PostgreSQL Universal Database

Oleg Bartunov

Moscow University,
Postgres Professional

Dec 5, 2017, PGCONF.ASIA, Tokyo

When I started using Postgres

- No UTF-8, even no 8-bit
- No WAL
- No MVCC
- No replication
- No usable non-scalar data types
- No subselects, no window functions, no CTE
- It was Postgres95

Oleg Bartunov, major PostgreSQL contributor since Postgres95

PROFESSIONAL
Postgres

How to choose a right database ?

- People usually choose a database looking on
  - Functionality, Performance
  - Availability - License, price
  - Local expertise, Personal experience
  - Compatibility to existing environment
  - Support
  - Other
- People start to think about one important feature of database after the project needs
  - Need new functionality, Better performance
- Project is in production, no way to change database
  - Starting to use various ugly «solutions»
  - System works, but looks pretty strange

Your system with wrong database

PostgreSQL Universal Database

- Any project could start with PostgreSQL
- PostgreSQL is a reliable and stable database with rich functionality and long history
- PostgreSQL has liberal BSD license, cross platform (~30)
- Developed by international community, no vendor lock

- PostgreSQL is **EXTENSIBLE**, this is the very important feature, which people miss ! It allow database to support
  - New workloads
  - New functionality
  - New environment
  - Often without restarting a server, no need core programmer.

Extendibility of PostgreSQL

"The main design goals of the new system are to:
2) provide user extendibility for data types, operators and access methods,"

Stonebraker M., Rowe L. A. The design of Postgres. – ACM, 1986. – T. 15. – №. 2. – C. 340-355.

"It is imperative that a user be able to construct new access methods to provide efficient access to instances of nontraditional base types"

Michael Stonebraker, Jeff Anton, Michael Hirohama.

Extendability in POSTGRES , IEEE Data Eng. Bull. 10 (2) pp.16-23, 1987

Extendibility of PostgreSQL

0. Ability to change executor and planner

1. Data types, operators, access methods, indexes without core developers, deploy online !

2. FDW, procedural languages

Need more:

1. API to pluggable storages
2. API to TM — XTM, eXtensible transaction manager
3. Type-specific compression

….more…

Case 1: Web

1996: Start using Postgres on Web, no 8-bit support — introduced locale

1999: World"s top-5 portal. We start with PostgreSQL 6.5.?

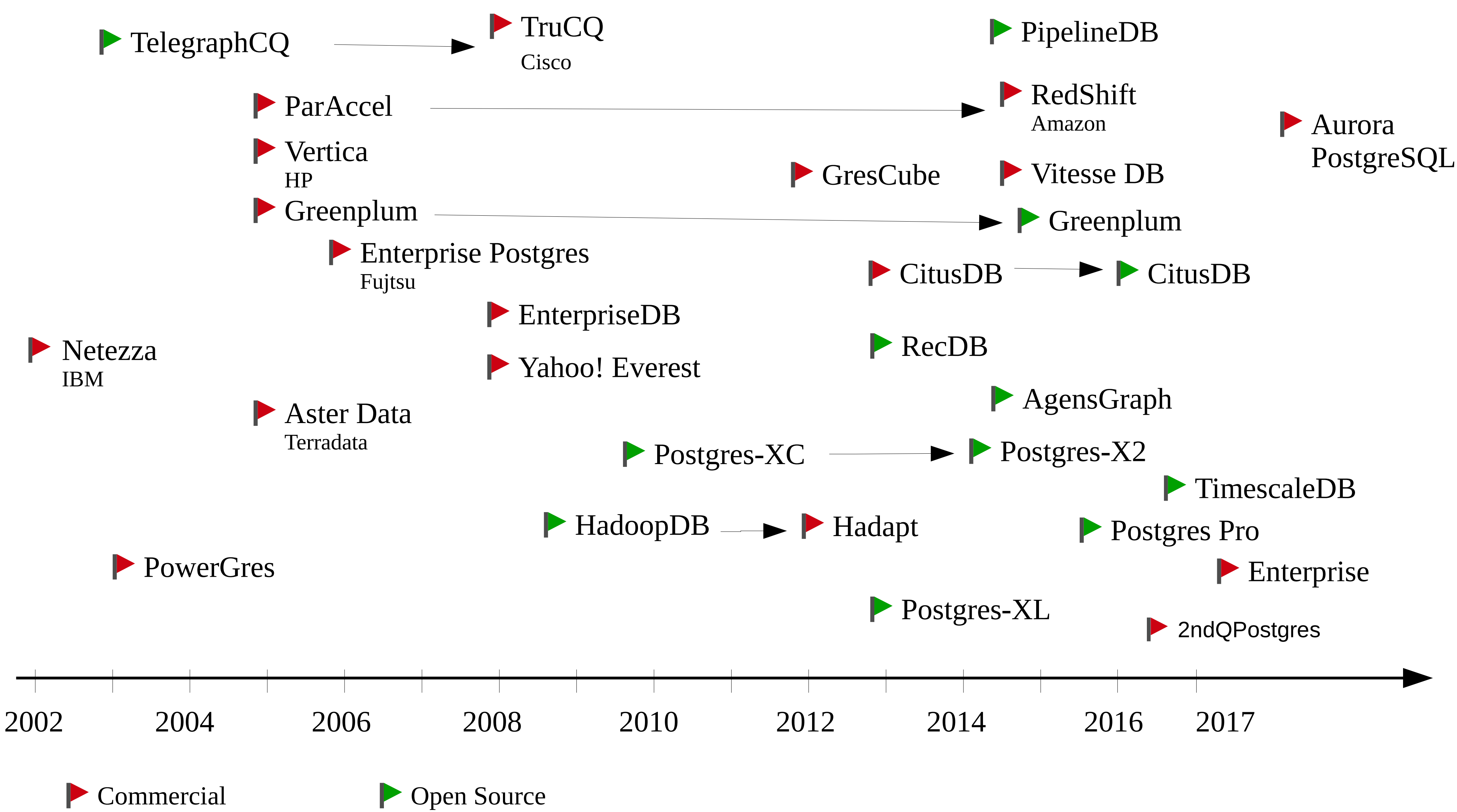
Requirement: on the hardware ~ my smartphone to support > 1 mln. users/day

- Quickly run out of resources
- Denormalize, use arrays -> slow -> discover GiST → improve GiST - intarray with indexes
- Need FTS, made tsearch using intarray and GiST indexes
- Need fast search on hierarchical data — ltree — GiST indexes
- Need flexible schema — hstore — GiST index
- Need faster FTS — GIN index for tsearch, hstore
- Need misprint search — pg_trgm — GiST/GIN indexes
- Compete to NoSQL - better/binary json - jsonb — GIN index
- Need faster FTS — RUM access methods

2017: STILL USING PostgreSQL !

Extensibility makes PostgreSQL Universal Database !

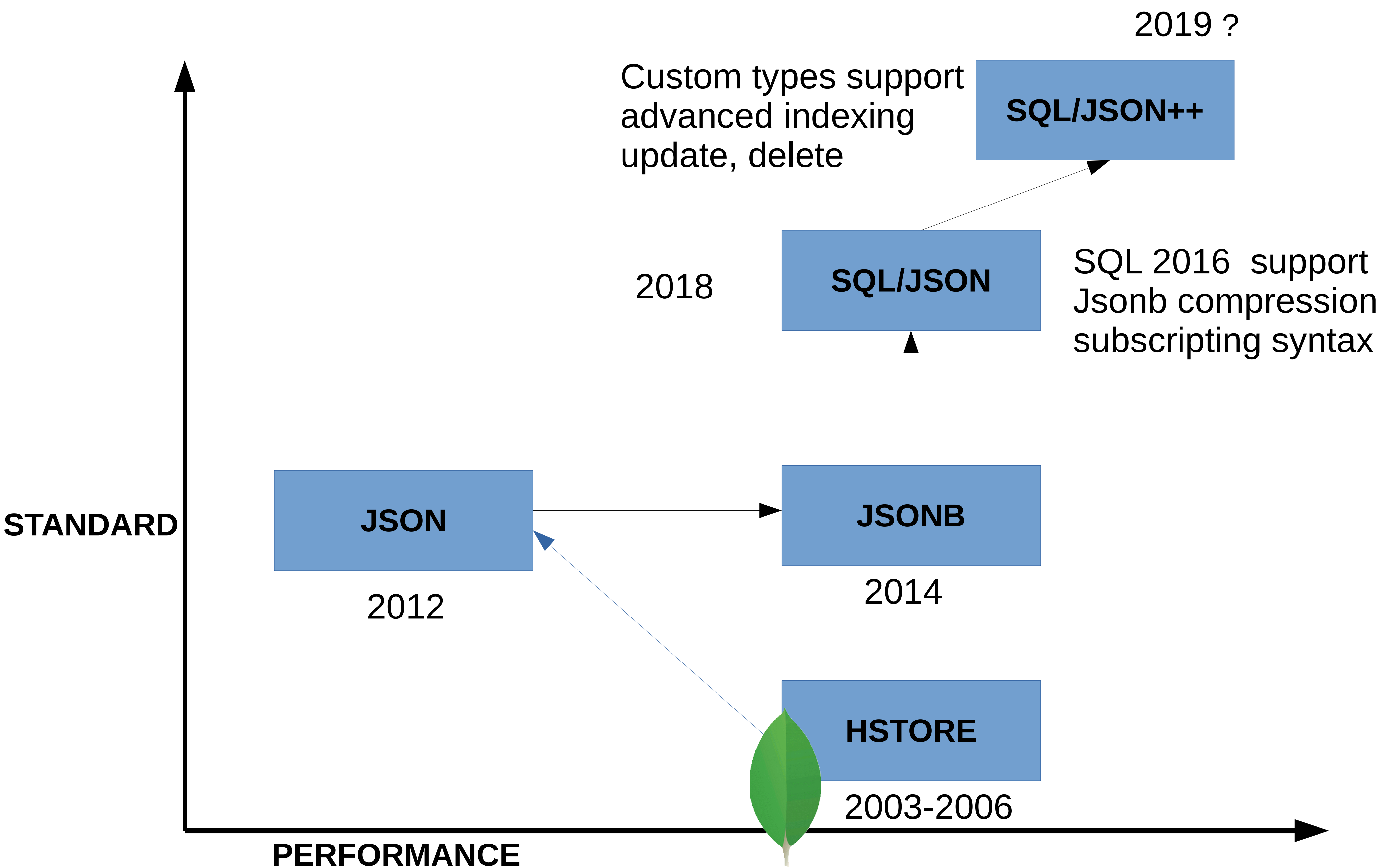
PostgreSQL Forks (we love forks!)



PostgreSQL Forks (we love forks!)

- PostgreSQL - OLTP, Web, Science
  extensions: GIS (PostGIS), time-series (TimeScaleDB), Stream data (PipelineDB), interactive analytics (Citus), vector ops, async mpp on GPU (PG-Strom)
- AgensGraph — Graph model, graph query language
- Postgres Pro Enterpise - Multimaster, block compression, 64-bit TX, partitioning, adaptive planning, scheduler, incremental backup
- Redshift — Cloud, column storage, compression, OLAP
- Greenplum - MPP, OLAP (complex reports)
- Postgres-XL — MPP OLAP (ACID)
- Vitesse DB — OLAP (JIT, column storage, threads)
- Aurora PostgreSQL — OLTP, S3 storage
- EDB Postgres Advanced Server - Oracle compatibility

