

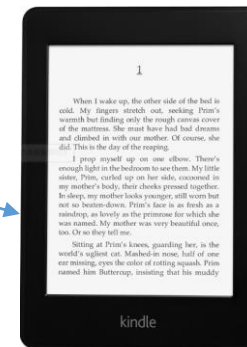
# 亿级SQL Server运维的最佳实践

宋运剑



1# Not Only SQL Server

---



# 关系数据库 VS NOSQL





# SQL Server不 应该是今天用 到的唯一DB

- SQL Server是典型的关系数据库
- 对于OLTP来说，对应的数据应该是价值密度高的数据
- SQL Server用于NoSQL、时序数据、全文检索、消息数据，图（2017中支持）、层级数据则有很多额外开销
- 关系数据库不适用Schema频繁变更的场景（变更Schema绝对是痛不欲生的事情）
- 读远远大于写的操作类型或昂贵操作应该放入缓存（关系数据库ACID特性会给读操作带来大量额外开销）
- 关系数据库难以横向扩展（Scale-Out）
- 大量低价值密度数据的分析（Spark等分布式模型都会比CUBE等有更好的性能和扩展性）

# 常用DB

全文检索	ElasticSearch
NoSQL	MongoDB
时序数据库	InfluxDB Graphite
K-V	Redis Memory Cache
消息队列	Kafka RabbitMQ
图数据库	Neo4j
宽列	Cassandra Hbase

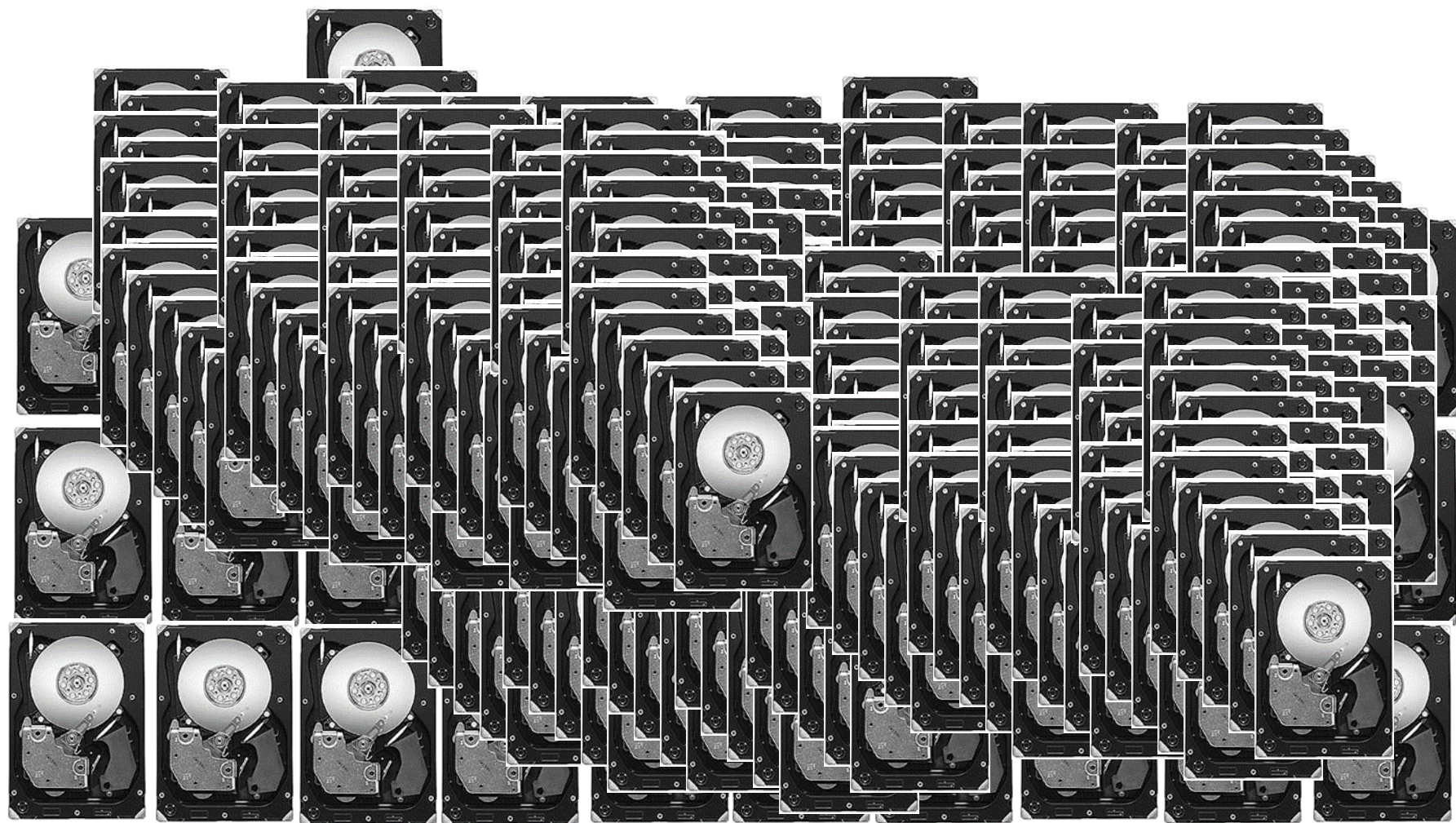
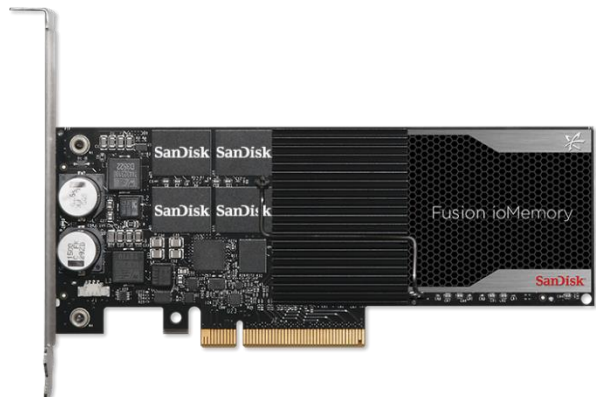
2# 硬件充足

