljubnem trenutku);
- implicitna deduplikacija - prihranek prostora;
- grafični vmesnik za upravljanje in kreiranje VDB.

Procesi

- Vsaka tabela je svoje opravilo.
- Opravilo je sestavljeno iz podopravil (zaporedno):
 - preverjanje sprememb v strukturi in generiranje spremenjene kode;
 - prenos iz produkcijskih tabel;
 - insert/update;
 - delete.
- Podopravila imajo korake (vzporedno).

Postopek

- insert/update
 - prenese se podatke od zadnjega ora_rowscn;
 - primerjajo se primarni ključi in hash (standard_hash) vsebine vrstice;
 - pri spremembi se stara vrstica “zapre”, doda se nova vrstica;
 - pri insertu se samo doda nova vrstica.
- delete
 - prepišejo se vsi primarni ključi iz izvorne tabele;
 - preveri se arhivska tabela - če kakšnega ključa ni, se zapis v arhivski tabeli zapre.

Generiranje kode

- PL/SQL koda ostane statična.
- Generira se samo SQL stavke.
- SQL stavki se izvajajo s pomočjo EXECUTE IMMEDIATE.
- Kjerkoli je možno, se uporabi paralelizm.
- Število paralelnih procesov se izračuna iz statistike in števila CPU-jev, ki so na razpolago.
- Baza je v času delovanja vedno maksimalno obremenjena.

Primer kode

```
INSERT /*+ gather_plan_statistics [gen_do_transfer_gtt:compare_data] */ INTO
gtt$rid
  (ts_rid, ar_rid, hash_equal)
SELECT /*+ rowid(src) use_nl (dest) index(dest I_AR_C_CNTR_UNIT_LOAN_U) */
      src.rowid, dest.rowid, decode (dest.src_hash, NULL, 'N', src.src_hash,
'Y', 'N')
FROM ETL_STAGE.TS_C_CNTR_UNIT_LOAN src
LEFT OUTER JOIN ETL_DATA.AR_C_CNTR_UNIT_LOAN dest
  ON dest.src_pk = src.src_pk
  AND dest.to_dt = :c_to_dt_null
WHERE src.ROWID BETWEEN :p_rid_from AND :p_rid_to
```