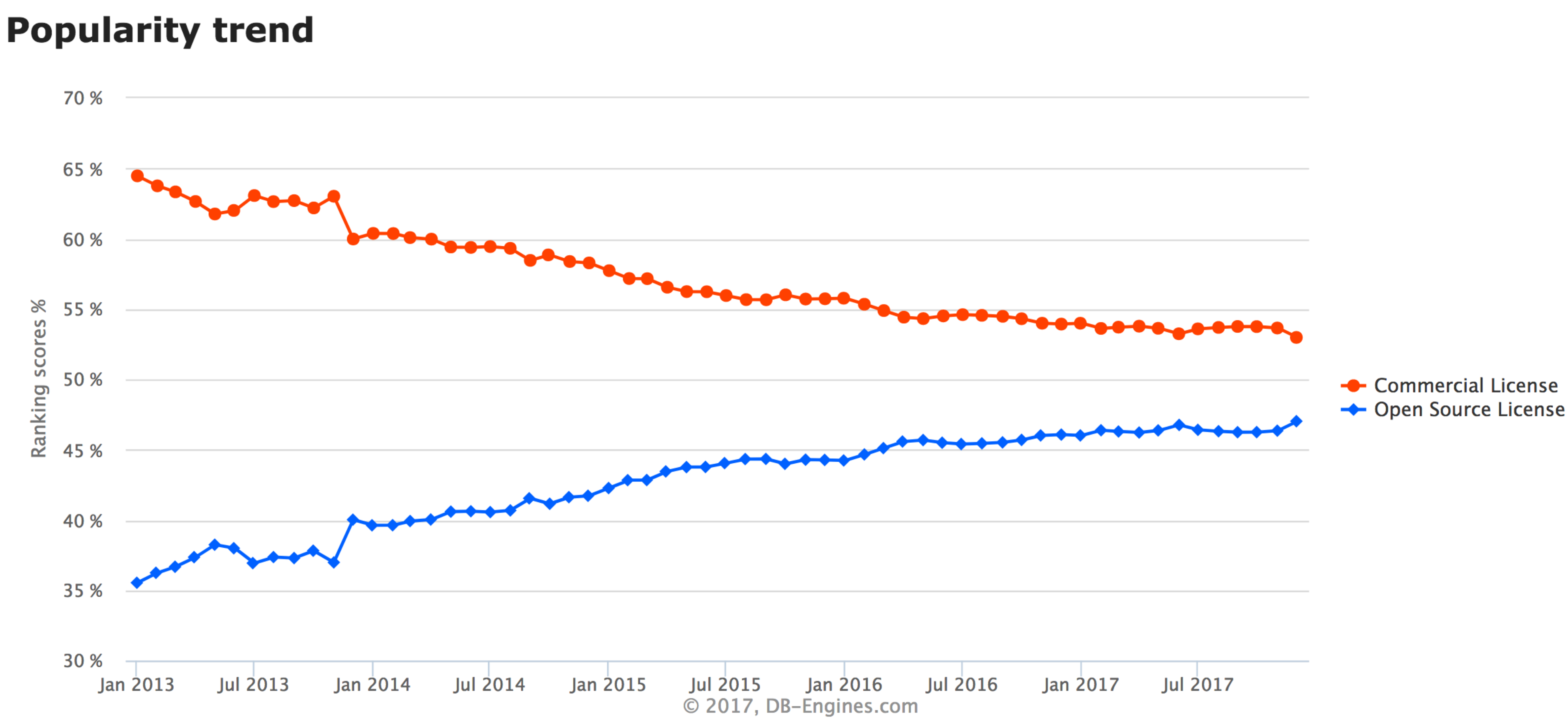
Case 2: NoSQL Postgres



PostgreSQL users

+BIG RUSSIAN Enterprise !

Open source vs commercial DBMS



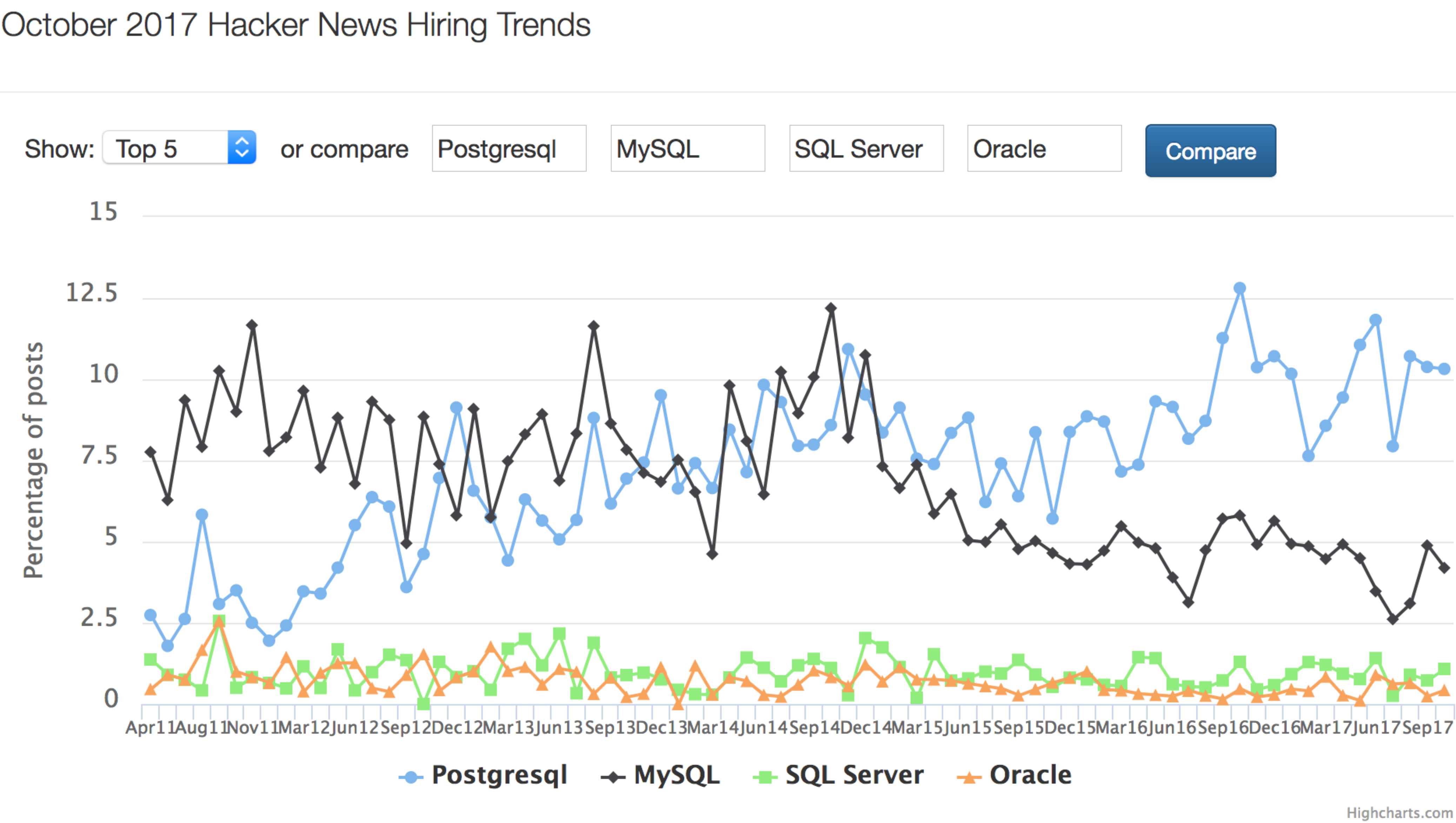
PostgreSQL is #4 !

337 systems in ranking, November 2017

Rank Nov 2017	Rank Oct 2017	Rank Nov 2016	DBMS	Database Model	Score Nov 2017	Score Oct 2017	Score Nov 2016
1.	1.	1.	Oracle	Relational DBMS	1360.05	+11.25	-52.96
2.	2.	2.	MySQL	Relational DBMS	1322.03	+23.20	-51.53
3.	3.	3.	Microsoft SQL Server	Relational DBMS	1215.08	+4.76	+1.27
4.	4.	4.	PostgreSQL	Relational DBMS	379.92	+6.64	+54.10
5.	5.	5.	MongoDB	Document store	330.47	+1.07	+5.00
6.	6.	6.	DB2	Relational DBMS	194.06	-0.53	+12.61
7.	7.	↑ 8.	Microsoft Access	Relational DBMS	133.31	+3.86	+7.34
8.	8.	↓ 7.	Cassandra	Wide column store	124.21	-0.58	-9.76
9.	9.	9.	Redis	Key-value store	121.18	-0.87	+5.64
10.	10.	↑ 11.	Elasticsearch	Search engine	119.41	-0.82	+16.84

