

NHN에서는 CUBRID라는 오픈소스 DBMS를 개발하고 있으며 사내외 여러 서비스에 적용해 안정적으로 운영하고 있습니다. 여기서 살펴볼 PostgreSQL도 오픈소스 DBMS며, 여러 나라의 다양한 개발자들의 자발적인 노력으로 개발되고 있습니다. 오픈소스 프로젝트로는 상당히 긴 역사인 15년 이상의 오랜 역사를 가지고 있고 최근에는 1년 여만에 새로운 버전인 9.2 버전을 출시했습니다. 이 글에서는 꾸준히 발전해 가는 PostgreSQL가 어떤 데이터베이스인지 알아보겠습니다.

PostgreSQL을 알아야 할 이유

PostgreSQL(<http://www.postgresql.org>)은 북미와 일본에서는 높은 인지도와 많은 인기를 얻고 있는 RDBMS다. 국내에서는 아직 잘 사용하지 않고 있지만, 기능과 성능면에서 매우 훌륭한 RDBMS이기 때문에 PostgreSQL가 어떠한 데이터베이스인지 시간을 들여 알아볼 필요는 있다.

PostgreSQL(포스트-그레스-큐엘 [Post-Gres-Q-L]로 발음)은 객체-관계형 데이터베이스 시스템(ORDBMS)으로, 엔터프라이즈급 DBMS의 기능과 차세대 DBMS에서나 볼 수 있을 법한 많은 기능을 제공하는 오픈소스 DBMS다. 실제 기능적인 면에서는 Oracle과 유사한 것이 많아, Oracle 사용자들이 가장 쉽게 적응할 수 있는 오픈소스 DBMS가 PostgreSQL이라는 세간의 평 또한 많다.

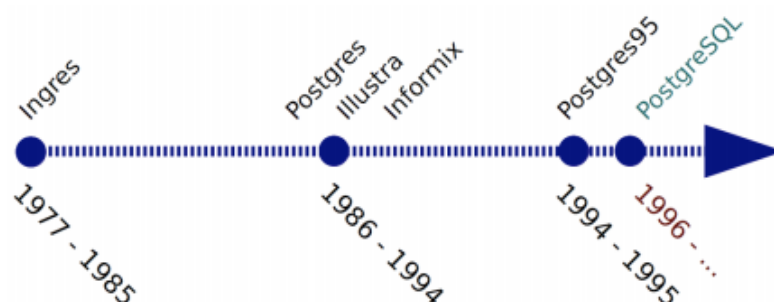
역사

PostgreSQL의 조상에는 여러 제품이 있는데, 그중 Ingres(Interactive Graphics REtrieval System)는 PostgreSQL의 시조라고 할 수 있다. 이 Ingres는 현재까지도 활발하게 활동하고 있는 데이터베이스 계의 거장 [Michael Stonebraker](#)가 시작한 프로젝트다.



그림 1 Michael Stonebraker(원본 출처: <https://www.facebook.com/michael.stonebraker>)

이 Ingres는 1977년 미국 버클리대학에서 시작한 프로젝트다. 이후 Ingres를 잇는 Postgres(Post-Ingres)라는 또 다른 프로젝트 또한 그의 손을 거쳐 탄생됐다. 1991년 Postgres 버전 3이 출시되면서 많은 사용자를 확보했으나 사용자 지원에 대한 부담이 증가해 결국 프로젝트는 1993년에 종료됐다. 이후에도 Postgres는 현재의 Informix 제품에 상당 부분 영향을 준 것으로 알려져 있다(Postgres를 상용화한 Illustra는 1997년 Informix에 인수됐고 2001년에 Illustra는 IBM에 인수됐다).



하지만 Postgres 사용자와 학생들이 프로젝트 종료 선언과 별개로 Postgres의 개발을 계속 진행했고, SQL 지원과 구조 개선을 통해 Postgres보다 40% 정도 빠른 성능을 보여 주는 Postgres95를 만들어 냈다.

이 Postgres95는 1996년 오픈소스가 되면서 Postgres를 계승했다는 것과 SQL을 지원한다(Postgres는 SQL이 아니라 QUEL이라는 언어를 지원했다)는 것을 반영하기 위해 현재의 PostgreSQL로 이름을 변경한 후 1997년 최초 버전을 6.0으로 정해 PostgreSQL을 출시했다.

