

ELEKTRONSKI FAKULTET NIŠ

SEMINARSKI RAD

PREDMET:

Sistemi za upravljanje bazama podataka

TEMA:

Replikacija baza podataka

Profesor:

Doc. Dr Aleksandar Stanimirović

Student:

Tijana Spasić 1064

Niš, Maj 2020

Replikacija baza podataka	3
Kako funkcioniše replikacija?	3
Razvoj replikacija baza podataka	4
Replikacija baza podataka vs. mirroring	5
Alati za replikaciju baze podataka	5
SQL Server replikacija baze podataka	6
Arhitektura SQL Server replikacija - Komponente SQL Servera	6
Agenti	8
Tipovi MS SQL Server replikacija	10
Prednosti i nedostaci	15
Praktična primena transakcijske replikacije nad MS SQL Server bazom podataka	16
Literatura	25

Replikacija baza podataka

Replikacija baze podataka je učestalo elektronsko kopiranje podataka iz baze podataka na jednom računaru ili serveru u bazu podataka na drugom - tako da svi korisnici dele isti nivo informacija. Rezultat je distribuirana baza podataka u kojoj korisnici mogu brzo pristupiti podacima relevantnim za svoje zadatke bez ometanja rada drugih. Brojni elementi doprinose ukupnom procesu stvaranja i upravljanja replikacijom baze podataka.

Kako funkcioniše replikacija?

Replikacija baze podataka može biti ili jedna pojava ili tekući proces. Uključuje sve izvore podataka u distribuiranu infrastrukturu organizacije. Distribuirani sistem upravljanja organizacije koristi se za kopiranje i pravilnu distribuciju podataka između svih izvora.

Sistemi za upravljanje distribuiranom bazom podataka (DDBMS) rade na tome da osiguraju da se promene, dodavanja i brisanja izvršene nad podacima na bilo kojoj lokaciji automatski odražavaju na podatke na svim ostalim lokacijama. DDBMS je u osnovi naziv infrastrukture koja dopušta ili vrši replikaciju baze podataka - sistem koji upravlja distribuiranom bazom podataka, a koji je proizvod replikacije baze podataka.

Klasičan slučaj replikacije baze podataka uključuje jednu ili više aplikacija koje povezuju primarnu lokaciju za skladištenje sa sekundarnom lokacijom koja je često van lokacije. Danas su te primarne i sekundarne lokacije najčešće pojedinačne baze podataka - kao što su Oracle, MySQL, Microsoft SQL i MongoDB - kao i skladišta podataka koja objedinjuju podatke iz ovih izvora, nudeći usluge skladištenja i analitike na većim količinama podataka. Skladišta podataka često se nalaze u oblaku.

Postoji nekoliko načina kopiranja baze podataka. Različite tehnike nude različite prednosti, jer se razlikuju u temeljnosti, jednostavnosti i brzini. Idealan izbor tehnike zavisi od toga kako kompanije skladište podatke i u koje će svrhe poslužiti ponovljene informacije.

Što se tiče vremena prenosa podataka, postoje dve vrste replikacije podataka:

- Asinhrona replikacija je replikacija kod koje se podaci šalju na server modela - server sa kojeg replike kasnije uzimaju podatke - od klijenta. Zatim server modela upiše klijenta uz potvrdu da su podaci primljeni. Odatle se nastavlja kopiranje podataka u replike neodređenim ili nadziranim tempom.
- Sinhrona replikacija je replikacija kod koje se podaci kopiraju sa klijent servera na server model, a zatim se dalje vrši kopiranje na sve servere replike pre nego što je klijent obavešten da su podaci replicirani. Provera traje duže nego kod asinhrona replikacije,

međutim, za razliku od asinhronne replikacije postoji uvid da su svi podaci uspešno kopirani pre nego što se nastavi sa daljim radom.

Asinhrono kopiranje baze podataka nudi fleksibilnost i jednostavnost upotrebe, jer se replikacije dešavaju u pozadini. Međutim, postoji veći rizik da će podaci biti izgubljeni bez klijentovog znanja, jer potvrda dolazi pre glavnog procesa replikacije. Sinhrono kopiranje je nefleksibilnije i traje duže, ali verovatnije je da se obezbedi uspešno kopiranje podataka. Klijent će biti upozoren ako se kopiranje nije izvršilo uspešno, pošto potvrda dolazi nakon završetka čitavog procesa.

Postoji nekoliko tipova replikacije baze podataka na osnovu vrste arhitekture servera. Izraz vođa (engl. leader) će se koristiti u ovim tipovima da označi istu stvar kao model u prethodnim asinhronim nasuprot sinhronim primerima:

- *Single-leader* je arhitektura koju čini jedan server koji prima podatke od klijenata i replike crpe podatke odatle. Ovo je najčešća metoda. To je sinhronizovana metoda, ali pomalo nefleksibilna.
- *Multi-leader* arhitekturu čini više servera koji mogu da primaju podatke i služe kao model za replike. To je korisno kad se replike raspodele i vođe moraju biti blizu svakoj kako bi se sprečilo kašnjenje.
- Arhitektura bez vođe je arhitektura u kojoj svaki server može primati podatke i poslužiti kao model za replike. Ovo je pokrenuo *Amazon-ov DynamoDB*. Iako nudi maksimalnu fleksibilnost, on predstavlja izazove za sinhronizaciju.

Razvoj replikacija baza podataka

Rani slučajevi replikacije baze podataka obično su opisani kao *master-slave* konfiguracije, ali opisi danas imaju tendenciju da sadrže terminologiju kao što su *master-replika*, *master-sledbenik*, *primarna-sekundarna* i *server-klijent*.

Tehnike replikacije usredsređene na sisteme za upravljanje relacionim bazama podataka proširile su se pojavom virtualne mašine i distribuirale računarstvo u oblaku, tako da uključuju nerelacione tipove baza podataka. Opet, metode replikacije razlikuju se među takvim nerelacionim bazama podataka kao što su Redis, MongoDB i slično.

Iako je replikacija udaljenih kancelarijskih baza podataka možda godinama bila kanonski primer replikacije, sigurnosne šeme sigurnosnih kopija baze podataka takođe su se pojavile kao pokretači aktivnosti replikacije - kao što su horizontalno skaliranje konfiguracija distribuirane baze podataka, kako u prostorijama tako i na platforme za računarsko korišćenje u oblaku. Detalji replikacije variraju između takvih relacionih sistema kao što su IBM Db2, Microsoft SQL Server, Sibase, MySQL i PostgreSQL.

U svim slučajevima, dizajn replikacije podataka postaje akt uravnoteženja između performansi sistema i doslednosti podataka.

Replikacija baza podataka vs. mirroring

Iako se refleksija (engl. mirroring) podataka ponekad postavlja kao alternativni pristup replikaciji podataka, to je zapravo oblik replikacije podataka. U relacionoj refleksiji baze podataka, kompletne sigurnosne kopije baza podataka održavaju se za upotrebu u slučaju da primarna baza podataka ne uspe. Refleksija, u stvari, služe kao baza podataka u stanju pripravnosti. Refleksija podataka našla je znatnu upotrebu u Microsoft SQL Server zajednici.

Kod replikacije baze podataka fokus je obično na skali podataka baze podataka za upite - zahteve za podacima. Refleksija baze podataka, čini osnovu za inkrementalna ažuriranja baze podataka sa glavnog server i obično se implementira da bi se obezbedile mogućnosti pripravnosti ili oporavka od katastrofe. Jednostavno rečeno, refleksija je fokusirana na stvaranje rezervne kopije onoga što postoji, a replikacija je fokusirana na poboljšanje operativne efikasnosti u celini - što uključuje održavanje sigurnih sigurnosnih kopija podataka korišćenjem refleksije.

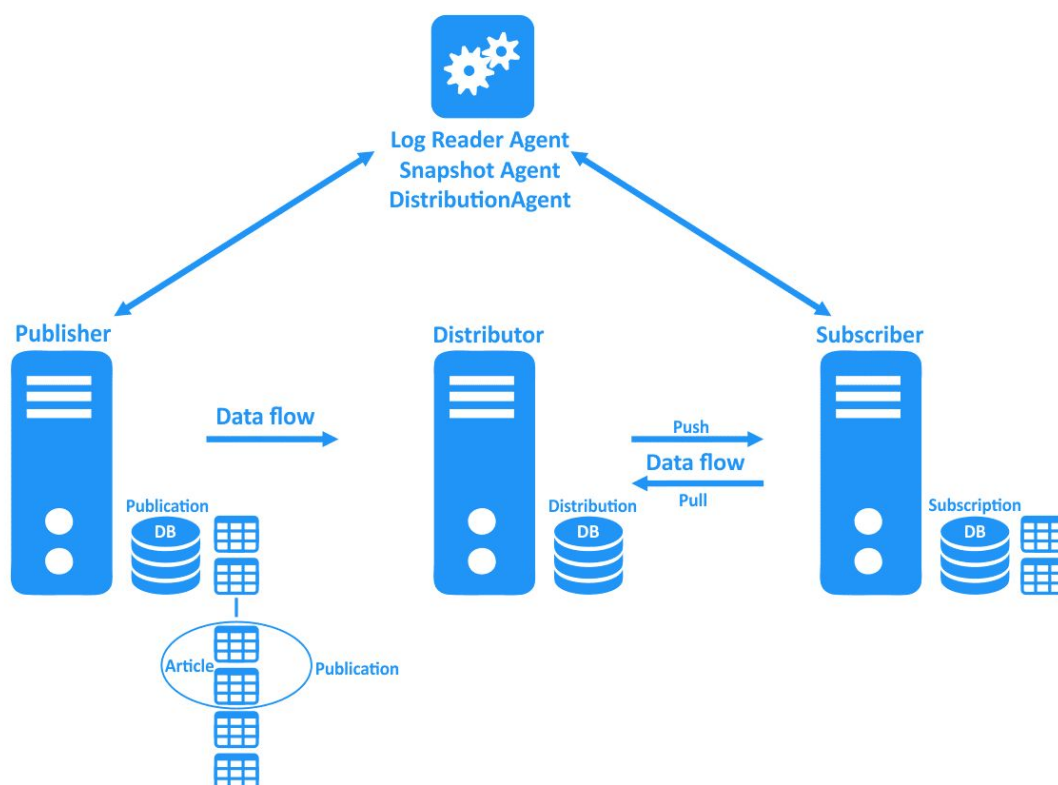
Alati za replikaciju baze podataka

Kompanije mogu ili koristiti alat za replikaciju baze podataka koji nudi njihov dobavljač softvera za baze podataka ili uložiti u alate za kopiranje trećih strana (engl. third-party) za izvršavanje i upravljanje procesima replikacije baze podataka.

