讲师：平民软件楼方鑫

主题：MySQL深层定制之OneSQL

About Me

- More than 15 years DBA experience, an old man!
  - Wrote SQLULDR2
  - Wrote MyDUL (Later renamed to AUL)
- 5 Years at eBay
  - Staff DBA
- 6 Years at Alibaba(AliPay)
  - DBA Manager
  - Data Architecture
- Now working at onexsoft.com
  - Providing Better Architecture Units/Components.
  - One good choice is better than many choices.

Three Periods

- Before 2004
  - Application Developer
- Be an Oracle Expert (Before 2008)
  - Client & Server Programing
  - Oracle Data File Format
  - Oracle Log File Format
  - Oracle Direct Memory Access
- Be an Data Arch & System Engineer
  - Scale out and up the payment system
  - User/Trade/Account/Deposit system split
  - No Single Point of Failure in payment system
- Be an ?
  - How to scale out and scale up easily?
  - Why every company has their own customization?

At eBay

- Very Large Business Volume.
- Have 20000 sessions for each User database.
- How can we reduce the sessions?
- Decision
  - System is different from Oracle
  - Expert in Oracle is not enough

At AliPay

- Split & Split & Split ……
- We cannot do anything at database side with Oracle, just ask developers to change the business, optimize the code.
- It’s a little hard, and finally application developers learns everything, but I got nothing (joke)
- We make it, but we cannot export it to others.
- Why copy and paste is not easy!

At OneXSoft

- Focus on Business
  - Fewer Requirements for Operation Team.
  - Fewer Requirements for Develop Team.
- Not Every Company can afford a Team.
- Huge Resource Wastage if all is different.
- Focus on Data Arch Components.
  - Cache、KV Store、MySQL、PostgreSQL etc.
  - Data Access Layer、Database Firewalls
  - Not only SQL, But SQL is the basic components.

Key Points

- Performance Degrade by Lots of Sessions.
- Resource Isolation for different operations.
- Make transactions faster and faster.
- We need a robust logical data copy.
- Spend too much resource on application code.

Sessions

- Resource redundancy for Business Growth
- Always have load spikes
- Application Code wrote in Java
- Network cross different switch/routers/IDC
- Database may not scale well as declared
- MTS is not so good enough.

Isolation

- Bad SQL always exists!
- Bad Logic Implementation always exists!
- Always big agent/customer/account
- GMV is very important
- Application is easy to be grouped, while data is very hard to be grouped.
- There is no good idea on Oracle, resource manager?

Faster Transaction

- Bidding on one item.
- Operation on one account.
- No fixed hot item or account, hard to pre-evaluate them.
- Faster CPU/IO is not enough, need some code level change!

Logical Copy

- How faster an Oracle switch?
- How to automate the Oracle switch?
- We need read/write split for read traffics!
- We need read/write split for no single point of failure of read traffics!
- We need save the cost of read traffics!
- We wait too long for Oracle ADG!

Application Rewrite

- Transparent to Application Developers!
- Very hard to copy/paste/change the developers frameworks.
- Bound to one programing language may not be a good thing, and it need long time and huge resource.
- Protocol level design, like network switches/routers, too heavy for all client side implementation.
- Cannot get anything about Net 8 Protocol internals.

Two Levels

- Enhance the MySQL
  - OneSQL branch
- Create the OneProxy
  - OneProxy for MySQL
  - OneProxy for PostgreSQL
  - OneProxy for Oracle [need Net 8 Protocol]

OneSQL

- Stable TPS for 100/500/1000/2000/4000/8000 concurrent sessions.
- Dedicate Thread (resource) Pool for queries and DMLs, and even more.
- Auto commit/rollback on success/failure patches.
- Optimized Master and Slave for better logical replication performance.
- Different Data Protection Mode as Oracle's Data Guard. How is Oracle's ODG or logical standby?
- Optimized Steps for Master/Slave switch over!

