UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

Cloudera, Inc.,
Petitioner

v.

R2 Solutions LLC,
Patent Owner

Case IPR2024-00303

U.S. Patent No. 8,190,610

Petitioner's Request for Director Review

TABLE OF CONTENTS

I. Introduction 1

II. Legal Standard 1

III. Background 2

A. The ’610 Patent and the Pike and Chowdhuri References 2

B. The Board’s Institution Decision Adopted Patent Owner’s Narrowing “Data Groups” Arguments 3

IV. Argument 4

A. The Board Misapprehended the Scope of Claim 1 4

B. The Board Misapprehended the Teachings of Pike and Chowdhuri and the Petition 9

1. The Board’s Flawed Analysis of Pike 9

2. The Board’s Flawed Analysis of Chowdhuri 11

3. The Board Misapprehended Petitioner’s Obviousness Rationales 14

TABLE OF AUTHORITIES

CASES **PAGE(S)**

CUPP Computing AS v. Trend Micro Inc.,
53 F.4th 1376 (Fed. Cir. 2022) 9

DK Crown Hldg's. Inc. v. Diogenes Ltd.,
IPR2023-00268, Paper 14 (P.T.A.B. Feb. 21, 2024).....1

Markman v. Westview Instruments, Inc.,
52 F.3d 967 (Fed. Cir. 1995) 5

Leibel-Flarsheim Co. v. Medrad, Inc.,
358 F.3d 898 (Fed. Cir. 2004) 5

Ultratec, Inc. v. CaptionCall, LLC,
872 F.3d 1267 (Fed. Cir. 2017) 1

OTHER

USPTO, *Revised Interim Director Review Process*, § 2.B.....1

I. Introduction

The Director should grant review of the Board's Institution Decision ("Dec.") because the Board improperly narrowed the scope of the claims—at Patent Owner's urging—based on exemplary embodiments. In doing so, the Board adopted Patent Owner's arguments requiring the prior art to disclose several unrecited features. Patent Owner took the opposite tact in litigation, asserting infringement under a broader view and arguing the claims are not limited to the disclosed embodiments. The Board also misapprehended numerous aspects of the prior art and Petitioner's proposed combinations by adopting arguments that find no support in the record.

II. Legal Standard

A party may request Director Review if the Board's institution decision presents (a) an abuse of discretion or (b) important issues of law or policy. USPTO, *Revised Interim Director Review Process*, § 2.B. The Board abuses its discretion when its decision (1) is clearly unreasonable, arbitrary, or fanciful; (2) is based on an erroneous conclusion of law; (3) rests on clearly erroneous fact findings; or (4) involves a record without evidence on which the Board could base its decision. *Ultratec, Inc. v. CaptionCall, LLC*, 872 F.3d 1267, 1272 (Fed. Cir. 2017). Director Review is available when the Board misapprehends: a claim term, the scope and content of the prior art, or the Petition's mapping of the prior art. *DK Crown Hldg's. Inc. v. Diogenes Ltd.*, IPR2023-00268, Paper 14 at 2-4 (P.T.A.B. Feb. 21, 2024).

III. Background

A. The ’610 Patent and the Pike and Chowdhuri References

The ’610 patent concerns MapReduce—“a programming methodology to perform parallel computations over distributed . . . data sets.” Pet. 1, 3-6; EX1001, 1:4-27. The patent purportedly improves MapReduce by “treating an input data set as a plurality of grouped sets of key/value pairs.” Pet. 6-9; EX1001, 1:64-2:51. Claim 1 recites this data grouping by mapping and reducing input data comprising “a plurality of data groups,” each data group having its own “schema.” EX1001, 1:64-2:51. In prior litigation, the district court construed “data group” to mean “a group of data and a mechanism for identifying data from that group.” Pet. 9-10; EX-1007, 23-25; EX-1008, 8. The Petition applied this construction. Pet. 9-10.