https://db-engines.com/en/ranking

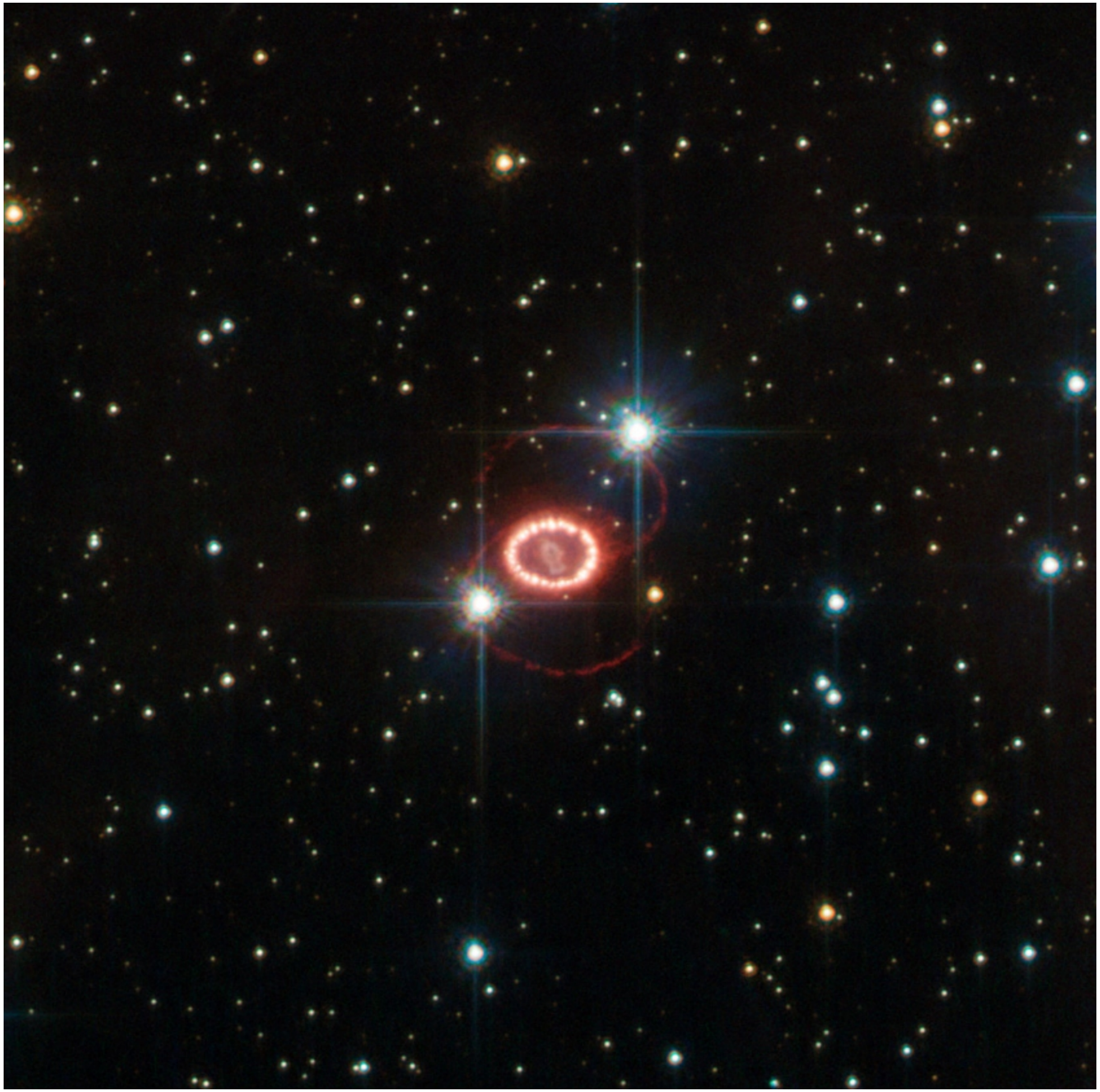
Hacker News Hiring Trends



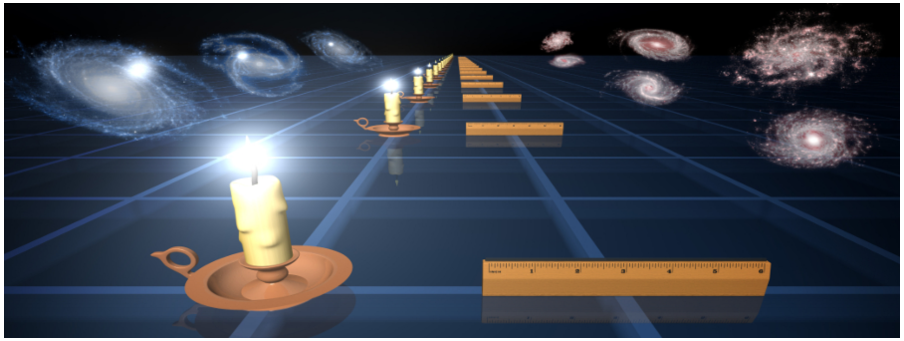
PostgreSQL in Russia is #1

Case 3: PostgreSQL and Supernovae

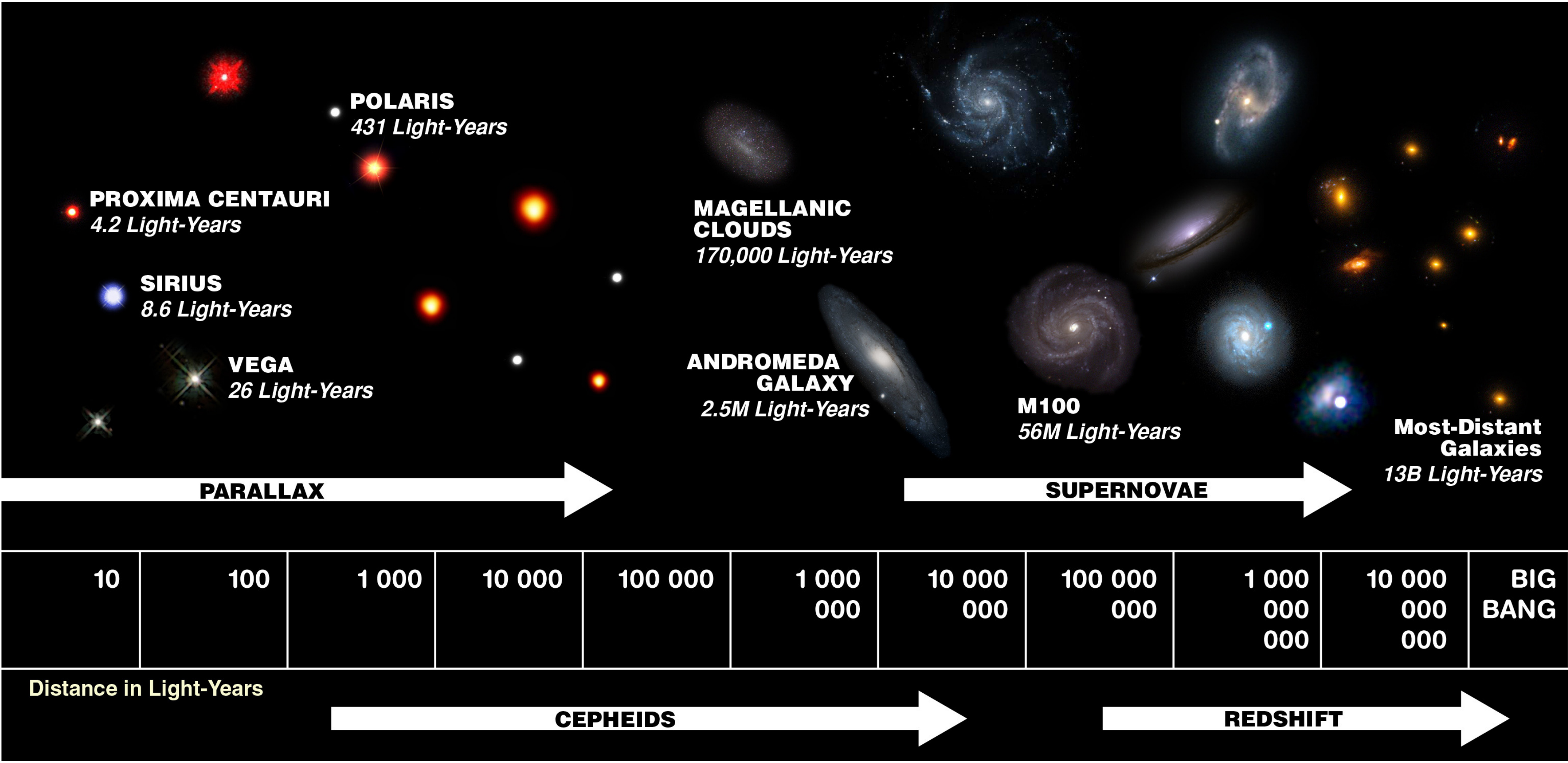
SN 1987a, Type II, +2.9, Tarantula Nebulae in LMC,
168 000 ly, progenitor: Sanduleak -69° 202,



The Scale of the Universe



- Supernovae(Ia) - «standard candles»
- Used to measure the distance to the host galaxy



What is the fate of the Universe ?

WRITTEN IN THE STARS

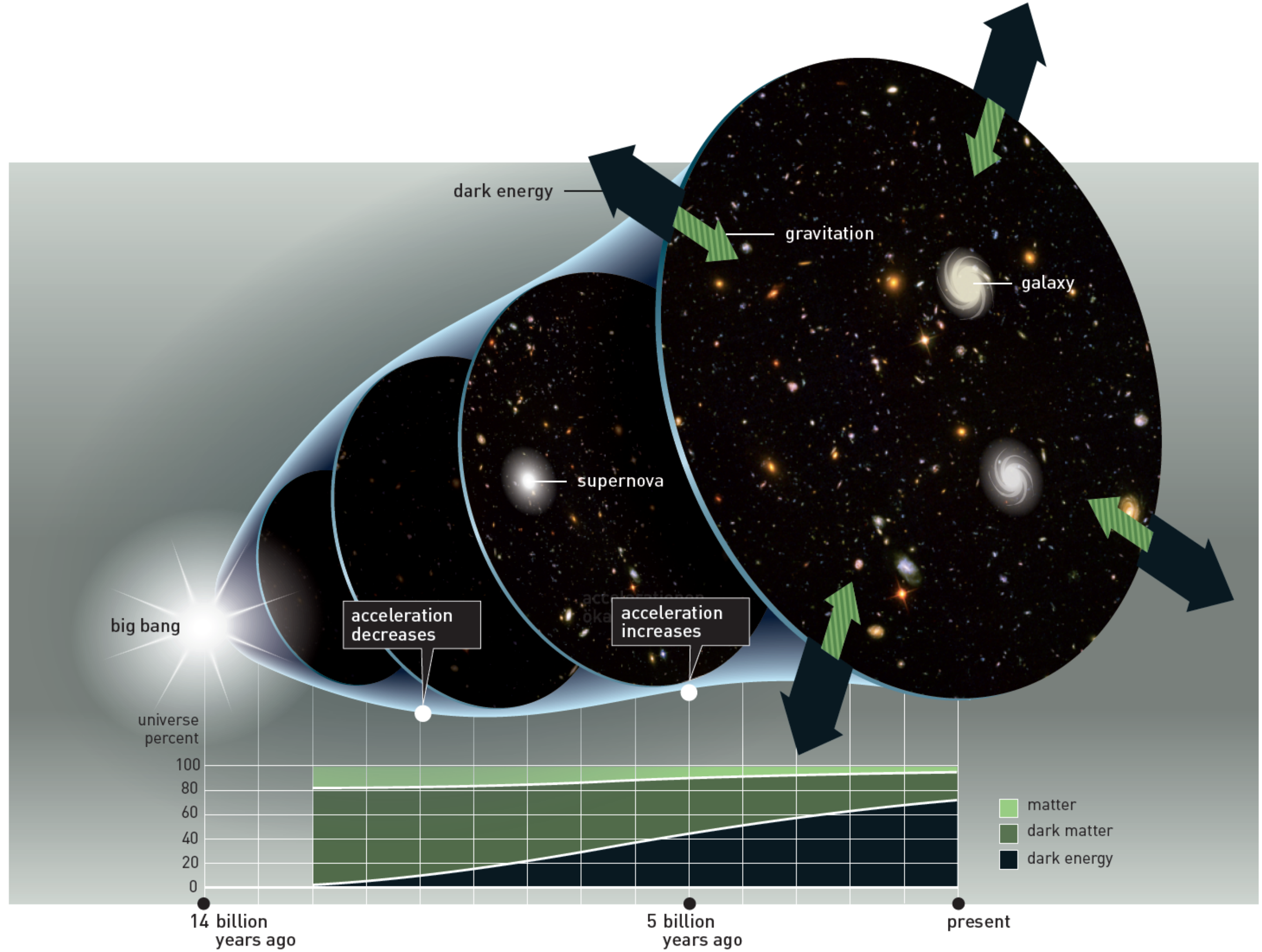


Figure 1. The world is growing. The expansion of the Universe began with the Big Bang 14 billion years ago, but slowed down during the first several billion years. Eventually it started to accelerate. The acceleration is believed to be driven by dark energy, which in the beginning constituted only a small part of the Universe. But as matter got diluted by the expansion, the dark energy became more dominant.

Oleg Bartunov

Edit

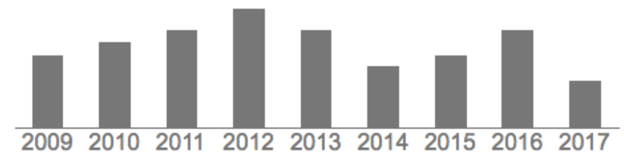
Follow

Lomonosov Moscow State University
astrophysics, databases, astroinformatics
Verified email at sai.msu.su - Homepage
My profile is public

Change photo

Google Scholar

Citation indices	All	Since 2012
Citations	1408	362
h-index	13	9
i10-index	17	9



Add co-authors

Sergei B. Popov

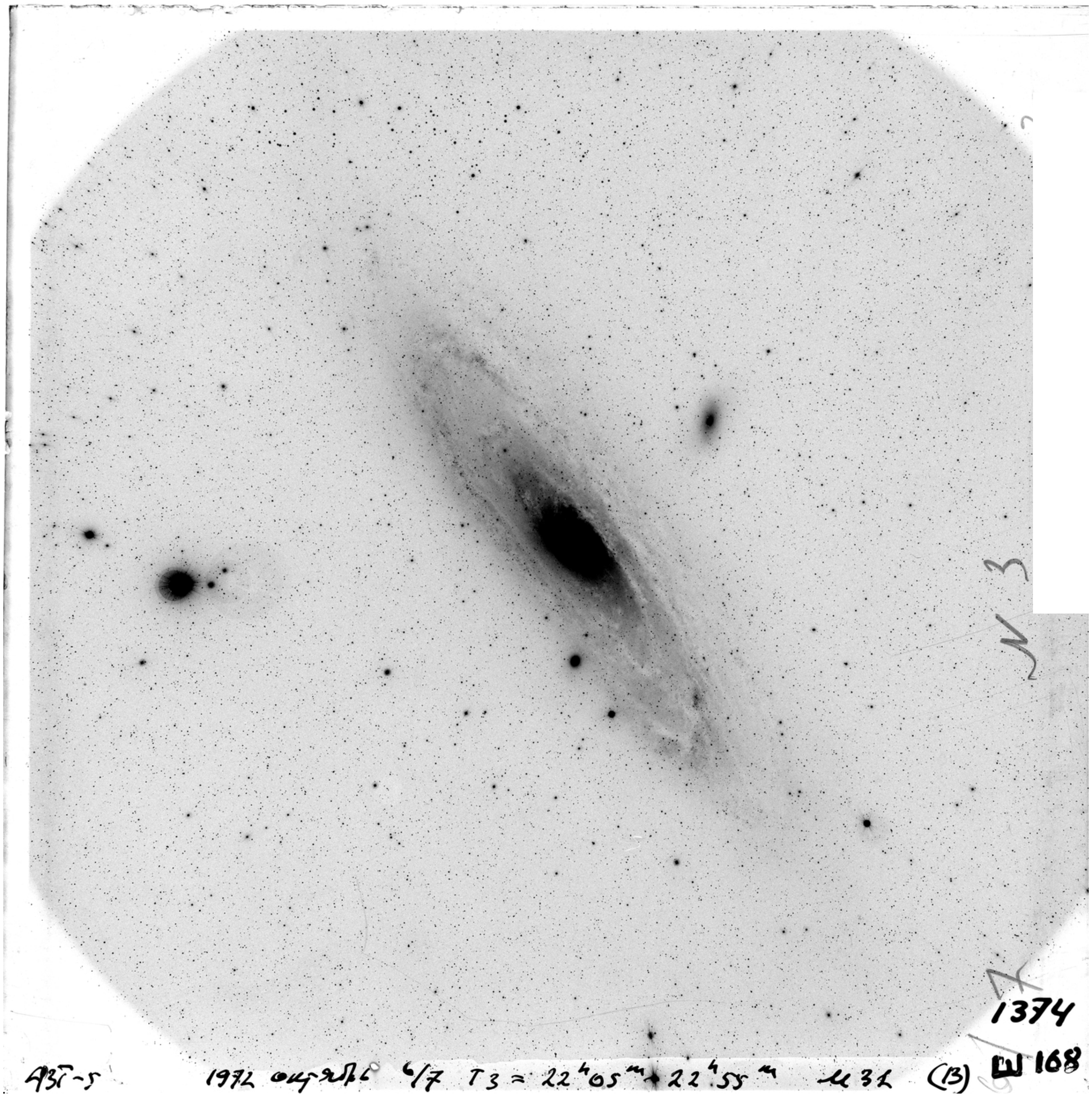
Co-authors Edit...