이후에도 PostgreSQL은 오픈 소스 커뮤니티에 의해 최근까지도 활발히 개발되고 있으며 2012년 10월 현재 9.2 버전까지 출시됐다.

또한 개방된 라이선스(BSD나 MIT 라이선스와 비슷하게 상용적인 사용과 수정을 허용하며 단, 사용 중 발생하는 문제에 대해 법적 책임이 없음을 명확히 함)의 특징으로 인해 20종 이상의 다양한 분기(fork)가 존재했으며, 그중 많은 수가 PostgreSQL에 다시 영향을 주거나 혹은 사라졌다.

PostgreSQL의 로고는 'Slonik'이라는 이름을 가진 코끼리다. 코끼리를 로고로 사용하게 된 정확한 근원은 찾을 수 없으나 과거의 흔적을 뒤져 보면, 오픈소스화된 직후 한 사용자가 아가사 크리스티의 소설 '코끼리는 기억한다'에서 착안해 제안한 것으로 보인다. 이후 모든 공식적인 자리에서 이 코끼리 로고는 빠지지 않고 등장하게 됐다. 코끼리는 크고 강하고 믿음직하며 기억력이 좋다는 이미지 때문에 Hadoop이나 Evernote도 코끼리를 공식 로고로 사용하고 있다.



그림 3 PostgreSQL의 로고*

기능 및 제한

PostgreSQL은 관계형 DBMS의 기본적인 기능인 트랜잭션과 ACID(Atomicity, Consistency, Isolation, Durability)를 지원한다. 또한 ANSI SQL:2008 규격을 상당 부분 만족시키고 있으며, 전부를 지원하는 것을 목표로 현재도 기능이 계속 추가되고 있다.

또한 PostgreSQL은 기본적인 신뢰도와 안정성을 위한 기능뿐만 아니라 진보적인 기능이나 학술적 연구를 위한 확장 기능도 많이 가지고 있다. PostgreSQL의 기능을 대략적으로 열거해 보더라도 상당히 많은 기능을 가지고 있음을 알게 된다.

- Nested transactions (savepoints)
- Point in time recovery
- Online/hot backups, Parallel restore
- Rules system (query rewrite system)
- B-tree, R-tree, hash, GiST method indexes
- Multi-Version Concurrency Control (MVCC)
- Tablespaces

- Procedural Language
- Information Schema
- I18N, L10N
- Database & Column level collation
- Array, XML, UUID type
- Auto-increment (sequences),
- Asynchronous replication
- LIMIT/OFFSET
- Full text search
- SSL, IPv6
- Key/Value storage
- Table inheritance

이외에도 엔터프라이즈급 DBMS의 다양한 기능과 새로운 기능을 자랑하고 있다.

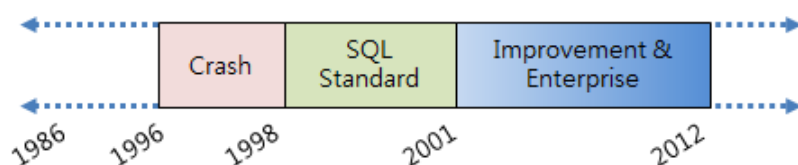
PostgreSQL의 일반적인 제한 사항은 아래와 같다.

표 1 기본 제한 사항

항목	제한 사항
최대 DB 크기(Database Size)	무제한
최대 테이블 크기(Table Size)	32TB
최대 레코드 크기(Row Size)	1.6TB
최대 컬럼 크기(Field Size)	1 GB
테이블당 최대 레코드 개수(Rows per Table)	무제한
테이블당 최대 컬럼 개수(Columns per Table)	250~1600개
테이블당 최대 인덱스 개수(Indexes per Table)	무제한

로드맵

2012년 10월 기준 최신 버전은 9.2이다. PostgreSQL의 발전 과정을 연도별로 간단히 정리하면 <그림 4>와 같이 나타낼 수 있을 것이다.



버전별 주요 기능은 다음과 같다.