---

充足的硬件或  
云资源能避免  
很多麻烦

- CPU
- IO
- 内存
- 网络





3# 缓存

---

# SQL Server调优 分为三个层面

- 应用程序设计
  - 缓存、数据归档
  - OLAP/OLTP分离、读写分离
- 数据库设计
  - 缺少索引、过多索引
  - 不适当的列类型、过度范式化、表分区
- 查询设计
  - 不好的查询语句（不适当的OR/LIKE/EXISTS操作、索引列转换）
  - 不恰当的使用视图/触发器/游标

大量用户

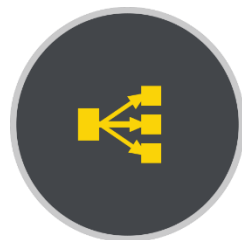
访问

负载均衡

应用服务器

每秒大量访问

数据库



过载！

大量用户

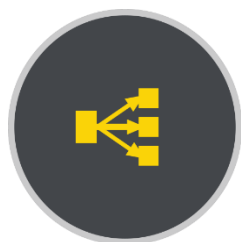
访问

负载均衡

应用服务器

缓存

数据库



可承载

只剩下少量访问

4# 监控

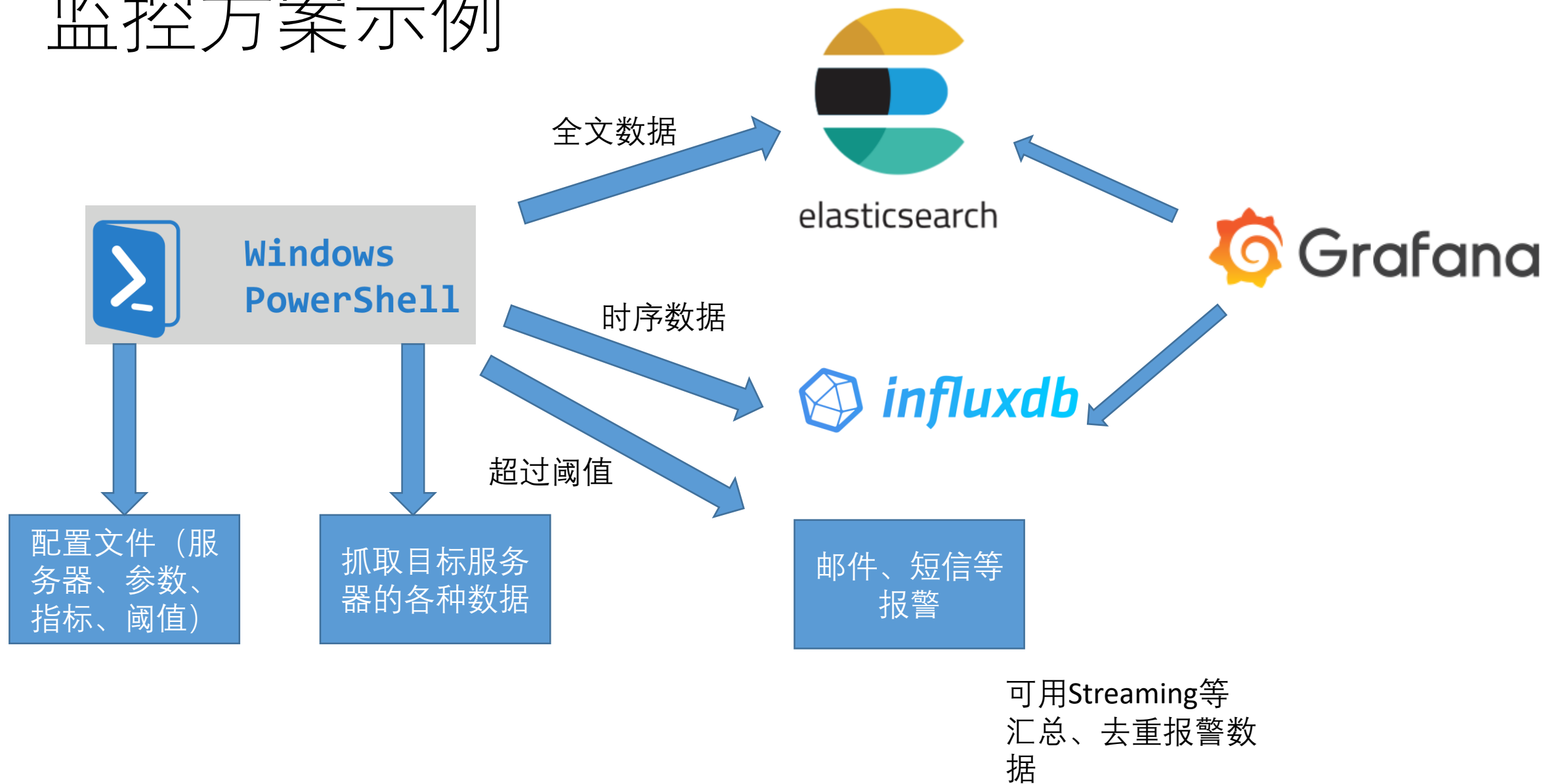
---

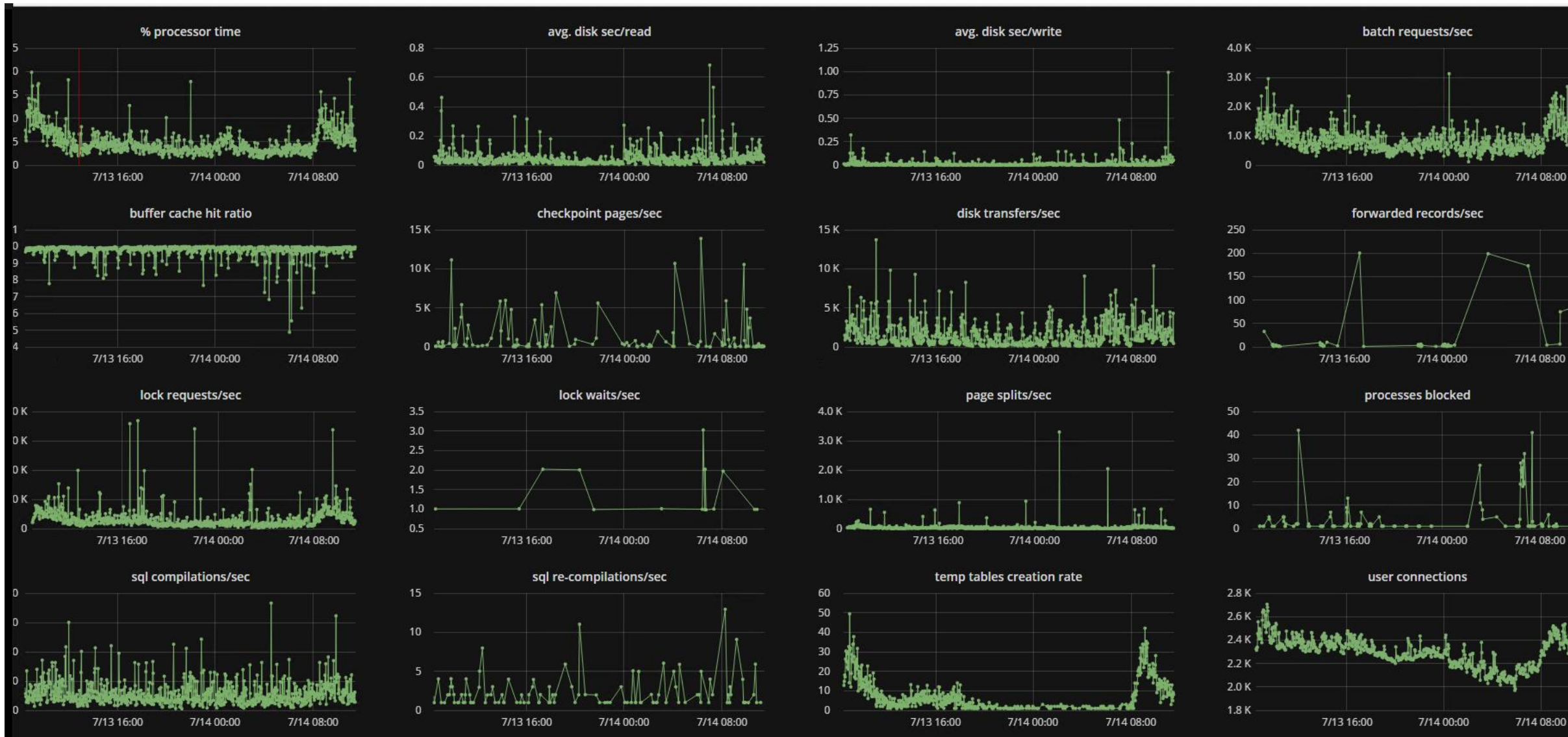
## 不仅仅是常规 监控

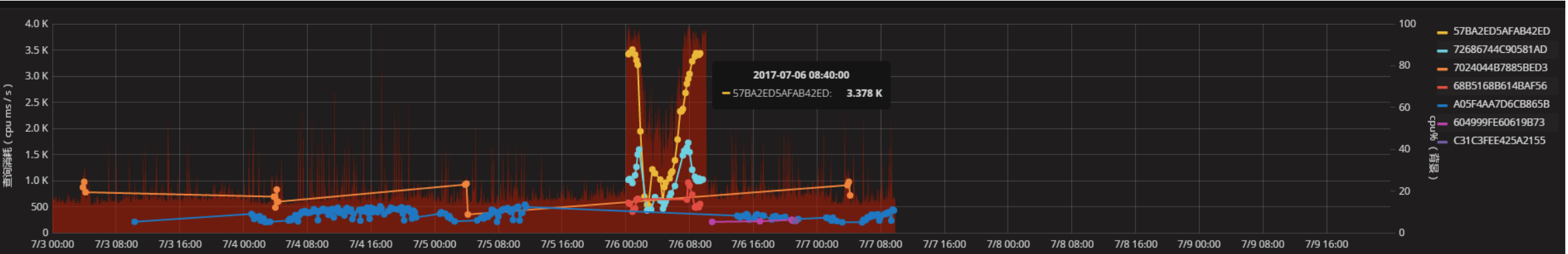
- CPU使用率、端口能否通、磁盘空间等（常规Zabbix等通用工具可完成）
- 各种性能计数器
- 高消耗语句
- 超时语句
- 异常作业
- DDL、DML审核



