

2019

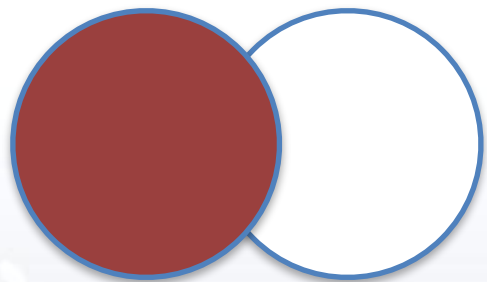
A low-angle, upward-looking shot of a modern glass skyscraper with a grid-like facade, set against a clear blue sky. The building's reflection is visible in the lower part of the frame.

企业级数据仓库实战

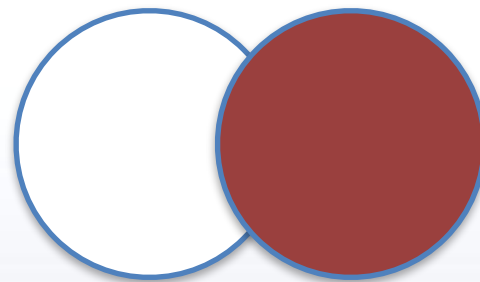
A light gray world map is centered in the background, composed of a grid of small dots. The map is partially obscured by the text and the building image above it.



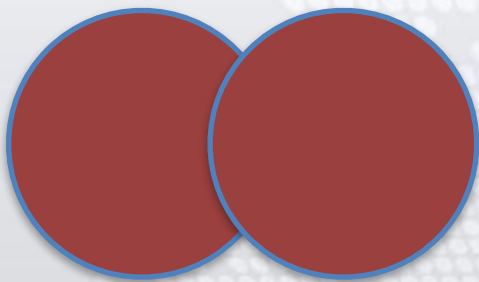
Hive 函数全攻略



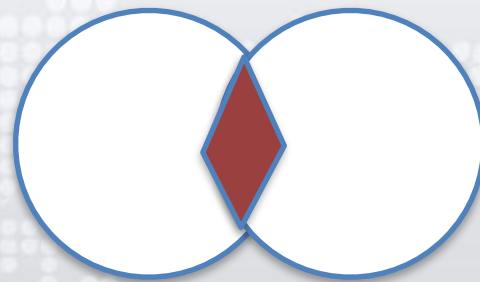
LEFT [OUTER] JOIN



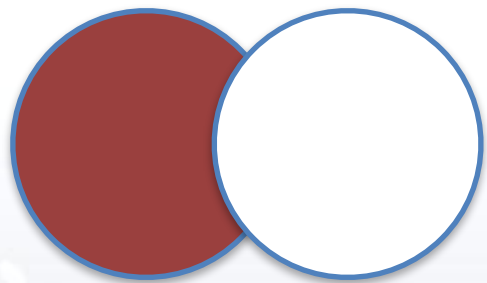
RIGHT [OUTER] JOIN



FULL (OUTER) JOIN



INNER JOIN



NOT IN



1、关系运算符(必会)

$A=B$ 、 $A>B$ 、 $A<B$ 、 $A\geq B$ 、 $A\leq B$ 、 $A \text{ IS NULL}$ 、 $A \text{ IS NOT NULL}$ 、 $A \text{ LIKE } B$ 、 $A \text{ RLIKE } B$

2、算术运算符(必会)

$A+B$ 、 $A-B$ 、 $A*B$ 、 A/B 、 $A\%B$

3、逻辑运算(必会)

$A \text{ AND } B$ 、 $A \text{ OR } B$ 、 $! A$

4、复杂类型操作(必会)

$A[n]$ 、 $M[\text{key}]$ 、 $S.x$



UDF(一个输入对一个输出)

数学函数: ROUND()、FLOOR()、RNAD()、ABS()

集合函数: SIZE()

类型转换: CAST()

日期函数: FROM_UNIXTIME()、UNIX_TIMESTAMP()、TO_DATE()、YEAR()、MONTH()、DAY()

DATEDIFF()、DATE_ADD()、DATE_SUB、ADD_MONTHS()、DATE_FORMAT()

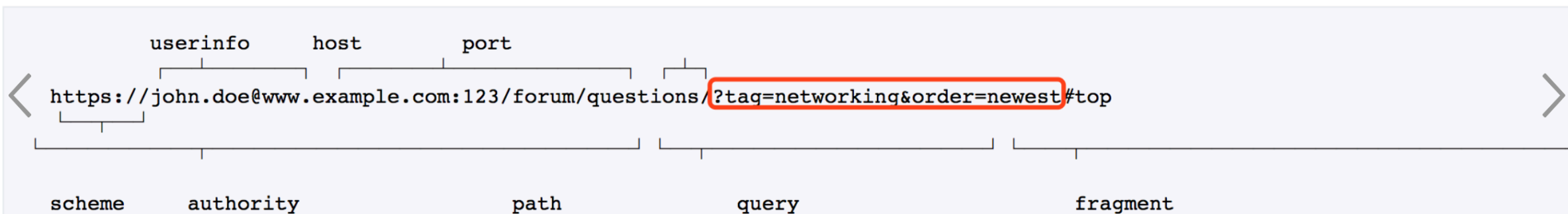
条件函数: IF()、NVL()、COALESCE()、CASE WHEN END

字符串函数: CONCAT()、CONCAT_WS()、GET_JSON_OBEJCT()、LENGTH()、LOWER()、
PARSE_URL()、REGEXP_REPLACE()、REPLACE()、SPLIT()、STR_TO_MAP()、SUBSTR()、
TRIM()



查阅相关文档https://en.wikipedia.org/wiki/Uniform_Resource_Identifier#Syntax

The following figure displays example URLs and their component parts.





UDAF(多个输入对一个输出)

COUNT()、SUM()、MIN()、MAX()、COLLECT_SET()

UDTF(一个输入多个输出)

EXPLODE()、JSON_TUPLE()、PARSE_URL_TUPLE



窗口函数

LEAD()、LAG()、FIRST_VALUE()、LAST_VALUE()

分析函数

ROW_NUMBER()



选择市面上任意一款移动出行APP作为研究对象，请分别从以下几个角度去思考如何构建一个数据仓库

- 1、数据分层模型
- 2、主题域划分
- 3、指标体系
- 4、中间层建设
- 5、数据架构选型
- 6、日常数仓运维

THANK YOU FOR YOUR GUIDANCE.

谢谢