표 2 버전별 주요기능

버 전	출시	주요 기능
0.01	1995	Postgres95 릴리스
1.0	1995	저작권 변경, 오픈소스화
6.0 ~ 6.5	1997 ~ 1999	PostgreSQL로 이름 변경 Index, VIEWS and RULEs Sequences, Triggers Genetic Query Optimizer Constraints, Subselect MVCC, JDBC interface,
7.0 ~ 7.4	2000 ~ 2010	Foreign keys SQL92 syntax JOINS Write-Ahead Log Information Schema, Internationalization
8.0 ~ 8.4	2005 ~ 2012	Microsoft Windows Native 버전 지원 Savepoint, Point-in-time recovery Two-phase commit Table spaces, Partitioning Full text search Common table expressions (CTE) SQL/XML, ENUM, UUID Type Window functions Per-database collation 복제, Warm standby
9.0	2010-09	Streaming 복제, Hot standby Microsoft Windows 64bit 버전 지원 Per-column conditional trigger
9.1	2011-09	기능 차별화 Synchronous 복제 Per-column collations Unlogged tables K-nearest-neighbor indexing Serializable isolation level Writeable CTE (WITH) SQL/MED External Data SE-Linux integration
9.2	2012-09	성능 최적화 linear scalability to 64 cores CPU 전력 소비량 감소 Cascade streaming 복제 JSON, Range Type Lock management 개선 Space-partitioned GiST index Index-only scans (covering)

이후 PostgreSQL 9.3 버전은 2013년 3분기 출시를 목표로 현재 개발이 진행 중이며, 주요 기능으로는 관리 기능 개선, Parallel query 지원, MERGE/UPSERT 지원, Multi-Master 복제, Materialized View 기능, 다중 언어 지원 개선 등이 계획되어 있다. 더 자세한 내용은 <http://wiki.postgresql.org/wiki/TODO> 에서 살펴볼 수 있다.

내부 구조

PostgreSQL의 프로세스 구조를 간단히 살펴보면 다음과 같다.