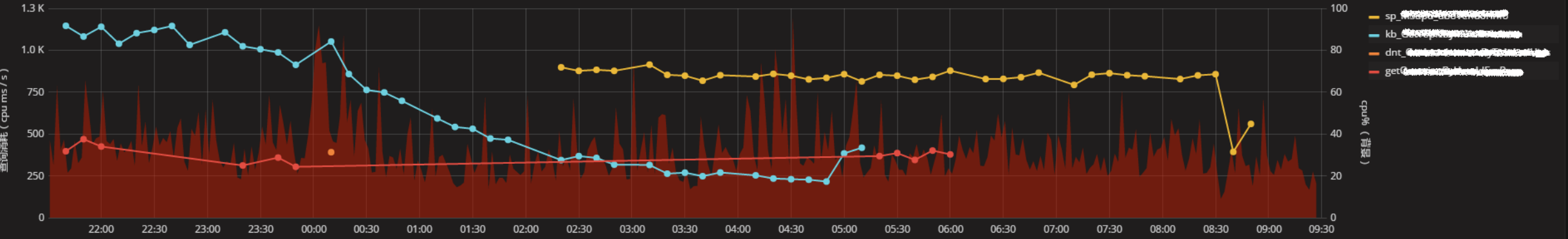
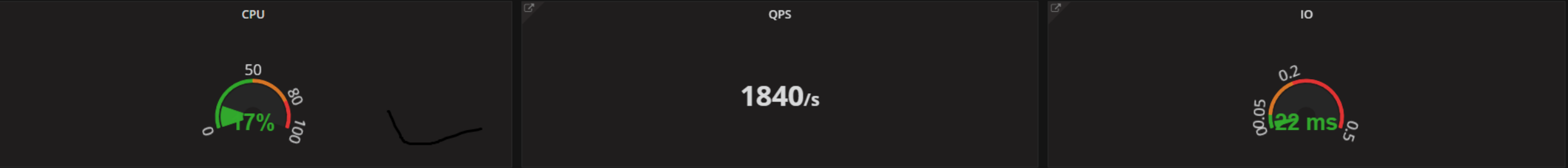
# 监控方案示例