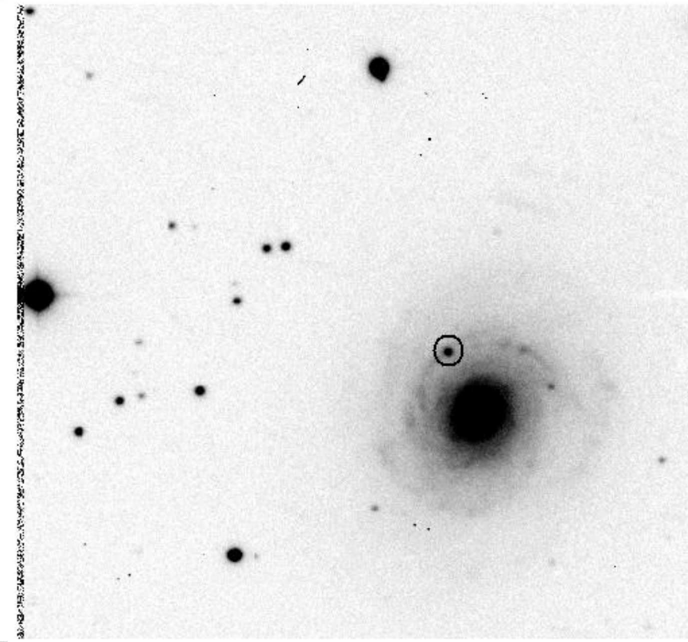
- Massimo Turatto
- Mario Hamuy
- Sergey Koposov
- Stefano Benetti
- Sergey Karpov
- Igor Chilingarian

Title | Add | More | 1–20 | Cited by | Year

Title	Cited by	Year
The rate of supernovae from the combined sample of five searches E Cappellaro, M Turatto, DY Tsvetkov, OS Bartunov, C Pollas, R Evans, ... arXiv preprint astro-ph/9611191	406	1996
Radiation hydrodynamics of SN 1987A. I. Global analysis of the light curve for the first 4 months S Blinnikov, P Lundqvist, O Bartunov, K Nomoto, K Iwamoto The Astrophysical Journal 532 (2), 1132	179	2000
A comparative modeling of supernova 1993J SI Blinnikov, R Eastman, OS Bartunov, VA Popolitov, SE Woosley The Astrophysical Journal 496 (1), 454	159	1998
The rate of supernovae. II. The selection effects and the frequencies per unit blue luminosity E Cappellaro, M Turatto, DY Tsvetkov, OS Bartunov, IN Makarova arXiv preprint astro-ph/9302017	118	1993
Non-equilibrium radiative transfer in supernova theory-models of linear Type-II supernovae SI Blinnikov, OS Bartunov Astronomy and Astrophysics 273, 106	102	1993
Distribution of supernovae relative to spiral arms and H II regions OS Bartunov, DY Tsvetkov, IV Filimonova Publications of the Astronomical Society of the Pacific 106 (706), 1276	100	1994

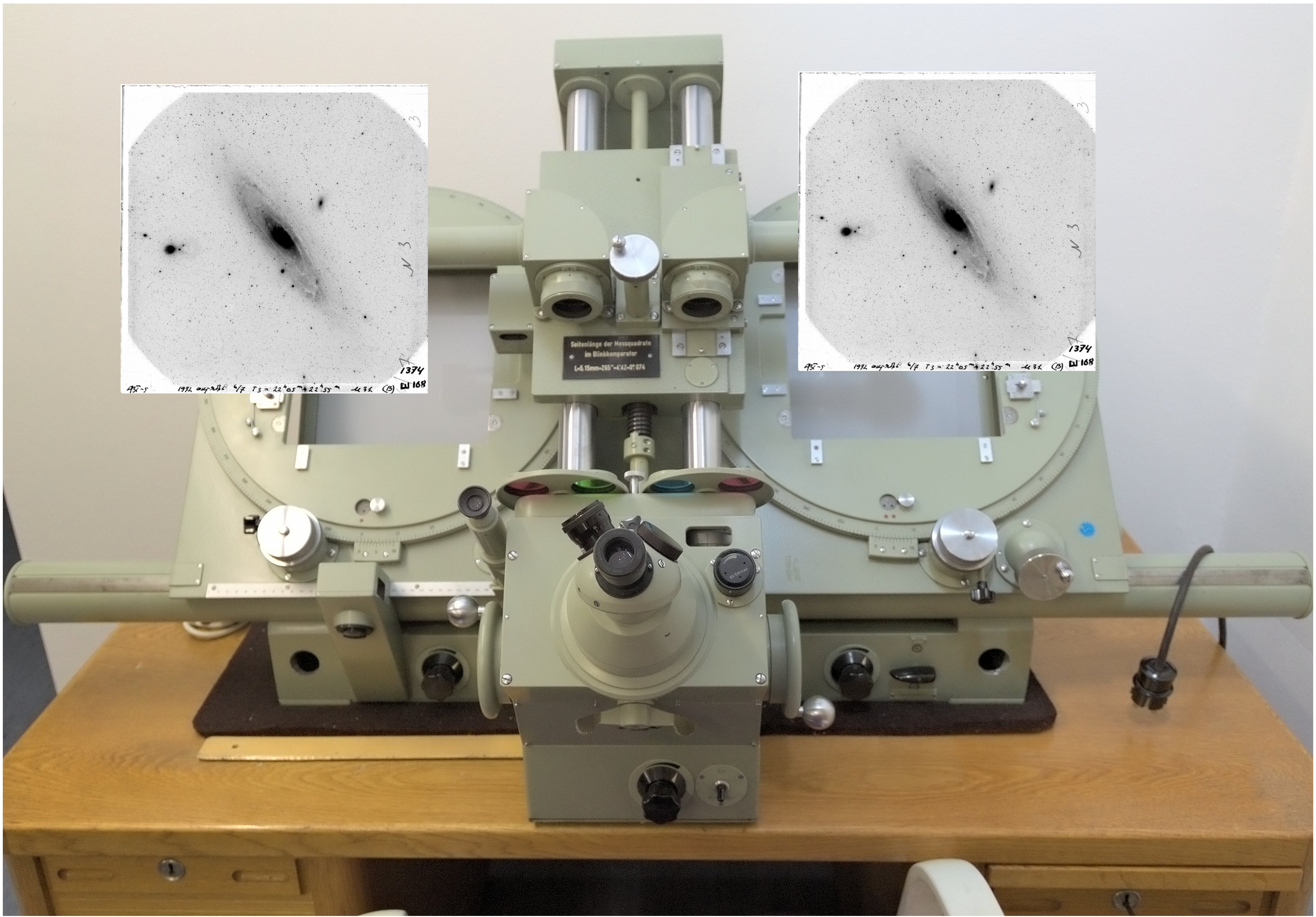
M31 (Andromeda), AZT-5 telescope



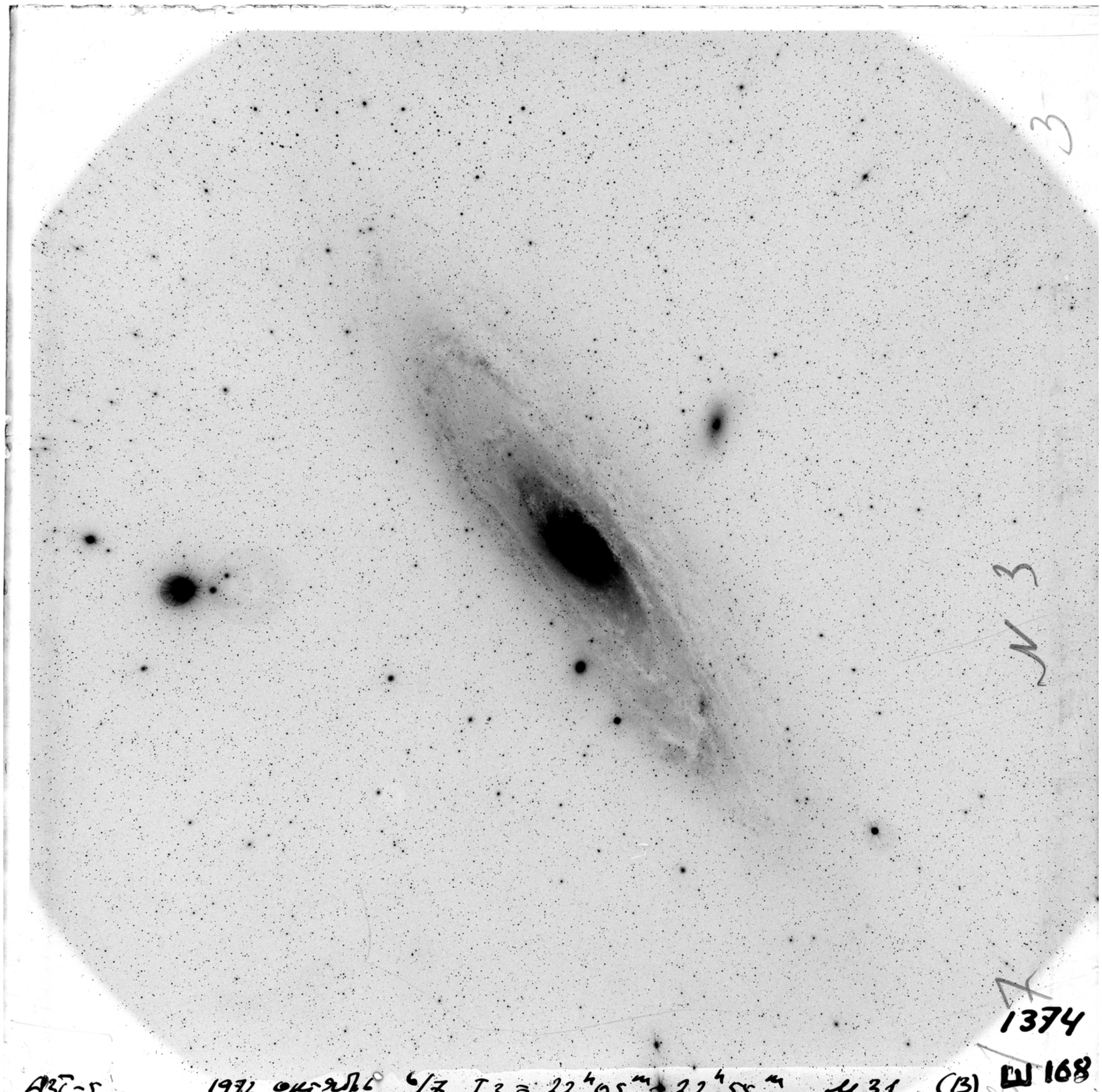


SN 2008fv in NGC 3147, Draco
Dmitry Tsvetkov, SAI MSU

Blink Comparator (manual discovery)
Many hours of hard work !



M31 (Andromeda), AZT-5 telescope



x-00001	1.1458447	-89.9186147
x-00002	1.3300139	-89.9332336
x-00003	3.2556022	-89.9641031
x-00004	3.6464625	-89.9060142
x-00005	6.3110253	-89.9523947
x-00006	6.6275517	-89.9279197
x-00007	7.8266025	-89.9129272
x-00008	9.0694378	-89.9714031
x-00009	9.6627953	-89.9244314
x-00010	10.0494292	-89.9705058
x-00011	10.4863922	-89.9699058
x-00012	11.0953692	-89.9016031
x-00013	11.3240233	-89.9344336
x-00014	11.7906064	-89.9070308
x-00015	12.0416581	-89.9300586
x-00016	12.0522308	-89.9002281
x-00017	12.2808536	-89.9107669
x-00018	13.0316142	-89.9214558
x-00019	13.8727033	-89.9577031
x-00020	14.6546639	-89.9191919
x-00021	18.3035981	-89.9447475
x-00022	18.5185631	-89.9446836
x-00023	19.8675597	-89.9836308
x-00024	20.9699533	-89.9226864
x-00025	21.6777744	-89.9256808
x-00026	23.3660669	-89.9036558
x-00027	24.2841308	-89.9516475
x-00028	24.3273161	-89.9202392
x-00029	24.5540458	-89.9246003
x-00030	24.5655172	-89.9122336
x-00031	26.3487519	-89.9460336
x-00032	26.5268008	-89.9311503
x-00033	26.6070808	-89.9271808
x-00034	27.4104919	-89.9768558
x-00035	27.8290442	-89.9304622
x-00036	28.5552036	-89.9199117
x-00037	29.4407347	-89.9762836
x-00038	30.5729608	-89.9377753
x-00039	30.7101131	-89.9105642
x-00040	33.2918250	-89.9106614
x-00041	33.4843678	-89.9442058

Spatial Join (Machine Discovery, < 1s)

