



# ODPS SQL

# 说明

- 本PPT适合以下用户：
  - 需要总体上了解ODPS SQL功能的用户
- 本PPT不涉及ODPS的详细使用说明。如需了解使用细节，请参考ODPS的说明文档。
- 使用说明以ODPS功能说明文档为准。



# 基本概念

- Project
  - ODPS的所有对象都隶属于项目空间
  - 计量计费及安全控制的基本单元
- Table
  - 数据存储存储在表中
  - 支持Partition表
  - 支持视图
  - 数据类型：Bigint, Double, Boolean, String, Datetime
- Task/Instance
  - 所有的ODPS计算任务都被转化为Instance来执行，用户可以查询Instance的查询状态及其他运行信息



# SQL

- DDL
  - Create/Drop/Alter操作
- DML
  - Insert overwrite/into
  - Select ... from .. where.. Group by|order by|distribute by | sort by | limit;
  - 子查询
  - Union all
  - Join/Mapjoin
- 内建函数


# SQL-DDL

- 创建表：
  - 示例：create table src\_part(key string, value string) partitioned by(pt string);
- 添加分区：
  - alter table src\_part add partition(pt='a');
- 删除表：
  - drop table src\_part;

# SQL-Insert

- 单路输出 ( Insert overwrite/Into ) :
  - 示例 : insert overwrite table src\_part partition(pt='a') select 'k', 'v' from dual;

s	i
a	1
b	3
b	2
a	4



s	i
a	1
b	3
b	2
a	4

# SQL-Insert

- 多路输出：

- 示例：

from dual

insert overwrite table src1 select s

insert overwrite table src2 select s;

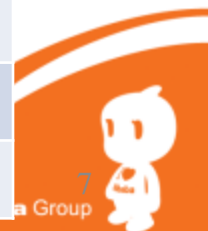
s	i
a	1
b	3
b	2
a	4



s	i
a	1
b	3
b	2
a	4




s	i
a	1
b	3
b	2
a	4



# SQL-Select

Select s, i from src distibte by s sort by i; --依照s的值，进行局部排序


s	i
a	1
b	3
b	2
a	4



s	i
a	1
a	4
b	2
b	3

Select s, I from tbl order by I limit 10; -- 全局排序，必须与limit联用

s	i
a	1
b	3
b	2
a	4



s	i
a	1
b	2
b	3
a	4

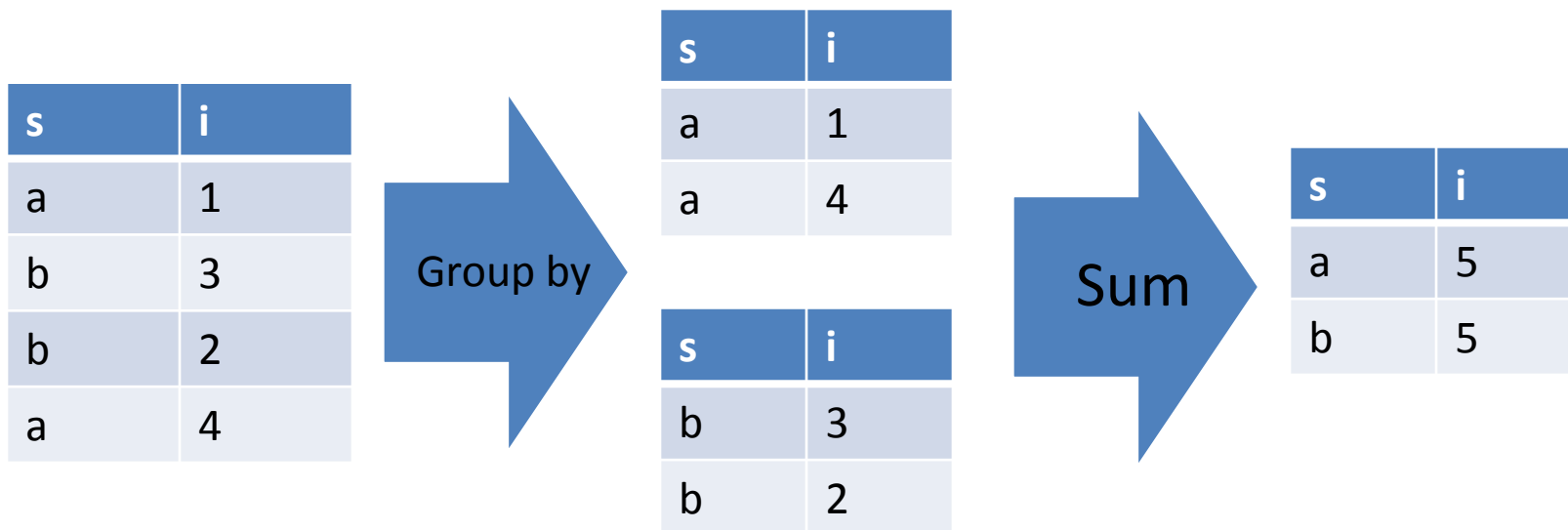




# SQL-Select

Group by:

- 通常与聚合函数联用
- 示例: `Select s, sum(i) from tbl group by s;`



# SQL-Union all

- Union all:合并多个查询的结果
  - 示例:select \* from (select s from src1 union all select s from src2) t;

s	i
a	1
b	3
b	2
a	4

s	i
a	1
b	3
b	2
a	4



s
a
b
b
a
a
b
b
a



# SQL-Join

- Join:将数据按照一定条件进行连接
  - 示例：select j\_left.key, j\_right.s, j\_left.value, j\_right.i from j\_left full outer join j\_right on j\_left.key=j\_right.s;
  - Mapjoin:join操作的一种优化，执行速度更快，在数据量较小时可以使用；

key	value
a	1
b	3

s	i
a	2
c	3



key	s	value	i
a	a	1	2
b	null	3	null
null	c	null	3

# SQL-内建函数

- 值函数
  - 数学运算函数
  - 字符串处理函数
    - 正则函数
  - 日期类型处理函数
- 窗口函数
  - count, avg, max, min, median, stddev, sum ...
- 聚合函数
  - count, avg, max, min, median, stddev, sum ...



# 应用限制

- 表中支持生命周期
- 支持视图（ View ）
- 应用限制：
  - Union all目前不能超过128路；
  - Where中的in判断中有子查询，目前返回结果不能超过1000个；
  - 目前最多支持16路join；
  - 多路插入操作目前最多支持128路同时插入；
  - 当使用select屏显时，目前最多显示1000行；但select作为子句时，没有这个显示，会将全部查询结果返回给上层查询；