SQL Server replikacija baze podataka

Kao i druge tehnike replikacije, replikacija SQL Servera ne distribuira celokupnu bazu podataka, već distribuira odabrane delove baze podataka samo kao tabele ili prikaze. SQL replikacija je prilično korisna i potrebno je znati kako je ispravno uraditi.

Arhitektura SQL Server replikacija - Komponente SQL Servera



Slika 2.1. Arhitektura SQL Server replikacija

Artikl je osnovna jedinica SQL servera koji se sastoji od tabela, prikaza i pohranjenih procedura. Pomoću opcije filtriranja, artikl u procesu replikacije SQL-a može se skalirati vertikalno ili horizontalno. Moguće je stvoriti više artikla u istom objektu uz određena ograničenja.

Nakon što je artikl uspešno kreiran i potrebno je promeniti neka svojstva, tada treba napraviti novi snimak replikacije. Ako artikl ima jednu ili više pretplata, onda ih sve treba ponovo inicijalizovati.

Publikacija je logična zbirka artikla. Omogućuje da definišemo i konfigurišemo svojstva artikala na višem nivou tako da se mogu naslediti od drugih artikala u grupi. Artikl se ne može distribuirati samostalno, ali treba ga objaviti.

Filter je skup uslova za artikl. Replikacija MS SQL Server-a omogućava upotrebu filtera i odabir prilagođenih entiteta za replikaciju, što rezultuje smanjenjem prometa, suvišnosti i količine podataka smeštenih u replici baze podataka. Na primer, mogu se odabrati samo najkritičnije tabele i polja koristeći filtere i izvršiti kopija samo ovih podataka.

U reprodukciji MS SQL baze podataka postoje tri glavne uloge: distributer, izdavač (engl. publisher) i pretplatnik (engl. subscriber):

Distributer je instanca MS SQL baze podataka konfigurirana za prikupljanje transakcija iz publikacija i za njihovo distribuiranje pretplatnicima. Distributer deluje kao baza podataka za skladištenje ponovljenih transakcija. Baza distributera može se istovremeno smatrati izdavačem i distributerom. U modelu lokalnog distributera, pojedinačna instanca MS SQL Server-a pokreće i izdavača i distributera. Model daljinskog distributera može se koristiti da se pretplatnici konfigurišu za upotrebu jedne instance MS SQL Server-a za dobijanje različitih publikacija (centralizovana distribucija). U ovom modelu, izdavač i distributer rade na različitim serverima.

Izdavač je glavna kopija baze podataka na kojoj je konfigurirana publikacija, čineći podatke dostupnim drugim MS SQL serverima koji su konfigurirani za upotrebu u procesu replikacije. Izdavač može imati više publikacija.

Pretplatnik je baza podataka koja prima replicirane podatke iz publikacije. Jedan pretplatnik može primiti podatke od više izdavača i publikacija. Model sa jednim pretplatnikom koristi se kada postoji jedan pretplatnik. Model sa više pretplatnika koristi se kada je više pretplatnika povezano u jednoj publikaciji.

Pretplata je zahtev za kopiju publikacije koja mora biti dostavljena pretplatniku. Pretplata se koristi za definisanje podataka o publikaciji koji moraju biti primljeni i gde i kada će ti podaci biti primljeni. Postoje dve vrste pretplate - *push* i *pull*.

- Push pretplata: promenjeni podaci nasilno se prenose od distributera u bazu podataka pretplatnika. Zahtev pretplatnika nije potreban.
- Pull pretplata: pretplatnik traži promenjene podatke napravljene od izdavača. Agent radi sa strane pretplatnika.

Baza pretplata je ciljna baza podataka u modelu replikacije MS SQL.

Metapodaci su podaci koji se koriste za opisivanje entiteta baze podataka. Postoji širok spektar ugrađenih funkcija metapodataka koje omogućavaju da se vratite informacije o instanci MS SQL Servera, instancama baze podataka i entitetima baze podataka.

U modelu *više izdavača - više pretplatnika*, izdavač može da deluje kao pretplatnik na jednom od MS SQL Servera.

Agenti

Agenti su komponente MS SQL Servera koji mogu poslužiti kao pozadinske usluge za sistem upravljanja relacionim bazama podataka i koriste se za planiranje automatskog izvršenja poslova, kao što su sigurnosna kopija i replikacija MS SQL baze podataka. Postoji pet vrsta agenata, kao što je to prikazano na slici 2.2: snapshot agent, agent za distribuciju, agent za čitanje logova, agent za spajanje i agent za čitanje redova.



Slika 2.2. Agenti SQL Server replikacija

Snapshot agent

- To je izvršna datoteka koja pomaže u pripremi datoteka snimka koje sadrže šemu i objavljene podatke tabele i objekte baze podataka.
- Obično čuva podatke u direktorijumu snimka i beleži sinhronizovane zadatke u distribuiranoj bazi podataka.

Distributivni agent

- Koristi se uglavnom za transakcijsku i replikaciju snimka.
- Primjenjuje snimak na pretplatnika i premešta transakcije iz distribuirane baze na pretplatnike.
- Pokreće se kod distributera da potpiše pretplate ili se pretplati na pretplatnika kako bi povukao pretplate.

Agent za čitanje logova

- Koristi se uz transakcijsku replikaciju koja pomera transakcije koje je potrebno kopirati iz logova, transakcija izdavača i distribuirane baze podataka.
- Svaka baza podataka ima svog agenta za čitanje logova koji radi na distributeru i može se povezati sa izdavačem.

Merge agent

- Koristi se sa datotekom snimka na početnom nivou i prenosi inkrementalne promene podataka koje se događaju.
- Svaka pretplata za spajanje ima ličnog merge agenta koji se može povezati i sa izdavačem i sa pretplatnikom.
- Može da unese promene uz pomoć okidača.

Agent za čitanje redova

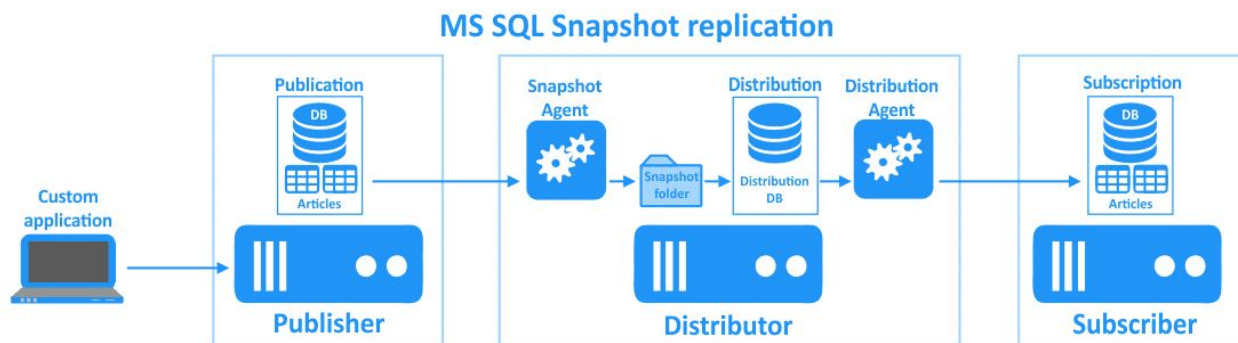
- Koristi se uz transakcijsku replikaciju zajedno s opcijom ažuriranja redova.
- Pokreće se kod distributera i prebacuje izmene izvršene kod pretplatnika nazad izdavaču.
- U slučaju ovog agenta, postoji samo jedna instanca koja servisira sve publikacije i izdavače za dodeljenu distributivnu bazu podataka.

Tipovi MS SQL Server replikacija

Replikacija MS SQL Servera je tehnologija za kopiranje i sinhronizaciju podataka između baza podataka kontinuirano ili redovno u zakazanim intervalima. Što se tiče pravca replikacije, MS SQL Server replikacija može biti: jednosmerna, jedan-prema-više, dvosmerna i više-na-jedan. Postoje četiri tipa replikacije MS SQL Server: snapshot replikacija, transakcijska replikacija, peer-to-peer replikacija i replikacija spajanja.

Snapshot replikacija

Snapshot replikacija se koristi za precizno kopiranje podataka onako kako se pojavljuje u trenutku kada je stvoren snimak baze podataka, slika 3.1. Ovaj tip replikacije može se koristiti kada se podaci retko menjaju; kada nije presudno da postoji replika baze podataka koja je starija od matične baze podataka; ili se izvrši velika količina promena u kratkom vremenskom periodu. Za snapshot reprodukciju ne vrši se praćenje promena. Na primer, snapshot replikacija se može koristiti kada se kurs ili satnica ažuriraju jednom dnevno i moraju se distribuirati s glavnog servera na servere u filijalama.

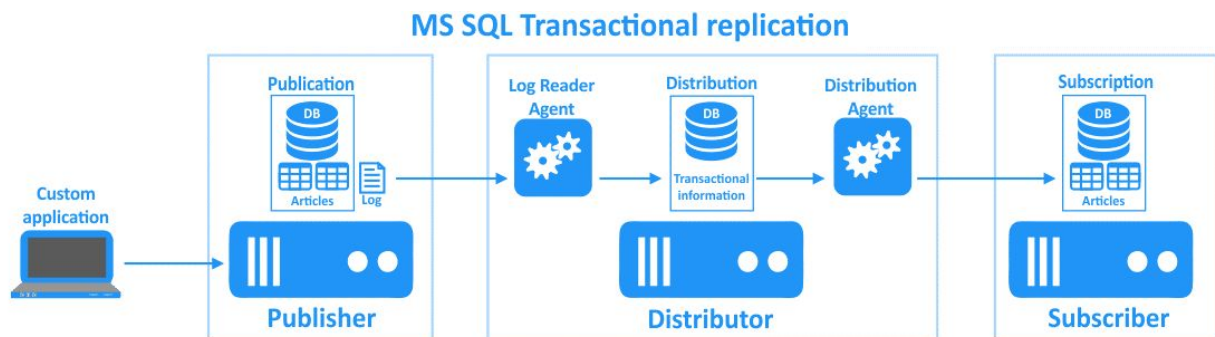


Slika 3.1.. MS SQL Snapshot replikacija

Transakcijska replikacija

Transakcijska replikacija je periodična automatizovana replikacija kada se podaci distribuiraju od glavne baze podataka do replike baze podataka u stvarnom vremenu (ili gotovo u stvarnom vremenu), slika 3.2. Transakcijska replikacija je složenija od snapshot replikacije. Ne samo da se replicira krajnje stanje baze podataka, već se izvršavaju i sve izvršene transakcije, što omogućava praćenje celokupne istorije transakcija na replici baze podataka. Na početku procesa

transakcijske replikacije, nad pretplatnikom se primenjuje snimak, a zatim se podaci posle promene kontinuirano prenose iz matične baze podataka u repliku baze podataka. Transakcijska replikacija se široko koristi kao jednosmerna replikacija.



