

Prirodoslovno-matematički fakultet Sveučilišta u Zagrebu

# **Snowflake**

## **Skladište i analiza podataka u oblaku**

Zagreb, 11.11.2024

Lea Markušić  
Petra Salaj

## Sadržaj

1. Uvod .....	3
2. Snowflake - općenito .....	3
3. Registracija .....	4
4. Funkcionalnosti .....	5
4.1. Snowflake Worksheets .....	5
4.2. Upiti na podacima koristeći radne listove .....	5
4.3. Povijest upita (Query History) .....	7
5. Izrada skladišta podataka u Snowflake-u .....	8
6. Upiti na skladište podataka .....	11
6.1. Dashboard .....	14
7. Podjela poslova i samoevaluacija .....	14
8. Zaključak .....	15
9. Literatura .....	15

## 1. Uvod

U ovome radu obradile smo Snowflake, alat za skladište i analizu podataka u oblaku, krenuvši od same instalacije pa sve do isprobavanja nekih njegovih funkcionalnosti. Koristile smo bazu podataka o sudjelovanju sportaša na olimpijskim igrama, do 2016. godine te smo ju na Snowflake-u pohranile kao staging bazu podataka, napravile skladište koje omogućava analizu uspjeha (osvojenih medalja) pojedinih nogometnih nacionalnih momčadi na zadnjih 10 olimpijskih igara i potom izvršile dva upita na izrađeno skladište.

## 2. Snowflake - općenito

Snowflake je platforma za pohranu podataka u oblaku, osmišljena kako bi pojednostavila proces pohrane, upravljanja i analize velikih količina podataka. Omogućava objedinjavanje podataka iz više izvora i različite načine njihove analize. Koristi se za podatkovno inženjerstvo, izradu jezera podataka, skladišta podataka, znanost o podacima, analize i vizualizacije podataka, dijeljenje podataka, a čak i za izradu aplikacija u Pythonu, Javi ili Scali i strojno učenje.

Sva tri sloja Snowflake-ove arhitekture (pohrana, izračunavanje i cloud usluge) postavljeni su i koriste se u potpunosti na nekoj (odabranoj) cloud platformi koje on podržava, a to su: Microsoft Azure, Amazon Web Services i Google Cloud Platform. Iako odabir cloud platforme ne bi trebao činiti razliku, neke funkcionalnosti (trenutno) nisu dostupne svima. Na primjer, korisnici koji odluče koristiti Microsoft Azure neće moći koristiti SnowSQL – komandno linijski alat za dodatne SQL operacije. No, najvažnije je da se podaci vrlo lako mogu i učitati u Snowflake sa bilo koje od tih platformi ili direktno iz interne pohrane računala.

Na Snowflake-u postoje četiri vrste korisničkih računa: *Standard*, *Enterprise*, *Business Critical* i *Virtual Private Snowflake*. Standardna verzija sadrži sve funkcionalnosti, sigurnost sa automatskom enkripcijom, dijeljenje podataka i optimizaciju pohrane, dok su preostale varijante namijenjene tvrtkama jer mogu pohraniti veću količinu podataka koji su bolje zaštićeni te postoji mogućnost oporavka, odnosno preusmjeravanje podataka među Snowflake računima za kontinuitet poslovanja. *Virtual Private Snowflake* omogućava sve kao prethodne, no račun je potpuno izoliran od svih ostalih Snowflake računa čime su podaci još sigurniji te za njega postoji poseban način aktivacije. Jedina zanimljiva razlika naprednijih verzija je ta što one sadrže i *Snowflake Feature Store* koji omogućava analizu podataka za strojno učenje i treniranje modela unutar samog Snowflake-a.

*Standard*, *Enterprise* i *Business Critical* verzije moguće je isprobati u probnom razdoblju, a nakon njegovog isteka, svaka verzija ima posebnu cijenu za nastavak korištenja ovisno o izabranoj cloud platformi i regiji u kojoj se korisnik nalazi. Za tri gore navedene verzije, naplaćuje se od \$2 do \$6.50 po kreditu, a za *Virtual Private Snowflake* cijena je na upit. Također, za podatke spremljene na platformi, naplaćuje se i \$20 - \$40 po terabajtu podataka mjesečno.

### 3. Registracija

Kako bi mogle isprobati ovaj alat, morale smo se registrirati. Snowflake pruža trideset dana besplatnog probnog razdoblja sa \$400 kredita na računu koji se koristi za pokretanje nekih funkcija. Probno razdoblje završava protekom trideset dana ili nakon što se potroši svih \$400, ovisno o tome što se od toga dogodi prije.

Samo pristupanje odvija se u nekoliko koraka prikazanih na donjim slikama.

Slika 1: Registracija – korak 1

Slika 2: Registracija – korak 2

Na samome početku (Slika 1) treba upisati osobne podatke. Drugi korak predstavlja odabir vrste Snowflakea (Slika 2): *Standard*, *Enterprise* i *Business Critical* (*Virtual Private Snowflake* nema mogućnost probnog perioda). Napravile smo dva računa, jedan *Standard* i jedan *Business Critical* te se i same uvjerile u to da su omogućene jednake funkcionalnosti (koje su nam za ovaj projekt potrebne) te nastavile koristiti *Standard* verziju. Također, može se odabrati na kojem oblaku će biti pohranjena baza podataka: Microsoft Azure-u, Amazon Web Services ili Google Cloud-u. Mi smo odabrale Google Cloud. Treba odabrati i regiju u kojoj se nalazimo, Microsoft Azure i Amazon Web Services mogu se koristiti u cijelom svijetu, dok se Google Cloud može koristiti samo u Europi i SAD-u. Zatim su postavljena tri pitanja, prikazana na iduće tri slike (Slika 3, Slika 4, Slika 5): zašto se prijavljujemo, za što će biti korišten i koji jezici su preferirani.

Nakon uspješne registracije, odmah su dostupni linkovi na kratki video, online tečaj te potpuna dokumentacija o Snowflake-u za lakši početak korištenja.