Observations: 10^5

```
x-00001|1.1458447|-89.9186147
x-00002|1.3300139|-89.9332336
x-00003|3.2556022|-89.9641031
x-00004|3.6464625|-89.9060142
x-00005|6.3110253|-89.9523947
x-00006|6.6275517|-89.9279197
x-00007|7.8266025|-89.9129272
x-00008|9.0694378|-89.9714031
x-00009|9.6627953|-89.9244314
x-00010|10.0494292|-89.9705058
x-00011|10.4863922|-89.9699058
x-00012|11.0953692|-89.9016031
x-00013|11.3240233|-89.9344336
x-00014|11.7906064|-89.9070308
x-00015|12.0416581|-89.9300586
x-00016|12.0522308|-89.9002281
x-00017|12.2808536|-89.9107669
x-00018|13.0316142|-89.9214558
x-00019|13.8727033|-89.9577031
x-00020|14.6546639|-89.9191919
x-00021|18.3035981|-89.9447475
x-00022|18.5185631|-89.9446836
x-00023|19.8675597|-89.9836308
x-00024|20.9699533|-89.9226864
x-00025|21.6777744|-89.9256808
x-00026|23.3660669|-89.9036558
x-00027|24.2841308|-89.9516475
x-00028|24.3273161|-89.9202392
x-00029|24.5540458|-89.9246003
x-00030|24.5655172|-89.9122336
x-00031|26.3487519|-89.9460336
x-00032|26.5268008|-89.9311503
x-00033|26.6070808|-89.9271808
x-00034|27.4104919|-89.9768558
x-00035|27.8290442|-89.9304622
x-00036|28.5552036|-89.9199117
x-00037|29.4407347|-89.9762836
x-00038|30.5729608|-89.9377753
```

$\bowtie\theta$

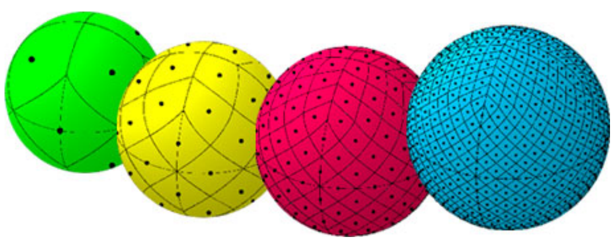
Catalog(s): 10^9

```
t-0000001|1.1458447|-89.9186147|0.015|0.028
t-0000002|1.3300139|-89.9332336|0.050|0.110
t-0000003|3.2556022|-89.9641031|0.050|0.050
t-0000004|3.6464625|-89.9060142|0.204|0.224
t-0000005|6.3110253|-89.9523947|0.114|0.050
t-0000006|6.6275517|-89.9279197|0.098|0.150
t-0000007|7.8266025|-89.9129272|0.025|0.021
t-0000008|9.0694378|-89.9714031|0.200|0.200
t-0000009|9.6627953|-89.9244314|0.000|0.000
t-0000010|10.0494292|-89.9705058|0.050|0.228
t-0000011|10.4863922|-89.9699058|0.200|0.200
t-0000012|11.0953692|-89.9016031|0.050|0.259
t-0000013|11.3240233|-89.9344336|0.050|0.050
t-0000014|11.7906064|-89.9070308|0.159|0.131
t-0000015|12.0416581|-89.9300586|0.216|0.050
t-0000016|12.0522308|-89.9002281|0.050|0.050
t-0000017|12.2808536|-89.9107669|0.050|0.050
t-0000018|13.0316142|-89.9214558|0.152|0.120
t-0000019|13.8727033|-89.9577031|0.050|0.121
t-0000020|14.6546639|-89.9191919|0.050|0.069
t-0000021|18.3035981|-89.9447475|0.139|0.440
t-0000022|18.5185631|-89.9446836|0.057|0.268
t-0000023|19.8675597|-89.9836308|0.050|0.120
t-0000024|20.9699533|-89.9226864|0.050|0.050
t-0000025|21.6777744|-89.9256808|0.055|0.105
t-0000026|23.3660669|-89.9036558|0.050|0.135
t-0000027|24.2841308|-89.9516475|0.213|0.050
t-0000028|24.3273161|-89.9202392|0.550|0.999
t-0000029|24.5540458|-89.9246003|0.160|0.086
t-0000030|24.5655172|-89.9122336|0.205|0.050
t-0000031|26.3487519|-89.9460336|0.050|0.095
t-0000032|26.5268008|-89.9311503|0.335|0.245
t-0000033|26.6070808|-89.9271808|0.050|0.075
t-0000034|27.4104919|-89.9768558|0.094|0.090
t-0000035|27.8290442|-89.9304622|0.017|0.019
t-0000036|28.5552036|-89.9199117|0.050|0.115
t-0000037|29.4407347|-89.9762836|0.635|0.265
t-0000038|30.5729608|-89.9377753|0.314|0.170
```

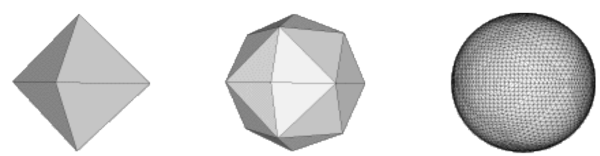
Astronomy meets database

Indexing the SKY with PostgreSQL

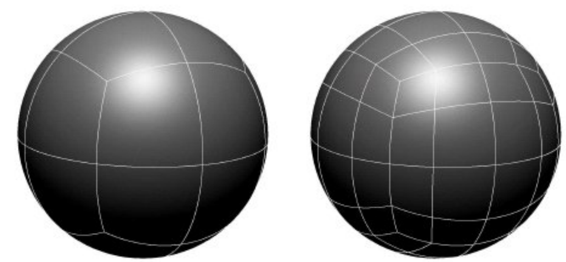
HEALPIX



HTM

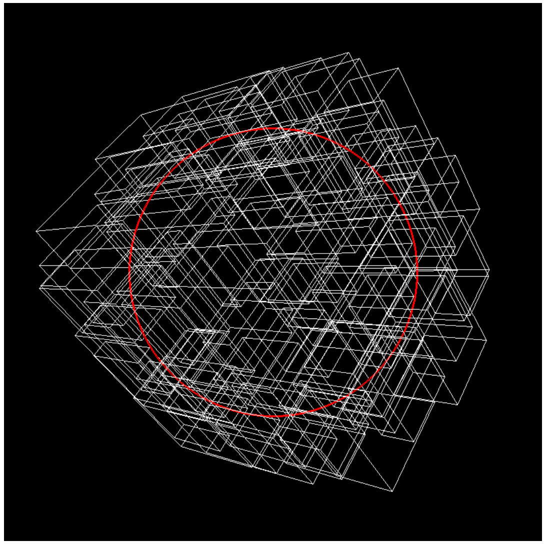


Q3C (PG)



The sphere segmentation in Q3C

PGSphere (PG)



WSDB Whole Sky DataBase

Database of astronomic catalogues in Cambridge University

~5 dbs,~ 40 users, up to ~ 10^7 queries per day, size 40Tb

pg 9.4 + q3c + hstore

Example of research:Koposov, S. E., Belokurov, V., Torrealba, G., & Evans, N. W. (2015). Beasts of the Southern Wild: Discovery of nine Ultra Faint satellites in the vicinity of the Magellanic Clouds. The Astrophysical Journal, 805(2), 130.

Gaia Alerts Database

Real time Detection of alerts in the Gaia

~10 dbs, 10 users, up to ~ 10^6 queries per day, size 30Tb

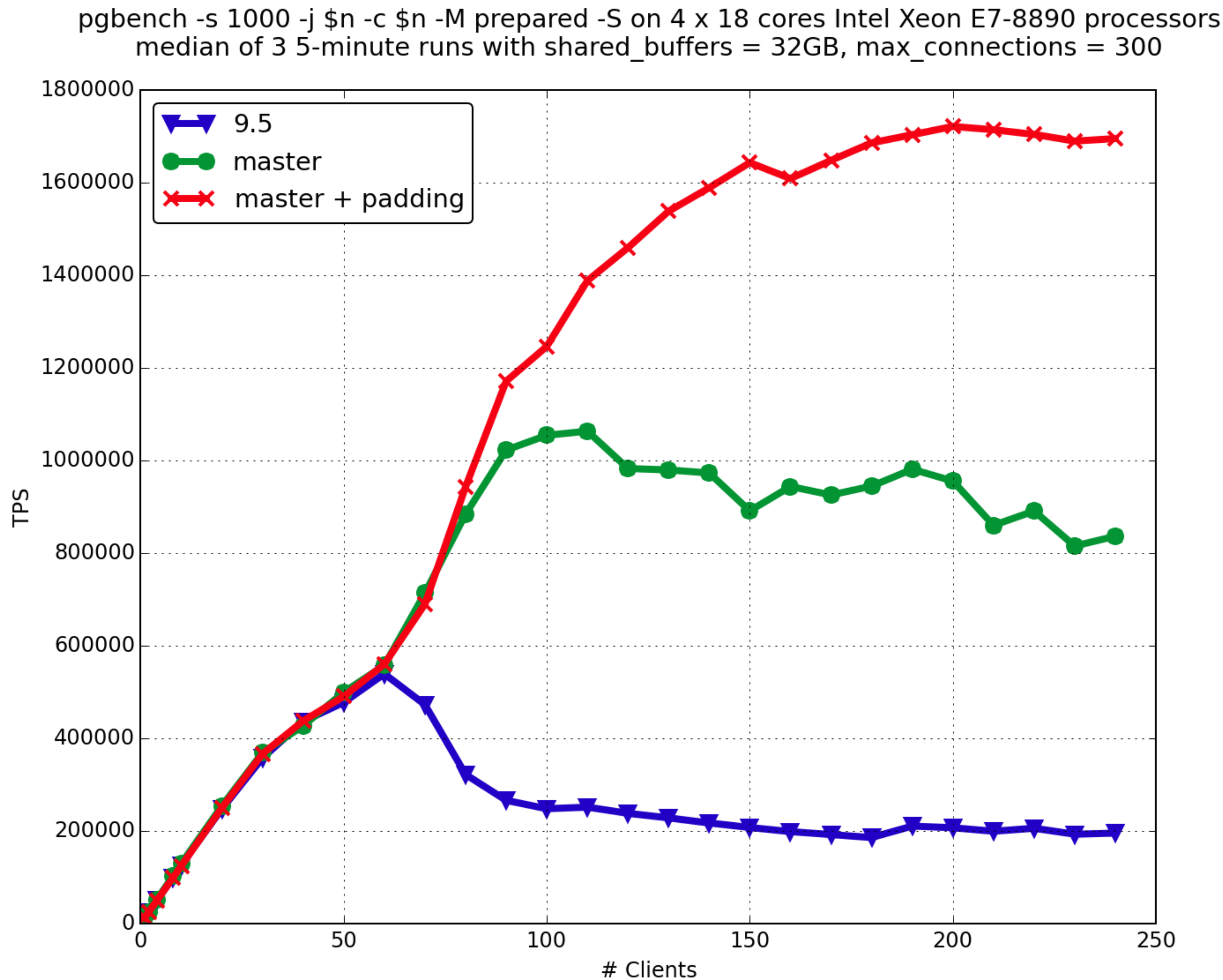
pg 9.3 + synchronous replication + q3c

Example of research: Campbell, H. C., Marsh, T. R., Fraser, M., Hodgkin, S. T., de Miguel, E., Gänsicke, B. T., ... & Koposov, S. E. (2015). Total eclipse of the heart: the AM CVn Gaia14aae/ASSASN-14cn. Monthly Notices of the Royal Astronomical Society, 452(1), 1060-1067.

MASTER database

- Robotic network telescopes by SAI MSU
- 8 observatories (5 in Russia, 3 outside)
- total size ~100TB
- pg 9.0-9.4 + pgsphere + replication
- See:
  - Lipunov, Vladimir, et al. "Master robotic net." Advances in Astronomy 2010 (2010).
  - Kornilov, Victor G., et al. "Robotic optical telescopes global network MASTER II. Equipment, structure, algorithms." Experimental Astronomy 33.1 (2012): 173-196.