The asserted Pike and Chowdhuri references disclose and render obvious mapping and reducing data groups of different schema. Pet. 10-14, 15-51. Pike teaches MapReduce, including mapping and reducing various input files such as text files and tables. *Id.*, 3-6, 10-12. Pike also discloses tracking input files through “application-specific” mapping and reducing processes to merge data based on a common key. *Id.*, 10-12, 28-33. Chowdhuri discloses relational database operations that mirror MapReduce, where different “customer” and “order” tables are scanned and joined to yield customers by state of origin. *Id.*, 12-14, 18-20. Chowdhuri discloses pseudo code with “o” and “c” identifiers for scanning, joining, and sorting

corresponding order and customer tables, just like the "emp" and "dept" identifiers and pseudo code in the '610 patent. *Id.*, 33-35; EX-1001, 4:6-59, 6:18-8:46, Fig. 4-5.

B. The Board's Institution Decision Adopted Patent Owner's Narrowing "Data Groups" Arguments

Patent Owner's Preliminary Response focused on the "data group" limitation. *See* Paper 8 ("POPR"), 34-68. While Patent Owner purported to apply the district court's construction, *id.*, 15-17, it instead argued for a much narrower application against the prior art by arguing the prior art fails to disclose "how to *use* 'data group' identifiers to deal with data of different schema." *Id.*, 39. For example, Patent Owner argued that the claimed "mapping functions" must be "specialized so that they can be 'selected' for a partition based on the 'data group' . . . and map the partitions 'differently' based on 'data group.'" *Id.*, 35-36. Patent Owner further argued that the "reducing" requires "distinguish[ing] the 'intermediate data' based on the parent 'data groups' to effectuate ***specialized*** processing on a per-'data group' basis," *id.*, 35-36 (emphasis added)—i.e., claim 1 requires reduce operators "designed to *use* an input file name to identify data from that file and ***adjust*** processing accordingly," *id.*, 55 (emphasis added).

Patent Owner's arguments rely on features not recited in claim 1. They also directly contradict positions Patent Owner took in district court. There, Patent Owner argued that the mapping and reducing functions ***need not*** be specific to a data group. *See* EX-1007, 16-18; EX-1008, 17-19. Patent Owner contended, for example, that

the mapping functions need not be specialized for each data group or schema because this would "import extraneous limitations [into the claims]." EX-1007, 16-18; EX-1008, 17-19 (same). Patent Owner also argued that processing "in a manner that is defined to correspond to that data group" does not require processing data groups differently. EX-1007, 27-29; *see R2 Sols. LLC v. Deezer S.A.*, 4:21-cv-00090, Dkt. 36 at 23-24 (E.D. Tex. Nov. 4, 2021) (arguing "the possibility that processing intermediate data of data groups, . . . , can be done in the same manner."). The district court agreed with Patent Owner. *See* EX-1007, 16-18; EX-1008, 17-19.

Petitioner apprised the Board of Patent Owner's previous positions, Pet. 10 (citing EX-1007; EX-1008), 42, but the Board nonetheless adopted Patent Owner's new and contradictory arguments to deny institution. Dec., 28-36, 41-49.[1]

IV. Argument

The Board abused its discretion by (1) improperly reading limitations in from the specification; and (2) misapprehending the Petition and the prior art teachings.

A. The Board Misapprehended the Scope of Claim 1

The Board committed legal error by adopting Patent Owner's inconsistent

[1] The Board went even further, requiring—as Patent Owner argued, POPR 41-44—that the references disclose "***form[ing]*** data groups characterized by different schema," Dec. 28 n.12 (emphasis added), rather than merely processing two groups with different schema, as required by claim 1.

interpretation of claim 1 and incorporating limitations from the specification when considering the prior art. While claims are read "in view of the specification, of which they are a part," *Markman v. Westview Instruments, Inc.*, 52 F.3d 967, 979 (Fed. Cir. 1995), it is "improper to read limitations . . . described in the specification . . . into the claims absent a clear indication in the intrinsic record," *Leibel-Flarsheim Co. v. Medrad, Inc.*, 358 F.3d 898, 913 (Fed. Cir. 2004). From the start of its analysis, the Board focused on the '610 patent's embodiments and a Patent Owner-provided "helpful summary of the invention" in annotated Figure 5. *See* Dec. 7-12 (citing POPR 9). From this, the Board distinguished the prior art under a view that a data group "identifier" must be used ***to program*** map and reduce functions for ***schema-specific*** processing. *See id.*, 28-36 (analyzing Pike), 36-49 (analyzing Chowdhuri). This is not compelled by the district court's "data group" construction. Nor is claim 1 limited by the specification, as Patent Owner asserted.