Slika 3.2. MS SQL Transakcijska replikacija

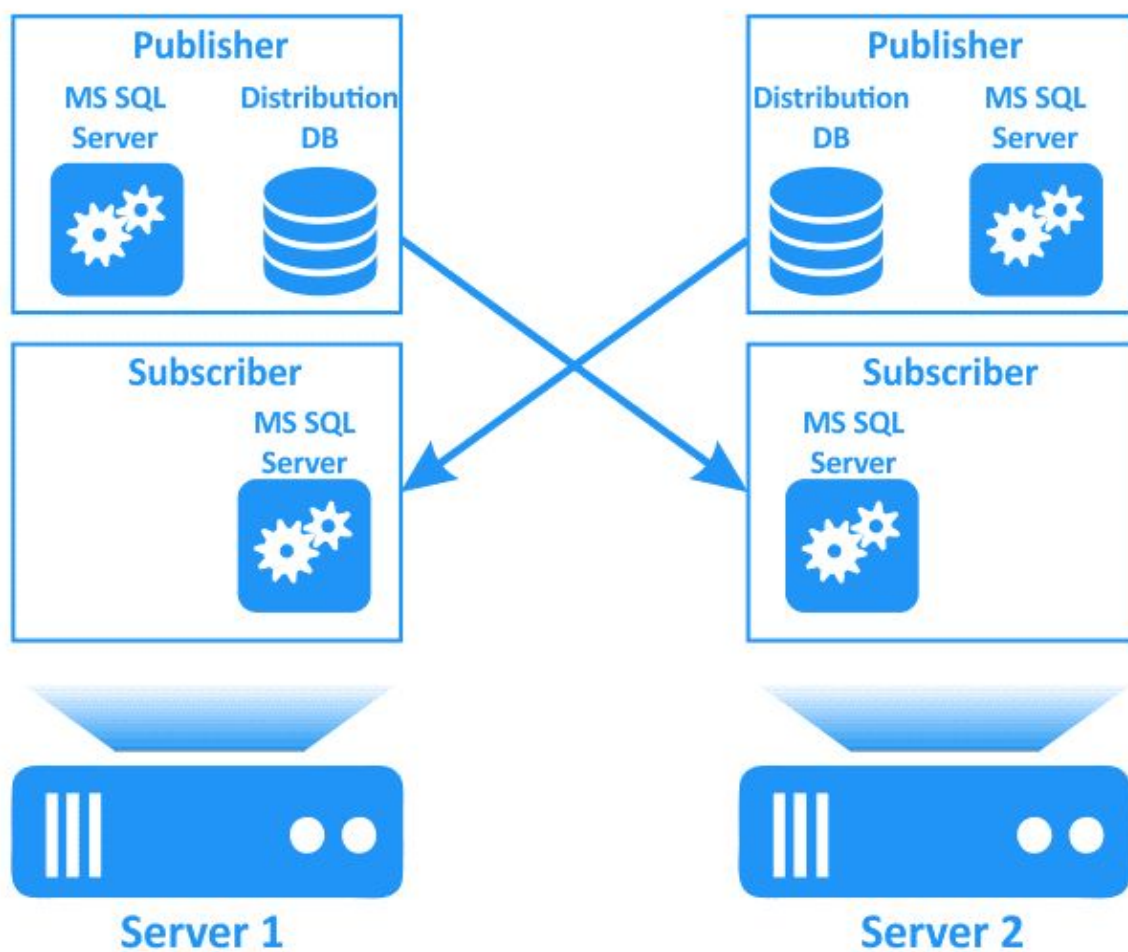
Primeri korišćenja transakcijske replikacije:

- Kreiranje servera baze podataka s replikom baze podataka koja se može koristiti za izvršavanje grešaka ako glavni server servera baze podataka ne uspe.
- Dobijanje izveštaja o operacijama izvršenim u filijalama korišćenjem više izdavača u filijalama i jednog pretplatnika u glavnoj kancelariji.
- Promene treba ponoviti što je pre moguće nakon što se dogode.
- Podaci u izvornoj bazi podataka često se menjaju.

Peer-to-peer replikacija

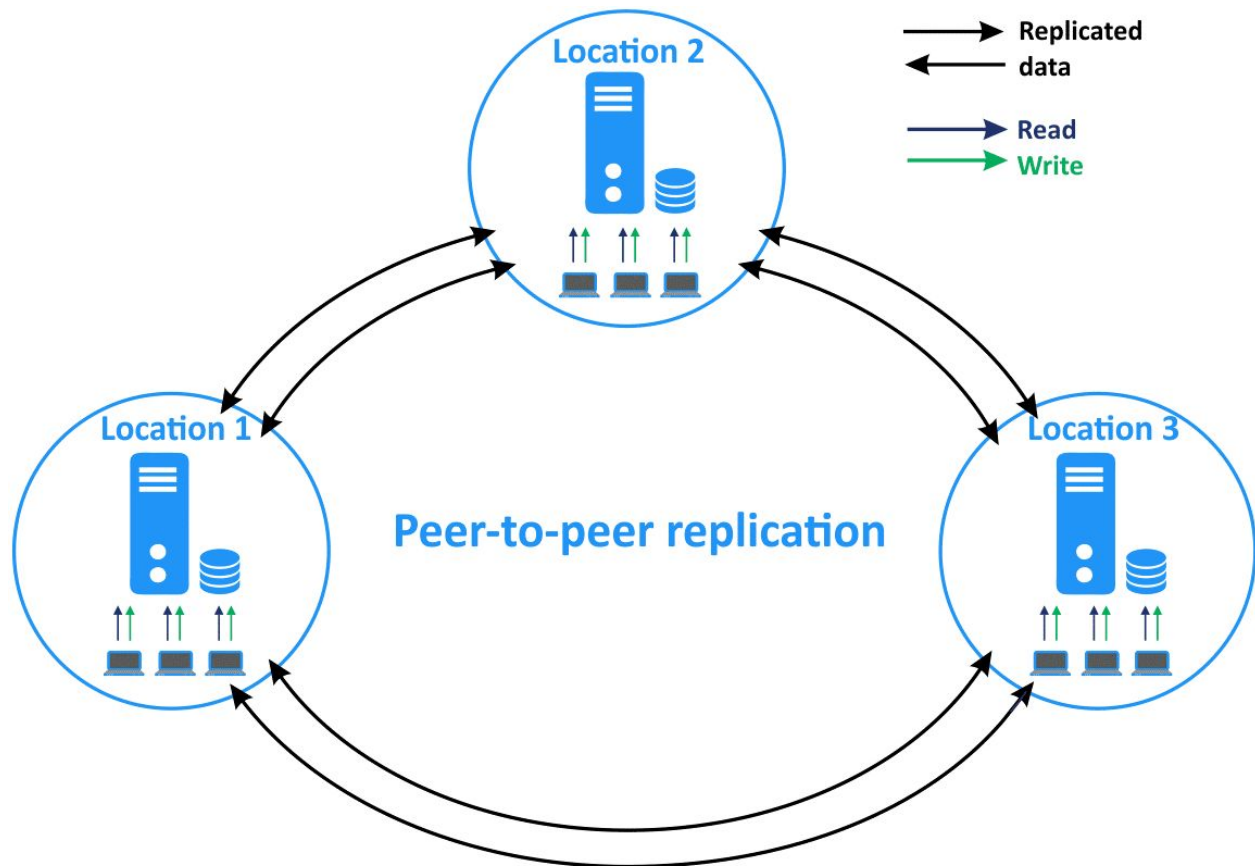
Peer-to-peer replikacija koristi se za kopiranje podataka iz baze podataka više pretplatnika istovremeno, slika 3.3. Ovaj tip replikacije MS SQL Server-a može se koristiti kada se serveri baze podataka distribuiraju širom sveta. Izmene se mogu izvršiti na bilo kom serveru baze podataka. Promene se šire na sve servere baze podataka. Peer-to-peer replikacija može pomoći u skaliranju aplikacije koja koristi bazu podataka. Glavni princip rada zasnovan je na transakcijskoj replikaciji.

MS SQL Peer-to-peer replication



Slika 3.3. MS SQL Peer-to-peer replikacija

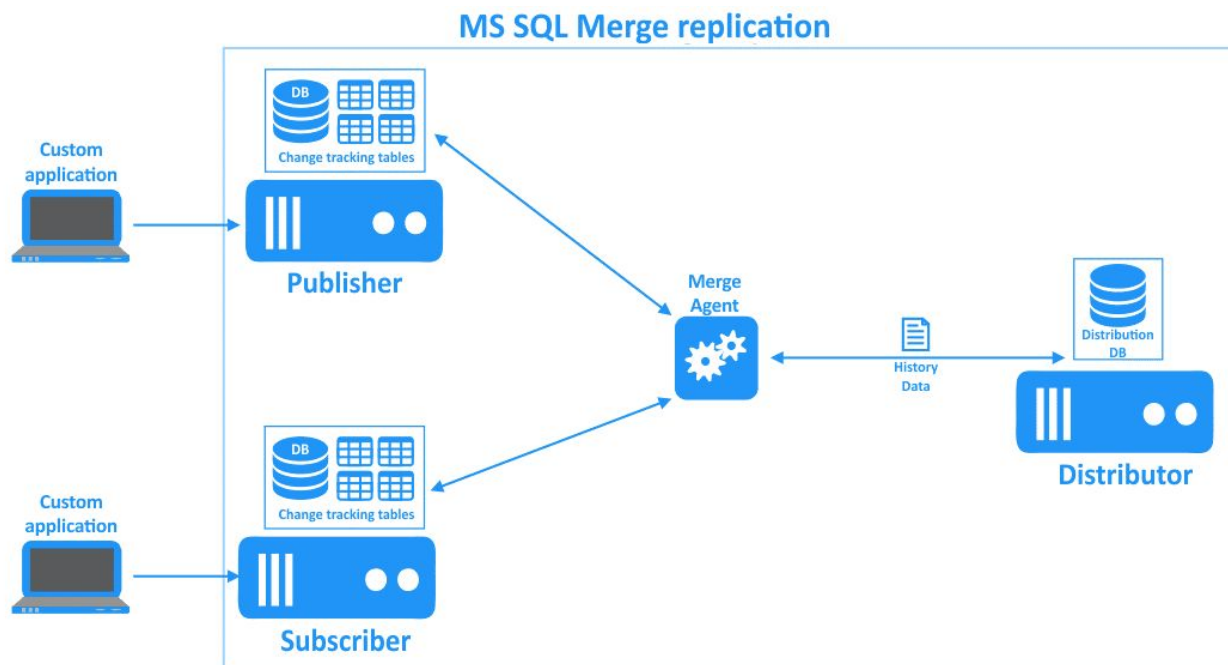
Na slici 3.4. se može videti kako se peer-to-peer replikacija MS SQL Server-a može koristiti između poslužitelja baza podataka koju su distribuirani širom sveta.