Slika 2: Pitanje 1

Slika 2: Pitanje 2

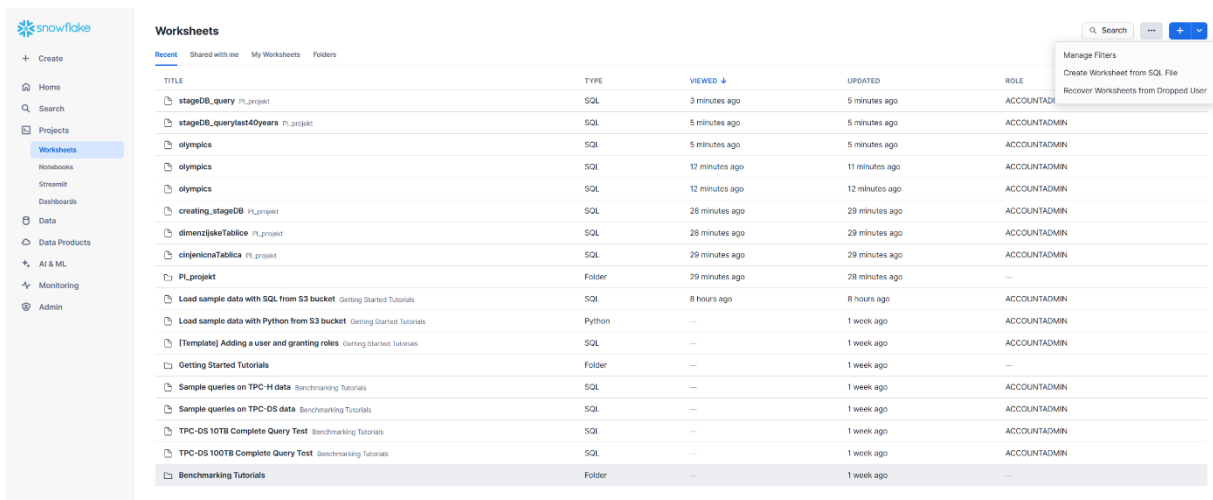
Slika 5: Pitanje 3

Slika 6: Uspješna registracija

## 4. Funkcionalnosti

### 4.1. Snowflake Worksheets

Snowflake worksheets (radni listovi) omogućuju korisnicima pisanje SQL ili Snowpark Python koda, izvršavanje upita, pregled rezultata i stvaranje vizualizacija. Korisnici mogu upravljati tim radnim listovima organiziranjem u mape, dijeljenjem s kolegama i čuvanjem povijesti verzija. Odabirom Projects → Worksheets na bočnom izborniku korisnici mogu vidjeti nedavne radne listove, dijeljene mape i one mape koje su sami stvorili. Novi radni listovi mogu biti kreirani za SQL ili Python. Također, mogu se kreirati učitavanjem SQL datoteke sa vlastitog računala. Na Slici 7 prikazane su spomenute opcije. (Izvor: <https://docs.snowflake.com/en/user-guide/ui-snowsight-worksheets-gs> )



The screenshot shows the Snowflake UI with a sidebar on the left containing navigation options like Home, Search, Projects, Worksheets, Notebooks, Streamlit, Dashboards, Data, Data Products, AI & ML, Monitoring, and Admin. The main area is titled 'Worksheets' and has tabs for 'Recent', 'Shared with me', 'My Worksheets', and 'Folders'. A table lists recent worksheets with columns for Title, Type, Viewed, Updated, and Role. A 'Manage Filters' dropdown menu is visible on the right side of the table.

TITLE	TYPE	VIEWED	UPDATED	ROLE
stageDB_query PL_project	SQL	3 minutes ago	5 minutes ago	ACCOUNTADMIN
stageDB_querylast40years PL_project	SQL	5 minutes ago	5 minutes ago	ACCOUNTADMIN
olympics	SQL	5 minutes ago	5 minutes ago	ACCOUNTADMIN
olympics	SQL	12 minutes ago	11 minutes ago	ACCOUNTADMIN
olympics	SQL	12 minutes ago	12 minutes ago	ACCOUNTADMIN
creating_stageDB PL_project	SQL	28 minutes ago	29 minutes ago	ACCOUNTADMIN
dimenzijskoTablica PL_project	SQL	29 minutes ago	29 minutes ago	ACCOUNTADMIN
cinjenionaTablica PL_project	SQL	29 minutes ago	29 minutes ago	ACCOUNTADMIN
PL_project	Folder	29 minutes ago	28 minutes ago	—
Load sample data with SQL from S3 bucket Getting Started Tutorials	SQL	8 hours ago	8 hours ago	ACCOUNTADMIN
Load sample data with Python from S3 bucket Getting Started Tutorials	Python	—	1 week ago	ACCOUNTADMIN
ITemplate Adding a user and granting roles Getting Started Tutorials	SQL	—	1 week ago	ACCOUNTADMIN
Getting Started Tutorials	Folder	—	1 week ago	—
Sample queries on TPC-H data Benchmarking Tutorials	SQL	—	1 week ago	ACCOUNTADMIN
Sample queries on TPC-DS data Benchmarking Tutorials	SQL	—	1 week ago	ACCOUNTADMIN
TPC-DS 10TB Complete Query Test Benchmarking Tutorials	SQL	—	1 week ago	ACCOUNTADMIN
TPC-DS 100TB Complete Query Test Benchmarking Tutorials	SQL	—	1 week ago	ACCOUNTADMIN
Benchmarking Tutorials	Folder	—	1 week ago	—

Slika 7: Worksheets (radni listovi)

Istovremeno je moguće imati otvoreno više radnih listova u različitim karticama, a s lijeve strane nalazi se prozor koji omogućuje istraživanje baza podataka i shema u Snowflake-u za vrijeme pisanja SQL upita. Položaj pomicanja (scroll position) sačuvan je u svakoj kartici što olakšava izvođenje usporedbi između radnih listova. Kartice radnog lista čuvaju se kroz sesije što omogućava nastavak rada tamo gdje se stalo prošli put.

### 4.2. Upiti na podacima koristeći radne listove

Snowflake podržava standardni SQL, uključujući podskup ANSI SQL:1999 i SQL:2003 analitičkih proširenja. Podržava neke dodatne značajke kao što su različite vrste JOIN-ova (npr. ASOF JOIN za analiziranje podataka vremenskih serija) i opcije za uklanjanje suvišnih JOIN-ova radi optimizacije performansi.

## 7 najboljih praksa za optimizaciju Snowflake upita

(Izvor: <https://www.chaosgenius.io/blog/snowflake-query-tuning-part1/> )

Učinkovitost upita izravno utječe na performanse, troškove i skalabilnost. Snowflake pruža niz alata i strategija koje omogućuju brže dohvaćanje podataka, niže računalne troškove i sveukupno bolje korištenje. Neke od njih su:

1. **Odabir prave veličine virtualnog skladišta:** Odabir odgovarajuće veličine skladišta ključan je za ravnotežu između troškova i performansi. Veća skladišta su bolja za složene upite koji zahtijevaju puno računanja, dok su manja prikladnija za jednostavnije zadatke. Preporuka je početi s manjom veličinom i pratiti performanse, povećavajući samo ako je potrebno kako bi izbjegli nepotrebne troškove.
2. **Korištenje predmemorije (cache):**

Snowflake sprema u predmemoriju (cache) i čuva rezultate upita za svaki izvršeni upit. Ovo se može iskoristiti s velikim učinkom za dramatično smanjenje vremena potrebnog za dobivanje odgovora. Obično se rezultati upita ponovno koriste ako su ispunjeni svi sljedeći uvjeti:  
(Izvor: <https://community.snowflake.com/s/article/Understanding-Result-Caching> )

  - Korisnik koji izvršava upit ima potrebne ovlasti pristupa za sve tablice korištene u upitu.
  - Novi upit sintaktički odgovara prethodno izvršenom upitu.
  - Podaci tablice koji doprinose rezultatu upita nisu se promijenili.
  - Trajni rezultat za prethodni upit još uvijek je dostupan.
  - Sve opcije konfiguracije koje utječu na način na koji je rezultat proizveden nisu promijenjene.
  - Upit ne uključuje funkcije koje se moraju procijeniti pri izvođenju (npr. `CURRENT_TIMESTAMP()`).
  - Mikroparticije tablice nisu se promijenile (npr. ponovno su klasterirane ili konsolidirane) zbog promjena drugih podataka u tablici.

