

ClickHouse Tech Excerpts

July 2020 ClickHouse Virtual Meetup

Alexander Zaitsev, Altinity

Alexey Milovidov, Yandex



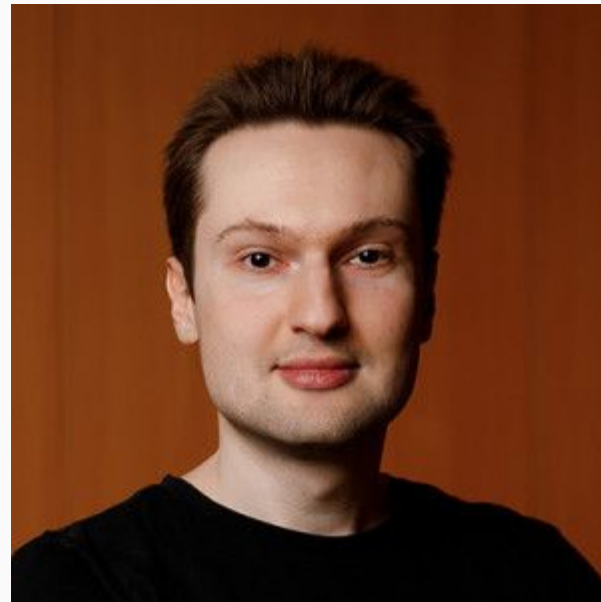
Altinity

www.altinity.com
info@altinity.com

Brief Intros



Alexander Zaitsev -- CTO, Altinity



Alexey Milovidov -- ClickHouse, Yandex

Brief Intros



www.altinity.com

Leading software and services
provider for ClickHouse

Major committer and community
sponsor in US and Western Europe



www.yandex.ru

Leading Russian IT company

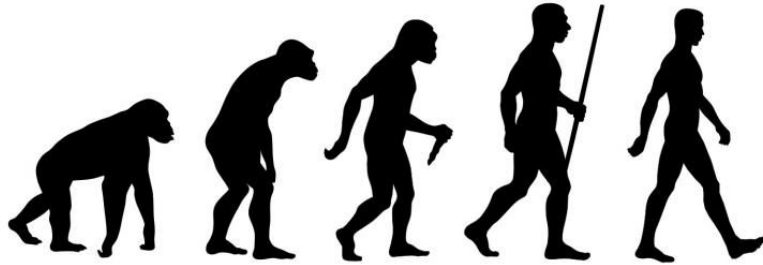
ClickHouse creator, major user and
community sponsor in Eastern Europe
and Asia

Happy Birthday to Open Source ClickHouse!