PostgreSQL scalability on 4x18 x86



~~When I started using Postgres~~ (Today)

Excellent scalablity on 4-socket X86 machines

Support a wide range of workloads

- High security data, relational, NoSQL, OLAP

Postgres-centric companies

- 2ndQuadrant, EDB, Postgres Professional (roadmaps)

Postgres groups

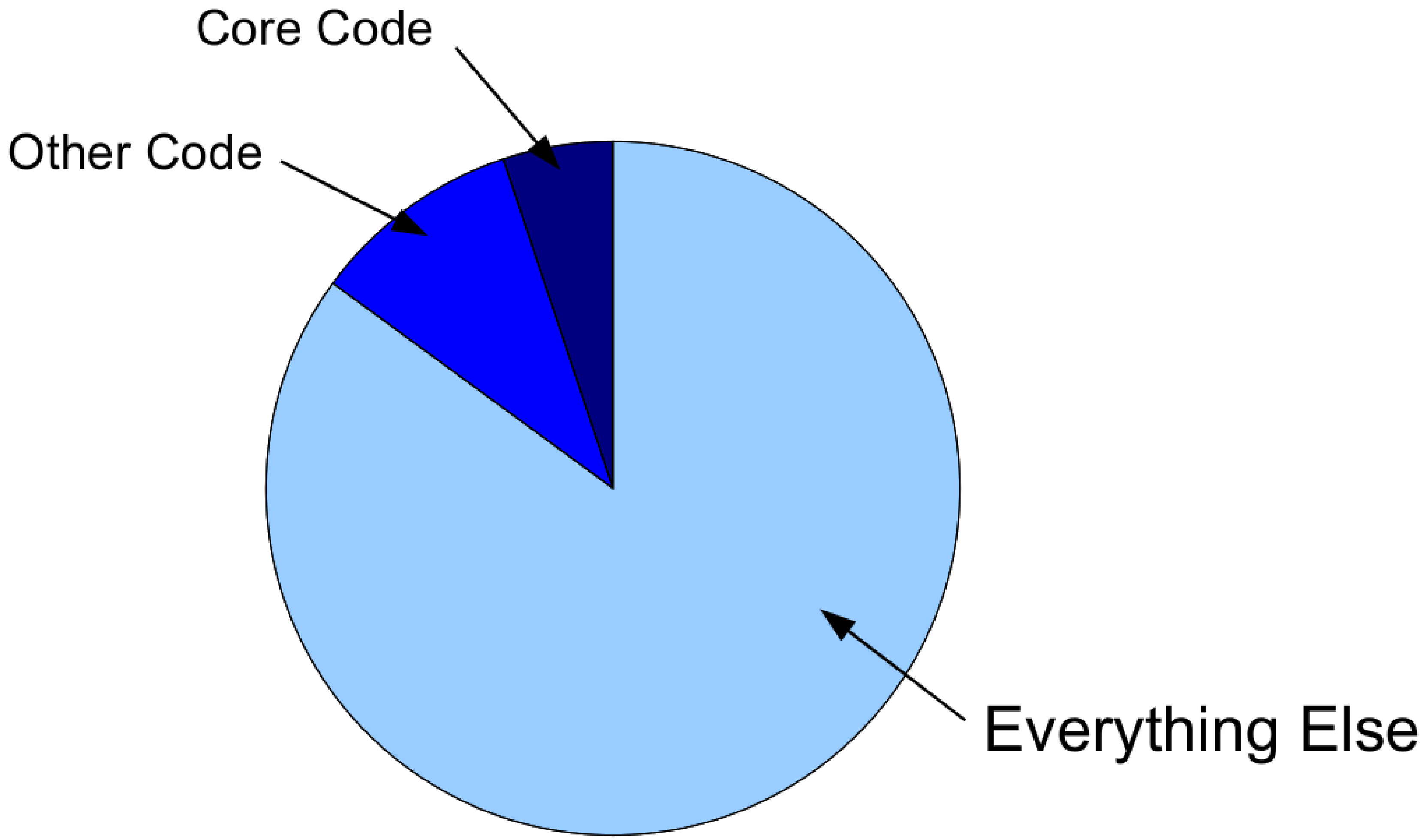
- NTT, Fujitsu, Amazon, Alibaba, Tencent,…

Postgres Development from

- Community driven → Business driven

Postgres is used in mission-critical enterprise systems

There are many ways to help Postgres



There are many ways to help Postgres

Core development
Development, review, testing, reporting bugs

Ecosystem
Extensions, drivers, ORM, monitoring tools… Postgres support in applications
Distributions, packages

Documentation
Improvement, translations, writing books, papers, ...blogging!

Meetings, Education
Creating of local communities, Conference, meetups, seminars, hackatons, educational and training courses

Use PostgreSQL!
Use Postgres in your company !

Sponsorship
Help development, sponsor community events.

For example: Improve documentation

- PG 11 — XML build !
- Why we don“t have figures yet ?!
  - Need community standards for diagrams, tools, draw diagrams
- Help beginners to learn PostgreSQL
  - Need friendly «PostgreSQL for beginners» with demo database

Figures in PostgreSQL doc.

Bookmarks

Chapter 65. GIN Indexes

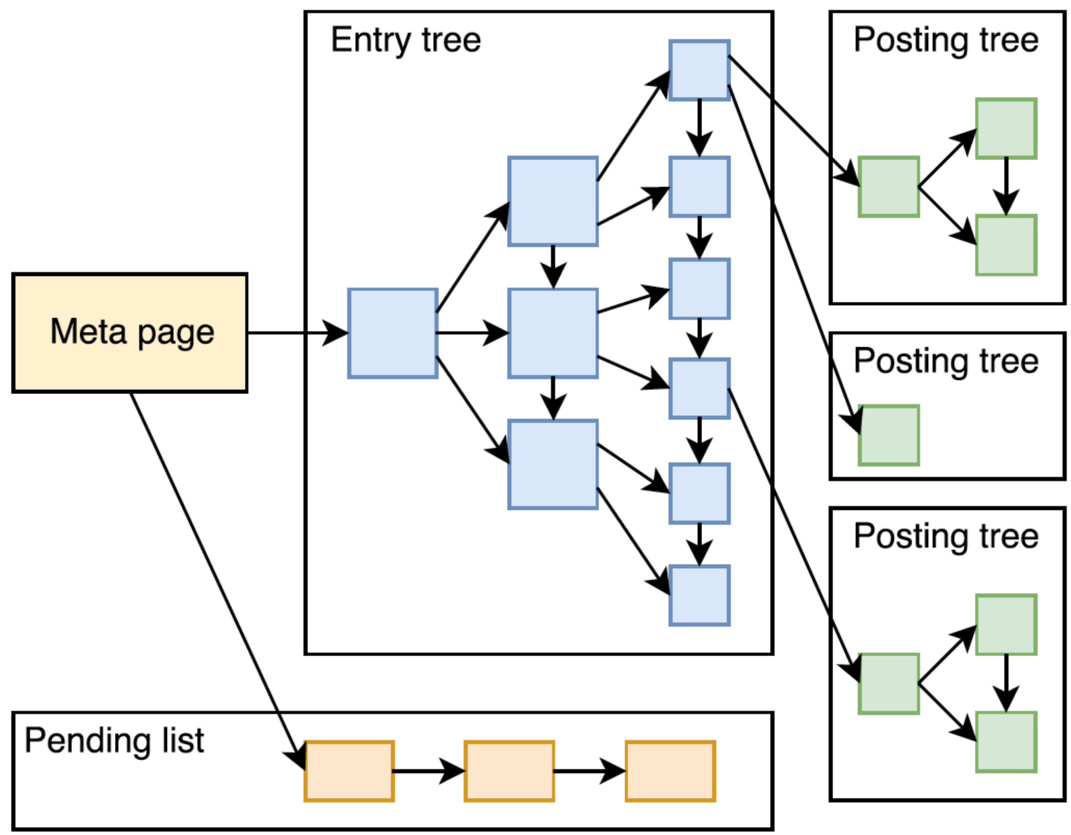
65.1. Introduction

GIN stands for Generalized Inverted Index. GIN is designed for handling cases where the items to be indexed are composite values, and the queries to be handled by the index need to search for element values that appear within the composite items. For example, the items could be documents, and the queries could be searches for documents containing specific words.

We use the word *item* to refer to a composite value that is to be indexed, and the word *key* to refer to an element value. GIN always stores and searches for keys, not item values per se.

A GIN index stores a set of (key, posting list) pairs, where a *posting list* is a set of row IDs in which the key occurs. The same row ID can appear in multiple posting lists, since an item can contain more than one key. Each key value is stored only once, so a GIN index is very compact for cases where the same key appears many times.

Figure 65.1. GIN index

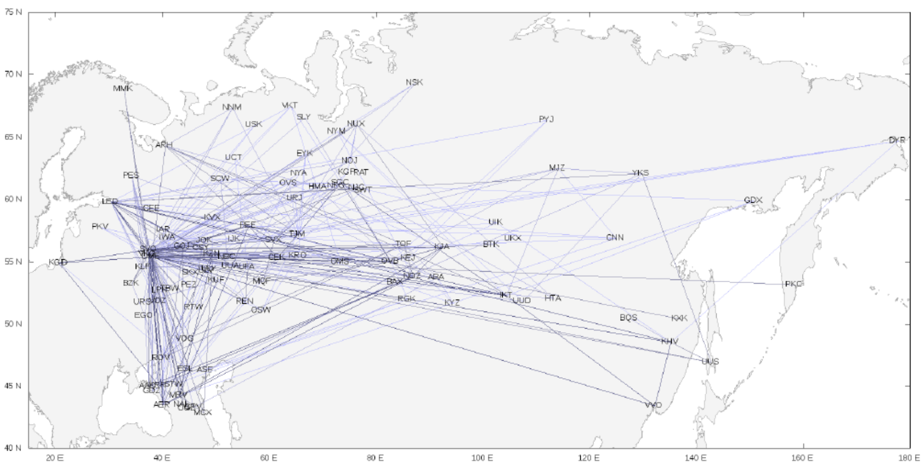


Sample Database in PostgreSQL doc

Appendix L. Demo Database “Airlines”

This is an overview of a demo database for PostgreSQL. This appendix describes the database schem which consists of eight tables and several views. The subject field of this database is airline flights Russia. You can download the database from *our website*. See Section L.1 for details.

Figure L.1. Airlines in Russia



You can use this database for various purposes, such as:

- learning SQL language on your own
- preparing books, manuals, and courses on SQL
- showing features in stories and articles

When developing this demo database, we pursued several goals:

- Database schema must be simple enough to be understood without extra explanations.
- At the same time, database schema must be complex enough to allow writing meaningful queries
- The database must contain true-to-life data that will be interesting to work with.

This demo database is distributed under the *PostgreSQL license*.

You can send us your feedback to *edu@postgrespro.ru*.

L.1. Installation

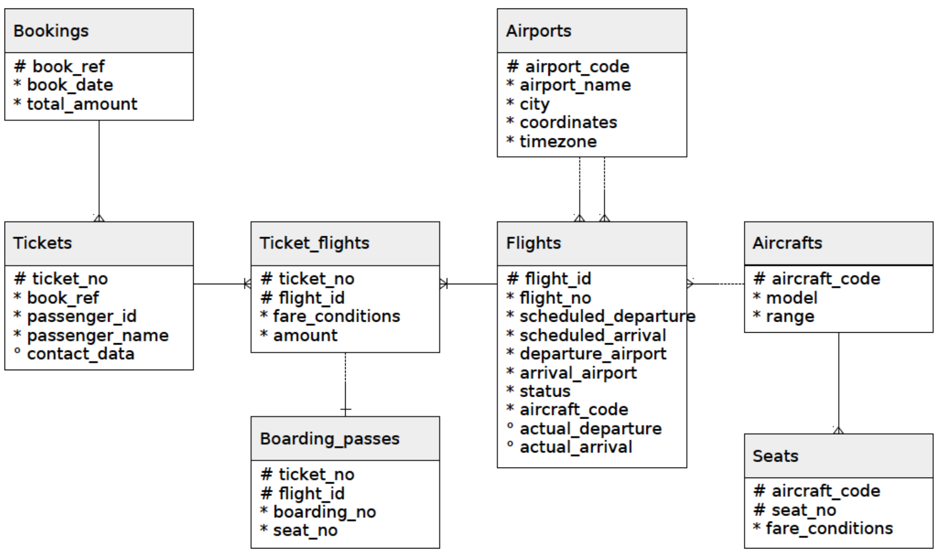
The demo database is available at *edu.postgrespro.ru* in three flavors, which differ only in the data si

- *demo-small.zip* (21 MB) — flight data for one month (DB size is about 300 MB)

The files include an SQL script that creates the `demo` database and fills it with data (virtually, it is a backup copy created with the pg_dump utility). Note that if the `demo` database already exists, it will be deleted and recreated! The owner of the `demo` database will be the DBMS user who run the script.