Moguće je u korisničkom sučelju Snowflake-a provjeriti je li upit koristio predmemoriju rezultata. Potrebno je odabrati opciju Monitoring → Query History i tamo se pojavi povijest svih izvršenih upita u zadnjih godinu dana. Odaberemo upit za koji želimo provjeriti te tada odaberemo opciju Query profile. Tamo bi se trebao nalaziti jedan čvor u kojem piše “QUERY RESULT REUSE 100%”.
3. **Korištenje materijaliziranih prikaza:** Materijalizirani prikazi pohranjuju unaprijed izračunate rezultate upita, omogućujući brži pristup podacima nego kod izravnih upita nad osnovnim tablicama. Korisni su za složene ili statične podatke jer izbjegavaju ponovne izračune i poboljšavaju izvedbu kod zadataka koji uključuju velika agregiranja ili ponavljajuće upite.
4. **Klasterizacija podataka i mikro-particioniranje:** Klasterizacija organizira podatke prema određenim stupcima (ključevima klastera), grupirajući slične retke unutar mikro-particija. Ovo omogućuje Snowflake-u da preskoči nepotrebne podatke (pruning particija), što poboljšava izvedbu smanjujući volumen skeniranih podataka, posebno u velikim tablicama.
5. **Korištenje ubrzanja upita u Snowflake-u:** Servis ubrzanja upita (Query Acceleration Service) privremeno dodaje računalne resurse za ubrzavanje velikih ili složenih upita. Kada je omogućena, QAS djeluje kao “pojačivač”, automatski dodajući snagu za zahtjevne upite, omogućujući bržu izvedbu bez prekomjernog trošenja resursa.
6. **Korištenje alata Query Profile:** Alat Query Profile u Snowflake-u pruža detaljne informacije o izvedbi izvršenih upita, pomažući korisnicima prepoznati područja u

kojima bi se mogla poboljšati učinkovitost. Prikazuje nedostatke poput neučinkovitih JOIN-ova ili prekomjernog skeniranja podataka, omogućujući korisnicima da optimiziraju upite kako bi postigli bolje performanse.

7. **Korištenje usluge optimizacije pretrage:** Dostupna u *Enterprise* izdanju, ova usluga gradi optimizirani put za pretragu radi brzog pristupa podacima, posebno za pretraživačke i analitičke upite. Strukturira metapodatke za učinkovitiju pretragu, omogućujući Snowflake-u da brzo odgovori na upite pristupajući samo potrebnim segmentima podataka.

Svaka od ovih metoda ima važnu ulogu u poboljšanju performansi Snowflake-a i upravljanju troškovima, pružajući fleksibilan pristup optimizaciji različitih vrsta zadataka.

### 4.3. Povijest upita (Query History)

Query History u Snowflake-u pruža detaljan uvid u povijest upita tijekom posljednjih 365 dana, omogućujući korisnicima praćenje performansi upita, korisničkih aktivnosti i korištenja sistemskih resursa. Dostupan je u ACCOUNT\_USAGE i READER\_ACCOUNT\_USAGE shemama te podržava praćenje po različitim dimenzijama, kao što su vrijeme, korisnik i skladište. Ključne metrike u ovom prikazu uključuju vrijeme izvršenja, broj skeniranih bajtova podataka, broj generiranih redaka te poruke o pogreškama za neuspjele upite. Prikaz također pokazuje iskorištenost resursa, poput vremena računalne obrade i korištenog prostora za pohranu. Korisnicima je Query History koristan za prepoznavanje problema i poboljšanje brzine i učinkovitosti upita, posebno kroz analizu gdje dolazi do zastoja ili ponavljanja pokušaja upita.

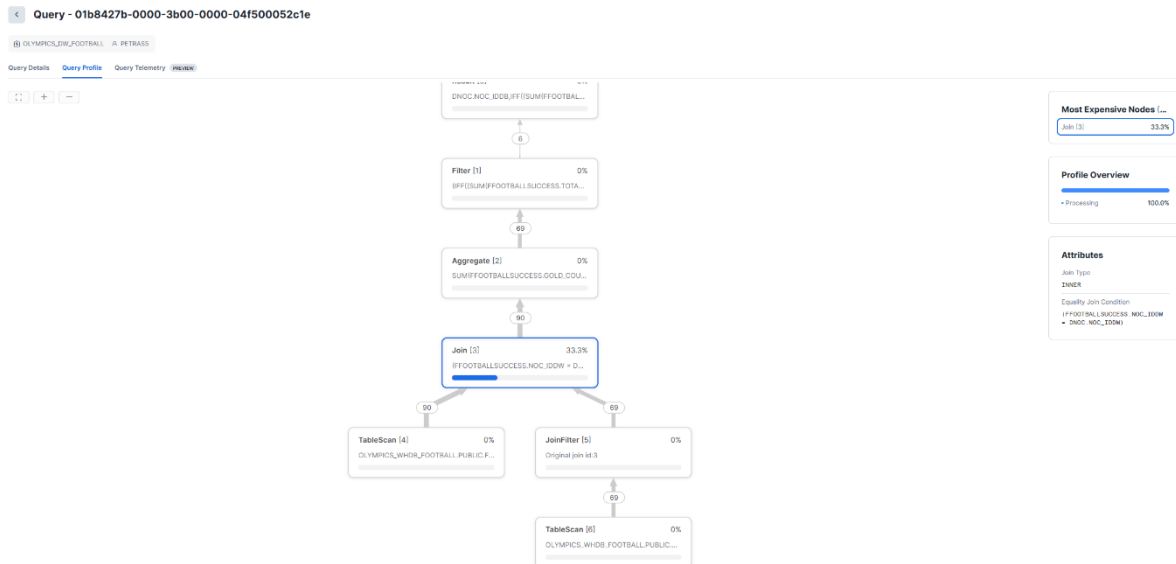
Odabirom nekog upita u Query Historyju otvaraju se njegovi detalji tj. Query Details kao što vidimo na Slici 8. Tamo je prikazan njegov SQL, rezultati, vrijeme izvršavanja upita, vrijeme i datum od početka i završetka izvršenja upita, veličina skladišta s kojim ima veze i mnoge druge informacije.

The screenshot displays the Snowflake Query History interface. The sidebar on the left contains navigation links: Create, Home, Search, Projects, Data, Data Products, AI & ML, Monitoring, Query History (selected), Task History, Dynamic Tables, Tables & Logs, Trust Center, and Admin. The main panel shows the details for a specific query with ID '01b8427b-0000-3b00-0000-04f500052c1e'. It includes tabs for 'Query Details', 'Query Profile', and 'Query History'. The 'Query Details' tab is active, showing a progress bar at 0%, the query ID, client name 'Ga 11.5', session ID '545033824370', and an empty query tag. Below the details is the 'SQL Text' section with a code editor containing a SQL query. At the bottom, the 'Results' section shows a table with 6 rows and 2 columns: 'NOCLIPPER' and 'GOLD\_PERCENT'.

NOCLIPPER	GOLD_PERCENT
1. ARG	0.000000
2. CAR	1.000000
3. FSA	1.000000
4. TCH	1.000000
5. MCX	1.000000
6. USA	0.000000

Slika 8: Query Details

Pod Query Profile možemo vidjeti kako je došlo do dohvata rezultata u obliku grafa i najvažnije koja je bila najskuplja operacija u upitu. Na upitu sa Slike 9 je to operacija JOIN.