Altinity

www.altinity.com
info@altinity.com



- Features
- Performance
- Usability
- Ecosystem

2016 - 2020



Altinity

www.altinity.com
info@altinity.com

ClickHouse Summer 2020 Highlights

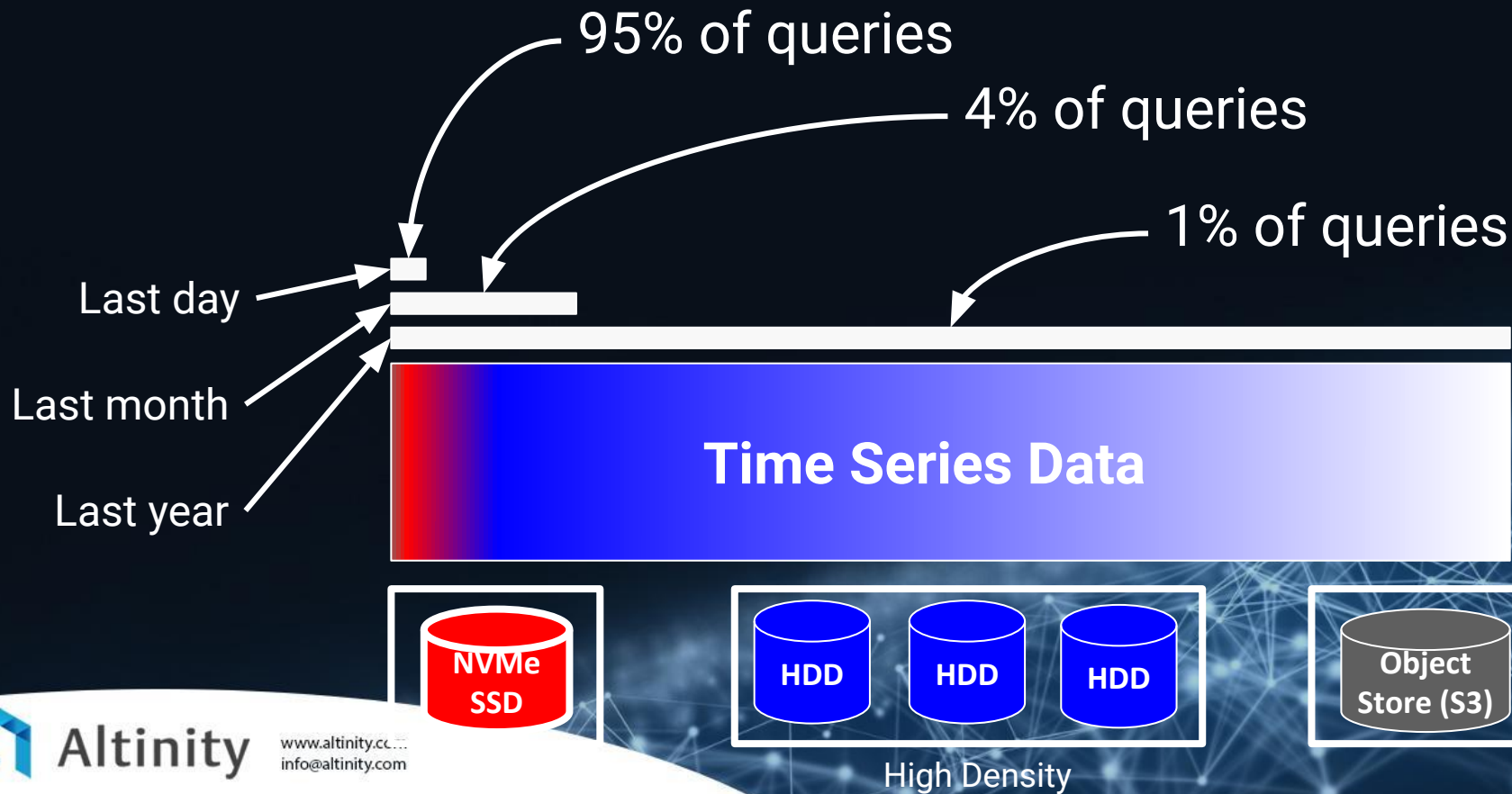
- Disk S3 for tiered storage
- Polymorphic MergeTree parts
- Cluster Improvements
- Performance optimizations



Altinity

www.altinity.com
info@altinity.com

Tiered storage matches storage to access



Altinity

www.altinity.com
info@altinity.com

S3 disk (20.6+)