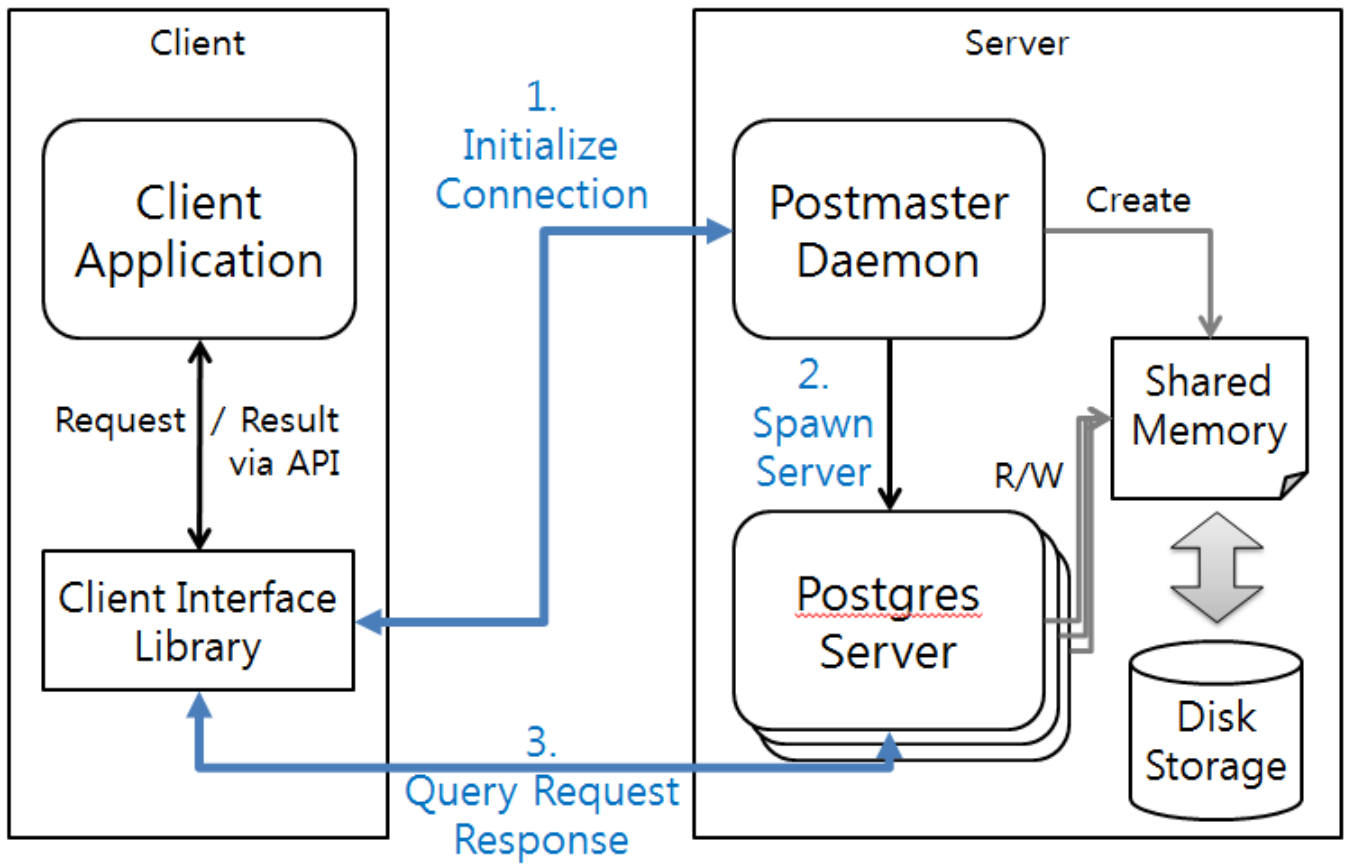


그림 5 프로세스 구조

클라이언트는 인터페이스 라이브러리(libpq, JDBC, ODBC 등의 다양한 인터페이스)를 통해 서버와의 연결을 요청(1)하면, Postmaster 프로세스가 서버와의 연결을 중계(2)한다. 이후 클라이언트는 할당된 서버와의 연결을 통해 질의를 수행(3)한다(그림 5).

서버 내부의 질의 수행 과정을 간단히 살펴보면 다음과 같다.