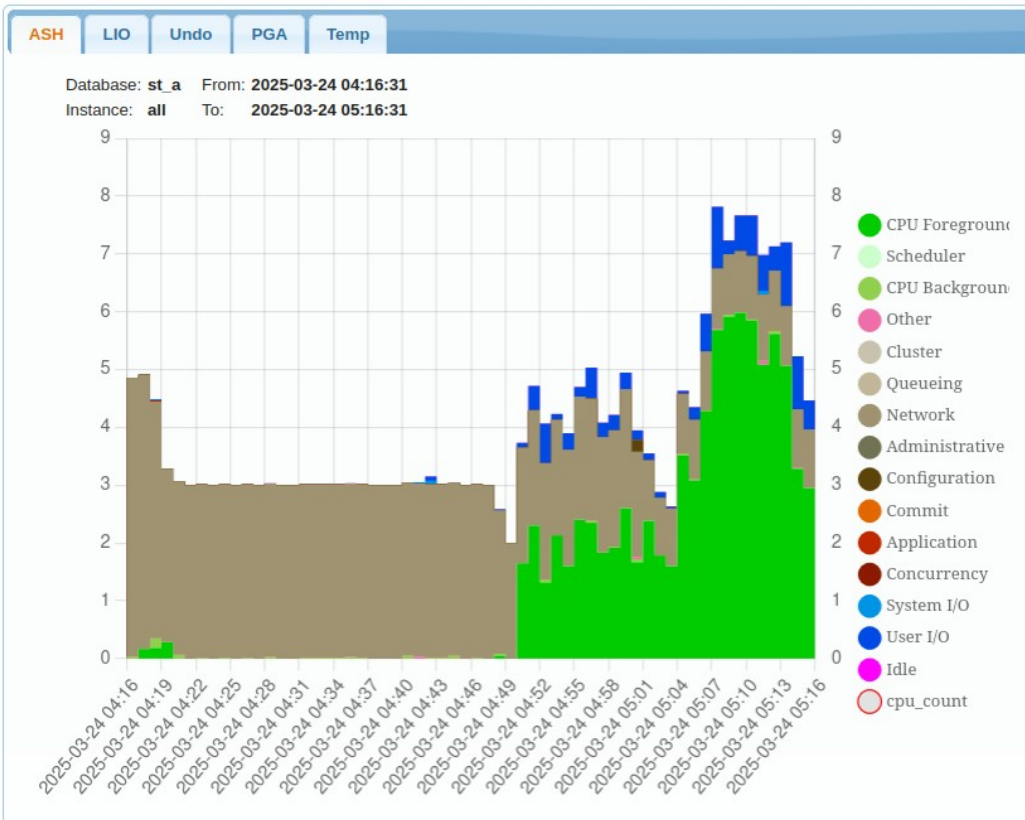
V primeru napak

- read only baza ni pripravljena;
- spremenjeni primarni ključi v izvornih tabelah;
- ponavljanje ETL procesov iz prejšnjih dni;
- izpuščanje dnevnega prometa.

Če se čas za ETL procese podaljša

- dodajanje strežnikov z arhivskimi podatki;
- dodajanje DejaVu strežnikov z read-only bazami.

Nadzor delovanja (Abakus APPM)



Top Table			
Activity	SQL ID	SQL Text	Duration
	fka6jj0vkuwth	INSERT /*+ append gather_plan_statistics [gen_do_e	00d 00:59:59
	2ku21ajs0uupn	INSERT /*+ append gather_plan_statistics [gen_do_e	00d 00:44:04
	fmhmr8z3nwshw	INSERT /*+ append gather_plan_statistics [gen_do_e	00d 00:31:57
	a5tmk24gy9mhp	SELECT "A1"."ID","A1"."PSD_ID","A1"."PFA_ID","A1".	00d 00:12:43
	67js7cvt483vd	SELECT "CUSTOMER","ID_CUSTOMER_7","SYS_DT","TO_DT"	00d 00:03:54
	0bquuhwhffrvx	SELECT "TABLE_NAME","COLUMN_NAME","HIDDEN_COLUMN"	00d 00:03:16
	cym6ysr45mtbb	SELECT "TAA_ID","VREDNOST" FROM "ETL_DATA"."AR_MET	00d 00:03:04
	33washjt34as1	SELECT "A1"."EXT_REF","A1"."APL_SRC","A1"."TBL_NM"	00d 00:02:25
	0sfs52ay91avy	SELECT "A1"."ID",MAX("A1"."ISK_ID"),MAX(GREATEST("	00d 00:02:05
	bcxpvwb6r0jtg	SELECT "TAA_ID","VREDNOST" FROM "ETL_DATA"."AR_OZP	00d 00:01:41

<https://apm.abakus.si/>

Nadzor delovanja (Abakus APPM)