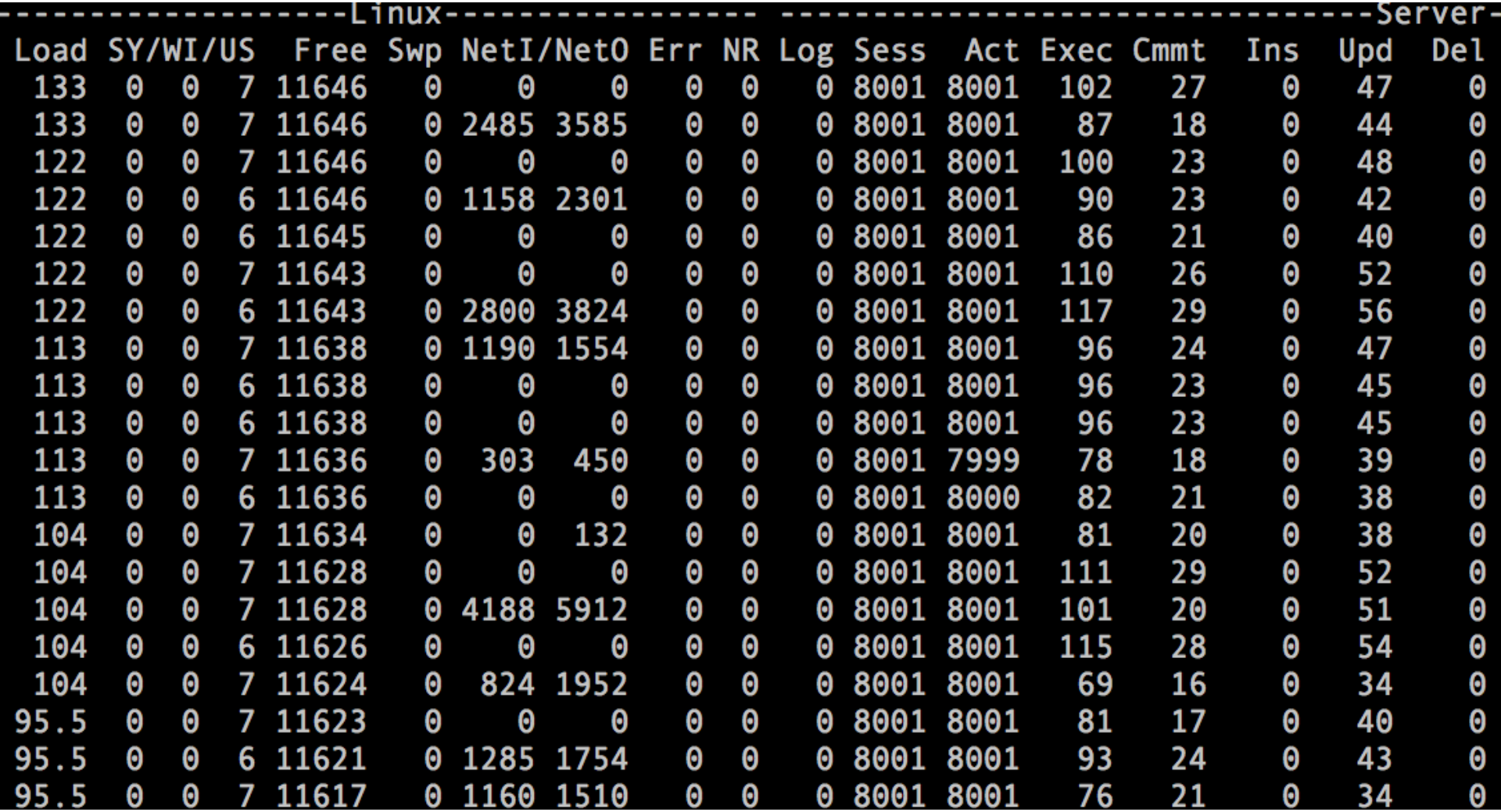
80% of Oracle

time	1m	5m	15m	usr	sys	idl	iow	si	so	ins	upd	del	sel	iud
13:46:48	3.01	2.28	2.02	45	16	38	0	0	0	34065	36311	0	93779	70376
13:46:49	3.01	2.28	2.02	45	16	37	1	0	0	32887	34923	0	96389	67810
13:46:50	3.01	2.28	2.02	46	16	37	1	0	0	34286	36688	0	93899	70974
13:46:51	3.09	2.31	2.03	46	16	37	0	0	0	33752	36082	0	93346	69834
13:46:52	3.09	2.31	2.03	45	16	38	1	0	0	32781	35001	0	93795	67782
13:46:53	3.09	2.31	2.03	46	17	36	1	0	0	34512	37044	0	93592	71556
13:46:54	3.09	2.31	2.03	46	16	37	1	0	0	33502	35826	0	94613	69328
13:46:55	3.09	2.31	2.03	45	16	38	1	0	0	33384	35485	0	94430	68869
13:46:56	3.40	2.39	2.06	45	17	37	1	0	0	33549	35832	0	94996	69381
13:46:57	3.40	2.39	2.06	46	16	37	1	0	0	33905	36276	0	93376	70181
13:46:58	3.40	2.39	2.06	45	16	38	1	0	0	33386	35544	0	94268	68930
13:46:59	3.40	2.39	2.06	46	17	37	0	0	0	34793	37170	0	93266	71963
13:47:00	3.40	2.39	2.06	44	15	39	1	0	0	31410	33404	0	96975	64814
13:47:01	3.53	2.43	2.07	46	17	37	0	0	0	34346	37023	0	93239	71369
13:47:02	3.53	2.43	2.07	45	16	38	1	0	0	32983	34837	0	95486	67820
time	1m	5m	15m	usr	sys	idl	iow	si	so	ins	upd	del	sel	iud

Test Case

```
option
  user test/test@172.30.12.4:3306:test
  log /dev/null
  time 1m
declare
  vid bigint 1  500
  vid2 bigint 500 1000
begin
  start;