Claim 1 more broadly recites MapReduce with two data groups of different schema and mapping and reducing those inputs so they are "identifiable" and "correspond" to the data groups. EX-1001, cl. 1; Pet. 3-9. These features do not require programming specific map and reduce functions,[2] nor do they require

[2] The mapping functions being "user-configurable" does not require these functions to be data-group- or schema-specific.

schema-specific processing. Claim 1 instead recites "providing each data partition to a selected one of a plurality of mapping functions" and "reducing the intermediate data," where the data groups "***ha[ve]*** a different schema."

Despite not engaging in a fulsome claim construction analysis, the Board adopted Patent Owner's narrowed scope of claim 1 when considering the prior art. Adopting Patent Owner's arguments, the Board found that the specification discloses: (1) "the 'data group' identifier [] ***call[s] particular programming objects*** (e.g., iterators) associated with specific 'data groups,'" Dec. 29-30 (quoting POPR 11-12) (emphasis added); and (2) that "[t]he '*data group*' identifiers . . . allow[] a programmer ***to program and execute*** MapReduce functions that correspond to a particular '*data group*,'" *id.* (citing EX-1001, Fig. 5) (bold emphasis added). Based on these "programming" *disclosures*, the Board distinguished Pike, which otherwise teaches the limitations of claim 1. *See id.*, 30 ("Petitioner does not cite a comparable disclosure, teaching, example, or suggestion in Pike . . . to use an input file name as a '*data group*' identifier ***for programming*** MapReduce functions.") (bold emphases added).

The Board also distinguished Pike by holding that the "mechanism" in the data group construction must "identify the data in each input file ***by schema (or data type)*** for separate [and schema-specific] map and reduce processing." *Id.*, 31 (emphasis added); *see also id.*, 33-34 (requiring Pike to "identify and partition input

file *data* ***based on a characteristic schema or data type***" and to produce "intermediate data '*corresponding*' and '*identifiable' **to a particular schema***") (emphases added). Identifying "the data" in a data group, as required by the district court's construction, does not require identifying a data group's "schema (or data type)." *See* EX-1007, 16-18. Nor does it require using the "mechanism" to perform schema-specific processing during the mapping or reducing. The Board erred by imposing these programming and schema-specific limitations on the prior art.

The Board continued its misapplication of claim 1 to Chowdhuri by requiring Chowdhuri to disclose unrecited "processing" steps. *See* Dec. 45-46. The Board agreed that Chowdhuri "discloses the example of processing two input data tables," such as its "customer and order tables," and that (like the '610 patent) each of these tables "has a different data type or 'schema' and is partitioned for parallel processing." *See id.*, 42. Yet the Board read in limitations by requiring Chowdhuri's table names to be "***carried through*** the mapping and reducing functions," *id.*, 45-46 (emphasis added), and its table names to "***guide processing*** [of] the data in a way that corresponds and is identifiable to a particular data table," *id.*, 46 (emphasis added). These features, like those purportedly missing in Pike, are not in the claims. They are examples in the specification. EX-1001, 3:65-4:59, 7:36-8:46, Figs. 4, 5.

The map functions in the embodiments of Figures 4 and 5, for example, describe schema-specific processing. These embodiments process input key/value

pairs and produce intermediate data "for a particular group identified with a group_id [in the map function]." *Id.*, 4:48-59. Two tables are joined using one map function for an "emp" group and another map function for a "dept" group. *Id.*, 7:36-8:14; Dec. 29-30. The patent provides pseudocode for these specific functions: "map('emp', 34, 'Smith')" and "map('dept', 34, 'Clerical')," and "reduce((34, ('emp', ('Smith', 'Robinson', 'Jasper')), ('dept', ('Clerical')))))." EX-1001, 7:36-67; Dec. 29-30. Claim 1 does not require this schema-specific processing.