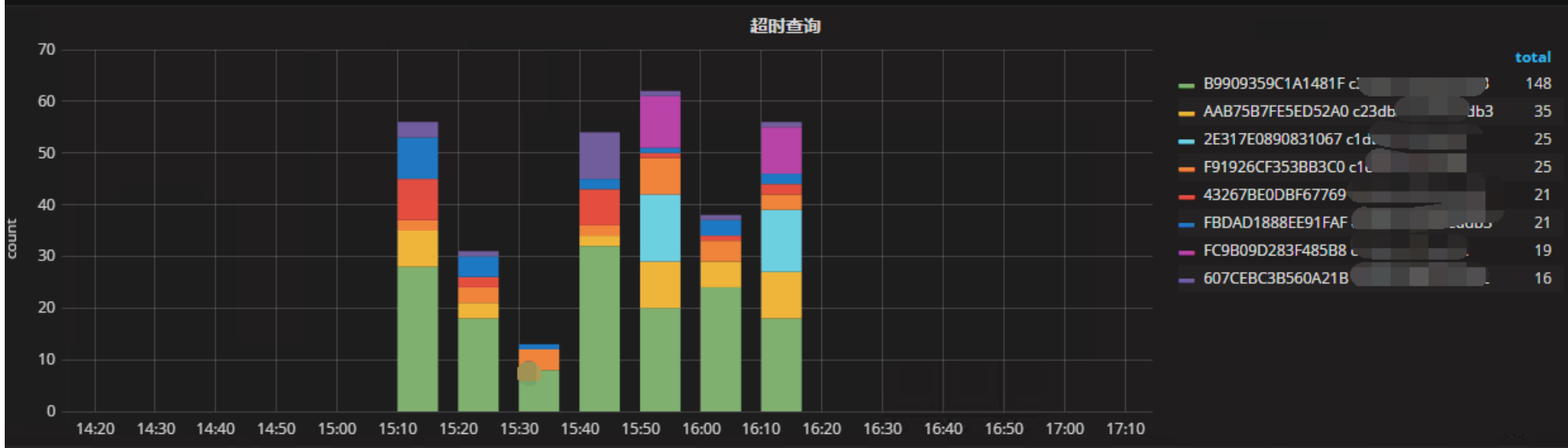




slow query							
logdatetime	sname	db_name	stmt	cpu_ms/s	lio/s	pio/s	cnt/s
2017-07-06 00:50:00	57BA2ED5AFAB42ED	MobileData	SELECT @iCount = COUNT(1) FROM queryRecord WITH(nolock) WHERE 1=1 AND CiytlD=@LocationID AND CslD=@CarSerialID	4 K	1 Mil	0	16
2017-07-06 00:30:00	57BA2ED5AFAB42ED	MobileData	SELECT @iCount = COUNT(1) FROM queryRecord WITH(nolock) WHERE 1=1 AND CiytlD=@LocationID AND CslD=@CarSerialID	3 K	1 Mil	0	16
2017-07-06 08:50:00	57BA2ED5AFAB42ED	MobileData	SELECT @iCount = COUNT(1) FROM queryRecord WITH(nolock) WHERE 1=1 AND CiytlD=@LocationID AND CslD=@CarSerialID	3 K	2 Mil	0	16
2017-07-06 09:20:00	57BA2ED5AFAB42ED	MobileData	SELECT @iCount = COUNT(1) FROM queryRecord WITH(nolock) WHERE 1=1 AND CiytlD=@LocationID AND CslD=@CarSerialID	3 K	2 Mil	0	16
2017-07-06 00:20:00	57BA2ED5AFAB42ED	MobileData	SELECT @iCount = COUNT(1) FROM queryRecord WITH(nolock) WHERE 1=1 AND CiytlD=@LocationID AND CslD=@CarSerialID	3 K	1 Mil	0	16
2017-07-06 09:00:00	57BA2ED5AFAB42ED	MobileData	SELECT @iCount = COUNT(1) FROM queryRecord WITH(nolock) WHERE 1=1 AND CiytlD=@LocationID AND CslD=@CarSerialID	3 K	2 Mil	0	15
2017-07-06 01:10:00	57BA2ED5AFAB42ED	MobileData	SELECT @iCount = COUNT(1) FROM queryRecord WITH(nolock) WHERE 1=1 AND CiytlD=@LocationID AND CslD=@CarSerialID	3 K	1 Mil	0	15



slow query						
logdatetime	sptime	db_name	stmt	cpu_ms/s	lio/s	pio/s cnt/s
2017-07-12 21:40:00	kb_GetTopicByMultiCondition	db_name	SELECT TOP ( @PageSize ) A.[Guid] FROM dbo.topics AS A WITH ( NOLOCK ) INNER JOIN dbo.topics AS B WITH ( NOLOCK ) ON A.id = B.id WHERE B.[is_active] = 1 AND B.[model_id] = @SerialId ORDER BY B.[id] DESC , B.[haowen_votes] DESC , B.[posts_count] DESC	1 K	395 K	0 5
2017-07-12 22:40:00	kb_GetTopicByMultiCondition	db_name	SELECT TOP ( @PageSize ) A.[Guid] FROM dbo.topics AS A WITH ( NOLOCK ) INNER JOIN dbo.topics AS B WITH ( NOLOCK ) ON A.id = B.id WHERE B.[is_active] = 1 AND B.[model_id] = @SerialId ORDER BY B.[id] DESC , B.[haowen_votes] DESC , B.[posts_count] DESC	1 K	400 K	0 5
2017-07-12 22:00:00	kb_GetTopicByMultiCondition	db_name	SELECT TOP ( @PageSize ) A.[Guid] FROM dbo.topics AS A WITH ( NOLOCK ) INNER JOIN dbo.topics AS B WITH ( NOLOCK ) ON A.id = B.id WHERE B.[is_active] = 1 AND B.[model_id] = @SerialId ORDER BY B.[id] DESC , B.[haowen_votes] DESC , B.[posts_count] DESC	1 K	399 K	0 5
2017-07-12 22:30:00	kb_GetTopicByMultiCondition	db_name	SELECT TOP ( @PageSize ) A.[Guid] FROM dbo.topics AS A WITH ( NOLOCK ) INNER JOIN dbo.topics AS B WITH ( NOLOCK ) ON A.id = B.id WHERE B.[is_active] = 1 AND B.[model_id] = @SerialId ORDER BY B.[id] DESC , B.[haowen_votes] DESC , B.[posts_count] DESC	1 K	401 K	0 5
2017-07-12 23:10:00	kb_GetTopicByMultiCondition	db_name	SELECT TOP ( @PageSize ) A.[Guid] FROM dbo.topics AS A WITH ( NOLOCK ) INNER JOIN dbo.topics AS B WITH ( NOLOCK ) ON A.id = B.id WHERE B.[is_active] = 1 AND B.[model_id] = @SerialId ORDER BY B.[id] DESC , B.[haowen_votes] DESC , B.[posts_count] DESC	1 K	400 K	0 5
2017-07-12 22:20:00	kb_GetTopicByMultiCondition	db_name	SELECT TOP ( @PageSize ) A.[Guid] FROM dbo.topics AS A WITH ( NOLOCK ) INNER JOIN dbo.topics AS B WITH ( NOLOCK ) ON A.id = B.id WHERE B.[is_active] = 1 AND B.[model_id] = @SerialId ORDER BY B.[id] DESC , B.[haowen_votes] DESC , B.[posts_count] DESC	1 K	390 K	0 5



logdatetime	servername	dbname	stmt
2017-07-16 16:15:23	c4dbsrv80\ucardb2		WHFRE UserId=@UserId AND C...
2017-07-16 16:15:23			s as b with(nolock) on a.LeadsID=b.LeadsID and b
2017-07-16 16:15:18		Autocare...	] as sr on sr.H5SId = s.H5SId where sr.cs_Id = @Sei
2017-07-16 16:15:18			SELECT a.carId FRC ...Time&@FuelT... 0 AND a.
2017-07-16 16:15:17			SELECT a.ca ...ID a.FuelType&@FuelType > (