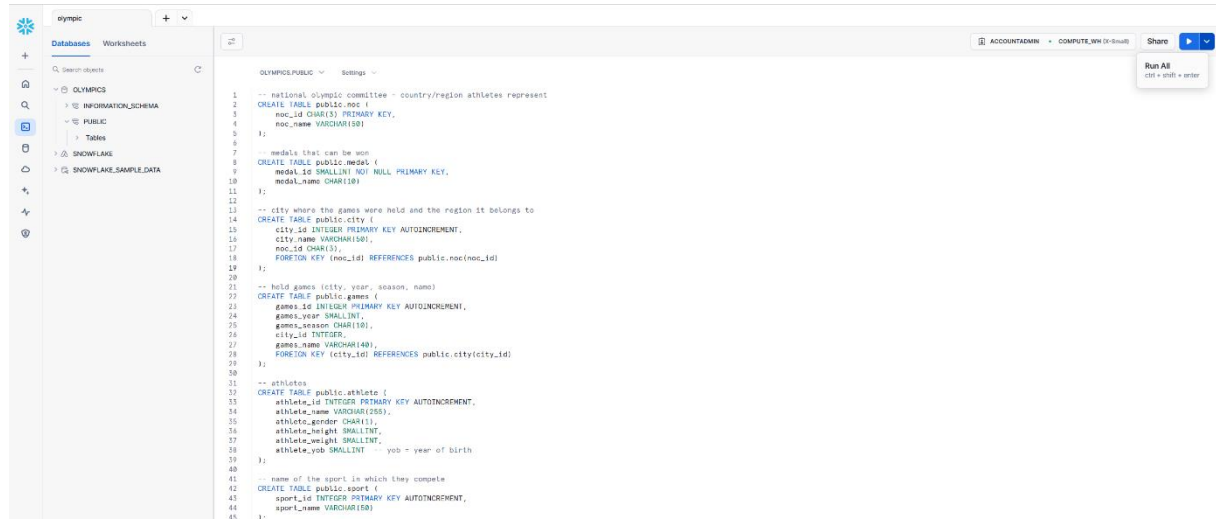


Slika 9: Query Profile

## 5. Izrada skladišta podataka u Snowflake-u

Za primjer izrade skladišta podataka odabrale smo bazu podataka o sudjelovanju sportaša na olimpijskim igrama, do 2016. godine.

Najprije smo odabrale opciju Data → Databases u bočnom izborniku gdje se nalazi opcija za kreiranje nove baze podataka. Potrebno je unijeti ime nove baze i opcionalno opis. Zatim smo u SQL worksheetu kreirale tablice dane baze podataka kao što je prikazano na slici 10.

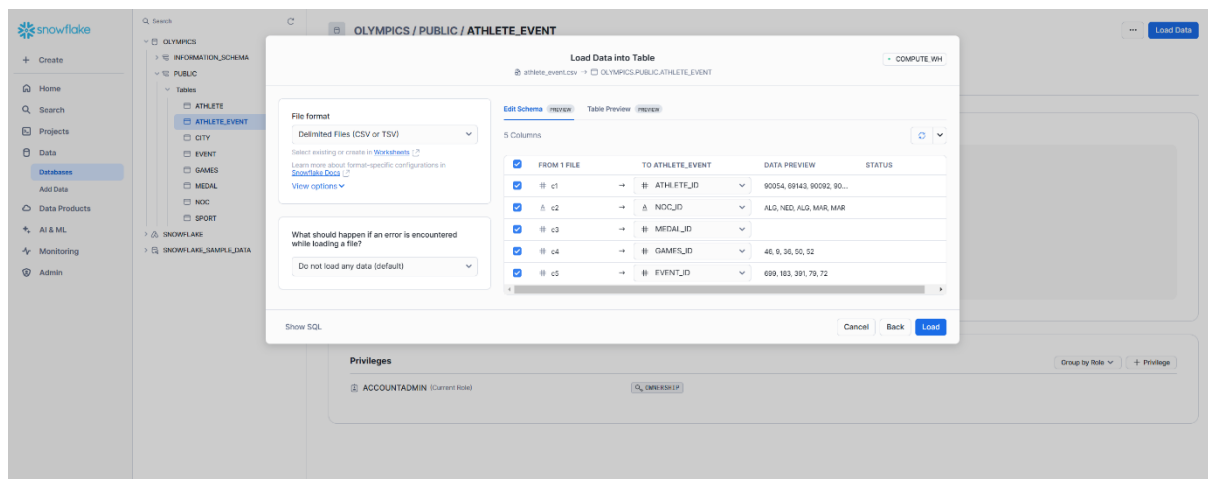


Slika 10: SQL za CREATE TABLE noc, medal, city, games, athlete, sport

Idući korak bio je punjenje novokreiranih tablica što je u Snowflake-u brzo i jednostavno. Odabirom Data u bočnom izborniku, izlista se popis naših postojećih baza podataka. Tamo pronađemo tablicu u koju želimo učitati podatke. Odabirom opcije Load Data ponude nam se 3 opcije; možemo povući i ubaciti datoteke (Drag and upload files) ili pretražiti datoteke na

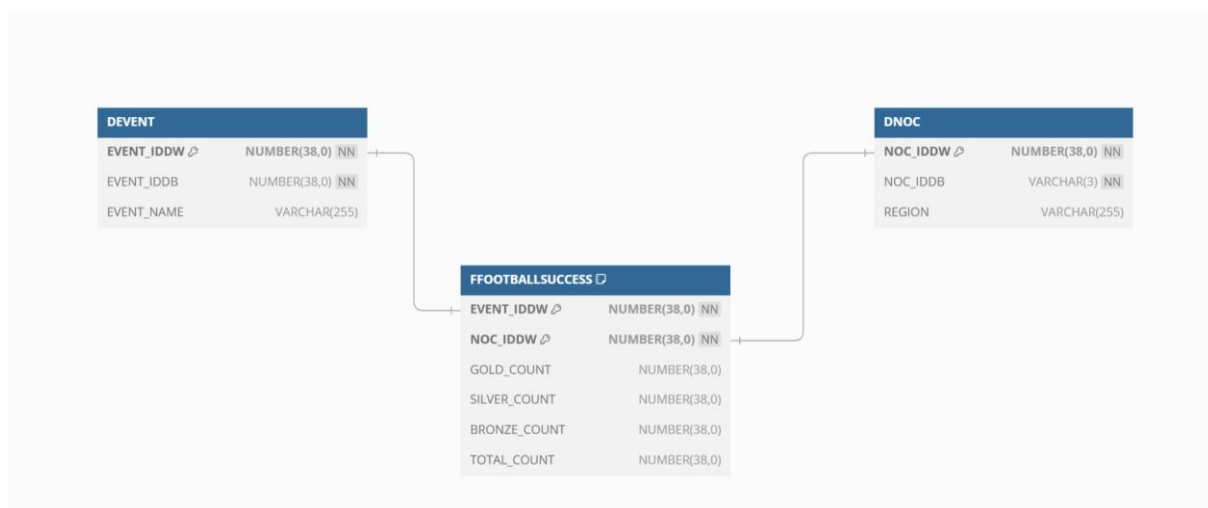


svom računalu (Browse) ili dodati iz staging baze podataka (Add from Stage). Podržani formati su: CSV/TSV, json, orc, avro, parquet, xml i maksimalna veličina datoteke je 250MB. Mi smo koristile dane CSV datoteke.