Claim 1 instead recites outputting "corresponding intermediate data" that is "identifiable" to a data group and processing it according to "data groups." EX-1001, cl. 1. This does not require *using* the data group mechanism in a particular way. The intermediate data need only be *identifiable* and *correspond* to an input data group through the mapping and reducing. This is how Petitioner applied Pike and Chowdhuri in the Petition. *See* Pet. 9-10, 15-51. In doing so, Petitioner explained that Pike teaches input file names as input to MapReduce, and tracking those through the map and reduce processing. *See id.*, 10-12, 15-51. Petitioner further explained that Chowdhuri teaches two data tables of different schema and merging those using a MapReduce-like process. *See id.*, 12-14, 15-51. Indeed, Chowdhuri *relies* on the data in its customer and order input tables to merge them into a logical output of customers by state of origin. *Id.*, 18-20, 25-28, 33-35, 38-42, 45.

Thus, the Board erred in adopting Patent Owners arguments improperly

limiting the claims to examples in the specification. Indeed, Patent Owners are not permitted to "shapeshift their claims through argument in an IPR." *CUPP Computing AS v. Trend Micro Inc.*, 53 F.4th 1376, 1383 (Fed. Cir. 2022). But that is what occurred here. The Director should grant review to focus the inquiry on claim 1, not the "claims the patentee now wishes it had secured." *Id*.

B. The Board Misapprehended the Teachings of Pike and Chowdhuri and the Petition

Separate from the Board's narrow reading of claim 1, the Institution Decision should be vacated because the Board misapprehended the teachings of Pike and Chowdhuri and the Petition's mapping of those references.

1. The Board's Flawed Analysis of Pike

Contrary to the Board's findings, Pike's input files disclose or at least render obvious "data groups" because Pike's input files and associated file names are groups of data with mechanisms for identifying the data from each group. Pet. 10-12, 16-18, 21-22, 35-38; EX-1002, 2:22-26, 9:56-10:7, Fig. 3; EX-1005 ¶¶88-93. In holding otherwise, the Board incorrectly found that Petitioner "refers to the '610 patent itself for the use of file names to identify Pike's otherwise unidentified 'tables.'" Dec. 30. This misapprehends the Petition. Petitioner explained what exemplary file names in Pike would include (e.g., a file name and extension), and that "[a] file name is an identifier associated with the data in the file, and typically includes information about the content, type, and structure of the data in each file."

Pet. 17-18 (citing EX-1005 ¶¶88-91). Petitioner further explained that the '610 patent describes "employee" and "department" tables that a POSITA would have understood would be stored as files with similar file name identifiers as Pike. *Id.* (citing (EX-1005 ¶92)). Explaining what exemplary file names would be in the prior art and in the context of the asserted patent is not "explicitly and implicitly rel[ying] on the teachings of the '610 patent" to assert obviousness. *See* Dec. 30-31.

The Board further misapprehended Pike by finding that its input file names "could not serve as the claimed '*data group*' mechanism" because Pike purportedly discloses data blocks that "can include data from more than one input file." Dec. 31-32. Petitioner did not rely on this embodiment of Pike, and the Board did not consider Pike's alternative embodiments that Petitioner did rely on. Specifically, Pike discloses "input files 202 hav[ing] a predefined maximum size (e.g., 1 GB), and the individual files [that] ***are*** the data blocks," EX-1002, 4:54-64 (emphasis added); Pet. 21-22. Pike also discloses that "the input files field may specify ***a portion*** of ***an input file*** (e.g., where the portion comprises a data block), *id.*, 6:55-62 (emphases added); Pet. 30-32. Pike alternatively teaching data blocks "from more than one input file" does not exclude Pike's teaching that each data block may correspond to a particular input file and tracking each of those data blocks/files through the map and reduce processing. Pet. 21-22, 30-33. The Board erred in holding otherwise.