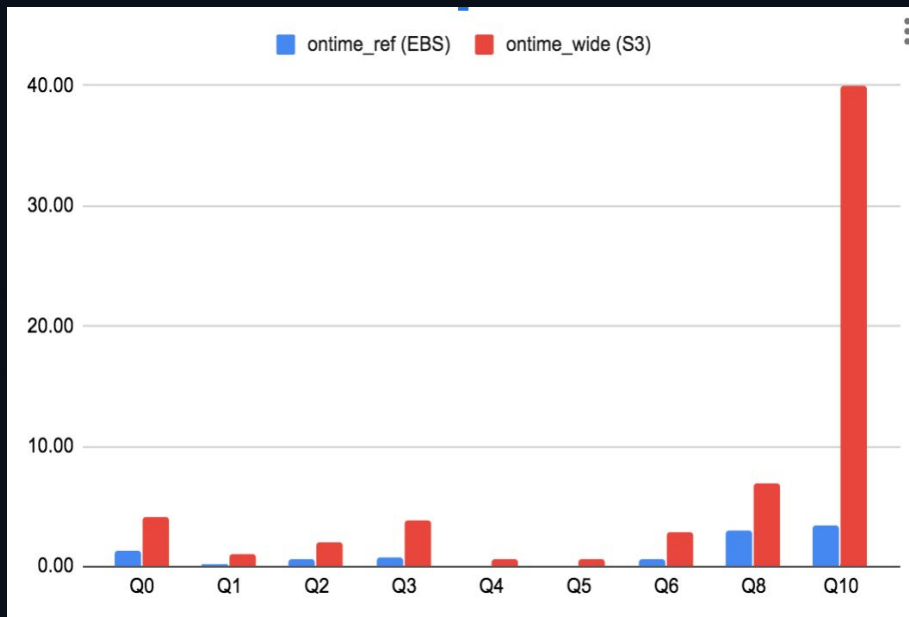
```
<storage_configuration>
  <disks>
    <s3>
      <type>s3</type>
      <endpoint>
        http://s3.us-east-1.amazonaws.com/altinity-qa/s3test2/data/
      </endpoint>
      <access_key_id>***</access_key_id>
      <secret_access_key>***</secret_access_key>
    </s3>
  </disks>
```



Altinity

www.altinity.com
info@altinity.com

S3 disk (20.6+)



- Local references to S3 data
- “Operations” are expensive
- Inserts are very slow
- Merge to reduce parts
- Parallelize

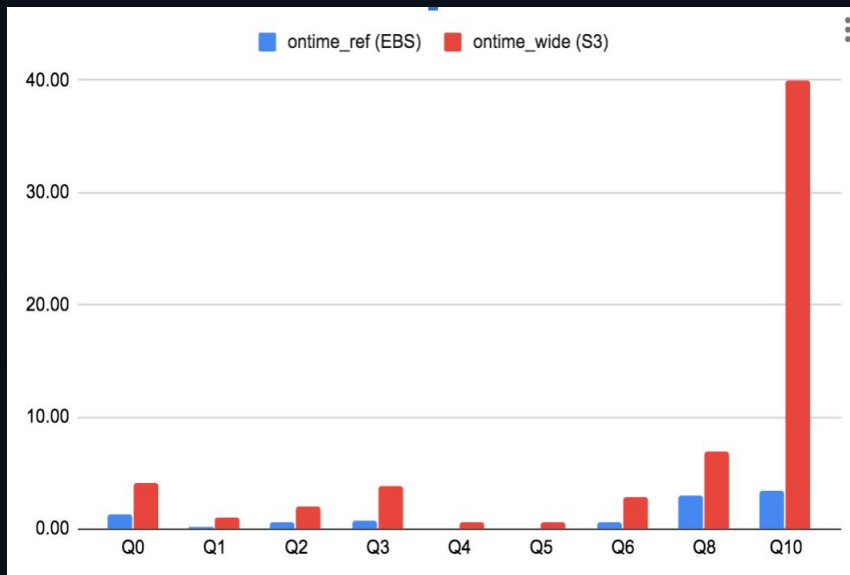
Queries from: <https://clickhouse.tech/docs/en/getting-started/example-datasets/ontime/>