  update t_binlog set col2=col2-1 where id = :vid;
  update t_binlog set col2=col2-1 where id = :vid2;
  commit;
end
```

MySQL 8K



Linux										Server					
Load	SY	WI	US	Free	Swp	NetI	NetO	Err	NR	Log	Sess	Act	Exec	Cmmt	Ins	Upd	Del
133	0	0	7	11646	0	0	0	0	0	0	8001	8001	102	27	0	47	0
133	0	0	7	11646	0	2485	3585	0	0	0	8001	8001	87	18	0	44	0
122	0	0	7	11646	0	0	0	0	0	0	8001	8001	100	23	0	48	0
122	0	0	6	11646	0	1158	2301	0	0	0	8001	8001	90	23	0	42	0
122	0	0	6	11645	0	0	0	0	0	0	8001	8001	86	21	0	40	0
122	0	0	7	11643	0	0	0	0	0	0	8001	8001	110	26	0	52	0
122	0	0	6	11643	0	2800	3824	0	0	0	8001	8001	117	29	0	56	0
113	0	0	7	11638	0	1190	1554	0	0	0	8001	8001	96	24	0	47	0
113	0	0	6	11638	0	0	0	0	0	0	8001	8001	96	23	0	45	0
113	0	0	6	11638	0	0	0	0	0	0	8001	8001	96	23	0	45	0
113	0	0	7	11636	0	303	450	0	0	0	8001	7999	78	18	0	39	0
113	0	0	6	11636	0	0	0	0	0	0	8001	8000	82	21	0	38	0
104	0	0	7	11634	0	0	132	0	0	0	8001	8001	81	20	0	38	0
104	0	0	7	11628	0	0	0	0	0	0	8001	8001	111	29	0	52	0
104	0	0	7	11628	0	4188	5912	0	0	0	8001	8001	101	20	0	51	0
104	0	0	6	11626	0	0	0	0	0	0	8001	8001	115	28	0	54	0
104	0	0	7	11624	0	824	1952	0	0	0	8001	8001	69	16	0	34	0
95.5	0	0	7	11623	0	0	0	0	0	0	8001	8001	81	17	0	40	0
95.5	0	0	6	11621	0	1285	1754	0	0	0	8001	8001	93	24	0	43	0
95.5	0	0	7	11617	0	1160	1510	0	0	0	8001	8001	76	21	0	34	0