The Board also erred in finding that Pike only uses "numerical identifiers to

locate and distribute input files for processing" instead of identifying "an original input file." Dec., 33 (reproducing EX-1002, Fig. 7A). The Board acknowledged the opposite is true by finding that "Pike discloses that '[t]he *input files field* [Fig. 7A] may specify ***a portion of an input file***.'" *Id.*, 33-34 (bold emphases added). Despite this, the Board held that Pike *distinguishes* its input file names from its Task Status Table "numerical identifiers," *see id.*, 34-35, and that Pike's "map-processed data *loses* its identification with its respective 'input files,'" *id.*, 34 (citing POPR 45-46). These findings lack support in the record. Pike discloses no distinction between its input files and what the Board referred to as "numerical identifiers," nor does it teach a loss of identification. As Petitioner explained, Pike's master process 320 "tracks the processing of ***each input file*** using identifiers through mapping and reducing," and Pike discloses "***the same input file*** ('12340')" in its Task Status Table for the map and reduce tasks. Pet. 30-32 (emphases added) (citing EX-1002, 6:55-7:12, 13:7-17, Fig. 7A). This is "indistinguishable from the '610 patent, which describes using 'emp' and 'dept' identifiers to associate intermediate data with its Employee or Department input tables." *Id.* The Board thus erred in finding that Pike does not track its input files through its map and reduce processing.

2. The Board's Flawed Analysis of Chowdhuri

The Board's adoption of Patent Owner's unsupported arguments continued with its analysis of Chowdhuri and its combination with Pike. *See* Dec. 41-49. As

Petitioner explained, Chowdhuri teaches MapReduce-like processing based on scan and join iterators. Pet. 19-20, 21-23, 26-28, 34-35, 40-42, 44-45, 49-50. Chowdhuri applies these processes to tables of different schema (e.g., "customer" and "order" tables) and programming parallel scan and join iterators to merge the tables based on content. *Id.* As Petitioner explained, the Pike-Chowdhuri system "would have processed Chowdhuri's tables using, for example, Chowdhuri's scanning, hashjoin, GroupBy, and sort iterators as map and reduce tasks in Pike, while employing Pike's fault-tolerant architecture to do so." *Id.*, 19-20; *see also id.*, 23-24, 26-28

The Board erred with respect to Chowdhuri in several ways. First, the Board found that Chowdhuri's scan operations are distinguishable from the claimed "mapping" because "Chowdhuri is not concerned with tracking a data source through data processing functions." Dec. 43. But Chowdhuri's scan iterators correspond to and render obvious the claimed "mapping" functions, and the Board was wrong to find otherwise. *See* Dec. 43. Chowdhuri's scan iterators work to scan specific inputs and produce output data that is consumed by another process, such as a hashjoin iterator. EX-1003 ¶130; *id.* (identifying source locations as "a disk file, a network connection, or even another iterator"). In doing so, "the iterator applies a predefined behavior to the data set being processed, manipulating the data according to the specification of that iterator." *See* EX-1003 ¶130. The Board misinterpreted these disclosures as the iterators not knowing "the source of the data stream." *See*

Dec. 43. But these disclosures make clear that, while Chowdhuri's iterators do not track the source ***location*** of the input data, they do understand the ***content*** of the input data. The Board misapprehended this fact and Chowdhuri's teachings.

Next, the Board found that Chowdhuri's scan operation does not output "intermediate data." Dec. 44, 47-48. This is incorrect. Chowdhuri's scan and join operations "produce data at its output which then becomes an input source of data to another operator." EX-1003 ¶128. Chowdhuri "pipelin[es] . . . ***intermediate results*** from one set of operators (i.e., producer tasks [like scanning]) and initiates another set of operators to process the next request (i.e., consumer tasks)," *id.* ¶134 (emphasis added), and "[o]nce these [input] tables are scanned, ***results are sent*** to the hash join iterator 303 to be matched," *id.* ¶131 (emphasis added). In the context of Chowdhuri's customer and order tables, "Chowdhuri's scan iterators select a state from its customer table and the order amount from its order tables (i.e., '[S]elect state, [S]um(order_item),' 'from customer c, order o')," and "[a] POSITA would have understood that an exemplary Chowdhuri ***value*** from its customer table would be 'Virginia' based on a customer key, and an exemplary ***value*** from its order table would be '$25' based on a ***customer key***." Pet. 33-34 (emphases added) (citing EX-1003 ¶¶126-27; EX-1005 ¶120). Chowdhuri therefore outputs intermediate data (e.g., "lists of values" for each key), and the Board misapprehended Chowdhuri to hold otherwise.