Last Active Sessions									
Session ID	Logon Time	Username	Machine	Event	Client Identifier	Client Info	Current SQL	Previous SQL	SQL Trace
1.2008.20250324023523	2025-03-24 02:35:23	ETL_USER	g-ora-px.gpsi.corp.local	SQL*Net message from dblink (driver id=1952673792, #bytes=1, =0)	ETP_RUN_ET:TS_OPEN_ITEMS	ETP_RUN_ET.RUN_ET	fka6jj0vkuwth.0	cmhzbxtat7pgt7.5	DISABLED
1.5373.20250324045006	2025-03-24 04:50:06	ETL_USER	g-ora-px.gpsi.corp.local	ON CPU (file number=44, first dba=2928128, block cnt=128)			1tm4y89nugk3y.1	1tm4y89nugk3y.0	DISABLED
1.1099.20250324050347	2025-03-24 05:03:47	ETL_USER	g-ora-px.gpsi.corp.local	ON CPU (count=1, where=2, timeout=0)			0sfs52ay91avy.0	33washjt34as1.1	DISABLED
1.2186.20250324050348	2025-03-24 05:03:48	ETL_USER	g-ora-px.gpsi.corp.local	ON CPU (count=2, where=2, timeout=0)			a5tmk24gy9mhp.1	bja1k7qzrq69g.0	DISABLED
1.2827.20250324050359	2025-03-24 05:03:59	ETL_USER	g-ora-px.gpsi.corp.local	ON CPU (driver id=1413697536, #bytes=8098, =0)			6uc1u0hsrrb7y.0	0ym0q3budakh0.0	DISABLED

Last Active Transactions									
Addr	Session ID	Username	Logon Time	Start Time	Status	Used UBLK	Used UREC	Used MB	
00083ccc740	1.4463.20250114152348	SYS	2025-01-14 15:23:48	2025-01-14 15:43:11	ACTIVE	1	1	.01	
000878cea938	1.2008.20250324023523	ETL_USER	2025-03-24 02:35:23	2025-03-24 02:35:23	ACTIVE	1	1	.01	

<https://appm.abakus.si/>

Nadzor delovanja (Abakus APPM)