8K Connection

Load	SY/WI/US	Free	Swp	NetI/NetO	Err	NR	Log	Sess	Act	Exec	Cmmt	Ins	Upd	Del
0.17	20 0 61	17849	0	0 0	0	0	0	8001	7840	100k	25k	0	50k	0
0.24	20 0 61	17840	0	18k 10k	0	0	0	8001	7809	103k	25k	0	51k	0
0.24	20 0 62	17832	0	87k 71k	0	0	0	8001	7774	102k	25k	0	51k	0
0.24	20 0 59	17823	0	0 0	0	0	0	8001	7779	103k	25k	0	51k	0
0.24	19 0 55	17814	0	0 0	0	0	0	8001	7736	101k	25k	0	50k	0
0.24	23 0 53	17806	0	66k 92k	0	0	0	8001	7836	101k	25k	0	50k	0
0.38	24 0 54	17795	0	0 0	0	0	0	8001	7762	102k	25k	0	51k	0
0.38	23 0 55	17785	0	0 0	0	0	0	8001	7763	100k	25k	0	50k	0
0.38	22 0 58	17776	0	219k 239k	0	0	0	8001	7696	102k	25k	0	51k	0
0.38	21 0 59	17766	0	62k 88k	0	0	0	8001	7738	101k	25k	0	50k	0
0.38	21 0 60	17756	0	0 0	0	0	0	8001	7804	103k	25k	0	51k	0
0.43	21 0 60	17748	0	46k 88k	0	0	0	8001	7753	102k	25k	0	51k	0
0.43	20 0 60	17739	0	3410 0	0	0	0	8001	7883	102k	25k	0	51k	0
0.43	20 0 61	17732	0	0 0	0	0	0	8001	7764	102k	25k	0	51k	0
0.43	20 0 61	17723	0	33k 52k	0	0	0	8001	7384	101k	25k	0	50k	0
0.43	19 0 59	17716	0	50k 36k	0	0	0	8001	7739	98k	24k	0	49k	0
0.55	19 0 59	17706	0	0 0	0	0	0	8001	7737	102k	25k	0	51k	0
0.55	20 0 55	17697	0	487k 585k	0	0	0	8001	7814	102k	25k	0	51k	0
0.55	22 0 55	17689	0	66k 65k	0	0	0	8001	7770	103k	25k	0	51k	0
0.55	25 0 54	17681	0	0 0	0	0	0	8001	7760	102k	25k	0	51k	0

16K Connections

```
------------------Linux----------------- -------------------------------Server-
 Load SY/WI/US  Free Swp NetI/NetO Err NR Log Sess  Act Exec Cmmt  Ins  Upd  Del
 0.26 20  0 54 18990   0  10m  12m   0  0   0  16k  14k  99k  24k    0  49k    0
 0.26 23  0 54 18967   0 133k 147k   0  0   0  16k  14k 100k  25k    0  50k    0
 0.26 24  0 54 18950   0    0    0   0  0   0  16k  14k  99k  24k    0  49k    0
 0.48 22  0 56 18933   0 168k 198k   0  0   0  16k  14k  99k  24k    0  49k    0
 0.48 21  0 59 18915   0    0    0   0  0   0  16k  13k 100k  25k    0  50k    0
 0.48 21  0 59 18899   0 201k 256k   0  0   0  16k  13k 100k  25k    0  50k    0
 0.48 20  0 61 18882   0    0    0   0  0   0  16k  14k 100k  25k    0  50k    0
 0.48 20  0 61 18868   0    0    0   0  0   0  16k  13k 100k  25k    0  50k    0
 0.76 20  0 61 18853   0  61k 119k   0  0   0  16k  14k 100k  25k    0  50k    0
 0.76 20  0 60 18839   0    0    0   0  0   0  16k  14k  99k  24k    0  49k    0
 0.76 19  0 62 18823   0  49k  45k   0  0   0  16k  13k 100k  25k    0  50k    0
 0.76 19  0 62 18809   0  18k  30k   0  0   0  16k  14k 101k  25k    0  50k    0
 0.76 19  0 58 18794   0    0    0   0  0   0  16k  13k  99k  24k    0  49k    0
 1.02 20  0 55 18779   0    0    0   0  0   0  16k  14k 100k  25k    0  50k    0
 1.02 21  0 52 18768   0    0    0   0  0   0  16k  13k  95k  23k    0  47k    0
 1.02 24  0 55 18754   0 555k 597k   0  0   0  16k  14k 100k  25k    0  50k    0
 1.02 22  0 56 18742   0    0    0   0  0   0  16k  13k  99k  24k    0  49k    0
 1.02 21  0 58 18728   0  58k  81k   0  0   0  16k  14k 100k  25k    0  50k    0
```