# 邮件报警示例

## 复制延迟

收件人: dbas

agent\_id : 344  
delay\_minute : 2882  
delay\_cmds : 5906536

agent\_id : 408  
delay\_minute : 2857  
delay\_cmds : 1114846

agent\_id : 454  
delay\_minute : 2885  
delay\_cmds : 69744

## 作业错误(死锁牺牲作业周累计)@ [redacted]

收件人: dbas

jobname : [redacted]  
endtime : 2017-07-09 13:32:25  
duration : 00:02:25  
schedule : every 10 minute(s) between 0:00 and 23:59  
message : 已以用户 SCOM\ClusService 的身份执行。 事务(进程 ID 427)与另一个进程被死锁在 锁 资源上, 并且已被选 (错误 1205)。该步骤失败。

jobname : [redacted]  
endtime : 2017-07-12 10:09:00  
duration : 00:28:00  
schedule : every 7 minute(s) between 7:00 and 15:00  
message : 已以用户 SCOM\cluservice 的身份执行。 事务(进程 ID 514)与另一个进程被死锁在 锁 资源上, 并且已被选 (错误 1205)。该步骤失败。

jobname : [redacted]  
endtime : 2017-07-12 10:09:00  
duration : 00:28:00  
schedule : every 34 minute(s) between 15:00 and 18:00  
message : 已以用户 SCOM\cluservice 的身份执行。 事务(进程 ID 514)与另一个进程被死锁在 锁 资源上, 并且已被选 (错误 1205)。该步骤失败。

## 发现长时间运行作业报警

### 长时间运行作业报警 [redacted]

收件人: 张 [redacted]; dbas

此邮件已经归档。 [查看原始项](#)

通知作业负责人  
抄送DBA

服务器: C5168B3F32 (C5168B3F32)

作业名称: [redacted]  
作业开始时间: 04/02/2017 03:00:01  
作业已执行时间: 3小时38分钟37秒  
作业历史平均执行时间: 1小时52分钟56秒  
作业描述: mailto:{[redacted]}@bitauto.com}

largelO@C/ [redacted] >> Ma [redacted] >> p\_Get\_M  
asterBandInSale, query\_hash:0xD913529F78EB3FD6, 每秒io:278096

收件人: dbas

此邮件已经归档。 [查看原始项](#)

```
SELECT DISTINCT m.* FROM MaiCheBase.dbo.CarMasterBrand_Full m WITH ( NOLOCK ) WHERE  
m.M [redacted] MaiCheBase.dbo.CarSource s WITH ( NOLOCK )  
JOIN MaiCheBase.dbo [redacted] Inner  
dealerInfo d ON ( [redacted] dealerId=d.dealerId INNER JOIN [redacted] Base.dbo.Dealer d  
ON ( [redacted] dealerId = d.DealerId WHERE [redacted] @cityId UNION SELECT  
cc.MasterBrand [redacted] WHERE  
cc.Bus [redacted] cc.Platform [redacted] Form )
```

# 5# 标准化

---



# 标准化参考

实例配置	最大并行度、并行开销阈值等
Windows配置	锁定内存页、执行卷维护任务、磁盘簇大小等
TempDB	个数和增长
数据库配置	数据大小和增长、兼容级别、日志延迟写、自动更新统计信息等
通用脚本	一些维护脚本、查看当前DB状态脚本、备份还原脚本等
扩展事件等	捕捉超时、捕捉慢查询、审核等
备份机制	根据RTO/RPO选择策略

## 标准化优势

- 避免遗漏设置导致的问题
- 避免没必要的人工成本
- 降低运维水平需求，初级运维人员可以利用现有脚本完成高级功能

# 标准化实现机制

- 虚拟化使用模板完成标准化（Hyper-V/Vmware）
- 实体机使用初始化脚本（PowerShell/Python等）
- 统一部署脚本将新实例的相关元数据存入统一优化流程

# 标准化脚本示例

dd hh:mm:ss.mss	session_id	sql_text	login_name	wait_info
107.19.56.40.000	176	<?query -- sp_server_diagnostics --?>	NT AUTHORITY\SYSTEM	(780ms)SP_SERVER_DIAGNOSTICS_SLEEP
001.05.35:13.963	827	<?query -- begin tran --?>	SQLSERVER\...	NULL
001.05.35:12.800	825	<?query -- begin tran --?>	SQLSERVER\...	NULL
001.05.35:12.780	816	<?query -- begin tran --?>	SQLSERVER\...	NULL
001.05.35:11.516	820	<?query -- begin tran --?>	SQLSERVER\...	NULL
001.05.35:11.206	822	<?query -- begin tran --?>	SQLSERVER\...	NULL
000.00.00:05.323	269	<?query -- backup database [f...] --?>	sa	(1121ms)ASYNC_IO_COMPLETION
000.00.00:03.476	239	<?query -- begin tran --?>	SQLSERVER\...	NULL
000.00.00:01.006	828	<?query -- begin tran --?>	SQLSERVER\...	NULL
000.00.00:00.806	823	<?query -- begin tran --?>	SQLSERVER\...	NULL
000.00.00:00.163	288	<?query -- --=====	SQLSERVER\...	NULL
000.00.00:00.093	449	<?query -- SELECT pid FROM sys.dm_exec_sessions --?>	SQLSERVER\...	(33ms)PAGEIOLATCH_SH:BitautoForum:1(*)
000.00.00:00.093	190	<?query -- SELECT pid FROM sys.dm_exec_sessions --?>	SQLSERVER\...	(28ms)PAGEIOLATCH_SH:BitautoForum:1(*)
000.00.00:00.056	629	<?query -- --=====	wmside	NULL

