

# Blink

## 洞察数据，洞悉未来

# Blink - 独享模式

UDF

- 网络及物理资源层面完全隔离
- 可开放UDF/DataStream等更底层的API

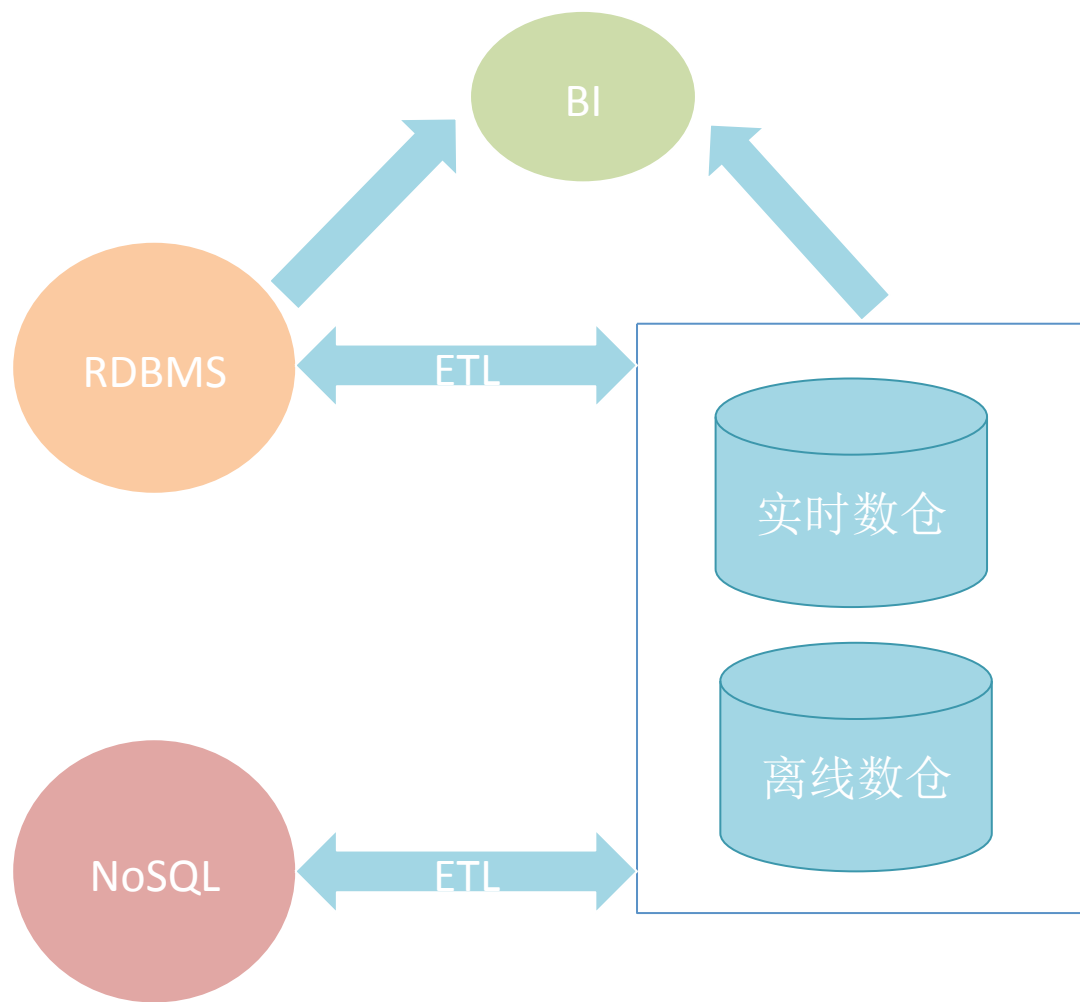
VPC

无缝对接用户VPC下的各种服务

Data Lake

- 数据分析
- 数据清洗
- 数据同步

# Data Warehouse



数据来源

应用或交易系统产生的结构化数据

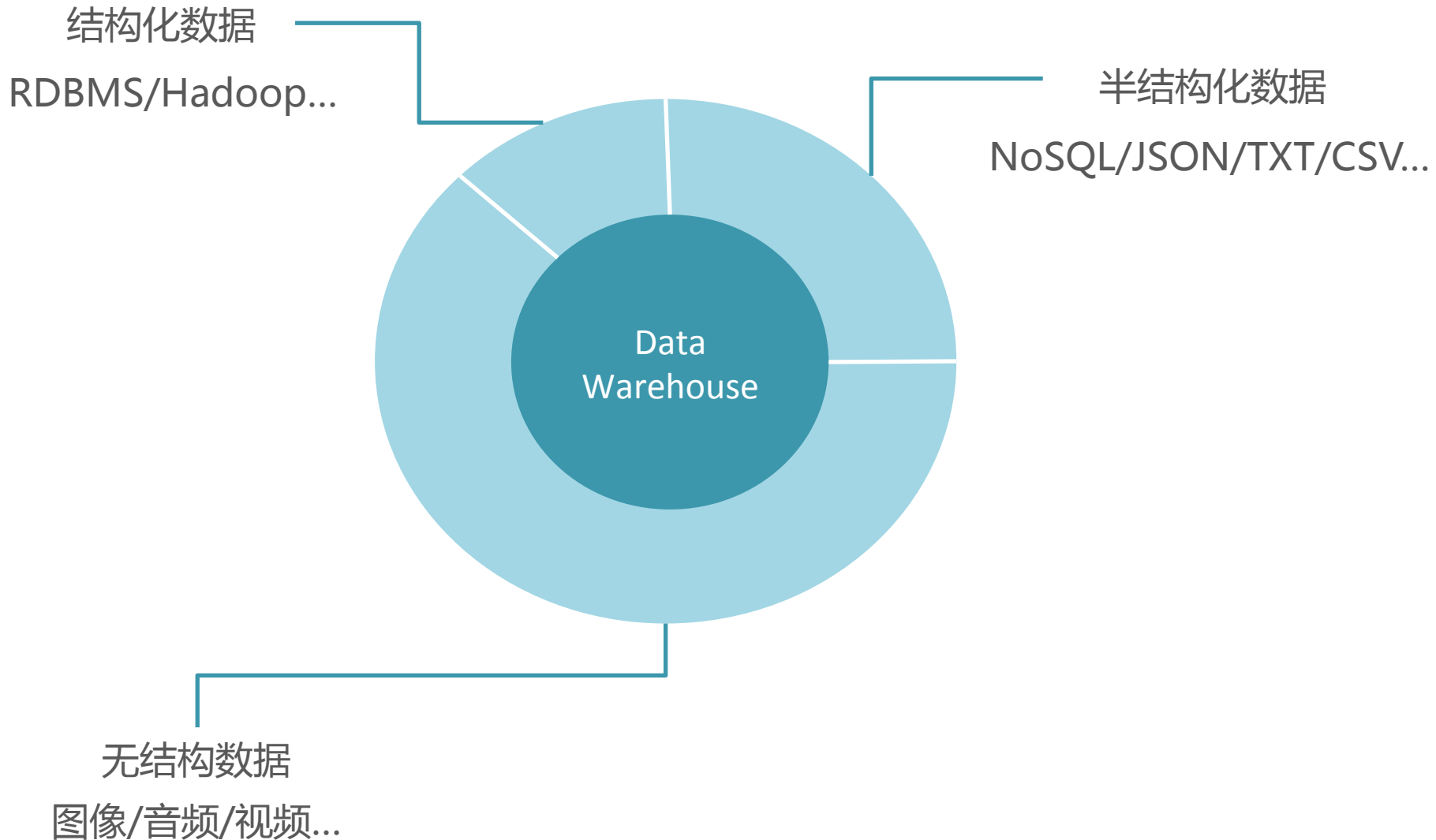
数据特点

- 结构固定
- 可提前定义
- 数据已经经过清洗，转换等ETL处理

数据作用

业务报表  
数据分析

# Data Lake



# Data Lake vs. Data Warehouse

## 数据存储多样化

- RDBMS : RDS/HybridDB/ADS...
- 数仓 : Hive/MaxCompute...
- 数据通道 : SLS/MQ/Kafka...
- 文件存储 : OSS/HDFS...

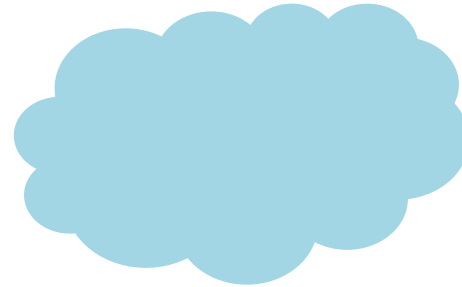
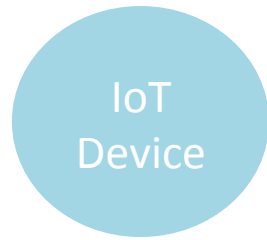
## 数据格式多样化