Slika 3.4. Peer-to-peer replikacija između poslužitelja

Merge replikacije

Replikacija spajanja, slika 3.5, je vrsta dvosmerne replikacije koja se obično koristi u okruženjima server-klijent za sinhronizaciju podataka na poslužiteljima baze podataka kada se ne može kontinuirano povezivati. Kad je mrežna veza uspostavljena između oba poslužitelja baze podataka, agenti za replikaciju spajanja otkrivaju promene napravljene na obe baze podataka i modifikuju baze podataka da bi uskladili i ažurirali njihovo stanje. Replikacija spajanja slična je transakcijskoj replikaciji, ali se podaci repliciraju od izdavača do pretplatnika i obrnuto.



Slika 3.5. MS SQL Merge replikacija

Ova vrsta replikacije baze podataka najsloženija je od svih vrsta replikacija MS SQL Server-a i retko se koristi. Na primer, replikaciju spajanja mogu koristiti više vršnjačkih prodavnica koje rade sa zajedničkim skladištem. Svakoju prodavnici je dozvoljeno da promeni podatke u bazi skladišta i istovremeno sve prodavnice moraju imati ažurirano stanje svojih baza podataka nakon otpreme robe ili isporuke zaliha u skladište. Replikacija spajanja može se koristiti u slučajevima kada ažurirane informacije moraju biti dostupne istovremeno za glavnu (ili centralnu) bazu podataka i baze podataka podružnica.

Prednosti i nedostaci

Replikaciju baze podataka često nadgleda administrator baze podataka ili replikacija. Ispravno implementirani sistem replikacije može ponuditi nekoliko prednosti, uključujući sledeće:

- *Smanjivanje opterećenja.* Budući da se replicirani podaci mogu raširiti na više servera, to eliminiše verovatnoću da će bilo koji server biti pretrpan korisničkim upitima za podatke.
- *Efikasnost.* Serveri koji su manje opterećeni upitima mogu da pruže poboljšane performanse manjem broju korisnika.
- *Visoka dostupnost.* Upotreba više poslužitelja s istim podacima osigurava visoku dostupnost, što znači da ako jedan server padne, i dalje se mogu osigurati prihvatljive performanse.

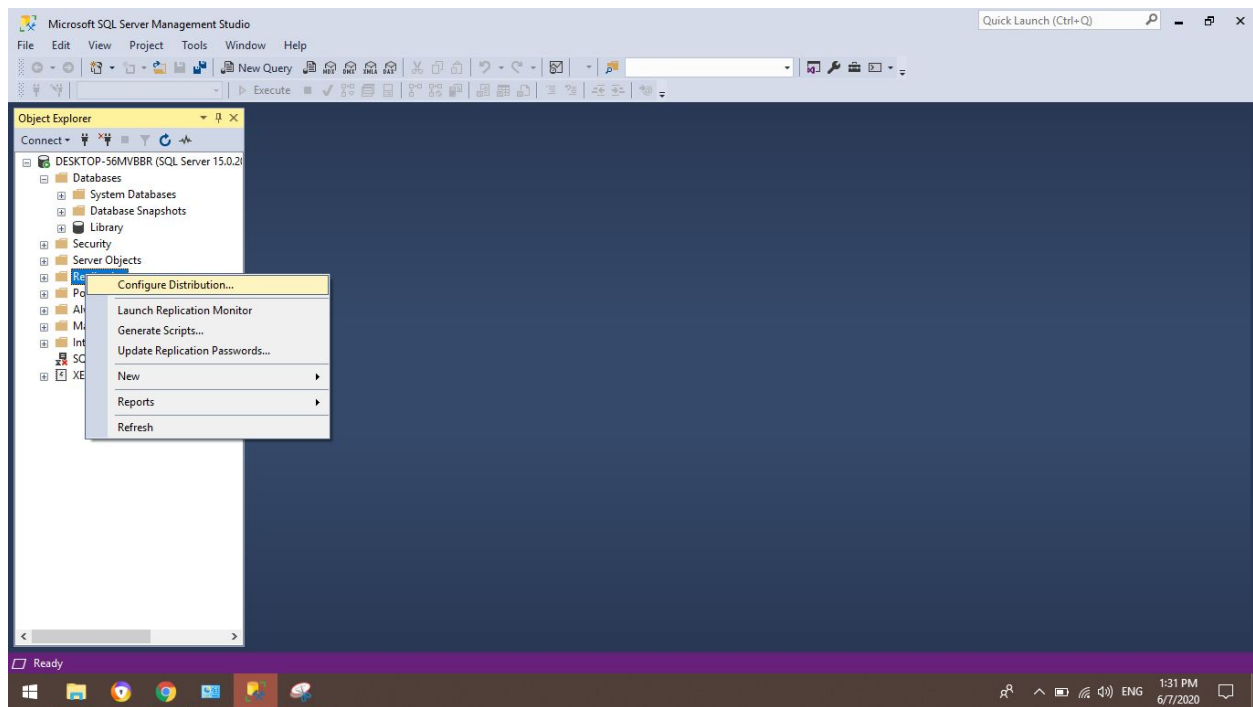
Mnogi nedostaci replikacije baze podataka proizilaze iz loših praksi upravljanja opštim podacima. Ovi nedostaci uključuju sledeće:

- *Gubitak podataka.* Gubitak podataka može se dogoditi tokom replikacije kada se kopiraju netačni podaci ili iteracije ili ažuriranja baze podataka, a samim tim se važni podaci brišu ili se ne uzimaju u obzir. To se može dogoditi ako primarni ključ koji se koristi za proveru kvaliteta podataka u replici nije u funkciji ili je pogrešan. Do toga može doći i ako su objekti baze podataka pogrešno konfigurisani unutar izvorne baze podataka.
- *Nedoslednost podataka.* Slično tome, pogrešne ili zastarele replike mogu izazvati da se različiti izvori ne sinhronizuju jedan sa drugim. To može dovesti do troškova skladišta podataka koji se troše nepotrebno na analizu i čuvanje nebitnih podataka.
- *Više servera.* Pokretanje više servera ima svojstvene troškove održavanja i troškova energije. Ako ih treća strana rešava, organizacija preuzima rizik od zaključavanja dobavljača ili problema sa uslugom van organizacije.

Praktična primena transakcijske replikacije nad MS SQL Server bazom podataka

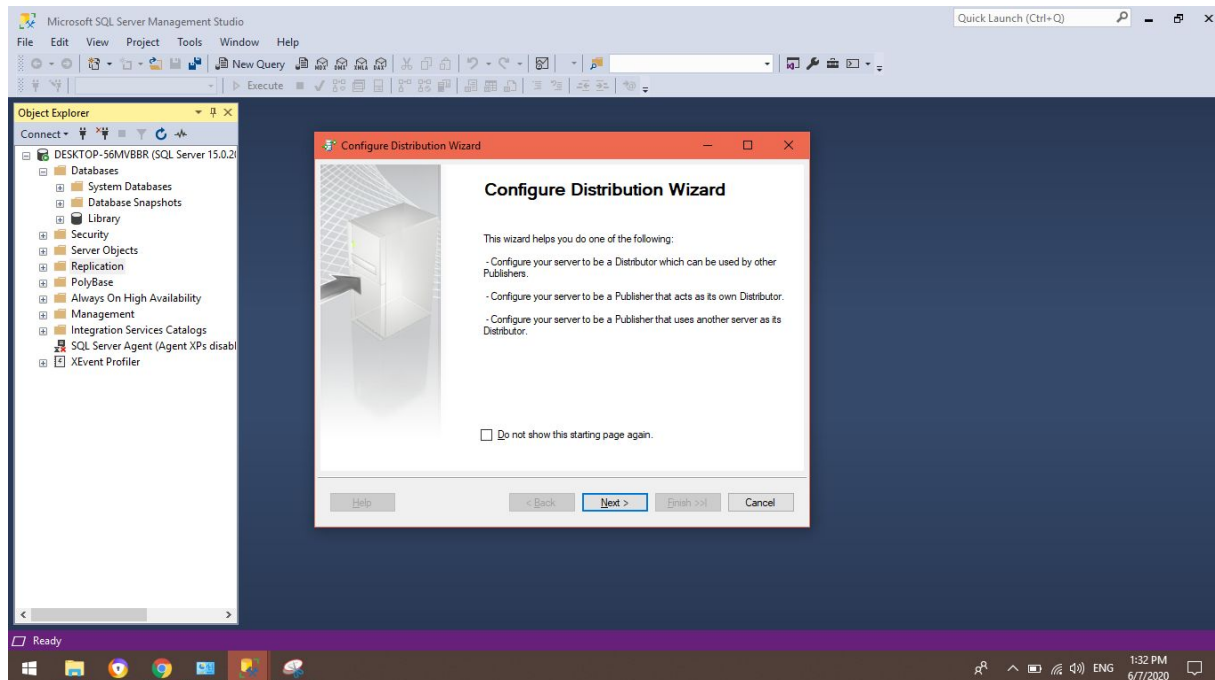
Kao što je to već pomenuto, replikacija baze podataka radi po principu master-slave, gde master server šalje podatke slave serveru, on vrši promenu u svojoj bazi podataka i vraća potvrdu masteru da je transakcija uspešno završena.

Na samom početku, na master serveru potrebno je konfigurisati distributera, kao što je to prikazano na slici 4.1.