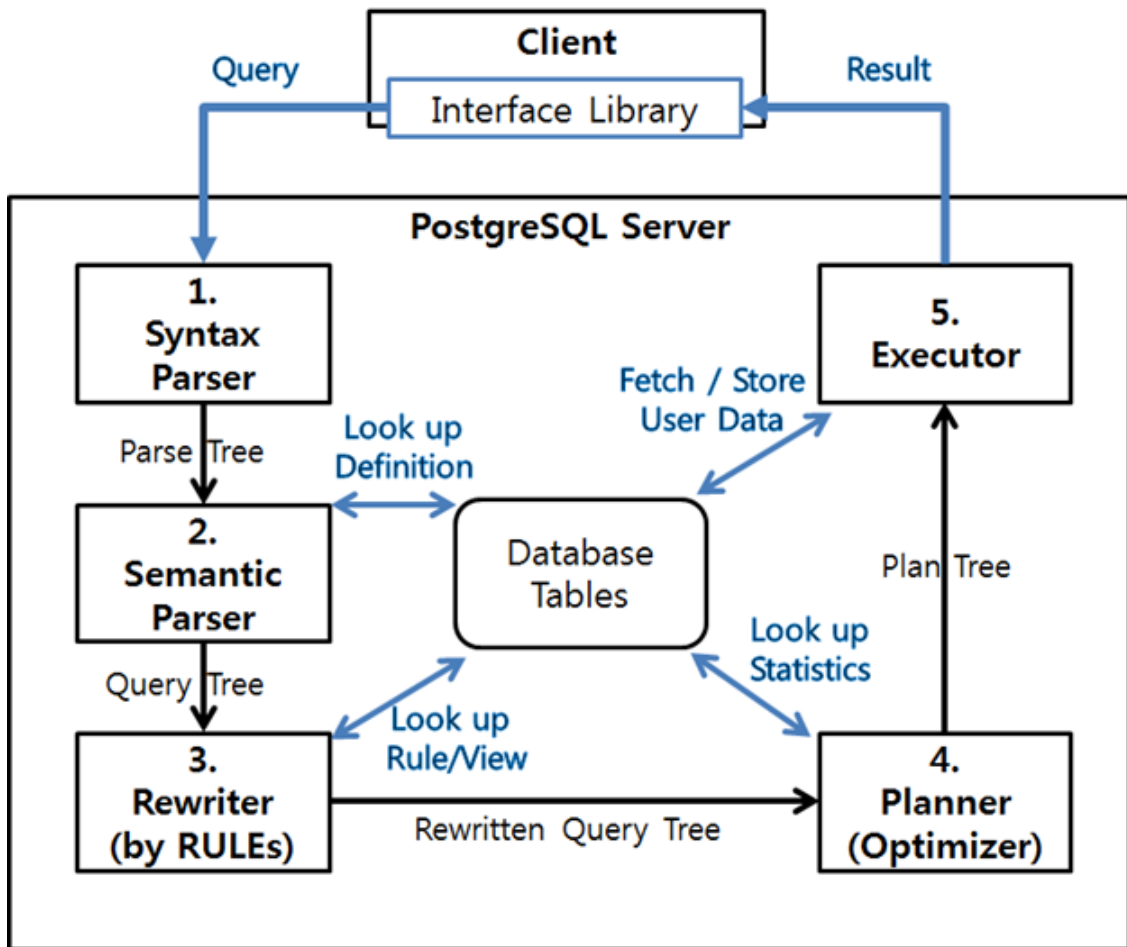


그림 6 쿼리 수행 절차

클라이언트로부터 질의 요청이 들어오면 구문 분석 과정(1)을 통해 Parse Tree를 생성하고 의미 분석 과정(2)를 통해 새로운 트랜잭션을 시작하고 Query Tree를 생성한다.

이후 서버에 정의된 Rule에 따라 Query Tree가 재 생성(3)되고 실행 가능한 여러 수행 계획 중 가장 최적화된 Plan Tree를 생성(4)한다. 서버는 이를 수행(5)하여 요청된 질의에 대한 결과를 클라이언트로 전달하게 된다 (그림 6).

서버의 쿼리 수행 과정에서는 데이터베이스 내부의 시스템 카탈로그가 많이 사용되는데, 사용자가 함수나 데이터 타입은 물론 인덱스 접근 방식 및 RULE 등을 시스템 카탈로그에 직접 정의할 수도 있다. 따라서 PostgreSQL에서는 이것이 기능을 새로 추가하거나 확장하는데 있어 중요한 포인트로 활용된다.

데이터가 저장되는 파일들은 여러 개의 페이지들로 구성되며, 하나의 페이지는 확장 가능한 slotted page 구조를 가진다(그림 7, 그림 8).