Finally, the Board found that Chowdhuri's "customer" and "order" table names are not "data groups" because they purportedly do not "serve to identify the *data* in each . . . table[]," and do not represent "an 'identifier (table name or otherwise) being, e.g., passed to specific functions to guide group-specific processing.'" Dec. 45-46, 46-47. This is wrong, as Chowdhuri's "customer" and "order" table names indicate that the tables include customer and order data. Pet. 18-20, 33-35, 40-42. Chowdhuri also processes these tables based on the table names and content by describing a scan and join of "o" (order) and "c" (customer") data. *Id.* Absent Chowdhuri's table name identifiers, Chowdhuri would be unable to logically merge its data based on customers by state of origin. *See id.* The Board's contrary finding appears to be predicated on Figure 5 of the '610 patent and the unclaimed requirement that "an 'identifier (table name or otherwise)" must be "***passed*** to specific functions to guide group-specific processing as in the '610 patent.'" Dec. 46-47 (quoting POPR 56-57). Claim 1 does not require reconstructing Figure 5 of the '610 patent. Even if this were a claim requirement, Chowdhuri teaches this by programming its iterators to scan and logically join data associated with "c" customer data with "o" order data. Pet. 12-14, 18-20, 33-35, 40-42.

3. The Board Misapprehended Petitioner's Obviousness Rationales

The Board also misapprehended Petitioner's obviousness rationales. Petitioner explained that it would have been obvious to track input files in Pike and

Pike-Chowdhuri, to the extent not already disclosed, because doing so would have enabled "post-processing operation[s]" for "rerun[ning] reduce tasks on a subset of intermediate data" without having re-map the initial data set. *See* Pet. 33, 35, 38, 42. The Board distilled this to "selectively reduc[ing]," without more, and held such reasoning was "grounded in benefits residing in the '610 patent." Dec. 35-36 (citing POPR 47). But the Board identified no such teaching or disclosed benefit in the '610 patent. The Board's finding therefore lacks adequate reasoning or bases.

The Board also erred in dismissing the reasons for combining Pike's task scheduling with Chowdhuri's scan and join iterators. *See* Dec. 48-49 (finding no "improvement in efficiency"). As Petitioner explained, a POSITA would have looked to Pike to map and merge Chowdhuri's tables using Pike's fault tolerant task scheduling. Pet. 20 (citing EX-1005 ¶97; Pike, 4:28-53, 6:37-54, 6:55-7:25, 11:57-13:3). Petitioner's expert explained that Pike's "fault tolerance is a major advance over existing processing methods" and is used to track input data through its processes and handle processing faults to "significantly reduce[] the time to complete large map-reduce operations." EX-1005 ¶97. Thus, Pike's data file tracking enables faults across a distributed system to be managed and addressed, in the event of individual process failure, thereby improving the system of Chowdhuri. *See id.*; Pet. 20. The Board did not fully address or consider Petitioner's rationales.

For the foregoing reasons, Petitioner respectfully requests Director Review.

Date: July 11, 2024

Respectfully submitted,

/Jeffrey A. Berkowitz/
Jeffrey A. Berkowitz, Lead Counsel
Reg. No. 36,743

CERTIFICATE OF SERVICE

The undersigned certifies that a copy of the foregoing **Petitioner's Request for Director Review** was served on July 11, 2024, via email to counsel of record for the Patent Owner at the following:

Carder W. Brooks
Brent N. Bumgardner
Nelson Bumgardner Conroy PC
3131 W. 7th Street, Suite 300
Fort Worth, TX 76107
Telephone: (817) 377-9111
carder@nelbum.com
brent@nelbum.com
R2Solutions-IPR@nelbum.com

Dated: July 11, 2024

/Mark A. Rosenberger/
Mark A. Rosenberger, Case Manager
Finnegan, Henderson, Farabow,
Garrett & Dunner LLP