Slika 11: Učitavanje podataka u tablice staging baze podataka

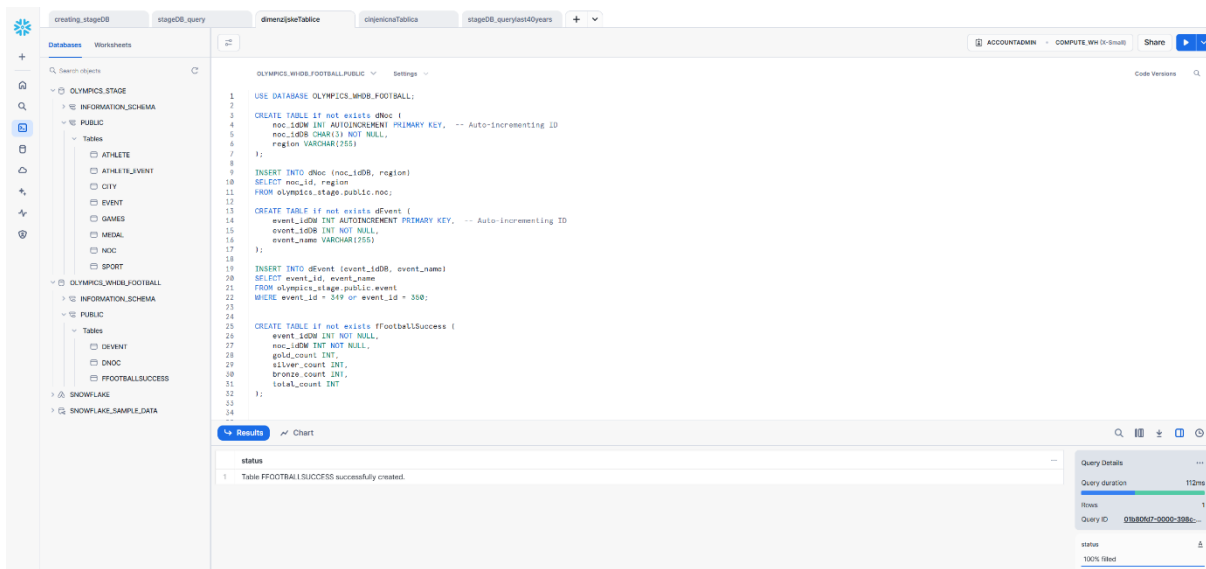
Sada kada imamo staging bazu podataka, idući korak je izrada skladišta u Snowflake-u. Naše skladište opisuje činjenicu: uspjeh pojedinih nogometnih nacionalnih momčadi na zadnjih 10 olimpijskih igara. Naš zvjezdasti model skladišta, prikazan na Slici 12, sastoji se od 1 činjenične tablice fFootballSuccess s primarnim ključem (event\_idDW, noc\_idDW) i mjerama gold\_count, silver\_count, bronze\_count i total\_count te 2 dimenzijske tablice dNoc i dEvent.



Slika 12: Zvjezdasti model skladišta s 1 činjeničnom i 2 dimenzijske tablice

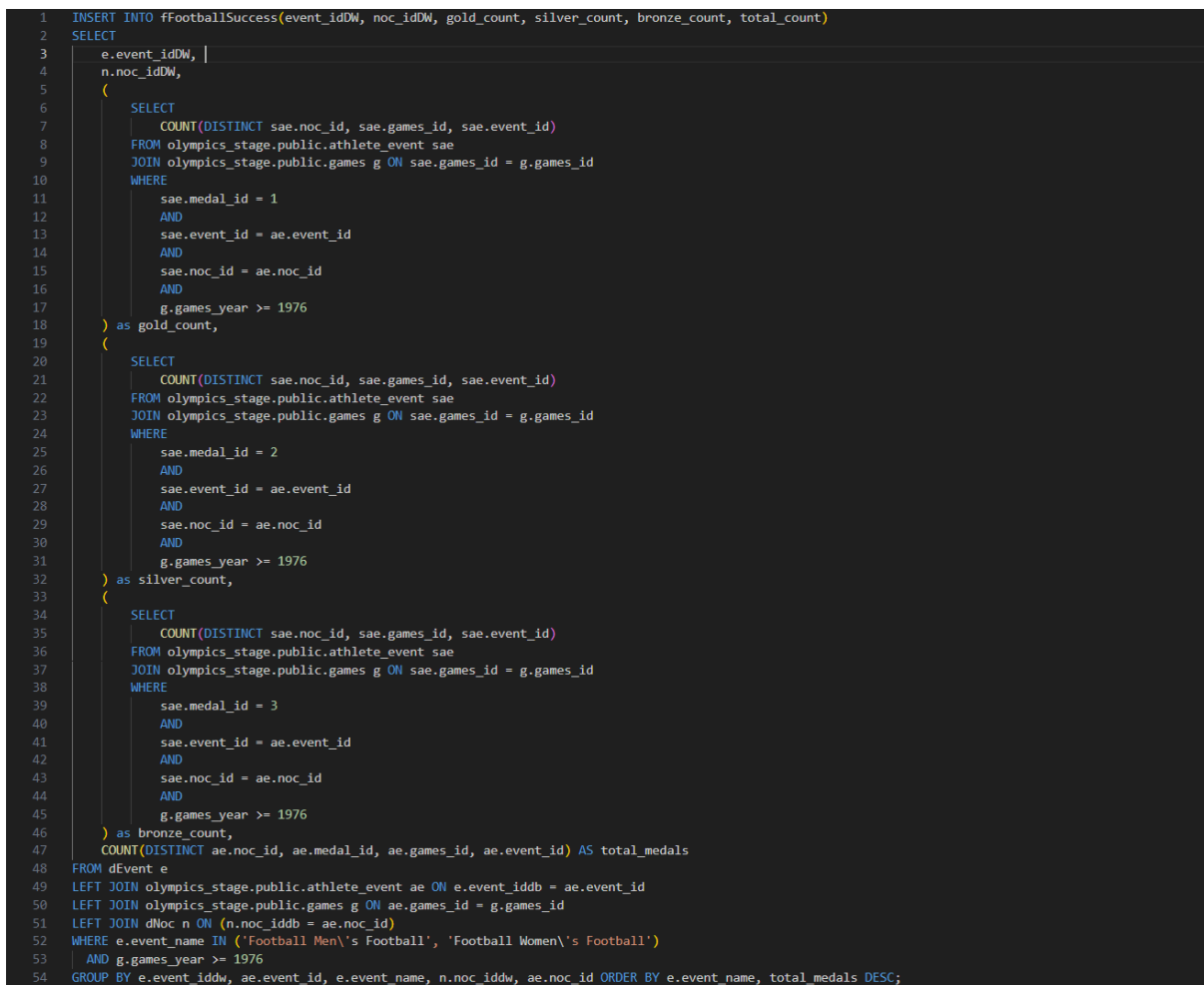
U Snowflake-u je potrebno stvoriti novo skladište odabirom opcija Admin → Warehouses u bočnom izborniku i klikom na +Warehouse. Tada je potrebno upisati ime novog skladišta, odabrati njegovu veličinu (ovisno o tome koliko kredita po satu troši; od X-Small za 1 kredit po satu do 4X-Large za 128 kredita po satu). Za naše potrebe dovoljno je bilo najmanje X-Small. Također, moguće je u naprednijim postavkama odabrati opciju Auto-resume kojom se automatski nastavlja rad skladišta kada se podnese bilo koji upit koji zahtijeva skladište i Auto-suspend koji automatski obustavlja skladište ako je neaktivno određeno vremensko razdoblje koje je potrebno upisati.