- RDBMS中结构化数据
- JSON/XML/CSV等半结构化数据
- 图像/音频/视频等非结构化数据

## 数据计算多样化

- 数据Schema不固定
- 异构数据源计算，无需将数据导入统一存储
- 支持SQL进行大数据分析，同时也可直接对数据进行全文检索/机器学习等。

# 场景分析 - IoT

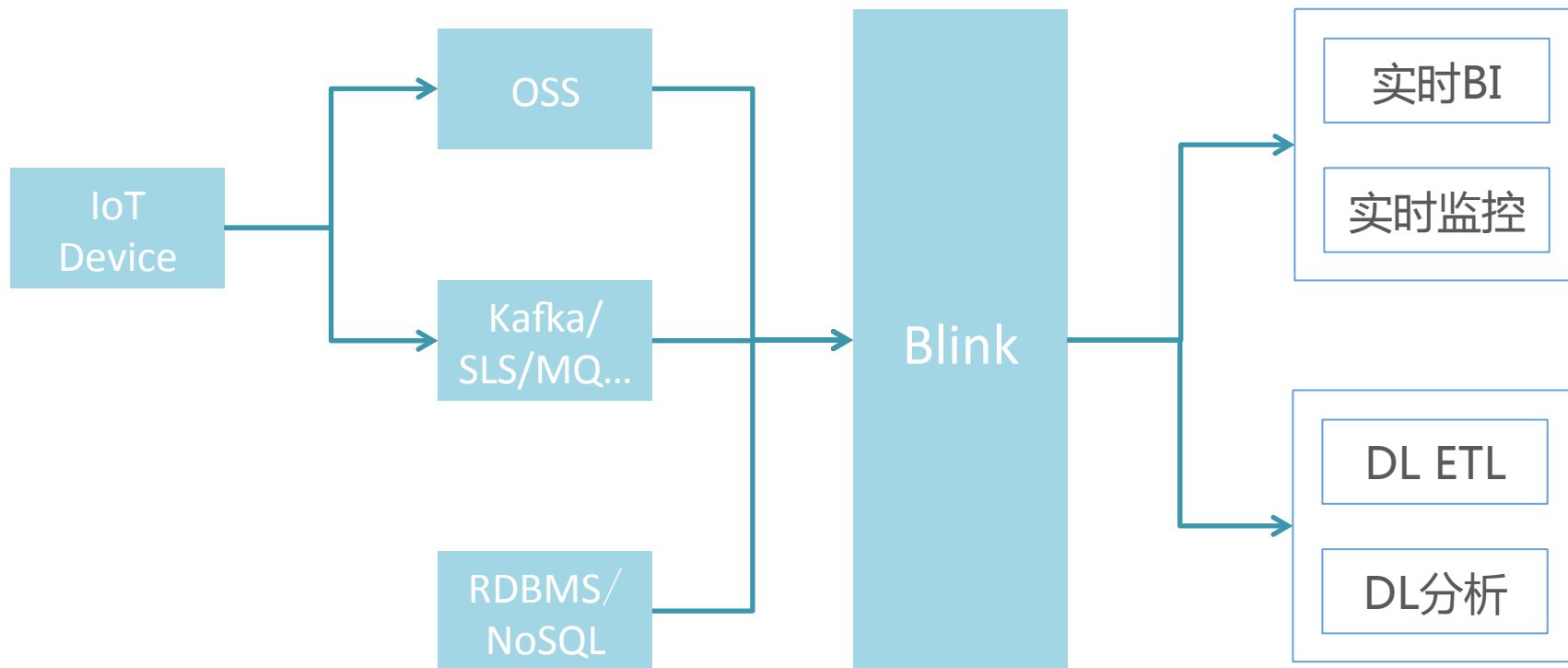


厂房中各点的