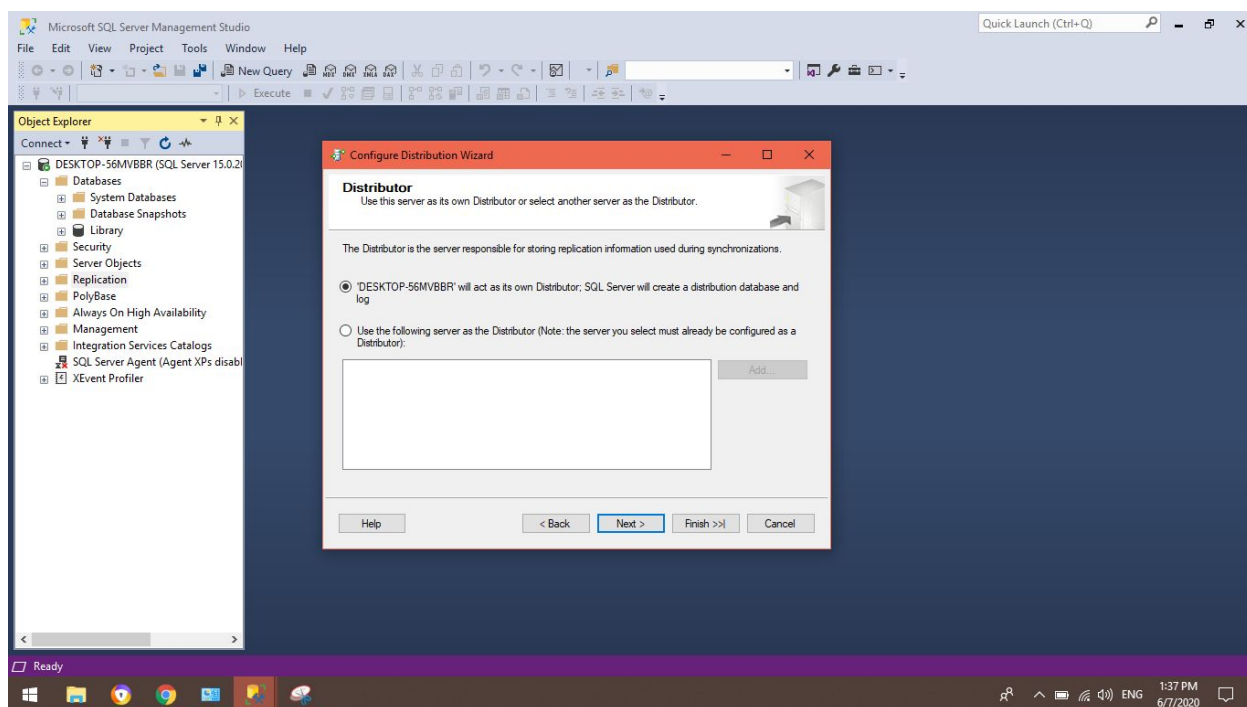


Slika 4.1. Konfiguracija distributera

Sam server je moguće konfigurisati na više načina, slike 4.2. i 4.3. U konkretnom slučaju server će se ponašati kao svoj distributer.

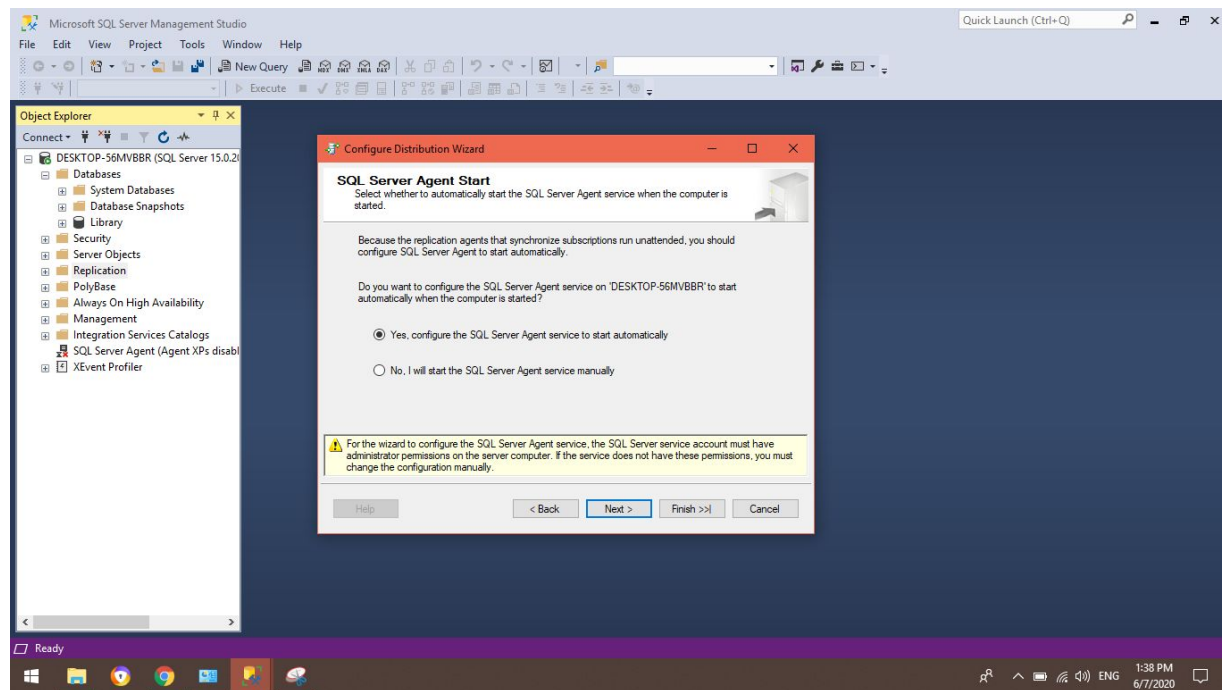


Slika 4.2. Konfiguracija distributera



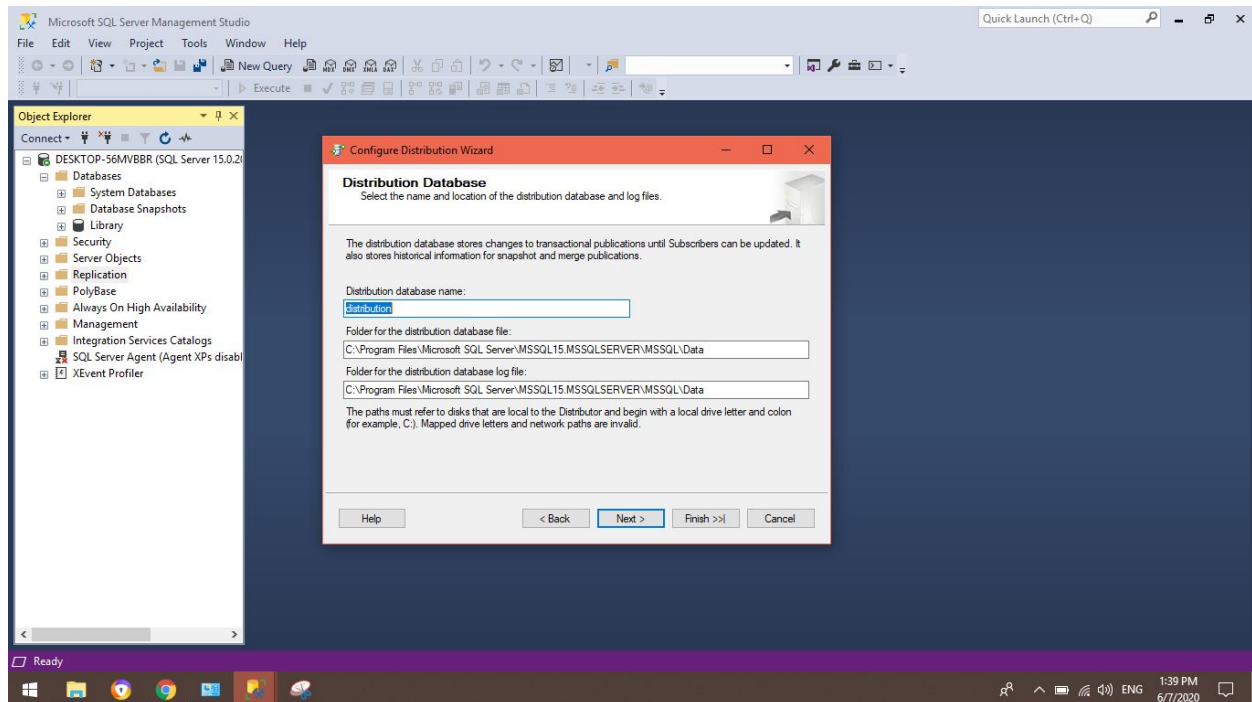
Slika 4.3. Konfiguracija distributera

Nakon odabira distributera potrebno je postaviti snapshot agenta, prikazano na slici 4.4.



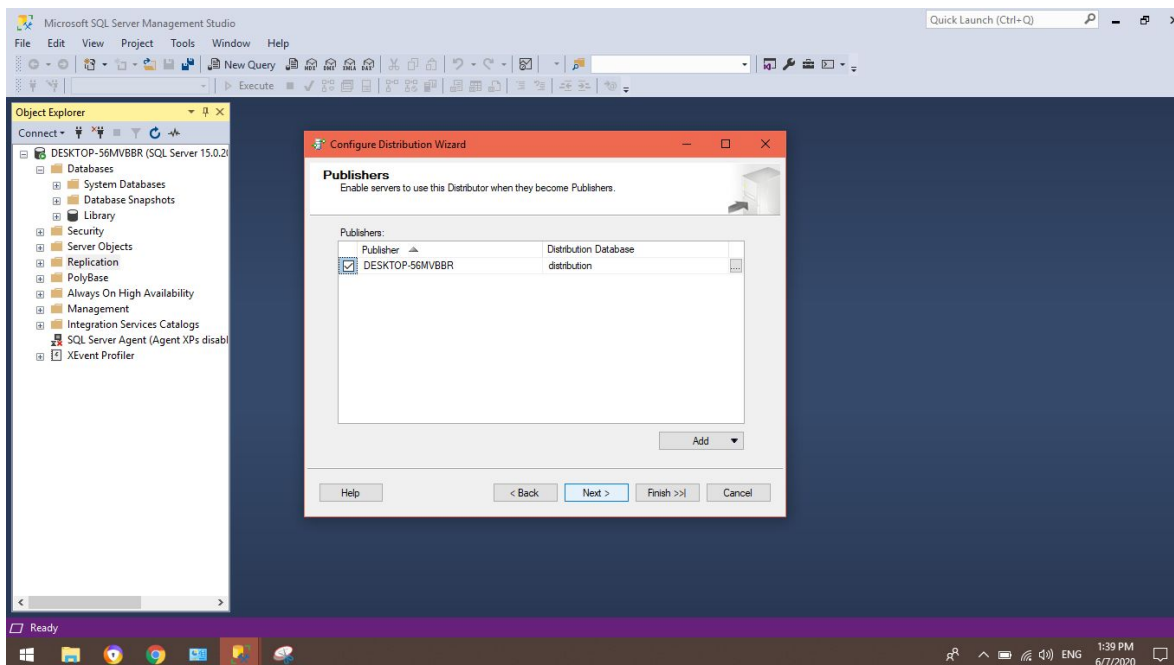
Slika 4.4. Snapshot agent

Zatim je potrebno kreirati bazu distributera, prikazano na slici 4.5.