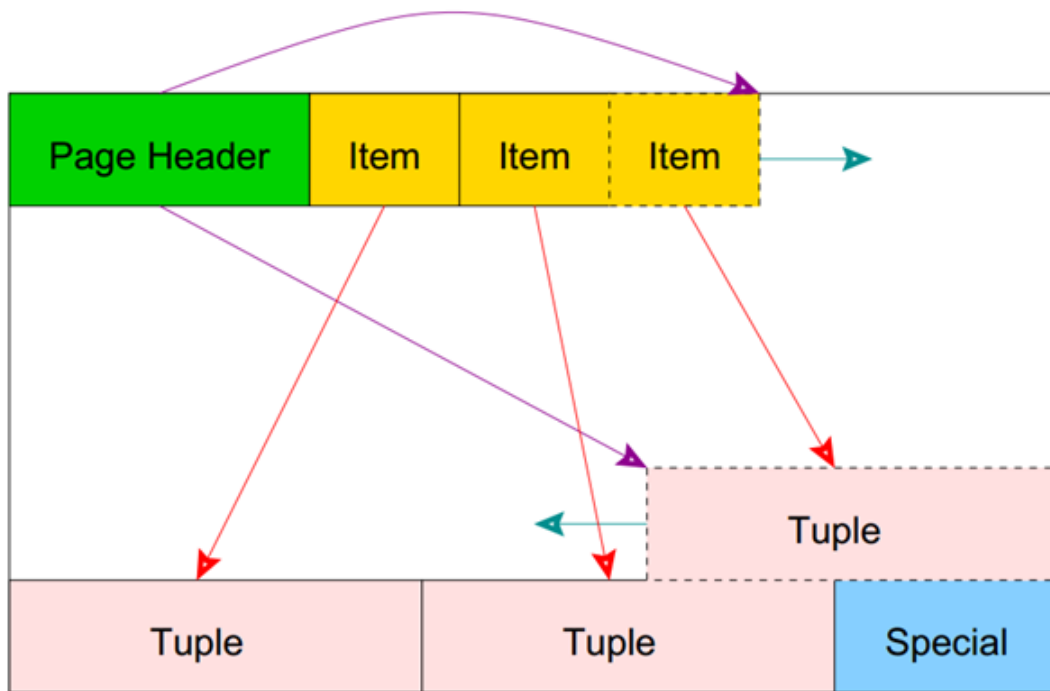


그림 7 데이터 페이지 구조

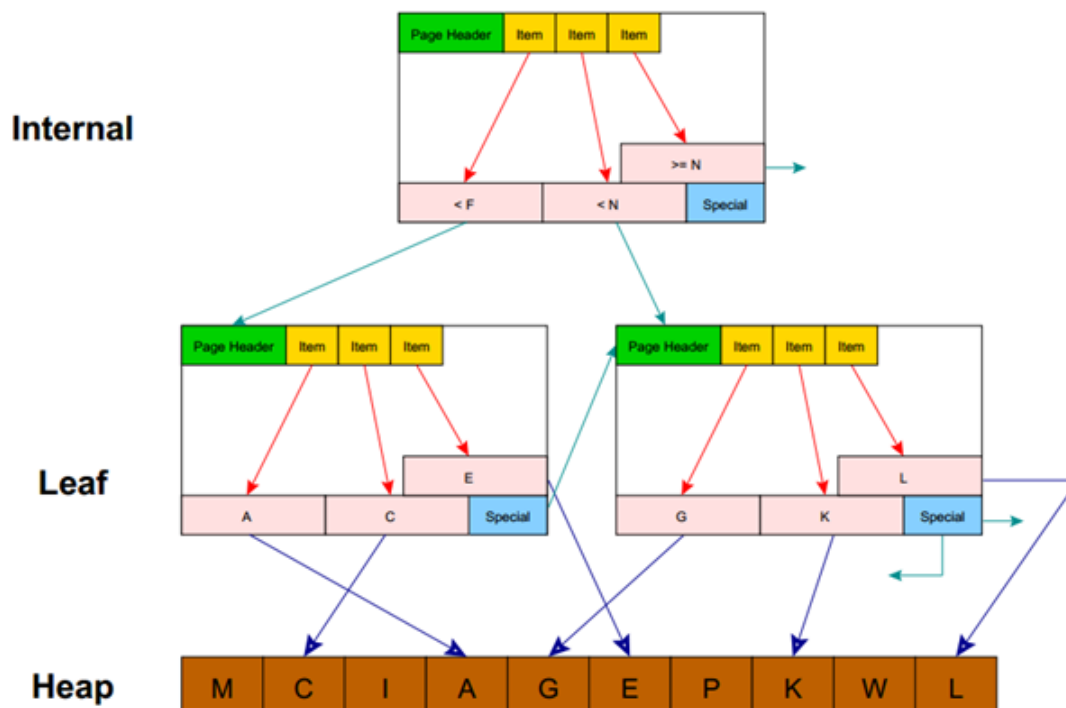


그림 8 인덱스 페이지 구조

개발 프로세스

PostgreSQL의 개발 프로세스 모델은 다음의 문장으로 설명할 수 있다.

"소수 주도의 커뮤니티 기반 오픈 소스 프로젝트"

이는 Linux, Apache, Eclipse 프로젝트와 같이 소수의 관리자와 다양한 개발자 그리고 다수의 사용자가 프로젝트 구성원을 이루고 있으며, 소수의 관리자 그룹(Core Team)은 다수의 사용자로부터의 요청과 피드백을 수집(우선순위를 정하기 위해 투표 방식을 활용하기도 한다. 자세한 내용은 <http://postgresql.uservoice.com> 를 참조한다.)하여 제품의 방향을 결정하고 코드의 최종 승인 및 릴리스 권한을 행사한다(MySQL 또는 JBoss와 같은 기업 관리형 개발프로세스와는 다른 모델로 분류됨).

개발자 그룹은 코드 커미터, 코드 개발자/기여자로 구성되어 있으며 이들은 미국, 일본, 유럽 등의 여러 나라에 분포되어 있다.

다양한 개발자들에 의해 개발된 코드는 다양한 리뷰 과정(Submission Review, Usability Review, Feature Test, Performance Review, Coding Review, Architecture Review, Review Review)과 Core Team의 승인을 거쳐 제품에 반영이 된다. 커뮤니티에서는 오래 전부터 사용해오던 메일링 리스트가 주로 사용되고 있으며, 매뉴얼을 포함한 다양한 문서들이 공식 홈페이지(<http://www.postgresql.org>)에서 잘 관리되고 있다.