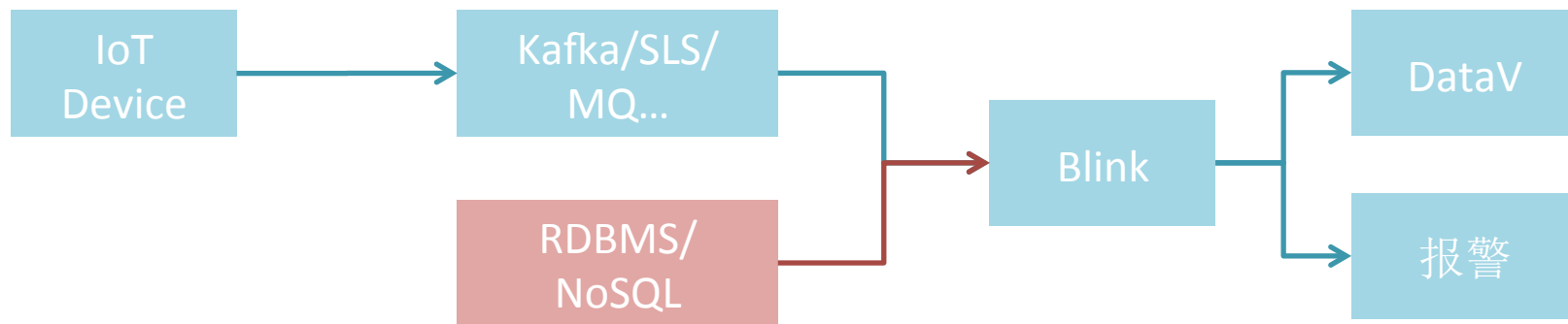
- 温度
- 湿度
- 压力等信息

- 环境监控
- 数据分析

# 场景分析 - IoT



# IoT场景 - 实时链路



实时BI

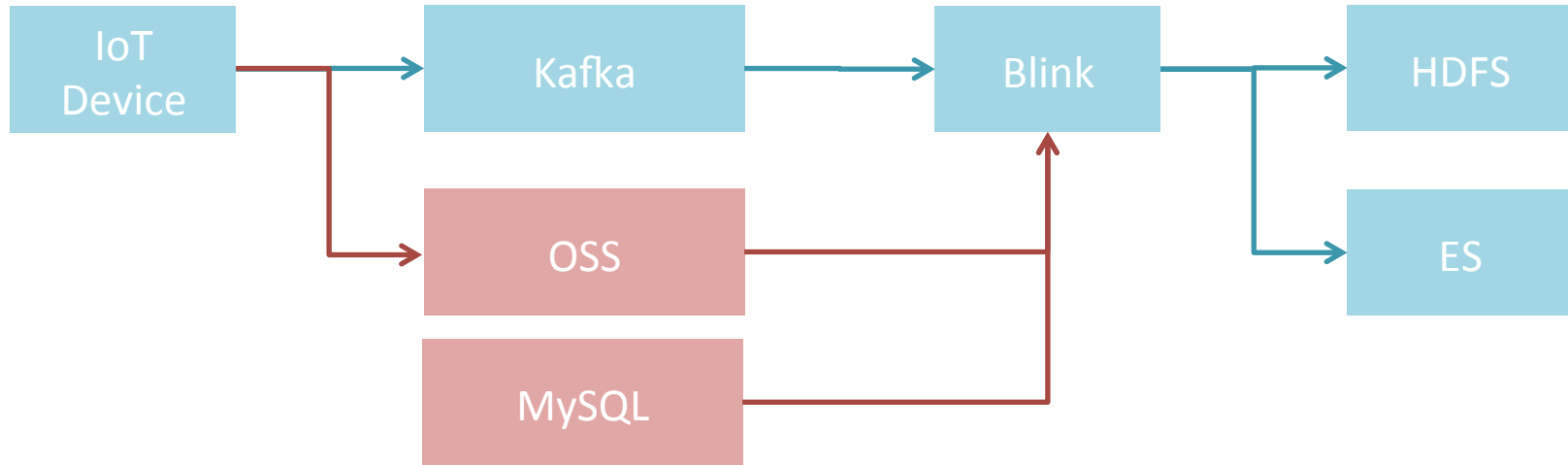
- 实时查看在线设备状态，设备个数
- 实时感知工厂各区域温度/湿度/压力等报警情况
- 重要设备的在线情况查看（流式维表Join）
- 工厂重要区域指标查看（流式维表Join）

