## **SQL TO PIG**

## **Cheat Sheet**

We know that lots of people come to Apache Pig from a relational database background, so we compiled this handy translation from SQL concepts to their Pig equivalents.

SQL CONCEPT	SQL	PIG
SELECT	<pre>SELECT column_name,column_name FROM table_name;</pre>	<pre>FOREACH alias GENERATE column_name,     column_name;</pre>
SELECT *	<pre>SELECT * FROM table_name;</pre>	FOREACH alias GENERATE *;
DISTINCT	<pre>SELECT DISTINCT column_name,column_name FROM table_name;</pre>	<pre>DISTINCT(FOREACH alias    GENERATE column_name, column_name);</pre>
WHERE	SELECT column_name,column_name FROM table_name WHERE column_name operator value;	FOREACH (FILTER alias BY column_name operator value)  GENERATE column_name, column_name;
AND/OR	<pre> WHERE (column_name operator value1    AND column_name operator value2)    OR column_name operator value3;</pre>	FILTER alias BY (column_name operator value1  AND column_name operator value2)  OR column_name operator value3;
ORDER BY	ORDER BY column_name ASC DESC, column_name ASC DESC;	ORDER alias BY column_name ASC DESC, column_name ASC DESC;
TOP/LIMIT	SELECT TOP number column_name FROM table_name ORDER BY column_name ASC DESC;	<pre>TOP(number, column_index, alias);</pre>
	SELECT column_name FROM table_name ORDER BY column_name ASC DESC LIMIT number;	<pre>FOREACH (GROUP alias BY column_name)    GENERATE LIMIT alias number;</pre>
GROUP BY	<pre>SELECT function(column_name) FROM table GROUP BY column_name;</pre>	<pre>FOREACH (GROUP alias BY column_name)    GENERATE function(alias.column_name);</pre>
LIKE	WHERE column_name LIKE pattern;	<pre>FILTER alias BY REGEX_EXTRACT(column_name,     pattern, 1) IS NOT NULL;</pre>
IN	WHERE column_name IN (value1,value2,);	<pre>FILTER alias BY column_name IN     (value1, value2,);</pre>

SQL CONCEPT	SQL	PIG
JOIN	<pre>SELECT column_name(s) FROM table1 JOIN table2 ON table1.column_name=table2.column_name;</pre>	<pre>FOREACH (JOIN alias1 BY column_name,     alias2 BY column_name)     GENERATE column_name(s);</pre>
LEFT/RIGHT/FULL OUTER JOIN	SELECT column_name(s) FROM table1 LEFT RIGHT FULL OUTER JOIN table2 ON table1.column_name=table2.column_name;	<pre>FOREACH (JOIN alias1 BY column_name    LEFT RIGHT FULL, alias2 BY column_name)    GENERATE column_name(s);</pre>
UNION ALL	SELECT column_name(s) FROM table1 UNION ALL SELECT column_name(s) FROM table2;	UNION alias1, alias2;
AVG	<pre>SELECT AVG(column_name) FROM table_name;</pre>	<pre>FOREACH (GROUP alias ALL) GENERATE   AVG(alias.column_name);</pre>
COUNT	<pre>SELECT COUNT(column_name) FROM table_name;</pre>	FOREACH (GROUP alias ALL) GENERATE COUNT(alias);
COUNT DISTINCT	<pre>SELECT COUNT(DISTINCT column_name) FROM table_name;</pre>	<pre>FOREACH alias {     unique_column = DISTINCT column_name;     GENERATE COUNT(unique_column); };</pre>
MAX	<pre>SELECT MAX(column_name) FROM table_name;</pre>	<pre>FOREACH (GROUP alias ALL) GENERATE    MAX(alias.column_name);</pre>
MIN	<pre>SELECT MIN(column_name) FROM table_name;</pre>	<pre>FOREACH (GROUP alias ALL) GENERATE   MIN(alias.column_name);</pre>
SUM	<pre>SELECT SUM(column_name) FROM table_name;</pre>	<pre>FOREACH (GROUP alias ALL) GENERATE   SUM(alias.column_name);</pre>
HAVING	<pre> HAVING aggregate_function(column_name)   operator value;</pre>	<pre>FILTER alias BY    aggregate_function(column_name) operator    value;</pre>
UCASE/UPPER	<pre>SELECT UCASE(column_name) FROM table_name;</pre>	FOREACH alias GENERATE UPPER(column_name);
LCASE/LOWER	<pre>SELECT LCASE(column_name) FROM table_name;</pre>	FOREACH alias GENERATE LOWER(column_name);
SUBSTRING	<pre>SELECT SUBSTRING(column_name,start,length) AS some_name FROM table_name;</pre>	<pre>FOREACH alias GENERATE SUBSTRING(column_name,     start, start+length) as some_name;</pre>
LEN	SELECT LEN(column_name) FROM table_name;	FOREACH alias GENERATE SIZE(column_name);
ROUND	<pre>SELECT ROUND(column_name,0) FROM table_name;</pre>	FOREACH alias GENERATE ROUND(column_name);

For more tips, download the full Pig Cheat Sheet: bit.ly/pigcheat

## **MORTAR**