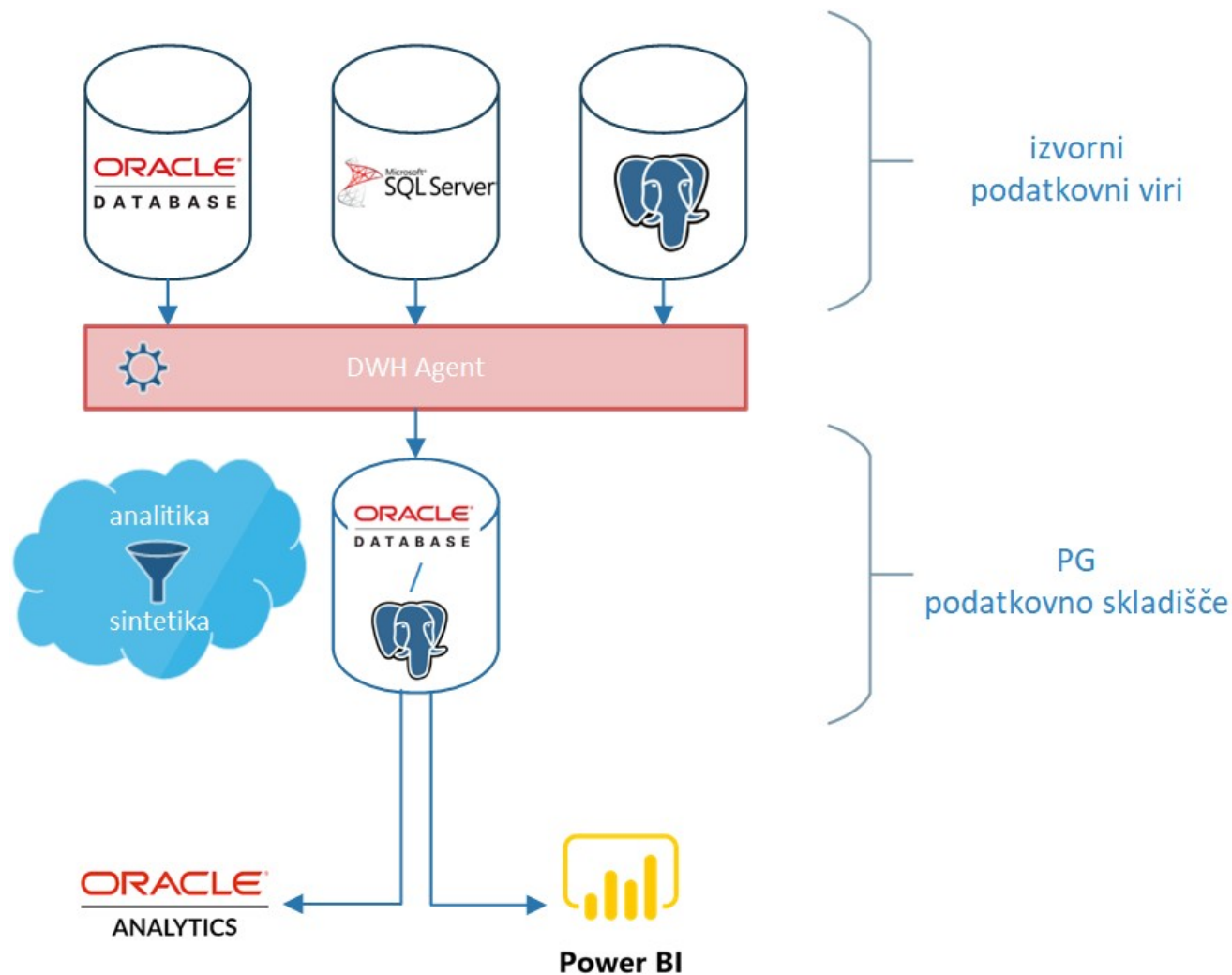
Last Longops										
Session ID	SQL ID / Plan Hash	Plan Hash	Start Time	Finish Time	Last Update	Operation	Elapsed	Remaining	Progress	Percent
1.2100.20250324045017	73u6tcwpp53ty / 2652974694	2652974694	2025-03-24 05:17:17	2025-03-24 05:17:35	2025-03-24 05:17:27	Table Scan	00d 00:00:10	00d 00:00:00	1978983/2002799 Blocks	99.00
1.2918.20250324050359	67js7cvt483vd / 1267478891	1267478891	2025-03-24 05:17:07	2025-03-24 05:17:59	2025-03-24 05:17:27	Table Scan	00d 00:00:20	00d 00:00:24	3128193/6805642 Blocks	46.00
1.5373.20250324045006	1tm4y89nugk3y / 578594736	578594736	2025-03-24 05:16:59	2025-03-24 05:17:45	2025-03-24 05:17:27	Table Scan	00d 00:00:28	00d 00:00:10	5049695/6805642 Blocks	74.00
1.2100.20250324045017	7stvkdgnccsf5 / 3240305401	3240305401	2025-03-24 05:16:42	2025-03-24 05:17:35	2025-03-24 05:16:50	Table Scan	00d 00:00:08	00d 00:00:00	953970/953970 Blocks	100.00
1.5373.20250324045006	1tm4y89nugk3y / 578594736	578594736	2025-03-24 05:16:26	2025-03-24 05:17:35	2025-03-24 05:16:58	Table Scan	00d 00:00:32	00d 00:00:00	6805642/6805642 Blocks	100.00
1.5373.20250324045006	1tm4y89nugk3y / 578594736	578594736	2025-03-24 05:15:55	2025-03-24 05:17:35	2025-03-24 05:16:25	Table Scan	00d 00:00:30	00d 00:00:00	6805642/6805642 Blocks	100.00
1.1099.20250324050347	0sfs52ay91avy / 2582739178	2582739178	2025-03-24 05:15:45	2025-03-24 05:17:35	2025-03-24 05:16:35	Hash Join	00d 00:00:50	00d 00:00:00	589465/589465 Blocks	100.00
1.2827.20250324050359	6uc1u0hsrrb7y / 3714826231	3714826231	2025-03-24 05:15:34	2025-03-24 05:17:58	2025-03-24 05:17:27	Table Scan	00d 00:01:53	00d 00:00:23	10930661/13139393 Blocks	83.00
1.5373.20250324045006	2rx5j7hdkncr0 / 1761388929	1761388929	2025-03-24 05:15:17	2025-03-24 05:17:35	2025-03-24 05:15:24	Table Scan	00d 00:00:07	00d 00:00:00	344394/344394 Blocks	100.00
1.5373.20250324045006	2mx9fad297u5s / 2461262687	2461262687	2025-03-24 05:15:01	2025-03-24 05:17:35	2025-03-24 05:15:08	Table Scan	00d 00:00:07	00d 00:00:00	344394/344394 Blocks	100.00
1.1099.20250324050347	0sfs52ay91avy / 2582739178	2582739178	2025-03-24 05:14:50	2025-03-24 05:17:35	2025-03-24 05:15:45	Table Scan	00d 00:00:55	00d 00:00:00	7597434/7597434 Blocks	100.00