Na Slici 13 nalaze se SQL upiti za kreiranje i punjenje dimenzijskih tablica te kreiranje činjenične tablice u Snowflake-u.



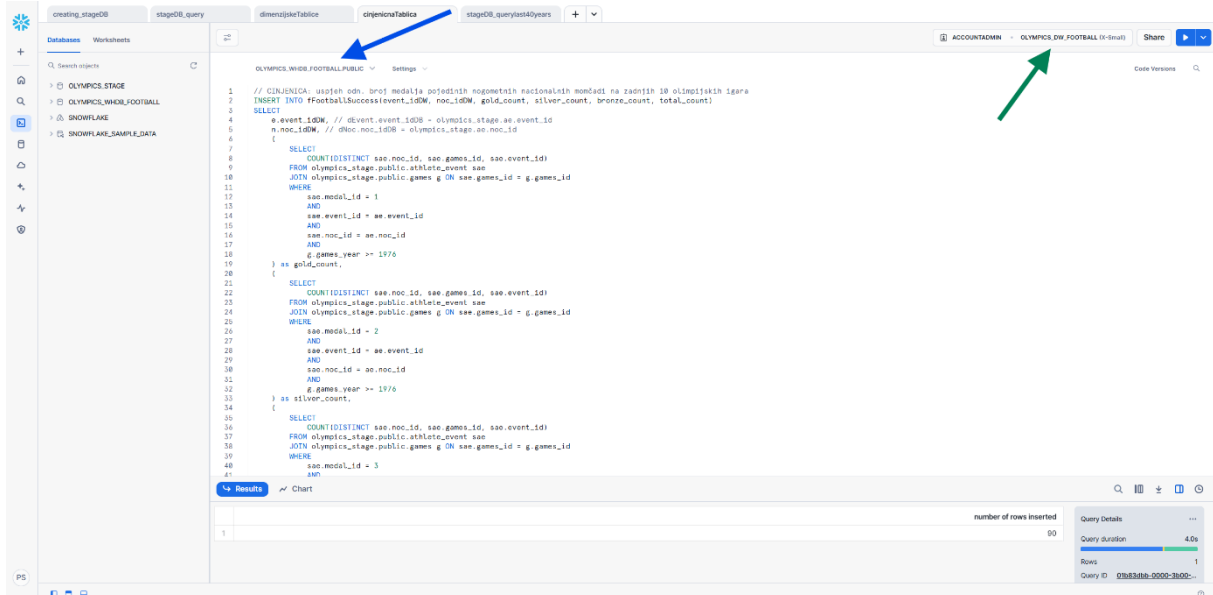
Slika 13: SQL naredbe za kreiranje i punjenje dimenzijskih te kreiranje činjenične tablice

Zadnji korak je punjenje činjenične tablice fFootballSucess, SQL naredba za punjenje prikazana je na Slici 14.



Slika 14: SQL za stvaranje činjenične tablice fFootballSucess

Kao što smo spomenuli kod kreiranja novog skladišta, u Snowflake-u potrebno je označiti s kojim skladištem podataka SQL upit ima veze. Na donjoj slici zelenom je strelicom označeno gdje se odabere skladište podataka OLYMPICS\_DW\_FOOTBALL s kojim upit ima poveznicu, a plavom baza podataka OLYMPICS\_WHDB\_FOOTBALL u koju se činjenična tablica sprema. Također, potrebno je naglasiti da prilikom ubacivanja koristimo staging bazu podataka OLYMPICS\_STAGE pa ispred imena tablica iz te baze podataka pišemo 'olympics\_stage.public.imeTablice'. Na Slici 15 se vidi da je ovim upitom ubačeno u tablicu 90 redaka u vremenu od 4 sekunde.