경쟁 제품

PostgreSQL은 엔터프라이즈급 상용 데이터베이스와 비교되기를 원하나 주로 유명 오픈소스 DBMS가 비교 대상이 된다. 오픈소스 DBMS의 캐치프레이즈로 제품 특징만을 나열해 보면 다음과 같다.

- PostgreSQL: The world's most advanced open source database
- MySQL: The world's most popular open source database
- Firebird: The true open source database
- CUBRID: Open Source Database Highly Optimized for Web Applications
- SQLite: self-contained library, serverless, zero-configuration, transactional SQL database engine

캐치프레이즈만으로 제품을 비교하긴 어려우나 PostgreSQL은 진보와 개방을 표방하고 있음을 알 수 있다.

다음은 PostgreSQL이 경쟁 제품으로 내세우는 제품에 대한 간략한 비교다.

표 3 경쟁 제품 비교

Oracle	오랫동안 검증된 방대한 양의 코드, 다양한 레퍼런스. 그러나 비싼 비용이 단점
DB2, MS SQL	Oracle과 비슷함
MySQL	다양한 응용과 레퍼런스. 그러나 기업형 개발 모델과 라이선스 부담
다른 상용 DB	오픈소스 DBMS에 의해 세력이 기우는 중
타 오픈소스DB	프로젝트에 개발자를 끌어 들이기 위해 힘든 노력 중

이전부터 PostgreSQL 진영은 엔터프라이즈 시장에 진출하려는 시도를 계속 해 왔고, 2004년에는 PostgreSQL을 이용한 전문 기업인 EnterpriseDB(<http://www.enterprisedb.com>)가 생겨나 엔터프라이즈 DBMS 시장을 좀 더 확고히 다지기 위한 노력을 기울이고 있다. 대표적인 산출물이 Postgres Plus Advanced Server 제품의 출시다. 이는 오픈 소스 PostgreSQL에 Oracle 호환성 기능(PL/SQL, SQL 구문, 함수, DB Links, OCI 라이브러리 등 지원)과 관리 도구를 추가해 용이한 데이터 및 응용 마이그레이션과 Oracle 대비 20% 이하의 비용 절감을 주무기로 내세우고 있다(그림 10).

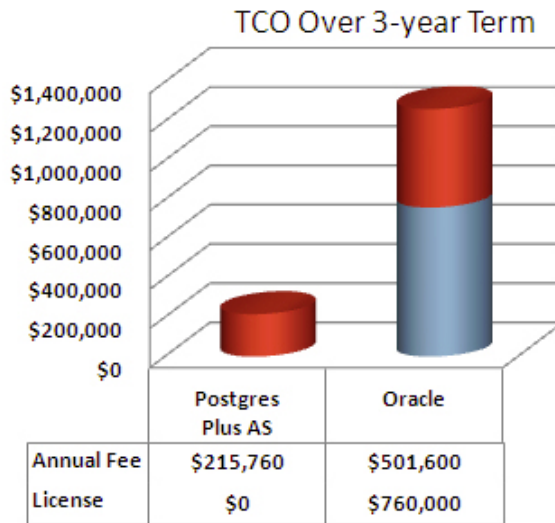


그림 10 Oracle 대비 비용 절감 효과(원본 출처)

또한 PostgreSQL 전문가의 교육, 컨설팅, 마이그레이션 및 기술 지원 서비스 등을 제공하는 등의 차별화된 서비스도 제공하며 각종 분야 약 300여 개의 레퍼런스 사이트를 통해 모든 업계에서 사용 가능한 데이터베이스임을 홍보하며 세계 각지에 사용자 층을 늘리고 있는 추세다.

현황, 동향

대부분의 PostgreSQL 사용기를 읽어 보면 알 수 있듯이 PostgreSQL을 사용하는 사람들은 대부분 개발자적인 성향을 가지고 있으며 제품에 대한 애정도와 충성도가 높은 편이다.

그도 그럴 것이 다른 제품과 비교하면 일반적으로 부족하지 않은 기능과 무난한 성능을 가지고 있고, 또 하나 중요한 것은 새로운 개발자를 끌어 들이기에 좋은 입문 조건들을 가지고 있다는 것이다.