实时监控 CEP

- 设备状态异常报警
- 厂区各监控指标(温度/湿度/压力)等报警



# IoT场景 – Data Lake ETL&Analysis



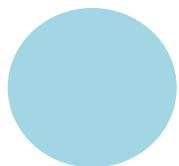
## Data Lake ETL

将IoT采集到的数据经过清洗，上传到数仓中做持久存储及分析

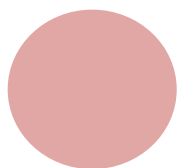
## Data Lake Analysis

- 通过Blink批处理功能可直接通过SQL查询OSS上日志文件
- 将OSS上设备运行状态日志与MySQL中数据关联并聚合，计算出关键设备某段时间的运行状态。

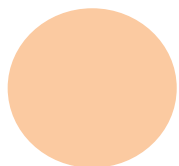
# Blink优势



实时/数据湖场景下ETL/分析/监控等功能集为一体



20+种connector，无缝对接各种类型数据源  
<RDS/Hbase/Redis/Kafka/Hive...>



异构数据源计算  
使用标准SQL，对异构数据源<RDBMS/NoSQL/Queue/Hive...>  
直接计算<JOIN/UNION/SELECT...>

# Blink优势

开发  
成本

一套SQL，实  
时/离线无缝切  
换

运维  
成本

全托管模式，  
保证高SLA情  
况下，减轻用  
户运维负担

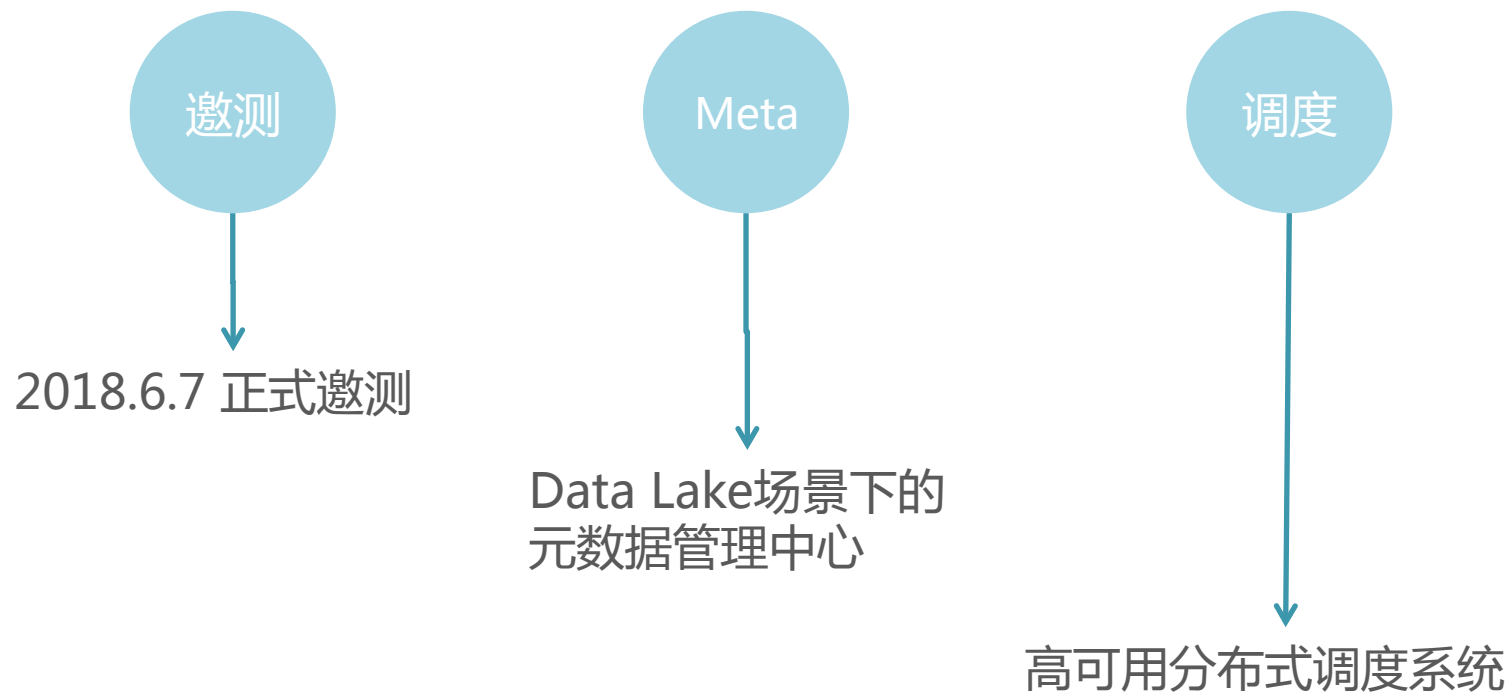
存储  
成本

直接分析异构数据  
源中，结构化/半  
结构化数据，对低  
频查询类场景，可  
大大节省存储成本

计算  
成本

一个产品，支持多  
种计算能力，降低  
用户计算成本

# Blink Data Lake Road Map



# Blink – 不止于流

更低的  
成本

更灵活  
的方式

更短的  
时间



洞悉数据  
洞察未来