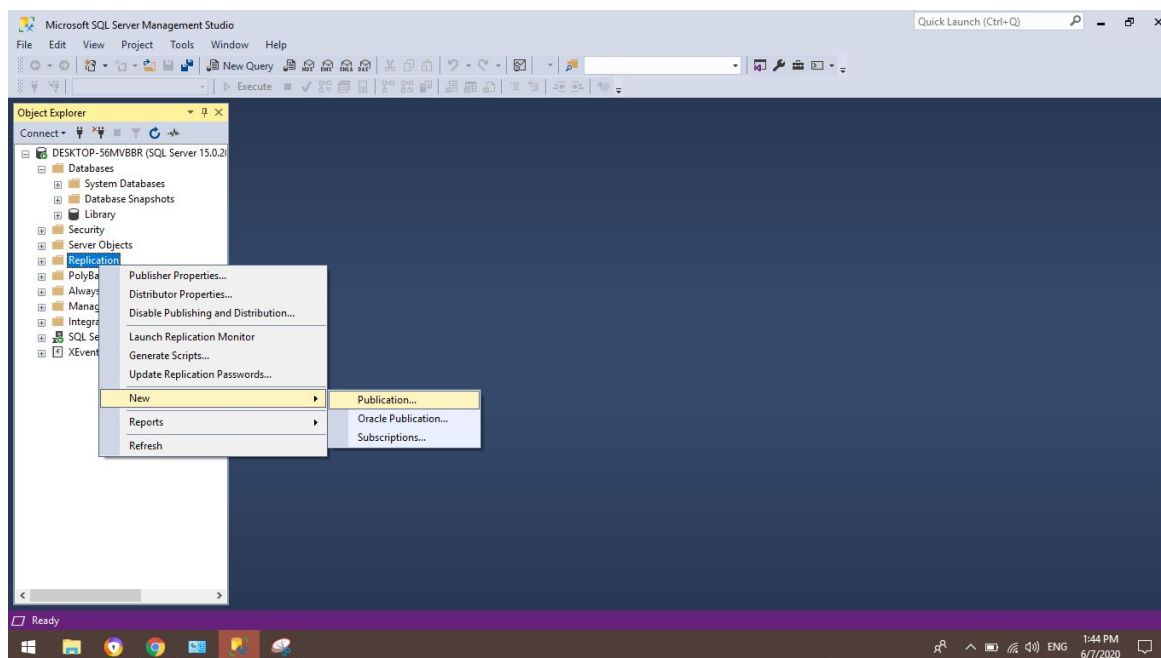
Slika 4.5. Kreiranje baze podataka za distributera

Na slici 4.6. potrebno je izabrati publisher-a, kao i bazu podataka za distributera koja će biti korišćena.



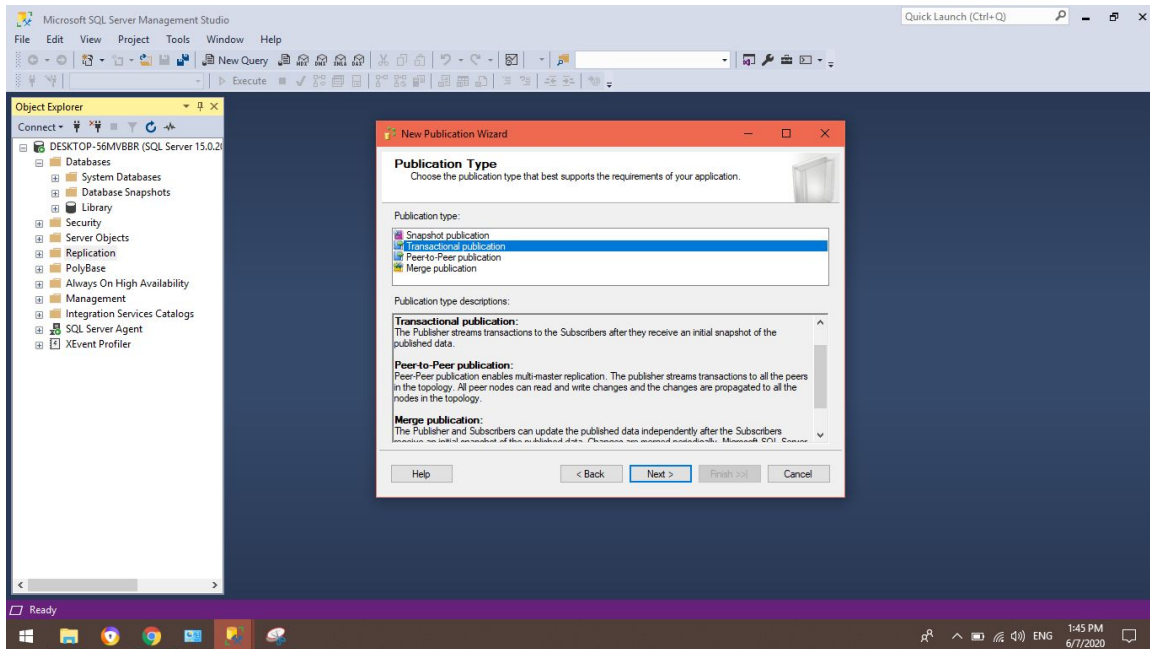
Slika 4.6. Odabir publisher-a

Nakon ovog koraka uspešno je postavljen distributer. Sada je potrebno napraviti publikaciju koju će subscriber-i koristiti.



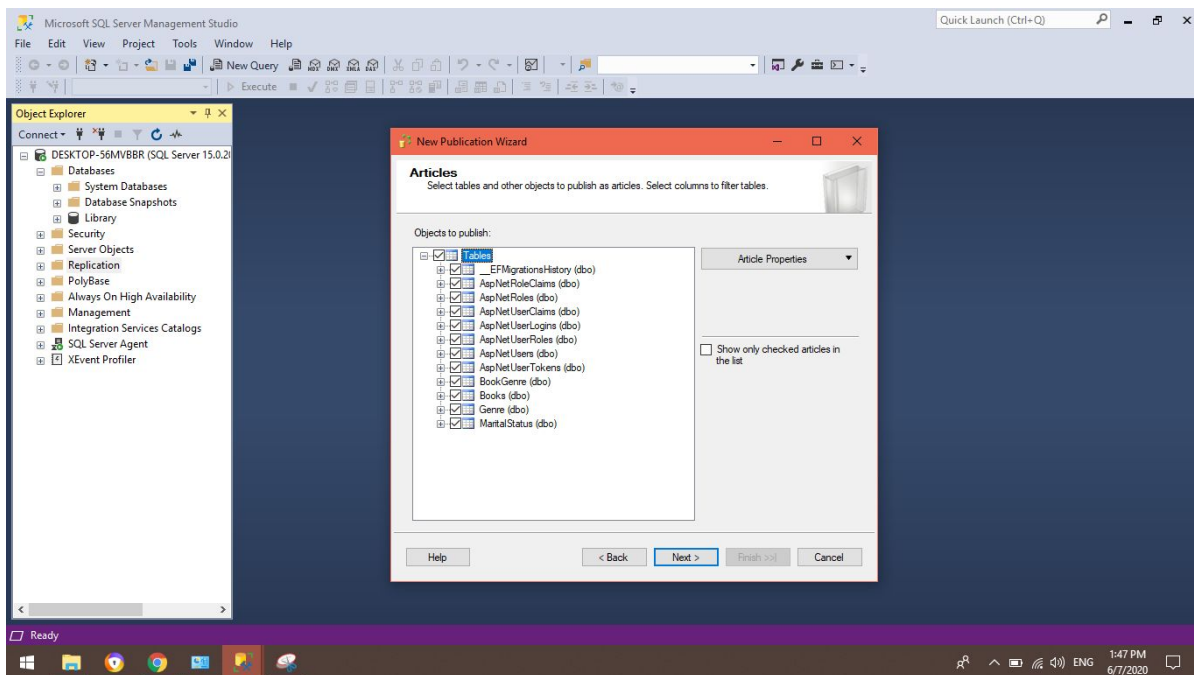
Slika 4.7. Kreiranje publikacije

Nakon izbora određene baze podataka koja će se koristiti u konkretnoj publikaciji, potrebno je izabrati tip same publikacije. U konkretnom primeru korišćena je transakcijska replikacija, slika 4.8.

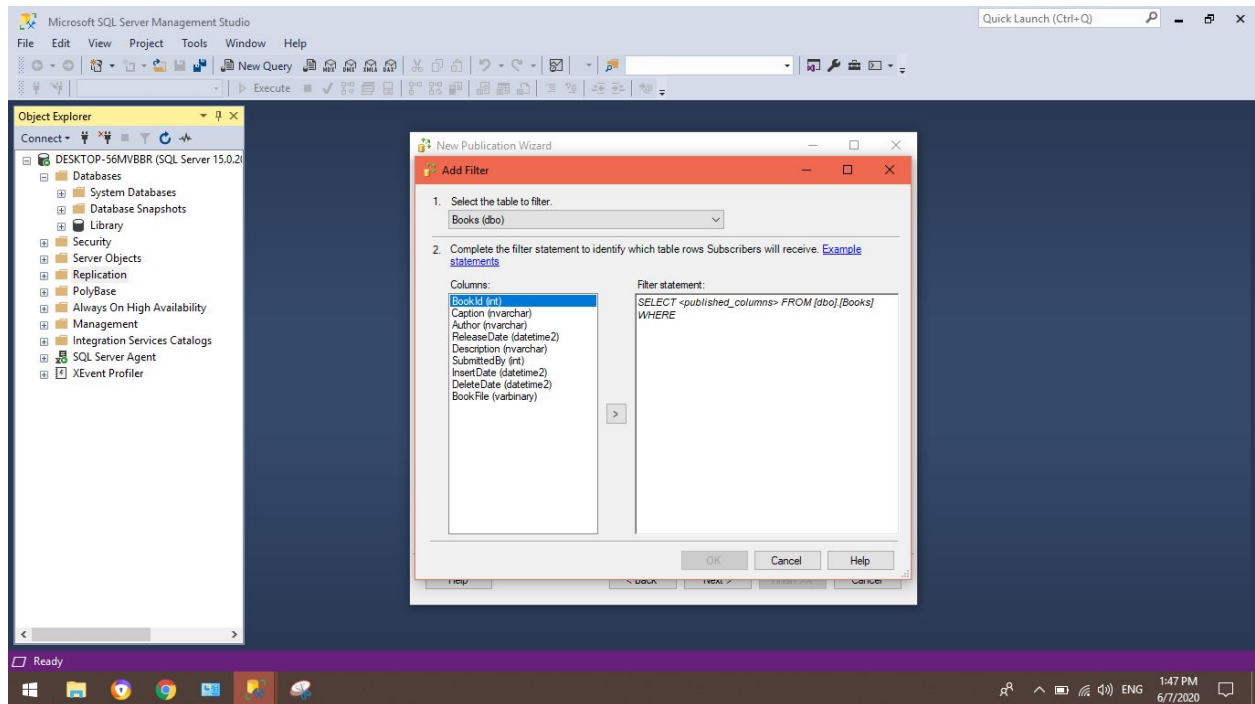


Slika 4.8. Odabir tipa publikacije

Zatim se vrši kreiranje artikala koje je dalje moguće filtrirati, slike 4.9. i 4.10.

