

### Scenario

Data: TPC-H (1 GB)  
 Query: Total revenue from european suppliers per region of customer

### Traditional Query Processing

```

SELECT rc.r_name AS CustRegion,
       SUM(l_extendedprice*(1-l_discount))
       AS Revenue
FROM   lineitem, orders, customer, nation nc,
       region rc, supplier, nation ns, region rs
WHERE  <foreign-key join>
       AND rs.r_name = 'EUROPE'
GROUP BY rc.r_name
ORDER BY cust_region
  
```

**60 min**

CustRegion	Revenue
AFRICA	\$5.76 bn
AMERICA	\$2.40 bn
ASIA	\$1.00 bn
EUROPE	\$0.66 bn
MIDDLE EAST	\$3.38 bn

### Approximate Query Processing with Derby/S

```

SELECT SOME rc.r_name AS CustRegion,
       ~SUM(l_extendedprice*(1-l_discount))
       AS Revenue
       INTERVAL (
         SUM(l_extendedprice*(1-l_discount)), 90
       ) AS Interval
FROM   lineitem, orders, customer, nation nc,
       region rc, supplier, nation ns, region rs
WHERE  <foreign-key join>
       AND rs.r_name = 'EUROPE'
GROUP BY rc.r_name
ORDER BY cust_region
  
```

**6 s**

CustRegion	Revenue	Interval
AFRICA	\$5.85 bn	\$0.082 bn
AMERICA	\$2.42 bn	\$0.048 bn
ASIA	\$0.95 bn	\$0.037 bn
EUROPE	\$0.62 bn	\$0.027 bn
MIDDLE EAST	\$3.52 bn	\$0.065 bn

### Sample Repository of Derby/S

```

CREATE SAMPLE revenueSample AS
  (SELECT *
   FROM lineitem, orders, customer,
        supplier
   WHERE l_orderkey = o_orderkey
        AND o_custkey = c_custkey
        AND l_suppkey = s_suppkey)
OF SIZE 60000 TUPLES
FOR AGGREGATIONS
  (SUM (l_extendedprice*(1-l_discount)))
MANAGED BY SYSTEM REFRESH IMMEDIATE
  
```

TableName	Type
SYSTABLES	S
SYS_COLUMNS	S
SYSLOGICALSAMPLES	S
SYSPHYSICALSAMPLES	S
LINEITEM	T
ORDERS	T
REVENUESAMPLE	A
REVENUESAMPLE_O	A

Fragment of SYSTABLES

PHYSICALSAMPLEID	p1	p2
LOGICALSAMPLEID	l1	l1
TABLERNAME	REVENUESAMPLE	REVENUESAMPLE_O
ALGORITHM	Reservoir Sampling	Outlier Index
PARAMETERS	-	-
OPTIMIZATION	NONE	Aggregation
OPTIMIZATION_ATTRIBUTES		SUM(l_extendedprice * (1-l_discount))

Fragment of SYSPHYSICALSAMPLES (pivoted)