OneProxy

- Protocol level switches/routers for databases!
- Cross system connection pool, never 20000 connection per database any more.
- Read failover or read/write split/balance with application redesign or code change.
- Split huge table into MySQL groups without application redesign or code change.
- Support parallel query like MPP system.
- Acting as a Performance schema.
- Working as a database firewall at the same time.
- High availability (HA) and high performance (40W QPS)

Transparent

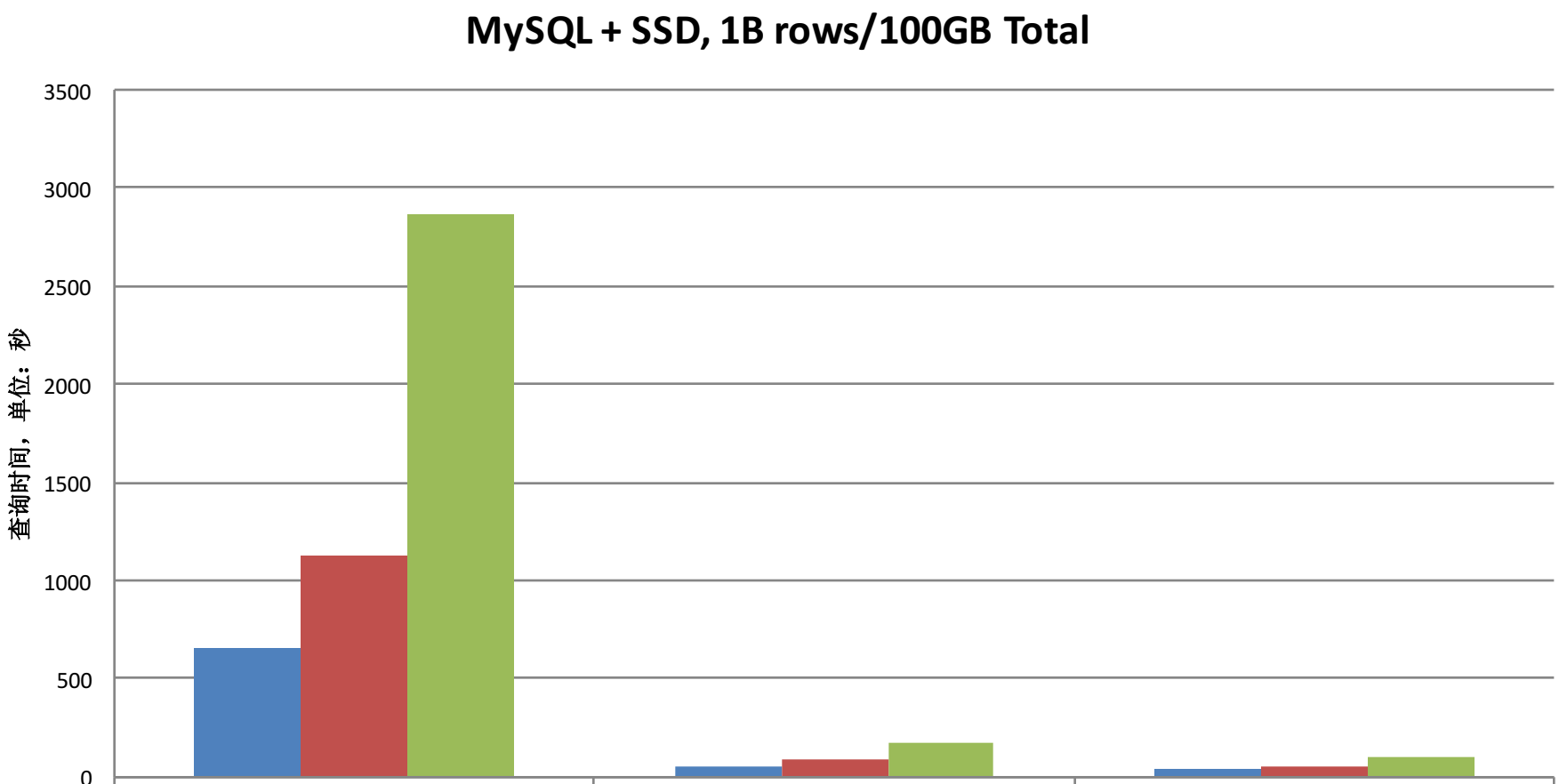
Traffic Policy

Policy	Query	DMLs
Master-only	master	master
Read-failover	master, slave if no master	master
Read-slave	slave, master if no slave	master
Read-balance	any	master
Big-slave	slave, master if no slave	master
Big-balance	any	master
Write-failover	any	master (auto failover)
Write-balance	any	any

Database Split

```
{
    "table"   : "my_range",
    "pkey"    : "id",
    "type"    : "int",
    "method"  : "range",
    "partitions":