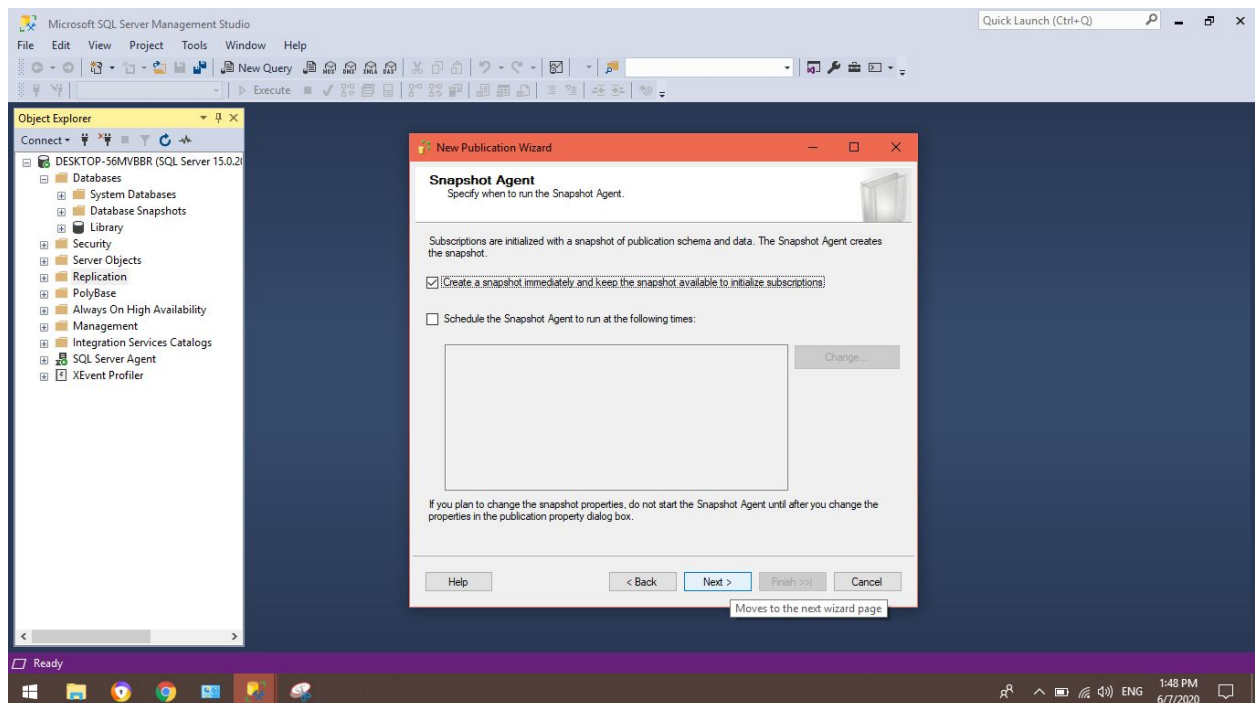


Slika 4.9. Izbor tabela i objekata za publikaciju u vidu artikala

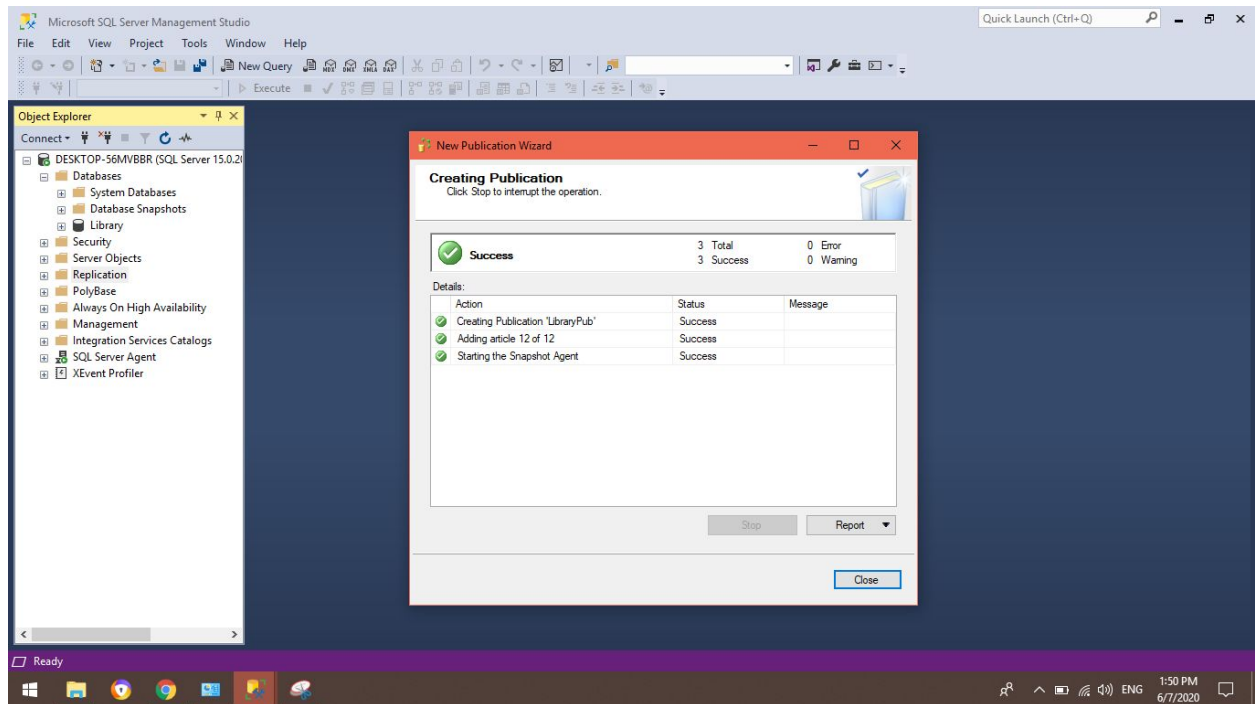


Slika 4.10. Filtriranje tabela

Zatim je potrebno postaviti snapshot agenta, slika 4.11.