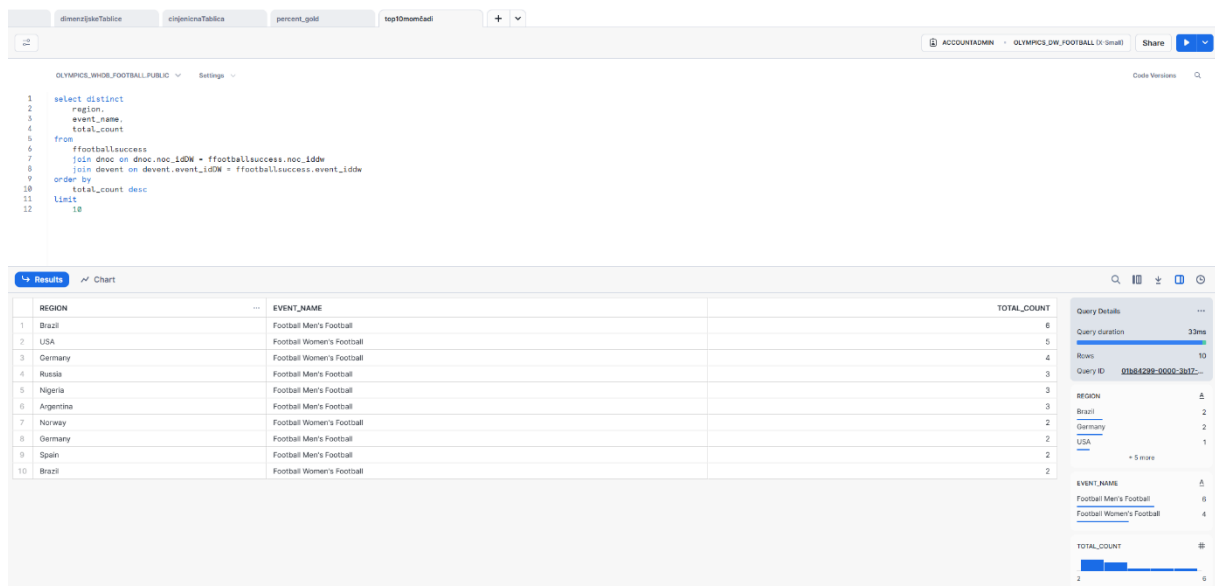


Slika 15: Unos podataka iz tablica staging baze u činjeničnu tablicu DW

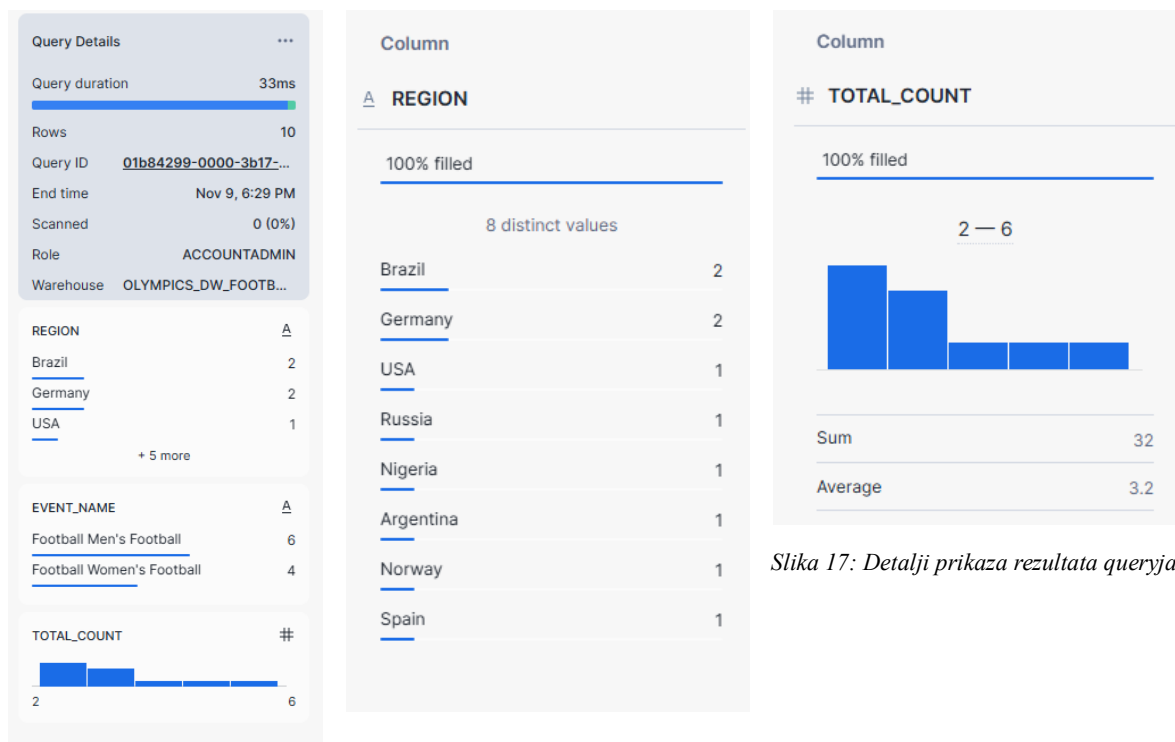
## 6. Upiti na skladište podataka

Na izrađenom skladištu isprobale smo sljedeća 2 upita:

1. 10 najuspješnijih momčadi po ukupnom broju medalja na zadnjih 10 olimpijskih igara  
Na Slici 16 vidimo kako nam se u worksheetu kad pokrenemo upit prikažu naši rezultati. Također, prikažu se i razni podaci koji pokazuju koliko je vremena trebalo da se upit izvrši, broj vraćenih redova, Query ID (po kojem je spremljen u Query History), vrijeme i datum završetka upita, podaci o skladištu s kojim upit ima veze te neke vizualizacije rezultata. Na svaku stavku kad se klikne prikažu se dodatne informacije kao što vidimo na Slici 17.

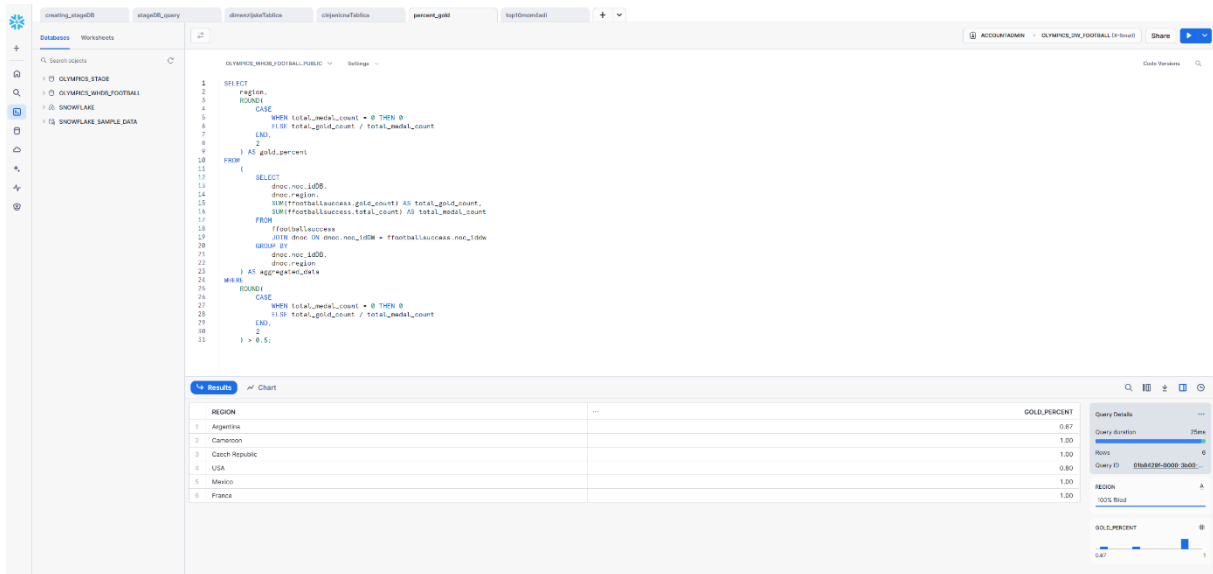


Slika 16: Worksheet za upit10 najuspješnijih momčadi na zadnjih 10 OI



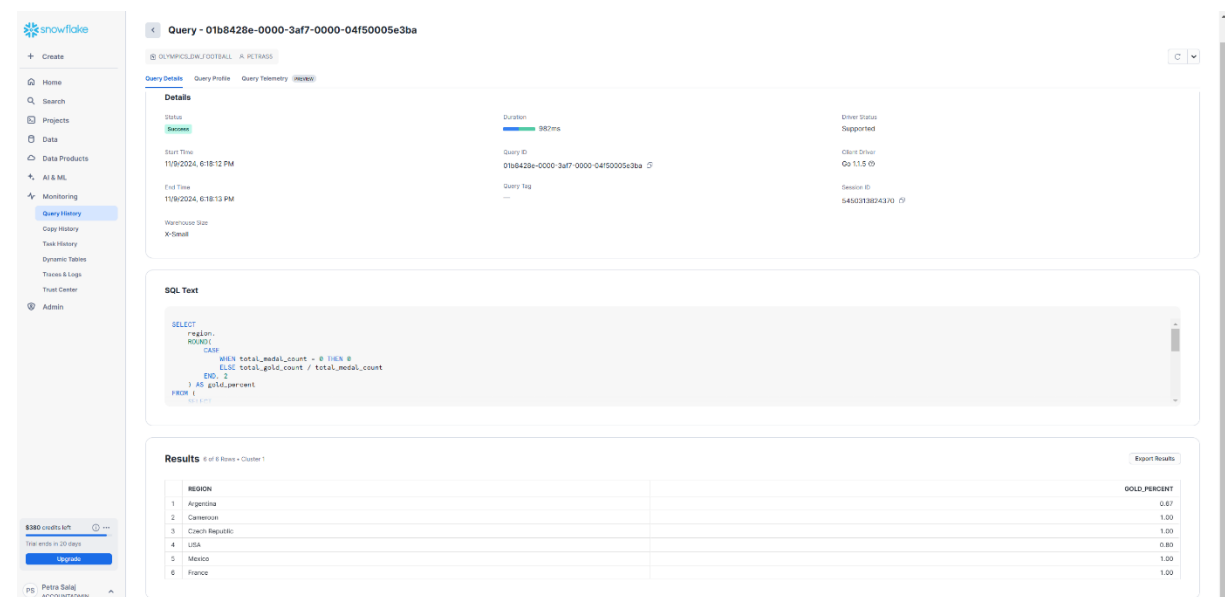
Slika 17: Detalji prikaza rezultata queryja

2. Popis država kojima je više od pola osvojenih medalja zlatno (omjer zlatnih medalja naprema njihovom ukupnom broju medalja veći od 0.5)  
Rezultati ovog upita prikazani su na Slici 18.