L.2. Schema Diagram

Figure L.2. Bookings Schema Diagram



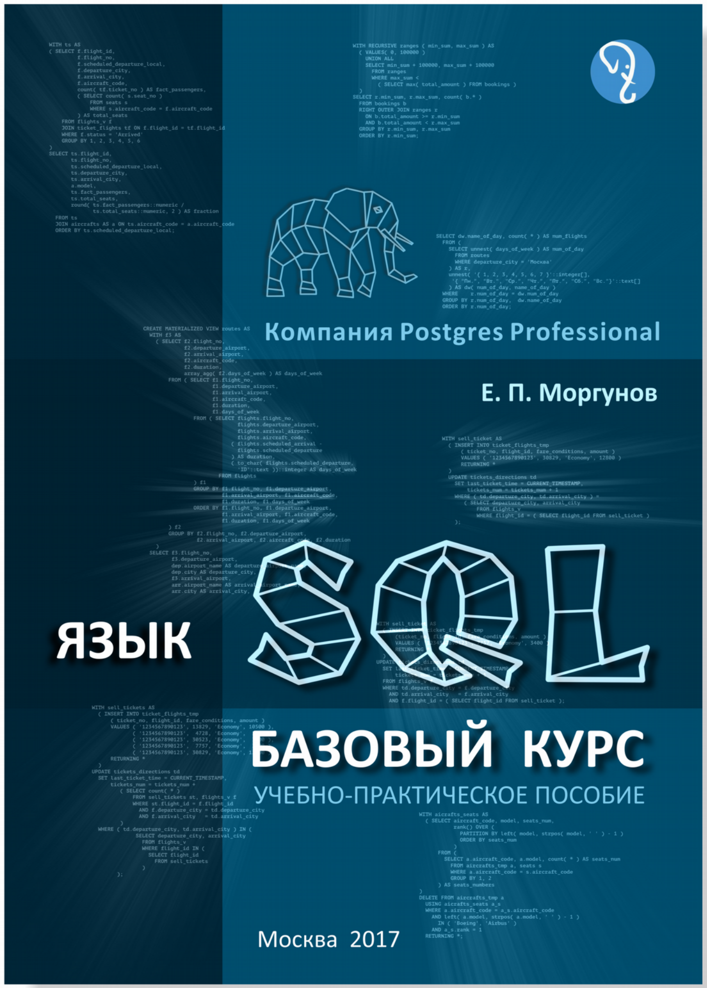
L.3. Schema Description

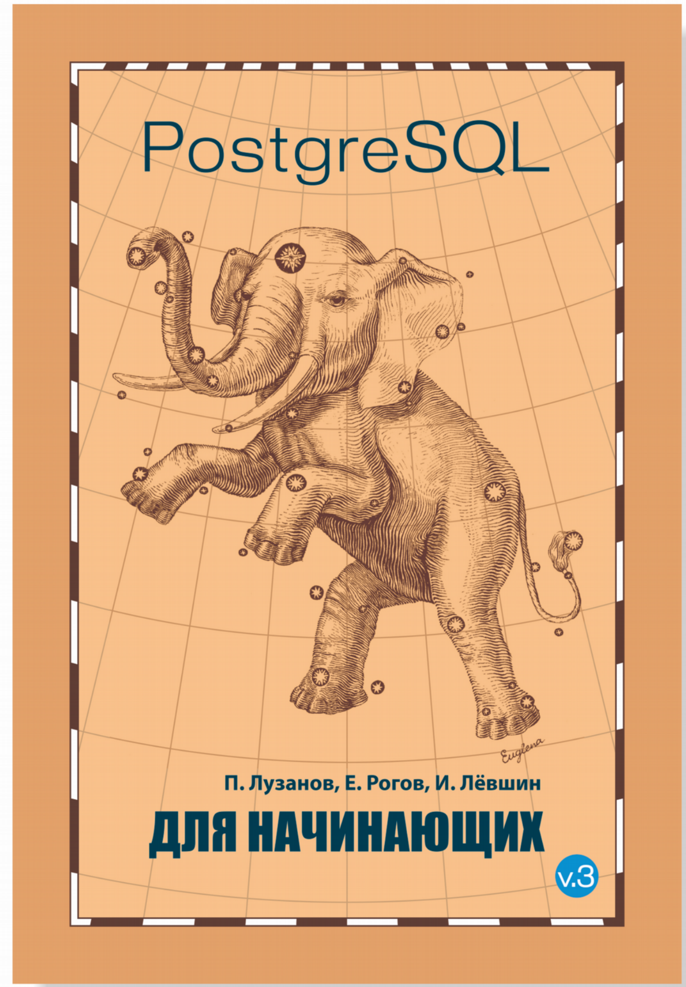
The main entity is a booking (`bookings`).

One booking can include several passengers, with a separate ticket (`tickets`) issued to each passenger. A ticket has a unique number and includes information about the passenger. As such, the passenger is not a separate entity. Both the passenger's name and identity document number can change over time, so it is impossible to uniquely identify all the tickets of a particular person; for simplicity, we can assume that all passengers are unique.

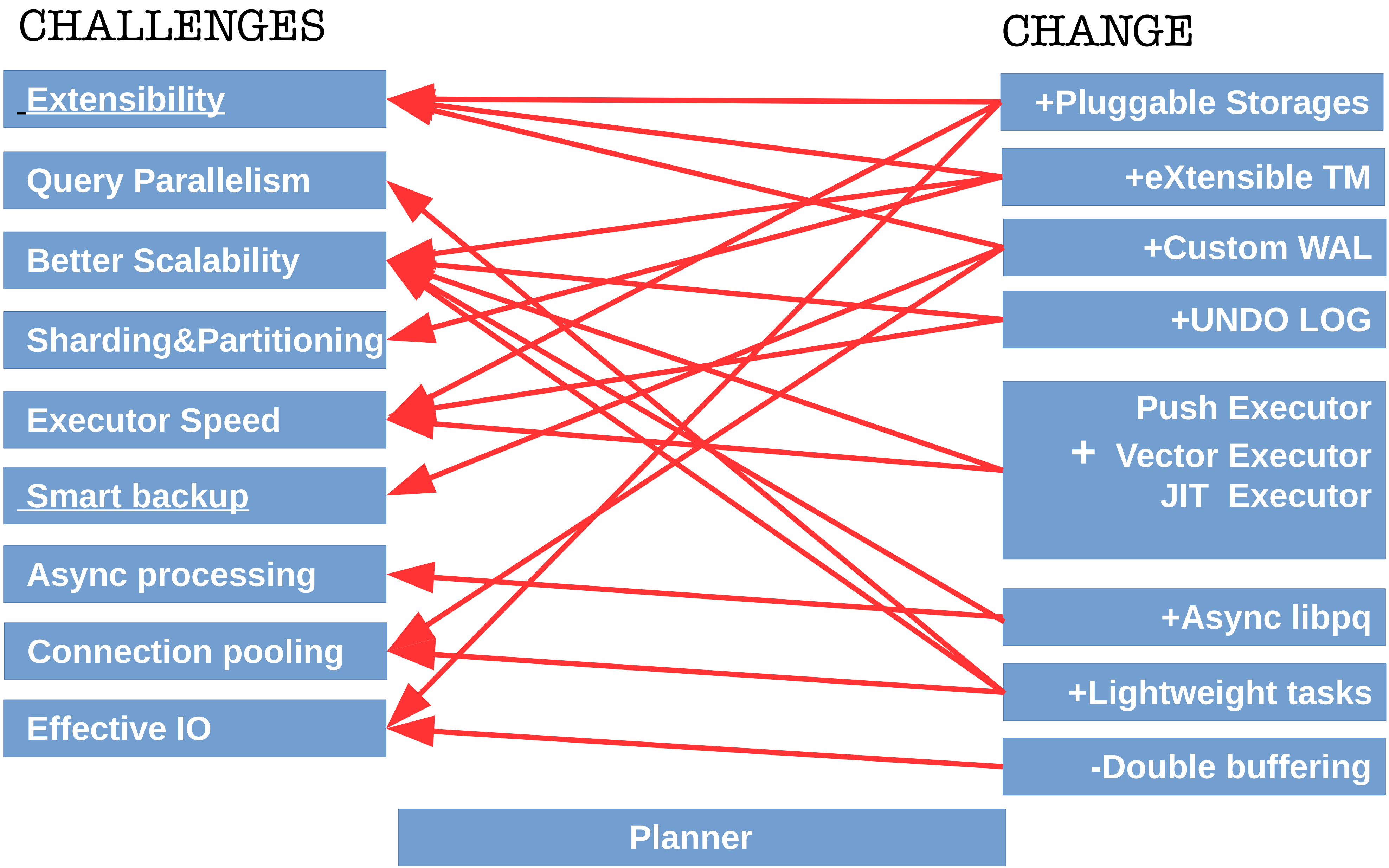
The ticket includes one or more flight segments (`ticket_flights`). Several flight segments can be included into a single ticket if there are no non-stop flights between the points of departure and destination (connecting flights), or if it is a round-trip ticket. Although there is no constraint in the schema, it is assumed that all tickets in the booking have the same flight segments.

Books about PostgreSQL
BSD License, translate them !





Wind of Change



Case 4: Trusted PostgreSQL

How to know
your database
is not compromised ?

Credereum uses digital signature to sign transactions and Blockchain to store signatures to solve this problem.

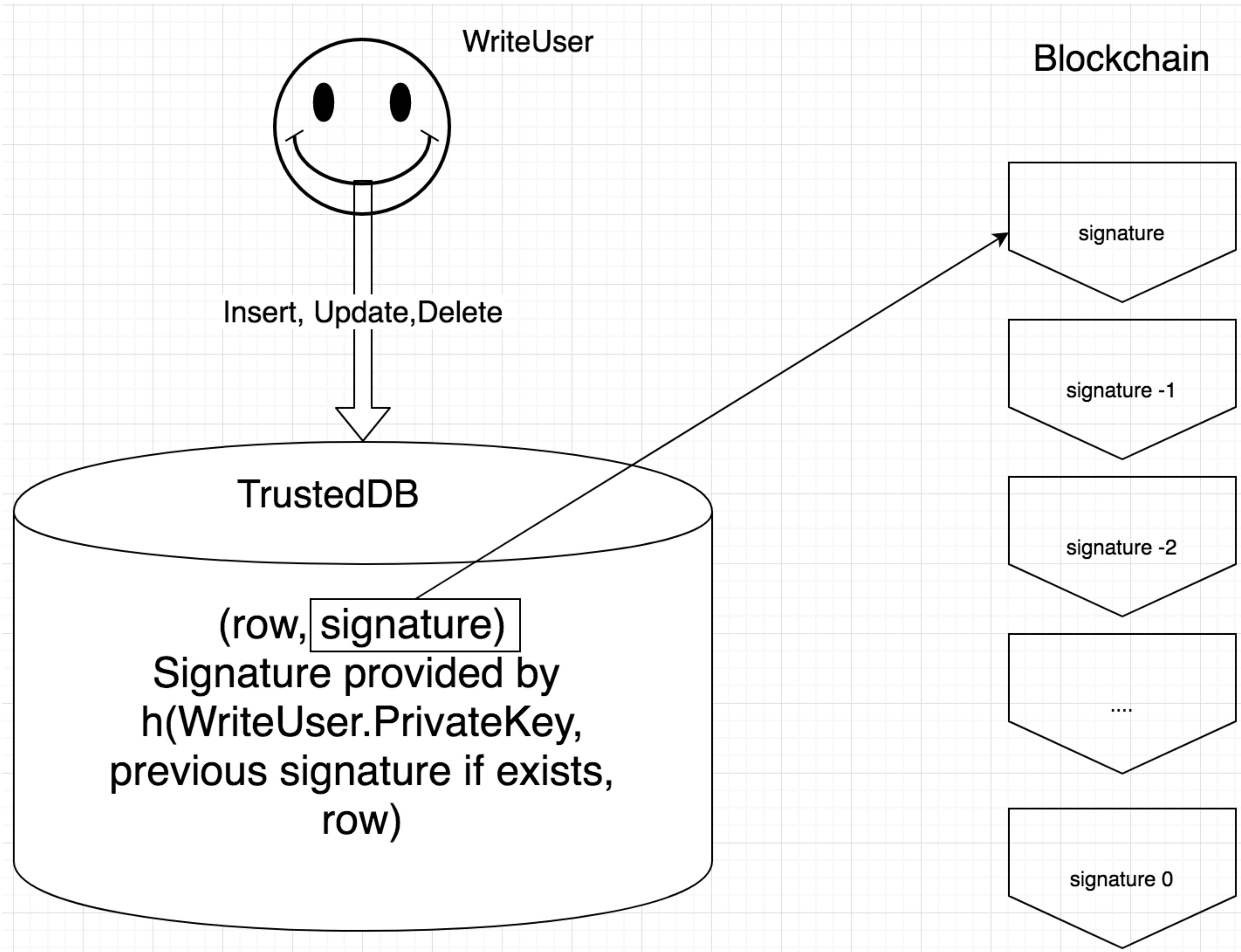
Let we have functions:

h(private key, previous signature, data) ⇒ signature

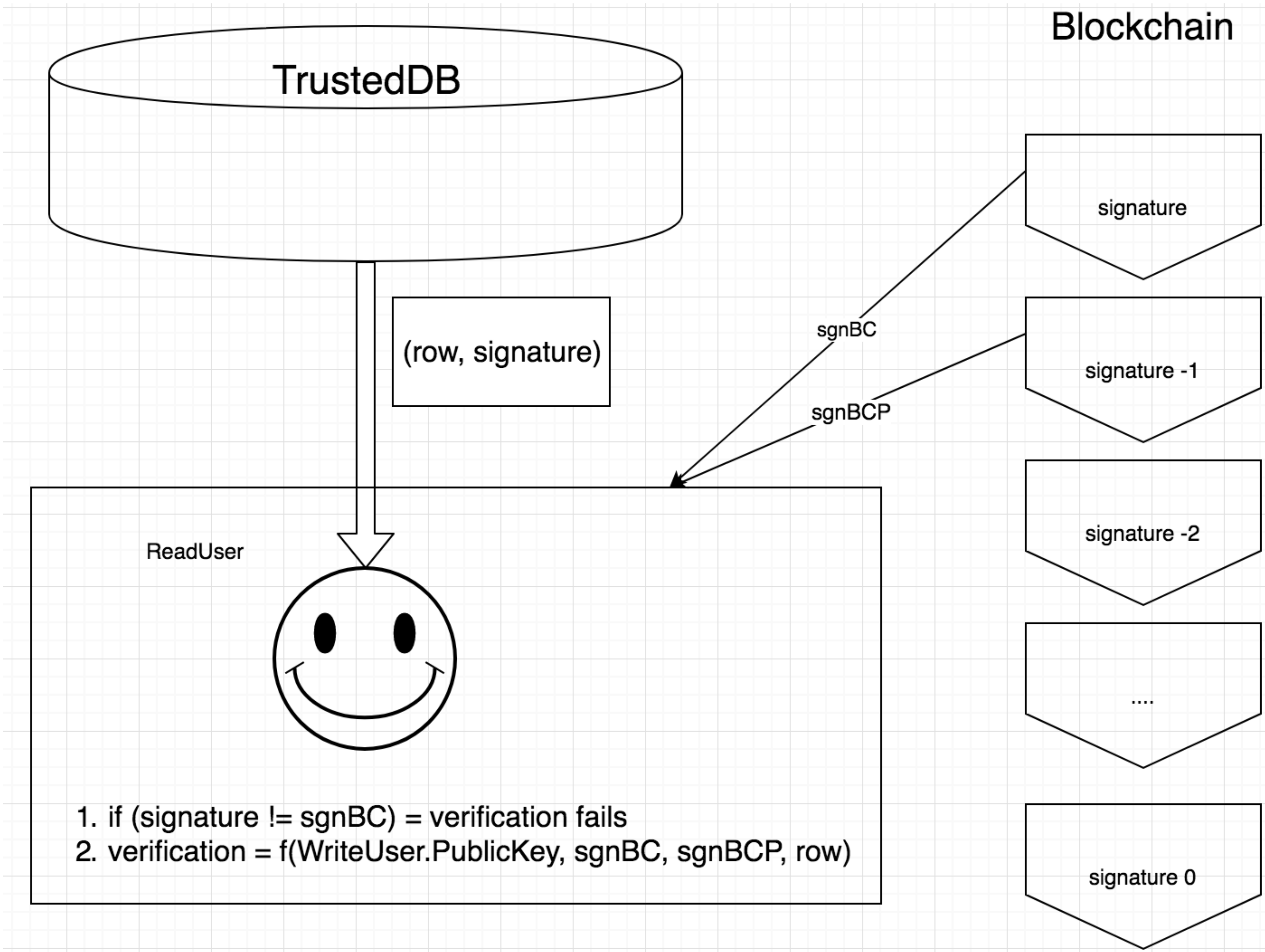
f(public key, signature, previous signature, data) ⇒ TRUE or FALSE

Roles

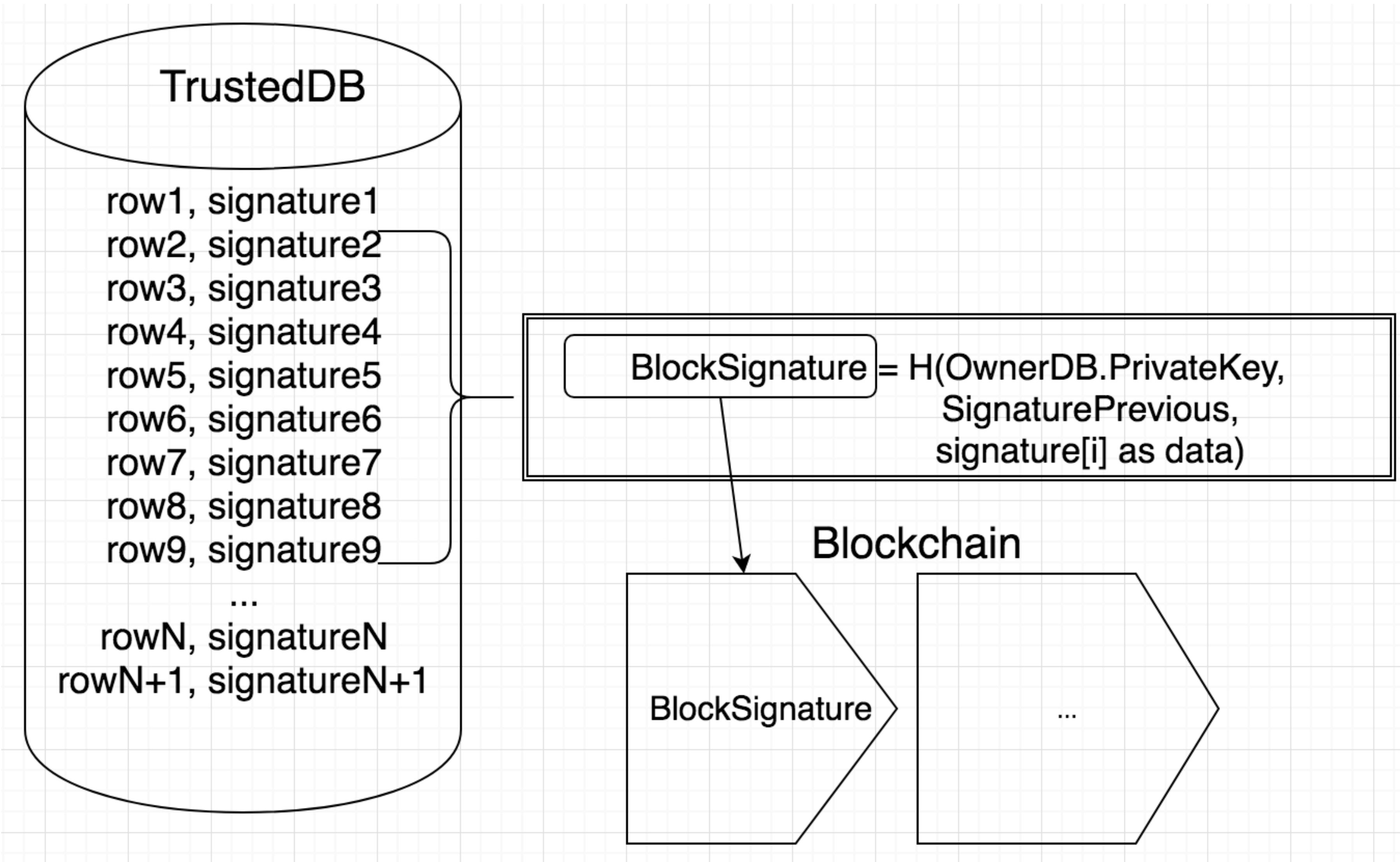
- Owner DB
- Write User
- Read User


WriteUser
Blockchain
Insert, Update,Delete
signature
signature -1
signature -2
....
signature 0
TrustedDB
(row, signature)
Signature provided by
h(WriteUser.PrivateKey,
previous signature if exists,
row)

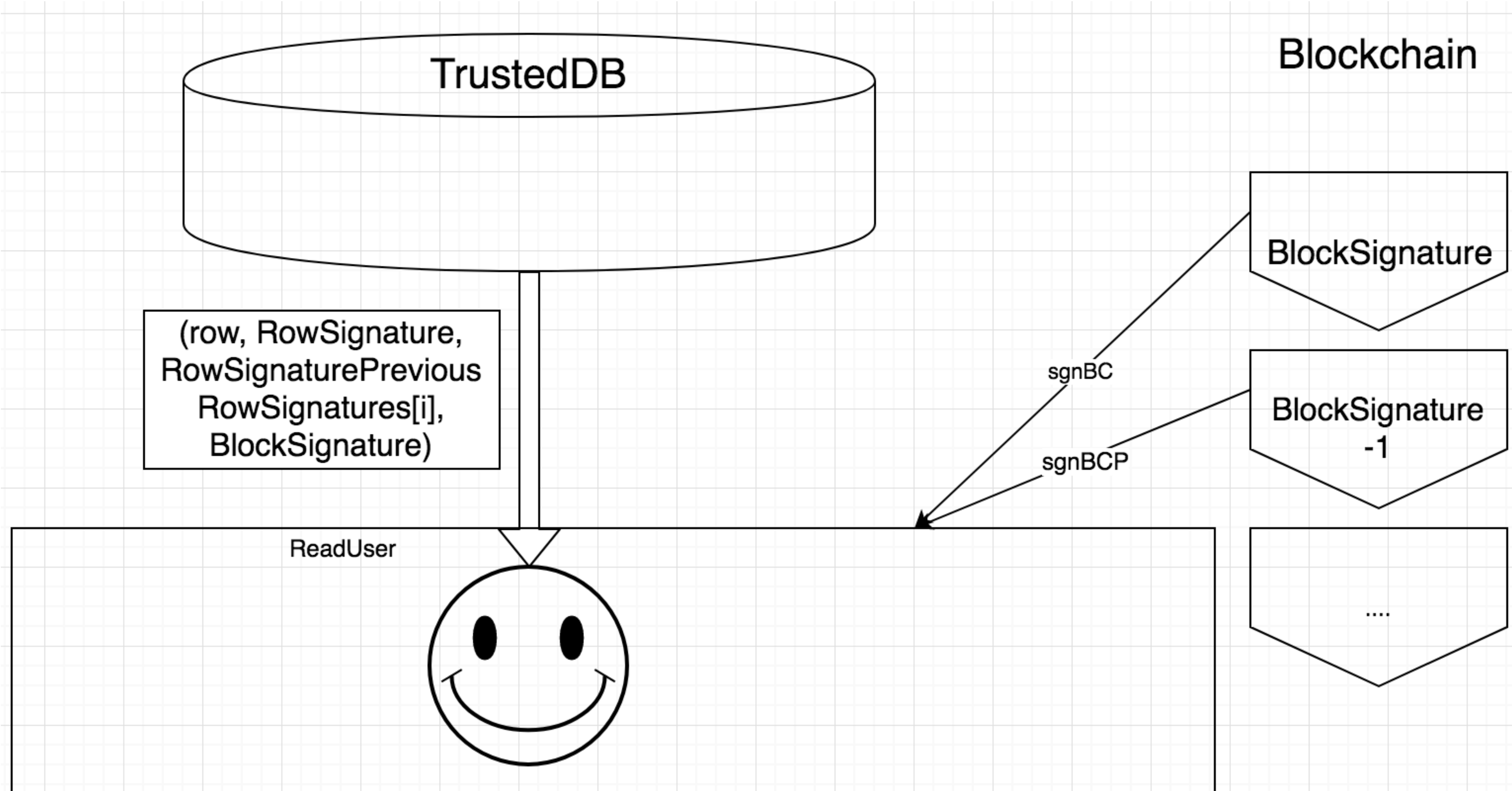
Verification



Optimize blockchain traffic



Verification



1. Check row <=> RowSignature (verificate row satisfied by RowSignature, i.e. they are both wrong or both truth, but choice of wrong/truth will done on next step)
    1. f(WriteUser.PublicKey, RowSignature,RowSignaturePrevious, row)
2. Block Verification (verification correctness of RowSignature)
    1. BlockSignature == sgnBC
    2. F(OwnerDB.PublicKey, sgnBC, sgnBCP, RowSignatures[i])

PostgreSQL Future

Several Postgres groups are working on

Postgres Distributed

Postgres Parallel

Postgres Asynchronous

Postgres Extendable+

Postgres NoSQL

Postgres Scalable

Postgres Trusted

Welcome to PGConf.ru 2018

February 5-7, Moscow

Lomonosov Moscow State University

ご清聴ありがとうございました