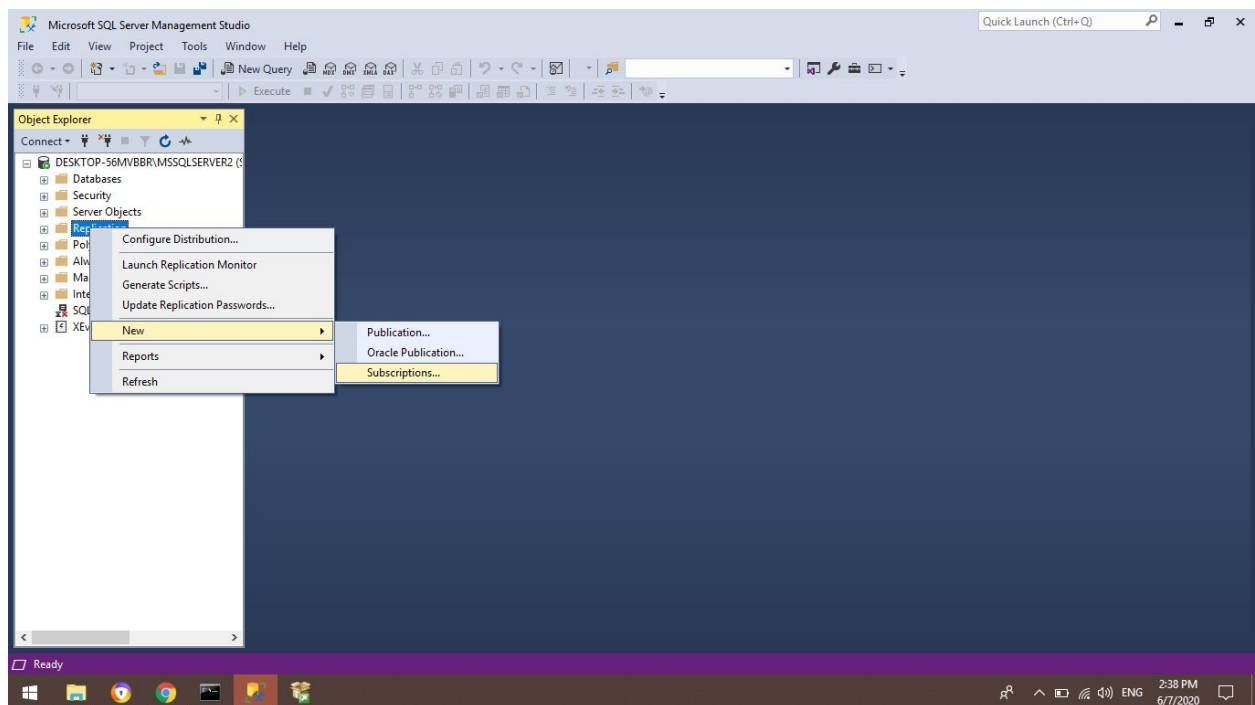


Slika 4.11. Konfiguracija snapshot agenta

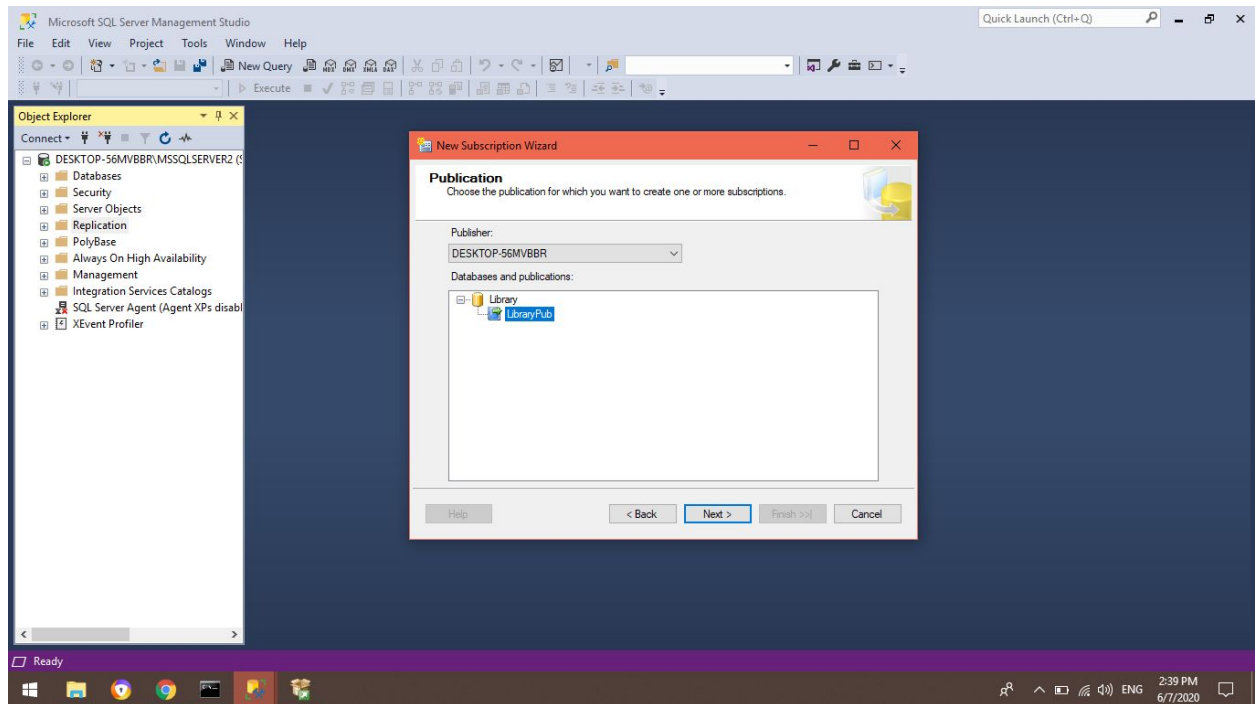


Slika 4.12. Kreirana publikacija

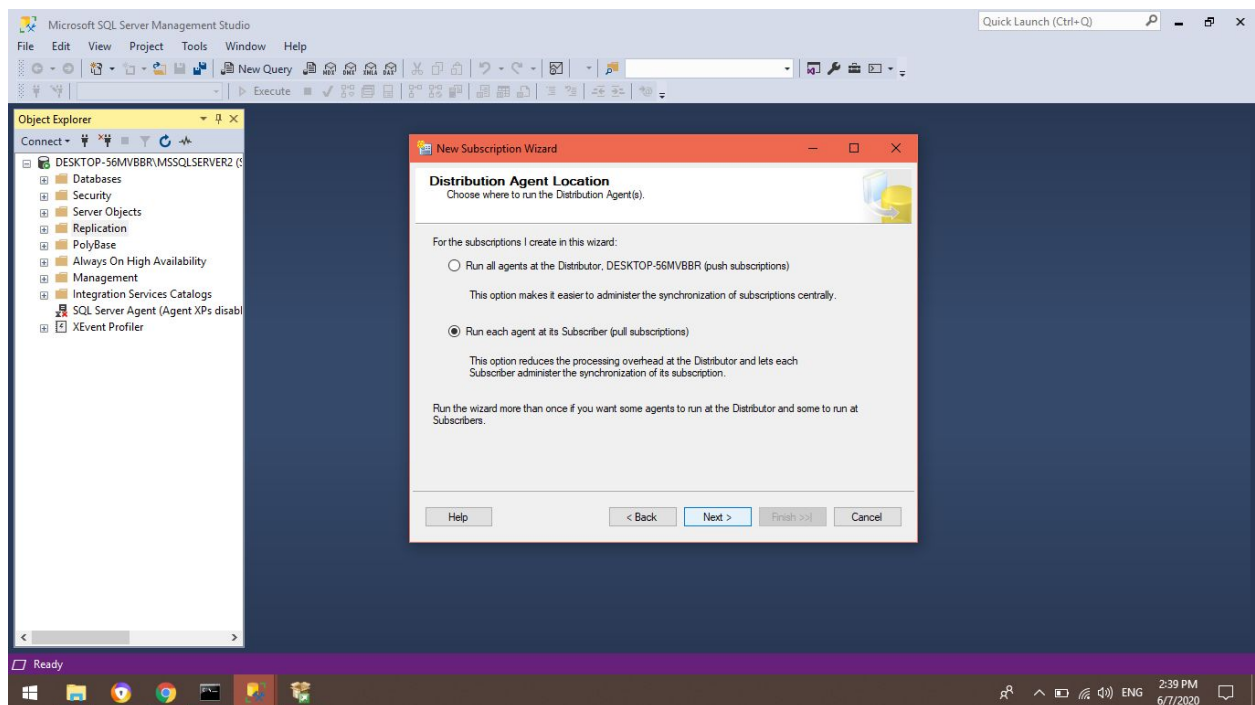
Nakon uspešno kreirane publikacije, potrebno je slave server konfigurisati kao subscriber-a, slika 4.13.



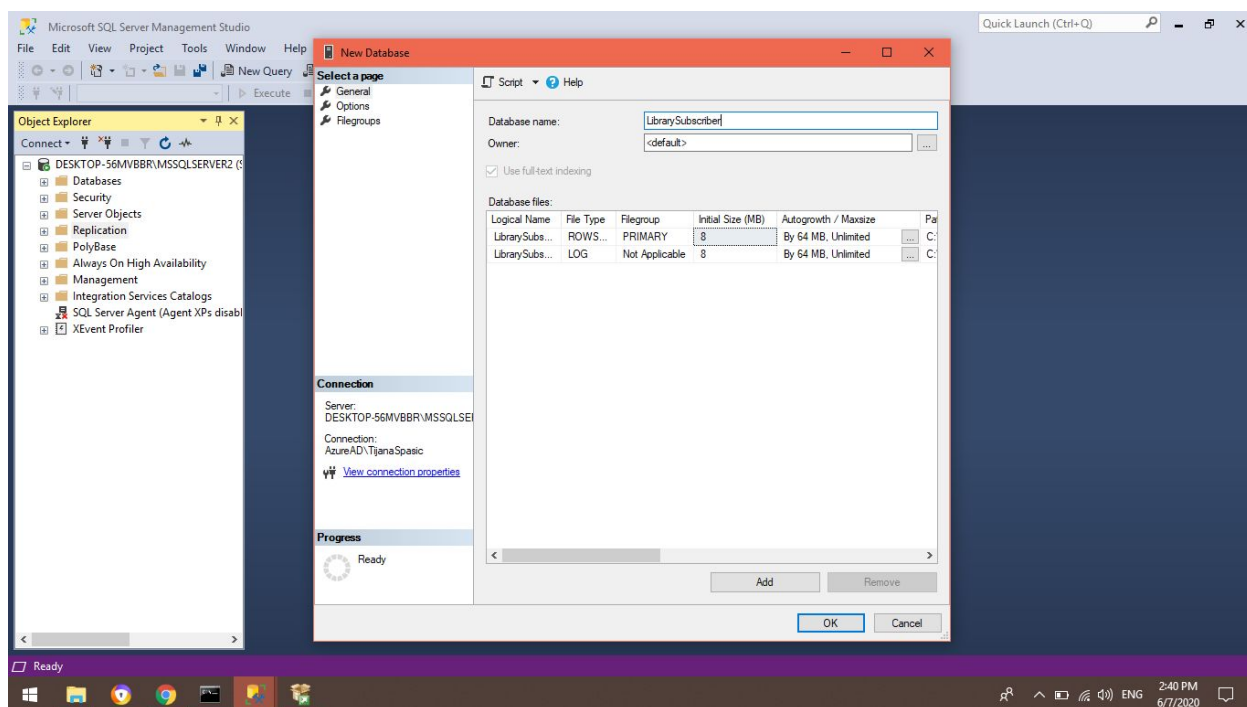
Slika 4.13. Kreiranje subscriber-a



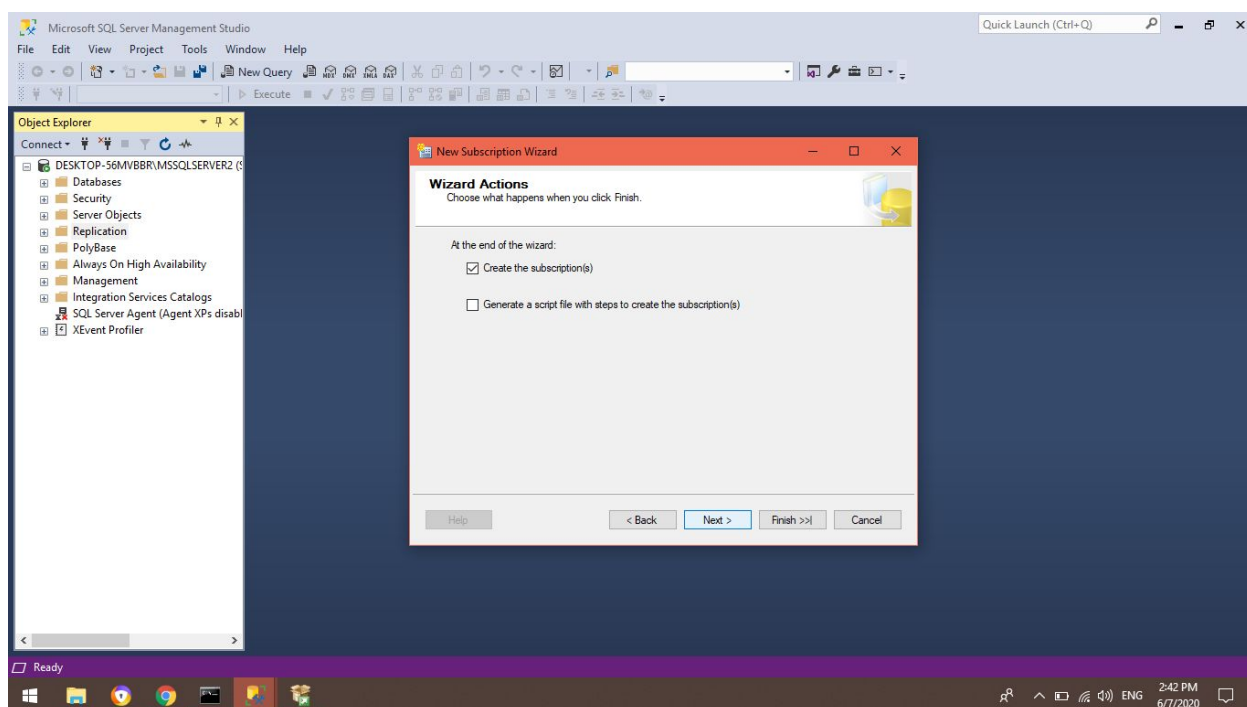
Slika 4.14. Izbor publishera i publikacije



Slika 4.15. Konfiguracija distributivnog agenta



Slika 4.16. Kreiranje baze podataka na strani subscriber-a



Slika 4.17. Uspešno kreiran subscriber

Literatura

[1] Michael Bose, “How to Configure MS SQL Server Replication: Complete Walkthrough”

Web-sajt

https://www.nakivo.com/blog/how-to-configure-ms-sql-server-replication-walkthrough/?fbclid=IwAR2ku5Yf0om4-JFvrUmle5v8d6rj-4LUZMzdMbWUrZUk__wOrBJ54xn8Jso

(pristupljeno maja 2020)

[2] Microsoft, “What is SQL Server Replication and How it Works?”

Web-sajt

<https://docs.microsoft.com/en-us/biztalk/core/high-availability-using-sql-server-always-on-availability-groups> (pristupljeno maja 2020)