1.1099.20250324050347	0sfs52ay91avy / 2582739178	2582739178	2025-03-24 05:14:26	2025-03-24 05:17:35	2025-03-24 05:14:50	Table Scan	00d 00:00:24	00d 00:00:00	2684311/2684311 Blocks	100.00
1.2827.20250324050359	28zfk5g021sd7 / 3714826231	3714826231	2025-03-24 05:14:24	2025-03-24 05:17:35	2025-03-24 05:15:19	Table Scan	00d 00:00:55	00d 00:00:00	13139393/13139393 Blocks	100.00
1.1099.20250324050347	33washjt34as1 / 3328094377	3328094377	2025-03-24 05:14:08	2025-03-24 05:17:35	2025-03-24 05:14:23	Sort Output	00d 00:00:15	00d 00:00:00	109299/109299 Blocks	100.00
1.2373.20250324050348	5z9g9h47zsk4m / 677869339	677869339	2025-03-24 05:14:02	2025-03-24 06:00:21	2025-03-24 05:17:27	Table Scan	00d 00:03:25	00d 00:42:46	5122131/69238317 Blocks	7.00
1.2918.20250324050359	67js7cvt483vd / 1267478891	1267478891	2025-03-24 05:14:00	2025-03-24 05:17:35	2025-03-24 05:15:17	Table Scan	00d 00:01:17	00d 00:00:00	6805642/6805642 Blocks	100.00
1.4826.19900101000000	grguh2jb2rgh / 1026777905	1026777905	2025-03-24 05:13:46	2025-03-24 05:17:35	2025-03-24 05:14:07	Table Scan	00d 00:00:21	00d 00:00:00	594824/594824 Blocks	100.00
1.1099.20250324050347	33washjt34as1 / 3328094377	3328094377	2025-03-24 05:13:40	2025-03-24 05:17:35	2025-03-24 05:14:03	Table Scan	00d 00:00:23	00d 00:00:00	940798/940798 Blocks	100.00
1.2918.20250324050359	gs3gkwhq4b1a / 366952323	366952323	2025-03-24 05:13:32	2025-03-24 05:17:35	2025-03-24 05:14:00	Table Scan	00d 00:00:28	00d 00:00:00	1505625/1505625 Blocks	100.00

<https://appm.abakus.si/>

Zaključek

- Produktijske baze so se razbremenile.
- Referenčna integriteta je zagotovljena z read-only bazami.
- ETL procesi navadno končajo pred 11 uro.
- Proces ETL se lahko ponovijo za 1 teden v preteklosti.
- Omogočena je skalabilnost z enostavnim dodajanjem novih strežnikov.

Kako naprej



ORA-03113: end-of-file on communication channel



Miha Petrič

Generali zavarovalnica d.d.
Specialist za podatkovna skladišča

Boris Oblak
Abakus plus d.o.o.

Kako naprej

- Univerzalni vmesnik za prenos iz različnih virov:
 - Oracle;
 - PostgreSQL;
 - MS SQL Server;
 - ...
- DWH arhiv v Oracle (SE ali EE) ali PostgreSQL bazi.