event_timestamp	client_hostname	client_app_name	database	object_name	sql_statement_completed	sql_text	username
2017-07-14 17:50:31.467		Microsoft SQL Server Management Studio	Adventureworks	dbo.tblProduct	use [Adventureworks]	exec sp_addarticle @Publication = ...	sa
2017-07-14 17:50:30.623		Microsoft SQL Server Management Studio	Adventureworks	dbo.tblProduct	use [Adventureworks]	exec sp_addarticle @Publication = ...	sa
2017-07-14 17:50:24.117	B...	Microsoft SQL Server Management Studio	Adventureworks	dbo.tblProduct	use [Adventureworks]	exec sp_addarticle @Publication = ...	sa
2017-07-11 08:05:7.833	N...	Microsoft SQL Server Management Studio - 查询	Adventureworks	dbo.tblProduct	CREATE TABLE [dbo].[tblProduct]	CREATE TABLE [dbo].[tblProduct]	sa
2017-07-06 16:28:28.457	J...	Microsoft SQL Server Management Studio - Query	Adventureworks	dbo.tblProduct	CREATE TABLE [dbo].[tblProduct]	CREATE TABLE [dbo].[tblProduct]	sa
2017-07-04 17:06:5.783	J...	Microsoft SQL Server Management Studio - 查询	Adventureworks	dbo.tblProduct	CREATE TABLE [dbo].[tblProduct]	CREATE TABLE [dbo].[tblProduct]	sa
2017-07-02 04:45:07.240		Net SqlClient Data Provider	Adventureworks	dbo.tblProduct	CREATE TABLE [dbo].[tblProduct]	CREATE TABLE [dbo].[tblProduct]	sa
2017-07-02 04:45:07.037		Net SqlClient Data Provider	Adventureworks	dbo.tblProduct	CREATE TABLE [dbo].[tblProduct]	CREATE TABLE [dbo].[tblProduct]	sa
2017-07-02 04:45:06.973		Net SqlClient Data Provider	Adventureworks	dbo.tblProduct	CREATE TABLE [dbo].[tblProduct]	CREATE TABLE [dbo].[tblProduct]	sa
2017-07-02 04:45:06.897		Net SqlClient Data Provider	Adventureworks	dbo.tblProduct	CREATE TABLE [dbo].[tblProduct]	CREATE TABLE [dbo].[tblProduct]	sa
2017-07-02 04:45:06.833	AA...	Net SqlClient Data Provider	Adventureworks	dbo.tblProduct	CREATE TABLE [dbo].[tblProduct]	CREATE TABLE [dbo].[tblProduct]	sa
2017-07-02 04:45:06.630	AV...	Net SqlClient Data Provider	Adventureworks	dbo.tblProduct	CREATE TABLE [dbo].[tblProduct]	CREATE TABLE [dbo].[tblProduct]	sa
2017-07-02 04:45:06.350	AV...	Net SqlClient Data Provider	Adventureworks	dbo.tblProduct	CREATE TABLE [dbo].[tblProduct]	CREATE TABLE [dbo].[tblProduct]	sa
2017-07-02 04:45:05.820		Net SqlClient Data Provider	Adventureworks	dbo.tblProduct	CREATE TABLE [dbo].[tblProduct]	CREATE TABLE [dbo].[tblProduct]	sa
2017-07-02 04:45:05.367	V...	Net SqlClient Data Provider	Adventureworks	dbo.tblProduct	CREATE TABLE [dbo].[tblProduct]	CREATE TABLE [dbo].[tblProduct]	sa

```

sp_restoreb tiredb

0% <

Messages

IF EXISTS(SELECT name FROM sys.databases WHERE name='TiredB')
BEGIN
RAISERROR (' 已有同名数据库, 如需要强制清除删除数据库可',20,1,1) with log
end
ALTER DATABASE tiredb SET SINGLE_USER WITH ROLLBACK IMMEDIATE;
restore database tiredb from disk=N'\\10.42.130.1\sqlf\backups\TiredB_20170626141310.trn' with NORECOVERY, REPLACE, STATS=10, Move 'TiredB_20170626141310.trn' to 'y:\data\tiredb\Move'
restore log tiredb from disk=N'\\10.42.130.1\sqlf\backups\TiredB_20170626141310.trn' with NORECOVERY, STATS=10
restore log tiredb from disk=N'\\10.42.130.1\sqlf\backups\TiredB_20170626141422.trn' with NORECOVERY, STATS=10
restore log tiredb from disk=N'\\10.42.130.1\sqlf\backups\TiredB_20170626182211.trn' with NORECOVERY, STATS=10
restore log tiredb from disk=N'\\10.42.130.1\sqlf\backups\TiredB_20170626200841.trn' with NORECOVERY, STATS=10
restore log tiredb from disk=N'\\10.42.130.1\sqlf\backups\TiredB_20170626210630.trn' with NORECOVERY, STATS=10
restore log tiredb from disk=N'\\10.42.130.1\sqlf\backups\TiredB_20170626220617.trn' with NORECOVERY, STATS=10
restore log tiredb from disk=N'\\10.42.130.1\sqlf\backups\TiredB_20170626230706.trn' with NORECOVERY, STATS=10
restore log tiredb from disk=N'\\10.42.130.1\sqlf\backups\TiredB_20170627000920.trn' with NORECOVERY, STATS=10
restore log tiredb from disk=N'\\10.42.130.1\sqlf\backups\TiredB_20170627021802.trn' with NORECOVERY, STATS=10
restore log tiredb from disk=N'\\10.42.130.1\sqlf\backups\TiredB_20170627041537.trn' with NORECOVERY, STATS=10
restore log tiredb from disk=N'\\10.42.130.1\sqlf\backups\TiredB_20170627061936.trn' with NORECOVERY, STATS=10
restore log tiredb from disk=N'\\10.42.130.1\sqlf\backups\TiredB_20170627091544.trn' with NORECOVERY, STATS=10
restore log tiredb from disk=N'\\10.42.130.1\sqlf\backups\TiredB_20170627101607.trn' with NORECOVERY, STATS=10
restore log tiredb from disk=N'\\10.42.130.1\sqlf\backups\TiredB_20170627121323.trn' with NORECOVERY, STATS=10
restore log tiredb from disk=N'\\10.42.130.1\sqlf\backups\TiredB_20170627141607.trn' with NORECOVERY, STATS=10
restore log tiredb from disk=N'\\10.42.130.1\sqlf\backups\TiredB_20170627161658.trn' with NORECOVERY, STATS=10
restore log tiredb from disk=N'\\10.42.130.1\sqlf\backups\TiredB_20170627181443.trn' with NORECOVERY, STATS=10
restore log tiredb from disk=N'\\10.42.130.1\sqlf\backups\TiredB_20170627190811.trn' with NORECOVERY, STATS=10