Slika 18: SQL worksheet za upit države kojima je većina osvojenih medalja zlatno

Klikom na Query ID otvara nam se Query Profile u kojem je prikazano kako je došlo do izračuna upita. Vidimo sa prijašnje slike da je vrijeme izvršavanja za ovaj upit bilo 25 ms, ali razlog tome je što je došlo do „QUERY RESULT REUSE 100%“ jer smo taj upit već ranije tog dana pokrenule. Međutim, ako pogledamo u Query History možemo lako pronaći koliko mu je trebalo prvi puta kada je bio pokrenut. Na Slici 19 vidimo da je tada izvršenje trajalo 982 ms i dohvat iz tablica skladišta je bio kompliciraniji. Na ovom primjeru dobro se vidi koliko je Snowflake-ova optimizacija izvršavanja upita značajna.

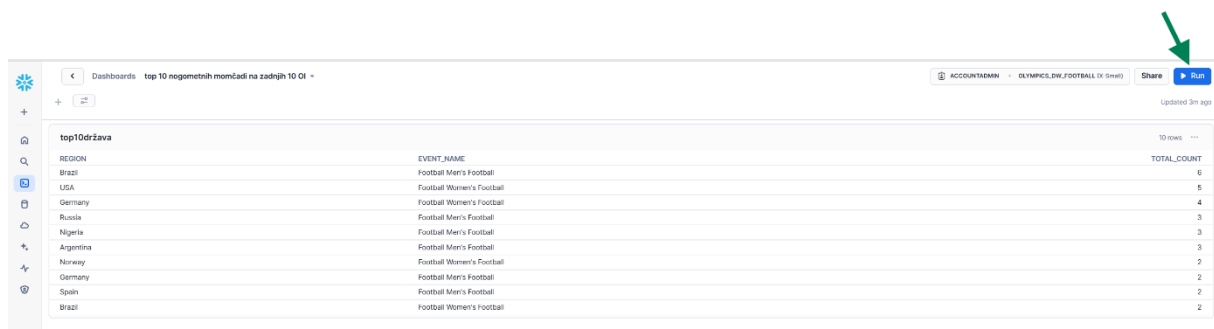


Slika 19: Performanse prvog izvršavanja upita 2

## 6.1. Dashboard

Snowflake nudi i opciju kreiranja Dashboarda. Iz radnog lista je jako lako napraviti Dashboard tako što prođemo mišem po njegovoj kartici, tada se pojave 3 točkice kojima otvaramo razne opcije. Jedna od opcija je Move to i tu možemo odabrati ili mapu ili Dashboard u koji želimo staviti taj radni list. Također, tu je i opcija Create Dashboard kojom stvaramo novi prikaz kojemu je potrebno samo dati ime.

Dashboardom dobivamo prikaz rezultata upita kao na donjoj Slici 20. Pritiskom na tipku Run označenu zelenom strelicom, u pozadini se ponovno izvršava upit kojim je dashboard dobiven te se osvježava tablica rezultata ako je došlo do promjene u skladištu podataka. Ovo je korisno da korisnik ne mora stalno pisati SQL za ovaj upit nego jednostavno, svaki put kada ga zanima je li se nešto promijenilo pokrene i pogleda na dashboardu.



REGION	EVENT_NAME	TOTAL_COUNT
Brazil	Football Men's Football	6
USA	Football Women's Football	5
Germany	Football Women's Football	4
Russia	Football Men's Football	3
Nigeria	Football Men's Football	3
Argentina	Football Men's Football	3
Norway	Football Women's Football	2
Germany	Football Men's Football	2
Spain	Football Men's Football	2
Brazil	Football Women's Football	2

Slika 20: Dashboard „top 10 nogometnih momčadi na zadnjih 10 OI“

## 7. Podjela poslova i samoevaluacija

Kostur ovog projekta osmislile smo i napravile zajedno. Dogovorile smo kako ćemo modelirati skladište i učitale podatke, napravile skladište i osmislile upite. Lein zadatak bio je istražiti općenito o Snowflake-u i objasniti postupak registracije dok je Petrin zadatak bio izvršiti upite te objasniti isprobane funkcionalnosti.

### 1. Lea Markušić

Ovaj projekt bio je zanimljiv način za upoznavanje s novim alatom. Najveći problem bili su nam gumbići na rubovima alata, za postavke prilikom kreiranja skladišta i pokretanja koda u radnim listovima, no čitanjem dokumentacije iz raznih izvora uspjele smo se snaći. Također, iako je dokumentacija jako opširna, njeno čitanje i pronalazak točno onoga što nas zanima zahtijevalo je dosta truda.

### 2. Petra Salaj

Problem u izradi ovog seminara bila mi je dokumentacija koja ima neke zastarjele verzije pa se sučelje Snowflake-a i slike u dokumentaciji ne poklapaju. Također, svaka stranica dokumentacije ima puno linkova na neke druge stranice pa se lako pogubiti u tako opširnoj dokumentaciji budući da se nismo susretali na studiju s puno ovako kompleksnih alata s puno mogućnosti. Međutim, ovaj seminar je bio koristan za učenje čitanja dokumentacija i upoznavanje s alatima za poslovnu inteligenciju.

## 8. Zaključak

Snowflake je sveobuhvatan alat za skladište i analizu podataka u oblaku. Isprobale smo kako napraviti tablice baze, ubaciti podatke, izraditi skladište, postavljati upite na skladište, izradu dashboarda i proučile koje sve detaljne informacije o tijeku izračunavanja postoje. Iako su izrada skladišta i postavljanje upita lako dostupni, nevješti korisnik može se izgubiti u moru njegovih mogućnosti i samoj njegovoj dokumentaciji. Smatramo da bi se daljnjim istraživanjem samog Snowflake-a, a onda i učenjem o podacima i što oni predstavljaju puno bolje upoznale sa svim njegovim funkcionalnostima te na koje načine se podaci mogu iskoristiti.

## 9. Literatura

- Overview of the Snowflake Cloud Data Platform (<https://youtu.be/9PBvVeCQi0w?si=GkeC3EmY-zdjl8hJ>)
- Snowflake dokumentacija (<https://docs.snowflake.com/>)
- <https://www.chaosgenius.io/blog/snowflake-query-tuning-part1/>
- <https://community.snowflake.com/s/article/Understanding-Result-Caching>