      [
        { "suffix" : "_0", "group" : “server1", "value" : 100000 },
        { "suffix" : "_1", "group" : “server2", "value" : 200000 },
        { "suffix" : "_2", "group" : “server3", "value" : 300000 },
        { "suffix" : "_3", "group" : “server4", "value" : null    }
      ]
}
```

MPP



	无并发	16并发	24并发
计数	651	58	42
汇总1	1129	89	59
汇总2	2861	174	106

More!

- We can do more!
- We need do more!
- We must do more!

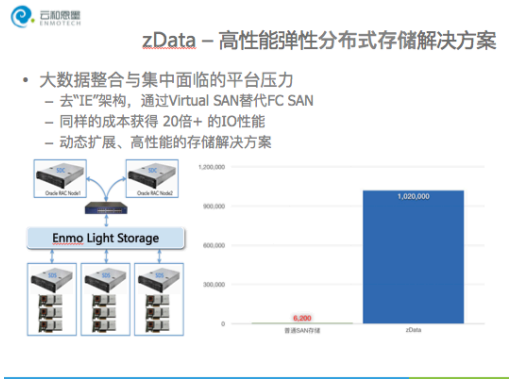
关注微信公众号 获取文档和更新

云和恩墨
（支持ORA错误自动查询）

恩墨学院

Oracle新闻

z3 – SQL审核



zData – 分布式存储

BayMax自动化巡检

特别感谢　合作伙伴

云和恩墨
ENMOTECH

Shannon Systems
宝　存　科　技

ACOUG
All China Oracle User Group
中国 Oracle 用户组

恩墨学院
ENMOEDU

Data Driven World
EN·CORE
恩核(北京)信息技术有限公司

OCMU
OCM联盟

SDOUG
Shandong Oracle User Group
山东Oracle用户组

ITPUB

ChinaUnix

IT168.com
www.it168.com

31会议

Mellanox
TECHNOLOGIES
Connect. Accelerate. Outperform.™

乐导科技
alidao.com

慕课网
imooc.com

清華大學出版社

Broadview®
www.broadview.com.cn

TURING
图灵教育

华章科技

THANKS