```

[illegible]

## 6# 还原策略

---

## 还原策略

- 注意到我说的不是备份策略
- RTO (Recovery Time Object)
- RPO (Recovery Point Object)

## 备份类型

- 完整备份
- 增量备份
  - 差异备份（基于变更数据，还原快，无法还原到指定时间点）
  - 日志备份（基于）
- 当RTO和RPO要求很高的时候，就不能再基于冷备



## 其他原则

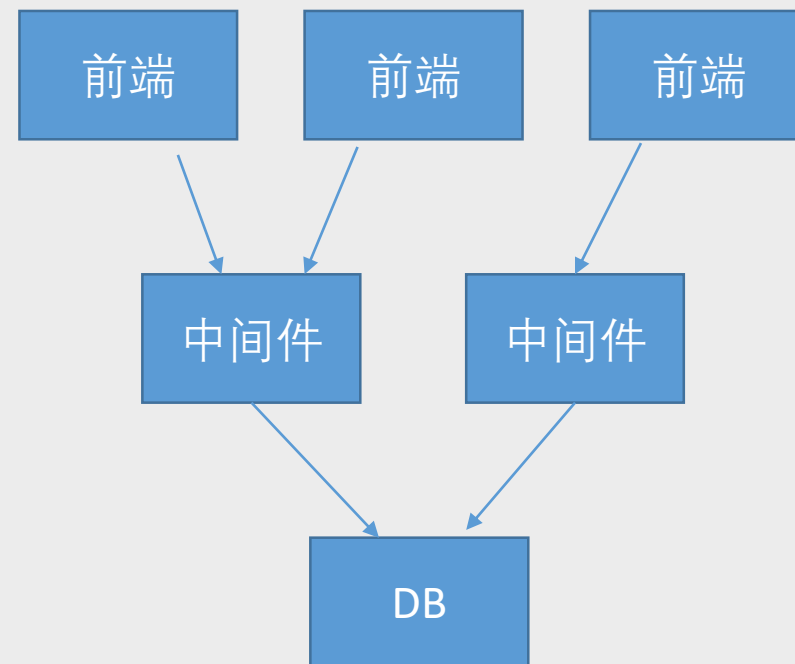
- 数据库之间弱耦合
  - 避免数据库之间直接进行数据同步
  - 依赖中间层
- 业务无关数据库尽量拆分到不同机器
  - 单台大机器N个数据库 vs 单大宿主机多个虚拟机
- 按照业务要求，定期数据归档
- 尽量微服务，一个业务实体不应该在数据库中调用另一个业务实体

## 7# SQL书写原则

---

# 书写原则

- 简单
- DB做简单的CRUD
- 业务操作在DB之外



# SQL 书写原则

General Guideline

## 参数化

数据类型统一，尽量避免对列做操作

避免使用任何ORM系统

简单、简单、简单，重要的事情说三遍

防止SQL注入

高并发下防止重编译（大量降低CPU）

# SQL 书写原则

General Guideline

## 参数化

数据类型统一，尽量避免对列做操作

避免使用任何ORM系统

简单、简单、简单，重要的事情说三遍

数据类型转换可能导致低效的执行计划

列上数据类型转换导致无法使用Seek

数据类型不一致导致隐式转换

# SQL 书写原则

General Guideline

参数化

数据类型统一，尽量避免对列做操作

避免使用任何ORM系统

简单、简单、简单，重要的事情说三遍

ORM对于跨数据库开发效率高

ORM往往导致复杂低效的SQL

# SQL 书写原则

General Guideline

## 参数化

数据类型统一，尽量避免对列做操作

避免使用任何ORM系统

简单、简单、简单，重要的事情说三遍

任何复杂SQL都应该分解为简单SQL

尽量避免相关子查询

存储过程中有逻辑分支是灾难

多表Join将中间结果存入带索引的临时表

如果可能，使用微服务架构，不同实体的数据  
不应该在DB层面有关联



# 写在最后

- 尽量工程化、标准化主动预防问题
- 在问题出现征兆之前，主动解决
- 能够避免很多午夜Surprise



Q A