프로젝트 페이지에 잘 정리된 매뉴얼, 문서와 300종 이상의 관련 서적, 그리고 세계 각국에서 개최되는 매년 10회 이상의 다양한 세미나와 컨퍼런스 등도 이를 뒷받침하고 있으며 최근에는 전문 잡지까지 등장했다. 이는 모두 왕성한 커뮤니티 활동의 산물이라 할 수 있다.

사용자가 꼽는 PostgreSQL의 대표적인 특징으로는 다음과 같은 것들이 있다.

- '신뢰도'는 제품의 최우선 사항
- ACID 및 트랜잭션 지원
- 다양한 인덱싱 기법 지원
- 유연한 Full-text search 기능
- 동시성 성능을 높여주는 MVCC 기능
- 다양하고 유연한 복제 방식 지원
- 다양한 프로시저(PL/pgSQL, Perl, Python, Ruby, TCL, 등)/인터페이스(JDBC, ODBC, C/C++, .Net, Perl, Python, 등) 언어
- 질 좋은 커뮤니티 지원 및 상업적인 지원
- 잘 만든 문서 및 충분한 매뉴얼 제공

그리고, 다양한 확장 기능과 확장 기능 개발의 용이성 등이 있으며 PostgreSQL만의 차별화된 확장 기능으로 다음과 같은 것들이 있다.

- GIS add-on 지원 (PostGIS)
- Key-Value 스토어 확장 기능 (HStore)
- DBLink 기능
- Crypto, UUID 등 다양한 함수, 타입 지원

이외에도 실용적이거나 실험적인 많은 확장 기능들이 있다.

이중 최근 들어 많이 회자되는 GIS(Geographic Information System) 기능에 대해 간략히 살펴보자. PostGIS(<http://postgis.refractory.net>) 는 PostgreSQL에 OpenGIS 규격 (<http://www.opengeospatial.org/standards/sfs>) 을 준수하며 Geographic object를 지원 가능하게 하는 미들웨어 형태의 확장 기능이다(그림 11). 2001년부터 개발됐으며 많은 기능과 성능 개선을 통해 오픈 소스로는 가장 많은 사용자를 확보하고 있다. 상용 제품으로는 Oracle Spatial, DB2, MS SQL Server 등도 있으나 비용 대 성능 측면에서 상용 제품은 그다지 환영 받지 못하고 있다.

게다가 PostGIS/PostgreSQL의 제공 기능이나 성능은 Oracle과 비견할만하다는 벤치마크 자료를 어렵지 않게 구할 수 있다.

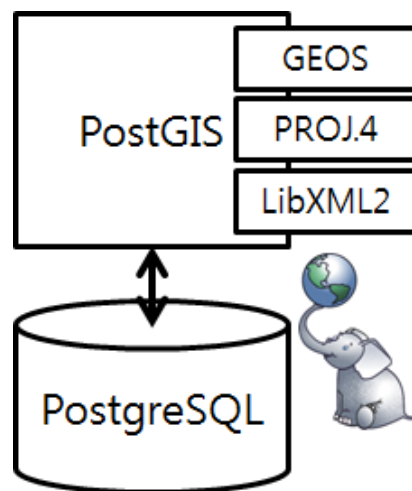


그림 11 PostGIS 구조

그리고 최근 동향을 살펴보면 GIS 분야 외에도 클라우드와 관련해서도 PostgreSQL이 많이 회자되고 있다. 최근 DbaaS(DB As A Service)를 제공하는 기업이 늘어나면서 비용과 라이선스 측면에서 유리한 PostgreSQL에 대한 수요가 증가했고 EnterpriseDB 사는 이에 맞추어 클라우드 시장을 겨냥하여 다음의 특징을 갖는 Postgres Plus Cloud Database 제품을 출시했다.

- Simple setup & web-based management
- Automatic scaling, load balancing and failover
- Automated online backup
- Database Cloning

이는 Amazon EC2, Eucalyptus cloud, Red Hat Openshift development platform cloud 등에서 사용되며 다른 Heroku, dotCloud 등의 많은 클라우드 서비스 업체에서도 PostgreSQL을 사용하는 서비스를 제공하고 있다.