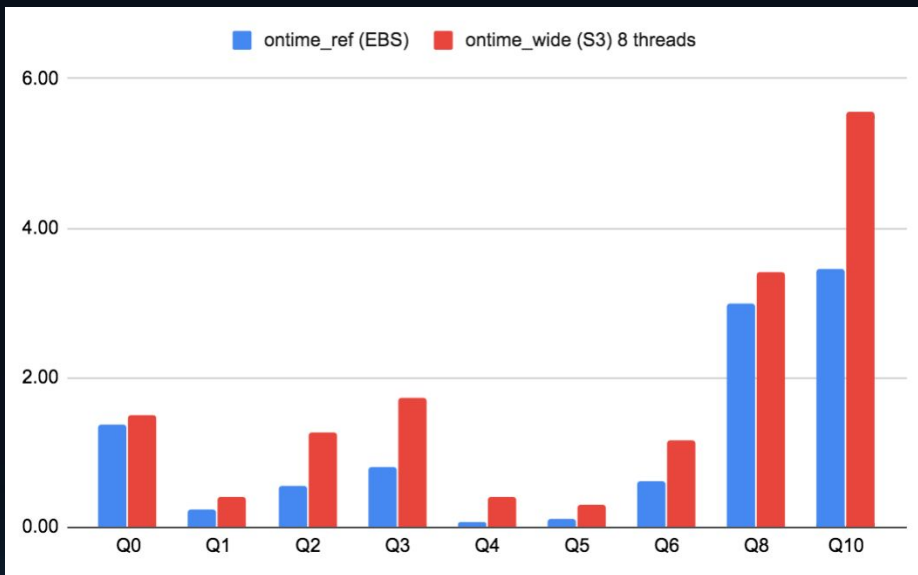
Altinity

www.altinity.com
info@altinity.com

S3 disk (20.6+)



Max_threads=2



Max_threads=8



Altinity

www.altinity.com
info@altinity.com

Compact parts for MergeTree (20.3)

“wide” parts (default)

“compact” parts (new)

Part



Part



Part



Part



Altinity

www.altinity.com
info@altinity.com

Compact parts for MergeTree

“wide” parts (default)

```
/var/lib/clickhouse/data/datasets/ontime_ref/2020_547_552_1/
.  
..  
ActualElapsedTime.bin  
ActualElapsedTime.mrk2  
AirlineID.bin  
AirlineID.mrk2  
AirTime.bin  
AirTime.mrk2  
ArrDel15.bin  
ArrDel15.mrk2  
ArrDelay.bin
```

224 files for 109 columns!

“compact” parts (new)

```
/var/lib/clickhouse/data/datasets/ontime_refc/2020_347_347_0/  
..  
..  
checksums.txt  
columns.txt  
count.txt  
data.bin  
data.mrk3  
minmax_Year.idx  
partition.dat  
primary.idx
```

8 files for 109 columns!



Altinity

www.altinity.com
info@altinity.com

Compact parts for MergeTree

- Single file but columnar inside
- Useful for small inserts
- Useful when file operations are expensive (e.g. S3)
- Threshold is controlled by merge_tree_settings:
 - min_bytes_for_wide_part
 - min_rows_for_wide_part
- Performance is still need to be improved
- **In-memory parts with WAL are coming!!!**



Cluster Improvements

- Distributed over Distributed (20.4)
- Parallel distributed INSERT SELECT (20.4)
- Performance improvements to distributed queries with respect to sharding_key (20.3-20.5)
- SYSTEM DROP REPLICA (20.5)
- DatabaseReplicated is coming!
- Writes with partitioned ZK is coming!!!



Some Other Performance Optimizations

- ❑ Parallel SELECT FINAL (20.5)
- ❑ GROUP BY and DISTINCT with respect to primary key (20.5, 20.6)
- ❑ ORDER BY optimizations (20.5)
- ❑ Much faster Kafka Engine (20.3, 20.5)
- ❑ Bulk mode in ODBC driver (1.1.8)



Q&A



Altinity

www.altinity.com
info@altinity.com



Read Our Blog:

<https://www.altinity.com/blog>



Contacts:

alz@altinity.com

skype: alex.zaitsev

telegram: @alexanderzaitsev



Altinity

www.altinity.com
info@altinity.com